FORMULATING HEALTHCARE EVIDENCE: CASE STUDIES IN MEDICAL TECHNO-PRACTICE IN THE UNITED KINGDOM, 1990 TO 2000

Alex Faulkner

PhD.

The University of Edinburgh

2004

I do not know which to prefer,
The beauty of inflections
Or the beauty of innuendoes,
The blackbird whistling
Or just after.

******

It was evening all afternoon.
It was snowing
And it was going to snow.
The blackbird sat
In the cedar-limbs.

Wallace Stevens
from Thirteen Ways of Looking at a Blackbird

C'mon I'll play it for ya
Lemme tell yuh about it
Lemme tell yuh about it
There were two trains
Two railroad tracks
-Click-clack click-clack-
One ah them leavin'-uh
'N the other one comin' back

Don Van Vliet
Click Clack

ACKNOWLEDGMENTS

I am indebted to the co-authors of the publications collected as part of this thesis, both for their contributions to the research and writing, and for giving their permission for the publications to be included here. A large number of named and anonymous referees provided insightful and helpful comments on pre-publication versions of the published papers, and I am grateful for their guidance.

I acknowledge the permission of the following publishers to reproduce previously published material: BMJ Publishing Group (two articles in the Journal of Epidemiology & Community Health); Elsevier (two articles in Social Science & Medicine); New York University Press; National Coordinating Centre for Health Technology Assessment; Oxford University Press (article from the European Journal of Public Health); Blackwell Publishers (chapter from Sociology of Health and Illness monograph no.3).

I am grateful to the Department of Social Medicine, University of Bristol for providing a stimulating environment during the 1990s while I was undertaking much of the work discussed in the published papers presented here. Similarly, I am grateful to Cardiff University School of Social Sciences, for providing subsequent encouragement and a conducive intellectual environment in which to formulate the overview essay.

I would like to acknowledge with gratitude the time I spent working with the late John Farndon, Professor of Surgery at Bristol, from whom I learned to think of 'patients with cancer' rather than the ubiquitous but unsympathetic 'cancer patients'.

I thank Edinburgh University for accepting me as a doctoral candidate, and Sarah Cunningham-Burley for providing constructive advice on the contents of the critical review essay included in the thesis.

Martha and Ruth Faulkner showed tolerance and amusement in having an 'artificial hipologist' as their father.

Last but not least, I thank Margaret Sleeboom for her generous and insightful support, and encouragement in completing this work.

CONTENTS

APPENDICES

DECLARATION

This thesis has been composed by Alex Faulkner.

It is my own original work except where indicated in the case of co-authored publications (Appendix 2).

The work has not been submitted elsewhere for any other degree or qualification.

Signed

Date 20.11.04

ABSTRACT

Formulating healthcare evidence: case studies in medical techno-practice in the UK, 1990-2000.

The development of multidisciplinary Health Services Research and Health Technology Assessment in the United Kingdom in the 1990s informs government R&D policy for a 'knowledge-based health service'. This thesis comprises health service case studies focused on medical techno-practice in a collection of eight publications and an original critical review essay. A wide variety of primary and secondary research methods are used in separate empirical studies. Perspectives from sociology applied to medicine and health care, science and technology studies (STS), and the multidisciplinary field of 'health services research' are combined to offer a detailed and reflexive account.

Health care is a field of policy and technical practice marked by complex, hybrid problems, as well as being associated with a wide variety of physical hazards and socially perceived risks. The case studies are drawn from three substantive fields of healthcare technology and practice affecting large proportions of the population: human implant technology (total hip prostheses); cancer detection (testing for early-stage prostate cancer); and outpatient care. The inter-related objectives of the work are, firstly, to analyse variability in patterns of health care. This enables, secondly, analysis of observed techno-practice variations in terms of their implications for effectiveness of interventions, social patterns of health care consumption, risks to patients and health care policy. Thirdly, explanations for variability in practices and policies can be offered, which suggests the need, fourthly, for sociologically-informed approaches to analysing policy for the introduction of 'new technology' into healthcare systems. The final objective is to make the case for, and contribute to a sociopolitical analysis of the advance of the new healthcare sciences and their articulation with public policy in contemporary advanced health care systems.

The results of the three case studies show the following. The proliferation of costly artificial hips evoked a national policy response that draws heavily upon health technology assessment evidence and processes. Early detection of prostate cancer presents dilemmas associated with surgical specialisation, and conflict between policy, clinical beliefs and practices. Outpatient care demonstrates tension between modernisation, evidentiality and the obduracy of shared socio-clinical practices. The common threads drawing the case studies together are the observed variability in patterns of health care delivery and practice; underlying patterns of medical beliefs, professional work organisation and socialisation; the shaping of healthcare risks for patient populations; and linkages between healthcare innovation, modernisation and counteracting controls.

The review essay builds upon the published work to develop a reflexive understanding of the activity of healthcare science as a policy-related enterprise, and argues that healthcare evidentiality as represented by the new healthcare sciences, its proponents and its institutional vehicles, should be considered as having legitimating and regulatory functions. Evidentiality should thus be considered one of the societal forces that must be embraced by a socio-politics of the dynamics of healthcare innovation and governance. The original critical review essay accompanying the published work contributes to this enterprise.

1. Introduction

1.1 Aims of the critical review

This review essay summarises the aims, methods, results, conclusions and contribution to knowledge made by eight submitted publications. A statement is given of my contribution to the multi-author publications presented (Appendix 2).

The majority of the publications present 'evidence' about the variability of different medical techno-practices, which I define as healthcare provided with the deployment of technical knowledge or material technology, or both. The essay analyses the disciplinary underpinnings of the work, and its foundations in the expansion of the new health care sciences geared toward high-priority health care policy problems during the 1990s in the United Kingdom. In this essay I combine frames of reference associated with multidisciplinary 'health services research' (HSR) and 'health technology assessment' (HTA) with more sociological perspectives associated especially with science & technology studies (STS) and the sociology of scientific knowledge. I treat HSR and HTA as examples of newly emergent healthcare sciences susceptible to the latter approaches.

Three case studies of medical techno-practice are presented in the published papers: artificial hip replacement, early detection of prostate cancer, and evaluation of outpatient services. These are chosen because they affect large proportions of the population and a wide cross-section of it. They have been of high priority within National Health Service research and development policy in the United Kingdom in the 1990s, and they demonstrate different patterns of the relationship between material technology and clinical practice (ranging from high salience of material technology in artificial hip replacement, through the clinical use and interpretation of blood testing technology in prostate cancer, to a complex mixture of technologies and shared clinical practices in outpatient services when considered at a high level of generality).

The common themes across the studies taken as a body of work are demonstrated in detail in this essay. They are summarised in section 1.3.2 below and are drawn out at the end of each case study chapter. By way of introduction, the first thread drawing the case studies together is the *observed variability* in patterns of health care delivery and practice, which can be interpreted as pointing to organisational 'inefficiencies' and possible inequalities in social access to healthcare services. Secondly, underlying observed variability, the case studies exhibit *patterns of shared, collective medical beliefs, professional work practices* and socialisation at a variety of levels of medical organisation and aggregation. Thirdly, patient populations experiencing these patterned techno-practices are exposed to particular shaping of healthcare *risks,* in other words risks to health and other aspects of citizenship that are attributable to the healthcare system itself. Finally, I argue that the case studies should be analysed as demonstrating linkages between the phenomena of healthcare *innovation, modernisation and counteracting controls*.

Further, I take the opportunity in this essay to extend the previously published work by reflecting in particular upon the role of the growth of 'evidentiality' and the new healthcare sciences as phenomena in themselves associated with the social legitimation and governance of contemporary public healthcare systems. I argue that it is important to formulate the societal phenomenon of varieties of scientific 'evidence' as one key discourse in the study of healthcare system innovation.

1.2 Disciplinary perspectives, perspectives on interdisciplinarity

In this introduction it will be useful to discuss the question of the disciplinary perspectives which I am drawing upon in the submitted publications and in the critical review presented here. The body of work presented in the publications is diverse both in subject matter and in the disciplinary perspectives that have shaped the research questions that have been addressed and the methodological approaches to analysis and interpretation that have been adopted. I suggest that to some extent

this is symptomatic of widespread trends in the institutions of a 'knowledge economy' and of practices of scientific research as manifest in contemporary health care systems. The interplay of disciplinary perspectives is a theme that I return to periodically and build upon in this essay.

The body of work presented and discussed here sits somewhat uneasily in the context of existing, conventional disciplinary and academic-institutional boundaries. It spans both social science and clinical science, and much (but not all) of the work could legitimately have been defined as falling within the remit of either. It is the case that these disciplines and boundaries carry with them a wide variety of different conventional assumptions and expectations about research methodologies, definition of research questions, appropriate styles and methods of analysis, and modes of academic presentation. Thus, for example, in this essay I will occasionally use the first person to describe aspects of my work, in the knowledge that this is a far more acceptable practice in 'social science' than it is in 'medicine', where third-person reportage in the passive voice is the norm. I do not dwell here upon the philosophical, political and methodological issues invoked by this particular difference.

This thesis embraces, in particular, the academic discipline of sociology, applied to medicine and health care, and the multidisciplinary areas of work commonly known now as 'health services research' (HSR) and health technology assessment (HTA) which typically in the 1990s was focused upon research agendas of effectiveness and cost-effectiveness of public healthcare. The latter I take to bring together in varying combinations the disciplines, sub-disciplines and practices of epidemiology; public health medicine; health economics; medical statistics; psychology; organisational analysis; medical/healthcare sociology; 'qualitative research'; and general practice and the many specialties of medicine and clinical science. Note the inclusion of sociology and qualitative research in my list. Not all conceptualisations of HSR/HTA in the 1990s would have incorporated these. For example a typical handbook of health service evaluation listed the required 'skills' as: "..biological sciences, clinical science, clinical practitioners, managers, health economists, statisticians and

information officers" " (St Leger et al, 1992[1]). And in Paper 1 analysing the disciplines and epistemology of Health Technology Assessment, I cited a prominent history of 'medical research' which noted the bringing-together of 'strange bedfellows, ranging from psychology and the social sciences to biomathematics' (Weatherall, 1995). HTA, nevertheless, can be characterised as having a rather stable core methodological programme (Woolf & Henshall, 2000). As these varying listings suggest, HSR and HTA have become fields of knowledge production marked by uneasy and unclearly-defined partnerships.

In this essay I will switch between a disciplinary, broadly sociological, perspective and a 'multidisciplinary' perspective, less clearly aligned with any one conventionally-defined academic discipline, which I take to be addressed to the evaluation of health care 'problems'. Health services research is a very good example of a problem-driven area of multidisciplinary research expertise.

At the time of writing (2003-4), the academic world has been witnessing for some years a notable increase in discussion of notions of multi-disciplinarity, inter-disciplinarity and trans-disciplinarity, and a variety of activity representing attempts to bring disparate disciplines together (e.g. Klein, 1996). The institutionalisation of this activity can be observed in the creation of a plethora of research institutes, centres, networks and partnerships in and around the academic world, typically with a form of matrix organisation around broad problem-themes. This development is evident in (and across) the physical, engineering and medical sciences as well as in the social sciences. The roots of this trend are themselves a topic worthy of the attention of social scientists. An analysis of them might point to the roles of science in a 'knowledge-based economy' (Ravetz, 2001); increasing concern about the articulation points between research-based knowledge and public policy ('evidence-based policy and practice', e.g. Davis & Nutley, 1999); increasing movement toward client and patient-oriented public services (e.g. Mead & Bower 2000; Greener 2003); the putative advent of a (global) 'risk society' (Beck, 1992; Beck 1999); the rise of the 'regulatory state' (Moran, 2001) and the movement from government to

[1] I use the Harvard system of referencing in this thesis.

governance (Rhodes, 1997); increasing efforts to address the complexity of sociotechnical issues such as 'the environment' and 'transport'; increasing recognition of the local and personal expertise of non-academic citizens (Irwin, 1995); increasing concern about the legitimisation of science (Fuller, 2000); and the effects of 'cognitive revolutions' such as biochemistry, molecular biology and genomics, and perceived needs to bring different sciences together to engage in practical problem-solving activity.

In the interdisciplinary field of science and technology studies and in the sociology of scientific knowledge, some of these trends in the contours of scientific disciplines and in the application of expert knowledge have been described - in widely quoted studies - in terms of "Mode 2 Knowledge Production" (Gibbons et al 1992), and "post-normal science" (Funtowicz and Ravetz, 1992).

Firstly, "Mode 2 knowledge production" is defined as a set of changes characterised by complexity, hybridity, non-linearity, reflexivity, heterogeneity, and transdisciplinarity in the cognitive domain. As suggested above, this form of 'knowledge production' involves new configurations of work, and a new social distribution of expertise amongst a wider constituency of multiple stakeholders than previously acknowledged as participants in science. In these circumstances previous organisational boundaries aligned with mono-disciplinarity start to blur, and underlying notions of competence may be redefined.

Secondly, "Post-normal science" can be understood as an attempt to conceptualise a set of developments as a reaction against reductionist and mechanistic assumptions about the partitioning of physical and social reality, the assumed value-neutrality of science, the social segregation of science from stakeholder and community participation, and the belief in science as a vehicle of precision and certainty. Postnormal science is associated with "unstructured" problems that embody complex cause-effect relationships crossing traditional domains of enquiry and thus promoting transdisciplinarity. Such problems typically exhibit a high plurality of social values

and factual knowledge in contexts where there is pressure to arrive at societally acceptable policy decisions.

Trends such as those outlined above, I would like to suggest, produce an increasingly institutionalised tension between mono-disciplinarity and cross-disciplinarity which is having far-reaching consequences in higher education and research institutions, both in relation to research and to teaching and learning. These tensions are also to be felt at the individual level. Strong arguments can be mounted in favour of cross-disciplinary approaches to tackling 'problems' in public policies and social practices. For example it has been argued in the case of understanding risk-related behavioural phenomena, that mono-disciplinarity has a *negative* impact because of its 'decontextualising and reductionist tendencies' (Horlick-Jones and Sime, 2004). Health care is a field of policy and practice marked by complex, hybrid problems, as well as being associated with a wide variety of physical hazards and socially perceived risks.

Thus it is reasonable to expect that the body of work presented and discussed in this review essay embodies some of the disciplinary tensions from which it arises. This section has highlighted a concern with the boundaries of disciplines. As well as embodying disciplinary tensions this essay seeks to challenge these boundaries in a manner that can engender dialogue, and point toward possibilities for the constructive development of multi-dimensional evaluation in healthcare organisation and policy processes. Such a challenge to conventional boundaries is echoed, for example, in calls to 'de-monopolise' the assessment paradigm of Health Technology Assessment (Webster, 2004) by making policy networks more open to societal participation and by seeking, in the terms of this essay, to make their assessments of innovative techno-practice more multi-dimensional.

1.3 Medical techno-practice: case study topics in health care and technology

In this section the substantive topics of the published papers are introduced, and the initial introduction to the relationships and common threads between them (sketched above in section 1.1) is further described.

In a total of eight research papers (referenced 1 to 8 below, see Appendix 1 for bibliographic listing), original contributions are made to three substantive health care topics: human implant technology (total hip prostheses, also referred to as 'artificial hips' or surgically as 'hip arthroplasty'); cancer detection and care (testing for early-stage prostate cancer); and outpatient care. This work forms part of the massive development in multidisciplinary Health Services Research and Health Technology Assessment which has been characteristic of government R&D policy in the United Kingdom of seeking to move toward a 'knowledge-based health service' (DoH 1993) during the 1990s. Part of this policy agenda was increasing awareness of a need for 'good evidence' on which to base an efficient deployment of resources to produce high-quality patient care. This phenomenon is analysed using approaches drawn from the sociology of scientific knowledge in one of the papers (Paper 1) which can be read as an introduction to the epistemology and methodologies of much of the content of the remaining seven pieces.

As noted above, the three substantive case study topics are all in areas which have been of high priority within National Health Service research and development policy in the United Kingdom during the 1990s. The detailed backgrounds to these topics are outlined in the introductions to the case study chapters 2 to 4 of this review essay, and are summarised briefly here for convenience.

Artificial hips

Total hip prostheses were by the 1970s a very successful high-technology surgical procedure, but the health services policy community identified a 'problem' in the 1990s. This was seen as a proliferation of new and often expensive designs of the

technology, and a parallel variation in patterns of clinical choice of models, for which the evidence on which to base decisions about clinical effectiveness was weak. Occasionally devices appeared to fail. This problem required analysis of the performance of different models of artificial hip, and in this author's view, analysis of the innovation environment which produced new technologies, and of processes of the production of 'evidence' and regulatory control. The methodology of HTA systematic review was required in order to assess and compare existing research on performance, while historical and sociological approaches were needed to conceptualise and analyse innovation processes, clinical practices and regulatory policy.

Detection of early-stage prostate cancer

Rates of mortality from prostate cancer, and the incidence of its detection, were rising dramatically during the 1980s and early 1990s internationally. However, it was clear that part of this trend could be attributed to increased rates of detection of the disease in its early stages, especially through a relatively new diagnostic test which became widely available to medical practitioners. For the NHS the key question was: should a population screening programme be introduced utilising this test? The policy response to this question required knowledge of the performance of the test in appropriate populations and of how the medical profession was using it in clinical practice. To address these questions required a combination of clinical science (urological) knowledge of patient profiles and the testing technology, together with identification of social and organisational 'variables' relevant to clinical practice in this field. Also helpful would be a sociological understanding of the importance to the construction of men's risk experience of interpretive practices in professional work. This is offered in this essay.

Outpatient service delivery

In the case of outpatient service delivery, there had been a strong feeling amongst some commentators and health care practitioners as early as the 1970s that a phenomenon of 'recycling' of patients through outpatient clinics was common. Sometimes described as a 'merry-go-round', and often attributed to young doctors

unable or unwilling to discharge patients, there appeared from these critical comments to be a degree of unjustifiable use of NHS hospital resources. The evidence for this phenomenon and its possible causes, however, was weak. Analysis of the institutional settings and conventional clinical practice in and through which outpatient care was delivered, was largely absent in the empirical research record. Multiple disciplines were again useful in tackling these questions and interpreting the resulting research.

It is worth noting that what I have described in the paragraphs above as a requirement to combine sociological perspectives and understandings with perspectives from HSR/HTA and clinical sciences was not necessarily recognised in this way by other research participants. My academic-institutional locale was primarily a department of epidemiology and public health, and my own experience was that it was difficult to claim (successfully) *as sociological* certain research questions, topics or agendas - for example what I formulate in this essay as 'shared clinical belief systems'. Indeed, in the analysis I present in Paper 1 in this thesis, I imply that the experience of sociology as a discipline in such environments was often one of what might be termed de-disciplinisation in which it lacked the authority to frame research questions and frequently became synonymous with 'qualitative research' (an unfortunate development in this author's view. However, I suggest that since the 1990s this situation has altered to some extent with the increased attention, especially, to patients' experience and the acceptability of healthcare innovations).

Having presented these thumbnail sketches of the three case study topics, and the methodologies required to address the research questions, I now turn to a synopsis of the aims of the thesis.

1.3.1 Summary of Aims and Objectives

Given this summary of the substantive themes of the work presented here, in the submitted papers I have the following objectives:

A. To apply rigorous, multidimensional, methods of health care evaluation to three substantive topics, drawing upon a variety of disciplinary perspectives;
B. To analyse variation in patterns of health care delivery at a variety of levels (in orthopaedic surgery, prostate cancer detection and outpatient services);
C. To analyse observed practice variations in terms of their implications for effectiveness of interventions, social patterns of health care consumption, and health care policy;
D. To offer explanations for observed clinical practices and health care policies;
E. To suggest alternative models for analysing policy for the introduction of 'new technology' into the health service, and to suggest specific changes to health care policies following from the study results;
F. To suggest implications of the case studies for future research – understood as the 'production of new health care knowledge';

In addition to the case study objectives, this essay has two further objectives:

G. To make the case for, and contribute to a sociopolitical analysis of the advance of the new healthcare sciences and their articulation with public policy in contemporary advanced health care systems;

H. To assess the roles and inter-relationships of different disciplines in producing knowledge relevant to healthcare.

1.3.2 Common themes introduced

As advertised above, a number of common threads draw together the submitted work when viewed as a whole. These are reprised briefly at the end of each case study section (chapters 2 to 4) and are discussed at greater length in the context of the advance of healthcare science during the 1990s, and of the need for a 'sociopolitics' of healthcare policy, in the final section. Here, I provide some amplification of these

inter-related themes in order to pave the way for a reading of the case study sections below, and the published works included in this thesis.

As noted above, Paper 1 presents an analysis of the phenomenon of the growth of healthcare technology assessment considered as a multidisciplinary problem-oriented scientific enterprise. HTA and its overlapping movement in healthcare science, health services research, have a primary concern to assess the quality, effectiveness and cost-effectiveness of healthcare interventions and systems. A key method of identifying the existence of possible ineffectiveness is the comparative method. This has been adopted widely in HSR and HTA. Early illustrations of this method, which might be termed the 'method of variations', to assess the delivery of health care can be seen in the work, for example, of Wennberg in the USA and McPherson in the UK (Wennberg and Gittelsohn, 1982; Wennberg, 1988; McPherson 1988). This has been and remains a widely used and highly influential approach to assessing the possible variations, and thus inequalities in health service provision and possible inefficiencies in modes of service delivery.

In the papers presented in this submission, effectiveness and quality of health care, represented in patterns of variations, are at issue in a number of dimensions of care delivery: in the design and composition of different material technologies; in aggregated professional practitioners' clinical beliefs and practices, and in patterns of clinical service rate variations within and across secondary care settings. Data and analysis of such variation is especially to the fore in Paper 3 (variations in performance of artificial hips), Paper 5 (clinical beliefs and practices in detection of early prostate cancer) and Paper 6 (variations in re-attendance rates in hospital outpatient care). Group and hospital-level variations in practice are metaphorically described later in this essay as 'institutional signatures' by which might be recognised systematic differences in medical practice at different levels of analysis.

Related to this focus upon care *delivery* is a concern with medical work organisation and professional development aspects of medicine that *underpin* observable variations in the delivery of care services. In particular, a number of the submitted

papers draw attention to the phenomenon of specialisation and sub-specialisation within the medical profession. This is evident in the case of orthopaedic surgeons who specialise in hip prostheses and who practice either in specialist orthopaedic centres or in general hospitals; in the case of urologists who specialise in prostate disease and whose work also may be organised through specialist centres or in general surgical care settings; and in the case of hospital outpatient care considered broadly, where specialisation and sub-specialisation are seen to be associated with patterns of variations in service delivery rates some of which are more readily explicable than others.

One of the important early contributions of medical sociology to the understanding of medical practice variations has been to note the 'conventional' and routinised nature of medical encounters, in which the clinical practices of health professionals employ 'recipes' in negotiating interactions with patients (e.g. Bloor, 1976). While the detailed analysis of outpatient assessment practices provided by Bloor provides an argument for the 'local' constructed nature of professional health beliefs, practices and behaviours, this may be contrasted with both the avowedly cosmopolitan nature of medical knowledge and professional orientation, and the identification of broadly-defined, shared patterns of institutional and professional practice demonstrated in the work presented in this thesis. As already noted, the published work included in the thesis provides examples of patterned variability in collective medical beliefs and practices in a variety of dimensions, locales and levels of aggregation. As the case studies show, this includes for example variability between consultant teams and between organisational units such as hospitals (outpatient services), between age-cohorts of clinicians across clinical centres (detection of prostate cancer), and between specialists and non-specialists in particular medical conditions (hip replacement).

Health hazards and the societal apprehension of 'risk' have assumed an extraordinarily large place in analysis of contemporary healthcare systems and public health (e.g. Lupton, 1995; Petersen 1996; Howson 1998; Robertson 2000) and of course in social theory more broadly, as noted above in the case of Beck's hypothesis

of the risk society (1992). The papers presented here share a concern with hazards and the social distribution of risk associated with healthcare itself, sometimes termed iatrogenic risk (and given its most extreme formulation, perhaps, in the early work of Illich (1975) but highly relevant to contemporary healthcare (Edwards, 1999)). As will be demonstrated, this thread runs through all three case study topics, evidenced in the differential performance and occasional failure of hip prostheses, in the differential exposure of men to the early detection of prostate cancer and to psychological risks of appraising the tests for the disease, and in the vagaries of patients' experience of hospital outpatient care systems. Alongside risks to health should be set risks to the healthcare system itself, as demonstrated in an analysis of the research agendas of healthcare science (see Paper 1), and thus the framing of some of the research questions in the papers presented here, which shows a major focus upon risks to healthcare budgets and the organisational efficiency of systems of healthcare delivery.

Technical innovation and innovative healthcare practice, at a general level of analysis, are central to the substantive topics investigated in the body of work presented here: innovation in the design and material composition of artificial hips, innovation in the testing technology available for the diagnosis of prostate cancer, and innovation in the organisation and culture of outpatient care. As the case studies show, and as I try to capture in using the term 'techno-practice', the social and the technical should be seen as closely interwoven in developing understandings of the dynamics of healthcare activity. Such activities are, broadly, 'sociotechnical' – as with other fields of human endeavour (cf. MacKenzie,1999). Of course, governments and healthcare executives espouse to some degree at least a doctrine of modernisation that promotes innovation in medical techno-practice. Innovation holds the potential to engender improved quality and effectiveness of healthcare services. But as May et al have pointed out (2001), this commitment in contemporary healthcare culture is often in conflict with the commitment to a strong evidence base for policy decisions. Indeed, it may be seen as somewhat paradoxical (as noted in Paper 1: 203) that a high level of innovation in the many facets of healthcare provision is central to the growth of the new healthcare sciences during the 1990s,

which brings a radical questioning of the evidential basis on which innovation decisions are made. This leads us to an analysis of healthcare innovation that accords the new healthcare sciences themselves a prime position in conceptualising the dynamics of state attempts to regulate and control innovation in the healthcare system. Thus I argue that 'evidence-based healthcare' should be conceptualised as a regulatory and credentialist phenomenon as well as a scientific movement. This point is developed and discussed further in the final discussion section of this essay.

This brief discussion has sought to draw attention to the main threads which interlink to draw together the body of work submitted in the published papers presented here. These interwoven threads can be summarised as analysing and theorising:

- observable variations in patterns of health care delivery and practice;
- patterned variability in medical professional clinical beliefs;
- variable patterns of medical professional work organisation and socialisation;
- the shaping of health risks associated with healthcare delivery and policy;
- healthcare innovation, 'evidentiality' and regulatory control;
- reflexive understanding of the activity of healthcare science as a policy-related enterprise manifest in Health Technology Assessment and Health Services Research.

1.3.3 Methodologies

Overall, the case studies combine substantive empirically-based analysis of selected aspects of healthcare technology, service delivery and policy, with interpretive analysis of the methodology and epistemology of health services research. The three case studies employ a variety of research methods appropriate to each. Each of the three case studies has its own detailed objectives which are outlined in the respective sections below. To take up the theme of interdisciplinarity again, it can be noted that in the text of the published papers the sociological dimensions of the subject-matter and the analysis are not conspicuous or explicit. This testifies to the fact that much of

this evidence-producing work was framed in the context of the generic multidisciplinary field of HSR and HTA, rather than from the standpoint of research agendas shaped by social or sociological interests and organisations. It is one of the tasks of the commentary on the case study chapters presented in this essay to remedy this by highlighting a sociological framing and interpretation of their subject-matter. In this respect I suggest that what is distinctively sociological is not whether research designs and methods of one type or another are used, but rather the institutional embedding of disciplinary perspectives and the concepts brought to bear, especially, in the framing of research questions and the interpreting of research results in terms of particular 'literatures'. Thus the argument I develop later in this essay reflexively builds a socio-political analysis *of* HSR/HTA as well as pointing to the social/sociological elements that were framed originally *within* HSR/HTA approaches.

2. Variability in orthopaedic technology: total hip replacement

2.1 Background

Total hip replacement became a major subject of critical attention in health policy communities during the 1990s, a development which was in many ways unlikely. By the 1990s this surgical procedure was hailed generally as one of the success stories of twentieth century technological medicine (Lefanu, 1999). By the mid-1990s about 40,000 artificial hips were being implanted annually in the United Kingdom. The general perception, therefore, was that it was a procedure that had reached maturity, with a very high success rate in terms of improving mobility and reducing pain for the many patients, suffering mainly from osteoarthritis or rheumatoid arthritis, who underwent the procedure. Yet at this time the national NHS R&D Health Technology Assessment programme rated it as one of the top ten priorities requiring the

production of robust research-based evidence. Why should total hip replacement have become such a cause for concern?

An (over-)simple answer to this question can be summarised as being that the procedure had become a 'victim of its own success'. The papers (Papers 2-4) which I have submitted here illustrate some aspects of this analysis, and I will elaborate upon it here. I would like to suggest here that a particular constellation of circumstances led to the "problematisation" (see e.g. Blume, 1992: 71) of artificial hip replacement as a controversial techno-practice within contemporary healthcare policy communities.

As noted above, the observation of patterns of variation in healthcare delivery has been one of the springboards for the upsurge in Health Services Research generally (Wennberg & Gittelsohn, 1982). Patterns of variation of use of different hip technologies became apparent both in the United Kingdom and between different national health care systems during the 1990s. A survey published in 1993 (Newman, 1993) showed that 70% of orthopaedic centres in the UK used both the more conventional 'cemented' and less conventional 'uncemented' modes of fixation, and that there was also significant geographical variation. As noted in Paper 4, uncemented models were more frequently used in the south of England. Thirty-percent of centres stocked at least two types of the (generally less frequently used) uncemented models. Thus, before the term 'postcode lottery' became current in public debate, it appeared that geographical factors would play at least some part in determining what type of prosthesis patients in a particular population might receive, and that this was to some extent independent of standards of clinically-defined need. Similarly, in spite of the internationally shared nature of many hospital surgical procedures, there were also variations in patterns of use of different types of hip prostheses between countries. One of the most conspicuous examples of this was the fact that in Finland over 50% of hips implanted were non-cemented, while in neighbouring Sweden and Norway the percentages were 4% and 15% respectively, strongly suggesting the influence of non-clinical forces shaping the orthopaedic marketplace.

At the same time, a certain amount of concern about the surgical procedure of hip arthroplasty inflamed the mass media imagination. It appeared that some junior surgeons had implanted some of the pieces of the implant 'the wrong way round'. This was reported in newspapers at the time (e.g. The Guardian, 'Shooting at the Hip', 23.2.93- 'Shoddy material and inept surgeons mean that up to 30% of hip replacements have to be redone'), and a television programme focused upon 'The High Price Of Hips' (BBC2, 1993), drawing attention to the relatively high cost of the procedure as well as surgical risks. Concern was also expressed by leading spokesmen (yes, men) within the orthopaedic profession itself, suggesting that there were potentially risky trends in hip implantation developing in some parts of the health service and the profession (Murray et al, 1995; Bulstrode et al, 1993). These trends were described derogatorily as trends toward 'designer hips', and seen as marked by commercially-driven innovations. As noted at the end of Paper 2 on the history of artificial hips, it is ironic that a technology that is invisible to the user should be shaped in part by social, or at least socio-medical, fashions. This raised the question of the effect of commercial orthopaedic engineering companies upon surgical practice, and in particular, in this author's view, pointed to a need to understand better the development of the technology itself and the routes by which different technologies reached clinical practice in the National Health Service. This indicated a need to draw upon perspectives from sociology and science & technology studies.

Costs of the apparently proliferating new models in particular were high, and occasionally devices appeared to fail. In 1997 the 3M 'Capital hip system' appeared to fail dramatically, and was recalled amidst public consternation, though not without the manufacturer claiming that in some centres the surgeons had used inappropriate procedures and that they had withdrawn the product anyway, prior to the controversy, for commercial reasons. My favourite quotation from the media focus upon this event came from a member of Parliament who was reported to have said that compared to orthopaedic device design and manufacturing 'quality control is

better on a lawnmower'! (The Capital incident is described in some more detail in Kent & Faulkner, 2002).

For total hip prostheses, therefore, the 'problem' as appraised by the health services policy community, was that there was a proliferation of new designs, and a parallel variation in patterns of clinical choice of models, for which the evidence on which to base decisions about clinical effectiveness was weak. The total hip replacement procedure itself, aside from the technology, was very expensive and with an ageing population increasingly large numbers of potential implantees appeared to be coming forward. Thus it appeared to the healthcare policy community that here was a technology whose design, adoption and diffusion were, essentially, out of control.

2.2 Objectives

Given this background, in Papers 2, 3 and 4 presented here the objectives are to:
- outline the history of innovation in the technology of hip prostheses from earliest experimentation to current developments (Paper 2);
- describe the range of hip prostheses available in the marketplace of the advanced (primarily Western) healthcare systems; and to submit the published evidence of the variation in their clinical effectiveness to critical review, using structured methods to appraise study quality and to summarise clinical outcomes; to make healthcare policy and research recommendations (Paper 3);
- to analyse the contribution made by clinical effectiveness studies to the National Health Service policy for total hip prostheses, to analyse the policy environment around artificial hips, and to propose further research which would shed light on the sociopolitical forces at play in that policy environment (Paper 4).

2.3 Methodology

The work on artificial hips comprises: Paper 2, providing an account of the history of the development of artificial hip technology, using mainly secondary written sources (a fuller account would require extensive archive research which was beyond the

scope of the study); this provides the historical context for Paper 3 which is a study taking a 'health technology assessment' (HTA) perspective to examine knowledge of the performance of the wide variety of hips available in the NHS and elsewhere, using a 'systematic review' methodology to examine existing studies. The systematic review method maps medical knowledge on the subject and subjects its methodology to critical appraisal; Paper 3 thus represents a 'classic' HSR/HTA research approach and methodology, but Paper 4 takes a more critical stance and argues that the HSR approach has certain limitations as a tool for analysing the constituent forces involved in policymaking about new health technologies and generating recommendations for health service policy. This paper, therefore, presents a sociological analysis of the policy context of the regulation of innovation in artificial hip technology in the UK and in the context of European Union medical device directives.

2.4 Results

The historical account and the systematic review provide evidence of the proliferation of innovations in the technology, and of the relative weakness of the evidence for clinical effectiveness as judged by strict HSR methodological criteria. Evidence for some models is better than for others, but there is a fundamental difficulty in attempts to use clinical evidence to evaluate *new* designs, which the HTA programme and NHS policy has focused on. The domination of policy discourse on artificial hips by the HTA agenda of 'clinical effectiveness' is interpreted, in the third paper, as a limitation of current policy development, and a sociological research framework is developed to address this bias, drawing especially upon theory in the sociology of science and technology, 'risk' and regulatory science. A notable feature of this analysis is that the knowledge, produced mainly by orthopaedic surgeons themselves, about the performance of artificial hips comes mainly from specialist surgical centres, rather than general hospitals, an observation that may suggest the existence of positive publication bias in the 'evidence base'. Thus the generality of results from published evaluations of the technology may

indicate a higher level of overall efficacy than would be found in non-specialist orthopaedic centres. This analysis of the distinctiveness of collective practice patterns is taken up again in the final discussion section in this essay.

2.5 Conclusions and contribution

These papers taken together as a case study focus to a large extent upon the *technology* of artificial hips/hip arthroplasty, and variations in its use and performance. This is important to emphasise because there are advantages and disadvantages entailed in this focus. Given an interest in understanding the processes whereby new technologies enter the healthcare system resulting in variations in healthcare delivery, I would like to suggest that a major advantage of this work is to highlight the fact that artificial hips are part of a large and fairly globalised orthopaedic products industry. This dimension, and any analysis of it, has been conspicuously absent in work within the conventional 'HSR/HTA' mould, as illustrated in Paper 3 here (though this work does allude in passing to some of the alliances to be observed between orthopaedic surgeons and the orthopaedic products industry). It is thus part of the contribution, especially of Paper 4, to focus attention on the healthcare products industries as a relatively 'invisible' actor shaping the patterns of variability in clinical healthcare delivery. However, bringing some topographical features into focus through one lens inevitably diminishes the focus on others. Thus the papers presented here have not concerned themselves with another key feature of the variations in hip replacement services, namely the socialisation and professional/working environments of surgeons themselves. Indeed, in reviewing the body of research represented by mainly clinical studies of performance of hip prostheses (Paper 3), mostly conducted by orthopaedic surgeons themselves, it was apparent that very few studies existed of the working practices, surgical procedures and prosthetic *choices* made within the sub-specialty. This would perhaps be methodologically difficult, but would certainly shed further light upon the observable patterns of hip implantation, and the relationship of these to the orthopaedic medical products industry. It is thus interesting to note that the systematic review method in HSR/HTA can also be regarded as a partial, though unintentional, sociology of

healthcare knowledge, in which the interests and knowledge-practices of stakeholders can be observed.

Thus the three papers presented have made contributions to UK health service policy regarding hip prostheses, to the field of knowledge of hip prosthesis performance and its sociopolitics, and to the conceptual broadening of health technology assessment as a healthcare science.

Developments in 'knowledge-based' health service policy

Total hip prostheses became a site for increasing attempts at regulatory activity in the UK during the 1990s. The historical account of the development of artificial hips (Paper 2) shows that continual innovation has been characteristic of this aspect of healthcare technology since its inception. Some commentators within the clinical and orthopaedic engineering professions themselves described a "trial and error culture" within contemporary orthopaedics (Huiskes, 1993).

During the 1990s it became clear that national HTA was being deployed as a sort of 'regulatory science' (e.g. Irwin et al, 1997; Lehoux and Blume, 2000), in other words the scientific production of evidence was becoming undertaken in an institutional context which involved negotiating control over the boundary from the scientific and R&D worlds of innovation into that of health service application. HTA organisations can be conceived of as intermediaries at the interface between, crudely, science and policy. Also at this interface, though acting most strongly on processes of control over the initial approval of the safety of new health technologies are the formal regulatory agencies - represented primarily at the time in question by the Medical Devices Agency, an agency of the Department of Health in the UK. Regulation of medical devices, of which hip implants are an example, became more controlled and Europeanised during the 1990s (see Paper 4; Altenstetter, 1996; and Kent & Faulkner, 2002 for further detail on medical device regulation in Europe). In contrast to medical device regulation, HTA has been concerned primarily with

control over the introduction and diffusion in the healthcare system of prostheses which have already been approved in terms of safety and biocompatibility.

Since the publication of Paper 3 in 1998 (and a companion report commissioned by the national HTA programme - Fitzpatrick et al, 1998) there have been a number of further developments in regulatory policy in relation to artificial hips. The National Institute of Clinical Excellence (NICE), introduced in 1999, is a major example in the UK of an institution combining processes of appraisal of scientific evidence with the formation of public policy, the attempted embodiment of the ideal of 'evidence-based policy'. Paper 3 contributed directly and explicitly to the knowledge on which was based NICE's guidance to the health service, issued first in March 2000. The primary guidance is that "the best prostheses (using long term viability as the determinant) demonstrate a revision rate (the rate at which they need to be replaced) of 10% or less at 10 years. This should be regarded as the current 'benchmark' in the selection of prostheses for primary Total Hip Replacement" (NICE, 2000). This guidance was in line with (though not explicitly cross-referenced to) the principle introduced in 1998 that revision of primary hip replacement due to aseptic loosening within 10 years of implantation should be reportable to the Medical Devices Agency (MDA, 1998). NICE's guidance was challenged on a variety of grounds by two orthopaedic engineering companies, and following the quasi-legal consideration of evidence presented, no change to the guidance was made. A further notable development is the closer working reported now between NICE and the NHS Purchasing and Supply Agency (PSA) which NICE depicts as enhancing the implementation of its guidance (NICE, 2000), and thus strengthening overall regulatory control.

The apparent failure of the Capital Hip was noted above in the introduction to this case study section. The Royal College of Surgeons (RCS) of England, following a massively detailed investigation, produced a report of the alleged failure of this model of hip (properly the '3M Capital Cemented Hip System') in 2001. It is worth noting that initial concern amongst clinicians had appeared around 1997-1998, and that due to this timing and the reliance upon published data, neither Paper 3 nor the

companion report by Fitzpatrick et al (1998) had included any published clinical evidence about this hip system in their reviews of evidence. Remarkably, though perhaps less remarkably given the key position accorded to scientific uncertainty in the sociology of scientific knowledge, the RCS report was unable to identify unequivocally the technical cause of the failure - of what was in fact one of four versions of the technology going under the name of the same hip system. The Capital hip was a variation of the classic 'Charnley' model of hip and had not required *de novo* clinical trial for marketing approval. The ensuing controversy led the British Orthopaedic Association, in new professional guidance (1999), to make a point of referring to the possible risks to safety associated with even small variations to existing models of implant. The Royal College of Surgeons' investigation, led by a clinical epidemiologist/statistician, can be understood as a good example of the enrolment of evidentiality into social processes of accountability and legitimation.

The British government has also concerned itself directly with artificial hips. The National Audit Office, the investigatory financial 'watchdog' of the British government, produced a wide-ranging report (not confined to, though highlighting, the issue of the performance of the technology) on many aspects of hip replacement in the NHS (2000). This report was followed by an update (2003). These reports note some progress but continuing weaknesses including in regard to systems of surveillance of implanted hips, NHS Trust policies for introduction of new prostheses, variations in performance across the NHS, concerns about the use of incentives by manufacturers, surgical training and expertise, conflict between waiting time targets and clinical urgency, under-reporting of adverse incidents, and compliance with NICE guidance in NHS Trusts.

Finally, it should be noted that there had been a variety of calls from various quarters, increasingly strong since the mid-1990s, for the introduction in the UK of a 'registry' system for the recording of all hip (and knee) replacements along the lines of that noted above for some of the Scandinavian countries (Norway, Sweden and Finland). The Capital controversy certainly added fuel to the debate about this. The Department of Health issued a formal consultation document (DoH, 2000) and a

system, after a great deal of debate and negotiation between the stakeholders involved, was announced in July 2001 and launched in April 2003 (reported in National Audit Office, 2003). It is a voluntary system, funded by a levy on NHS Trusts.

Thus innovation in artificial hip technology led to 'problems'. In such a widely used and relatively high-profile surgical procedure (waiting lists for orthopaedic surgery are always high on the health policy agenda in the UK) which employs high-tech material technology public confidence is of major concern in healthcare policy networks. Interestingly, the problems were due to technological success which led to some technological failure. The responses, in the shape of HTA research, professional body investigation, assessment by government bodies and regulatory guidance from the quasi-regulatory authority of the NHS have been extensive and searching. A re-legitimation of total hip replacement, with a major focus upon essentially modernist marshalling of evidence, has been sought. This response may be understood as in keeping with the putative cycle of 'reflexive modernisation' in which modernisation is interpreted as turning back on modernity itself (Beck et al 1994).

Broadening of HTA

It should be clear from the previous section that the papers on artificial hips presented in this submission have contributed to the development of a multi-faceted analysis, drawing upon different disciplinary perspectives, of the problem of total hip replacement in the NHS. The expansion of HTA/HSR approaches to embrace an approach that permits a sociopolitical analysis of innovation and regulation of new hip technologies has been suggested particularly in Paper 5, and also illustrated above. This type of analysis could be extended to the more recent developments in policy and evidentiality in relation to total hip replacement noted here. Paper 1 also has contributed to the development of approaches to the analysis of HTA as a part of the phenomenon of contemporary evidence-based healthcare policymaking that draw upon academic threads in the sociology of scientific knowledge, and science and

technology studies. Explicit reference to this work can subsequently be found, for example, in the important work of Lehoux and Blume (2000), May et al (e.g. 2001 and 2003) and Webster (2004), where further extension and theorisation of the field of HTA is to be seen.

I end this section by summarising the key threads of the analysis of this case study, threads which are picked up again in the following two case study chapters. There is clear evidence of *variability* in the performance of different hip technologies, implying inequitable population access to orthopaedic surgical care; this uncertain and variable supply of technologies and surgical expertise presents the service user, the potential implantee, with a healthcare environment in which personal health *risks* associated not with illness but with the medical system itself must be addressed; patterns of differential availability of different models of hip prostheses suggest that relatively stable, shared institutional practices exist within healthcare provider organisations and within their associated (orthopaedic) professional networks; the public, mass media and government attention to controversial aspects of the technology and service delivery *shapes the context* in which health service users may understand and meaningfully construe their health and healthcare; *regulatory policy* as a form of governance is seen to emerge from a complex of actors in a policy network including clinicians, commercial organisations, formal and quasi-regulatory bodies, citizens and other actors claiming to represent them. These themes from the study of the socio-technology of artificial hips may be seen to run through the case studies on prostate cancer detection and outpatient service delivery presented below.

3. Variability in cancer care: locating early prostate cancer

3.1 Background

Detection of prostate cancer has become an important topic, and in some senses a resource, in the development of healthcare science in the UK for several different reasons.

Rates of mortality from prostate cancer, and the incidence of its detection, had been rising dramatically during the 1980s and early 1990s internationally. It was also clear internationally that approaches to treating the disease varied dramatically. In particular in the USA a much more interventionist approach was evident compared to the policy most prevalent in the UK namely "watchful waiting", in other words active monitoring to detect signs of progression of the disease.

However, it was clear also that part of the upward trend in incidence could be attributed to increased rates of detection of the disease in its early stages, especially through a relatively new diagnostic blood test, the prostate-specific antigen (PSA) test. Various versions of the PSA test were used from the mid-1980s for *monitoring* prostate cancer progression, but in 1994 the Food & Drug Administration (FDA) in the USA, became the first regulatory agency in the world to approve a commercially-produced version of the test (the "Hybritech Tandem PSA Assay") for *detection* of the early stage disease - in conjunction with DRE (digital rectal examination).

In the early to mid-1990s in the UK, prostate cancer was the subject of mass media and public attention, and controversy, as the issue of screening for the disease became prominent. Media headlines such as "Rising fear of prostate cancer 'could cost the NHS £400m'" (The Guardian, 31.10.95) were frequently to be seen. It was common to hear about public figures who had contracted the disease, such as General Norman Schwarzkopf and the musician Frank Zappa. Such personalised references undoubtedly contributed to raising the public profile of the disease and to increasing the public perception of risks associated with it.

The fact that prostate cancer is a disease affecting men is also significant in its rise as a topic of healthcare science and policy. The absence of national screening programmes for men, compared to breast cancer and cervical screening for women in the UK was noted in public debate at the time. The perception of this gender disadvantage (whether it *is* a disadvantage to health in this instance is open to doubt) was reflected also by some academic work (Cameron & Bernardes, 1998).

Contemporary societies have been diagnosed as undergoing an institutionalisation of 'risk' (Giddens, 1991). Medical discourse and the incursion of the healthcare system into everyday life may be one instance of this form of social infiltration. Diagnostic and screening techno-practices can be considered to play a part in the shaping and experience of health risks. In sociologically-informed theoretical work on 'the new public health', Lupton has suggested that diagnostic testing is seen as offering people control in the face of the disorder represented by possible disease (1995:78). The knowledge provided by diagnostic testing in principle enables protective action to be planned. Thus the apprehension of risk and individual testing of asymptomatic people to determine if a disease is present are part of a process in which both are mutually supportive, constituting a process of personal knowledge generation. In a context of socially amplified risk and uncertainty, it is understandable that screening and early detection technologies are becoming not only an increasingly important part of modern healthcare services, but also a growing source of ethical and social concerns from public health perspectives (Stewart-Brown, 1997). Needless to say the advance of genetic testing further develops a trend in which social - personal, familial, and societal - apprehension of potential disease risks is exacerbated. Pressing the sociological analyses further, Howson (1998) has analysed cervical cancer screening as a healthcare site where the practices associated with risk and with population surveillance are brought together. Thus organised forms of screening for disease may increase the stock of knowledge of health and health behaviour held by health services and authorities, at the same time enhancing concerns about health risks in certain population groups. Howson has described the understanding of the 'relationships between the subjective articulation of risk and *the processes shaping*

those articulations' (my emphasis - 1998:210) as an essential task of the social scientific study of public health and disease prevention. Thus the emergence and uptake of PSA technology in health services, and the production of evidence about prostate cancer detection by the healthcare sciences, should be seen as part of a societal process in which the contours are being constructed within which men's personal appraisal of risk of prostate cancer is delineated.

Therefore, early detection and treatment of prostate cancer may be seen in retrospect to have many of the ingredients to attract the attention of the new healthcare sciences in pursuit of evidence-based healthcare. The issue of screening for prostate cancer was identified as one of the highest priorities in the first research agenda produced in the early deliberations of the UK national Health Technology Assessment programme. The importance of population screening issues in the NHS R&D agenda was evident in the fact that one of the Standing Group on Health Technology's (SGHT) advisory panels was assembled specifically to examine population screening (note: I am drawing here upon my own participatory experience during 1993-4 as 'Scientific Secretary' to the Acute Sector Panel of the HTA of the NHS SGHT). From an epidemiological point of view the prevalence and incidence of the disease were large and apparently growing. However, the disease was marked by great scientific and clinical uncertainties - the natural history of the disease was not clear or predictable, the best treatment was not known, and there was doubt as to whether the tests available for early detection of the disease could be used as population screening tools. This made for a considerable burden both on clinicians and on men who might have the disease or be worried about it. Healthcare science in the form of Health Technology Assessment was enrolled into the process of policy development, with the avowed aim of providing a solid basis for policy decisions. In other words, policy looked to science for legitimation.

In retrospect the HTA programme was described by the government as having 'a carefully structured portfolio of research designed to provide the evidence base for policy and practice'. Two systematic reviews (Chamberlain et al, 1997; Selley et al, 1997 - the current author was a co-author of the latter report) concluded that there

was inadequate evidence to support the introduction of mass screening using the PSA test and that the comparative effectiveness of alternative treatments for early prostate cancer was unknown. These conclusions, it was stated, were subsequently supported by HTA reports from several other countries and awareness of the conclusions 'helped to contain the uncontrolled dissemination of PSA testing' (http://www.dh.gov.uk, 09/2003).

Thus the crucial issue from the perspective of NHS policy was the question of the possible introduction of a formal population screening programme, and a threat of possible diffusion of testing in the face of a lack of evidence of public health benefit. As seen above, apart from the risk of the disease, there were also financial implications of a possible screening programme. A National Screening Committee for the UK first met in 1996 with the remit to advise the Department of Health by examining all existing and possible national screening programmes which might be of value. In keeping with the philosophy of the knowledge-based health service, its first report contained a strong statement about appropriate methodology by which strong evidence on which to base its conclusions and policy recommendations might be arrived at (NSC, 1998). The methodology of systematic reviews, the method by which the two HTA reports on prostate cancer had been produced, was highlighted. The degree to which a systematic approach was sought to forging a 'link between research evidence and the formulation of national screening policy', including a major role for the HTA programme, was described at the time in a strong statement of the aspirational role of healthcare science in policymaking by members of the NSC secretariat (Sherriff et al, 1998: 58).

Given that the evidence for the comparative effectiveness of treatments for the condition was deemed to be equivocal, and a policy for national screening programmes was not to be espoused, it was important, therefore, to know the extent of variation in then current clinical practice in using the diagnostic (PSA) test because this would have direct implications for the pattern and volume of different types of further diagnostic activity and of treatment undertaken. It would also be crucial to building the social and clinical contexts in which asymptomatic men would

appraise their wish to be tested for prostate cancer - in more sociological terms, to shape the context for men's subjective articulation of risk - and their subsequent experiences of testing and treatment processes. It was thus the aim of the study reported in Paper 5 to produce evidence about the variability in current clinical practice, to ascertain possible reasons underlying it, and to suggest implications for men who might be considered by medical interpretation or might consider themselves to be at risk from the disease.

Although not presented in this thesis because I was not the lead author of the publications, I contributed to several others arising from the same research. The aims of these related studies were primarily to assess the performance of diagnostic and screening technologies for early-stage prostate cancer (Selley et al, 1997) and to describe the *treatment* strategies envisaged by urologists in the face of diagnostic uncertainties and lack of convincing evidence about the comparative effectiveness of alternative treatments (Donovan et al, 1999).

3.2 Objectives

Given the background outlined above, the objectives of the study described in Paper 5 are to:

- describe possible variations in current (1995-6) clinical practice of consultant urologists in use of the prostate-specific antigen (PSA) test in detecting early-stage prostate cancer;
- to analyse practice variations in terms of potentially predictive characteristics of urologists' socialisation and hospital settings;
- to assess the implications of practice variations for men undergoing the PSA test, and for further research in the subject.

3.3 Methodology

Paper 5 thus reports a national postal UK questionnaire survey of urologists' views about the use of the PSA test. It was able to survey a large proportion of the urologists in the UK. It used a custom-designed questionnaire. The study might be criticised for relying upon a case vignette method rather than empirical observations. However, the internal coherence of the results suggested that it produced strong evidence that there was significant, systematic variation amongst urologists in their interpretation of PSA cut-off points, in other words the threshold above which further investigative action would be advised. The study did not include general oncologists, unless they were also urologists as defined by membership of the British Association of Urological Surgeons (BAUS). Plausible factors predicting this variation were identified by multi-variable analysis, though the interpretation of the implications for patients, on the basis of this study alone, remains speculative.

3.4 Results

The paper analyses UK urologists' views about the use of the blood test - the PSA test - to detect the disease in men without symptoms. Analysis shows that it is highly likely that aspects of the organisation of urological services, which vary between geographical areas and health care centres, and the socialisation of urologists, will lead to an unequal social patterning of detection of the disease amongst men in the UK. Centres where there was a urologist specialising in prostate cancer were more likely to use lower cut-off points in interpreting PSA levels, thus making it more likely that further action would follow in these centres. The extent of this variation has important implications for the use of the test by the medical professions, in primary and secondary care, and for the conduct of research in which the PSA test is used as a marker for the disease in asymptomatic men.

3.5 Conclusions and contribution

Subsequent to the early HTA work on prostate cancer, of which the paper discussed here formed a part, national healthcare science in the form of the HTA Programme in the UK has devoted further large amounts of resources (over £13 million) to research with the aim of producing evidence on which to base NHS policy.

The policy on PSA testing in Britain was conservative, and research needed to tackle the problem of low recruitment rates to clinical trials if the HTA approach was to produce the knowledge about treatment comparisons that was seen as necessary. This dilemma in research policy led to innovations in HTA methodology: 'New methodological approaches are required urgently to investigate this issue and to bridge the gap between clinical practice and the need to acquire evidence. Such approaches need to retain the essential principle of randomisation while incorporating more fully patients' perspectives and preferences' (Donovan et al, 1999). And in the meantime '....until more evidence accumulates, patients and urologists should use the information available from recent systematic reviews to reach shared decisions about treating localised prostate cancer and provide information that highlights uncertainties about the potential effects of such treatments on survival and quality of life' (Donovan et al, 1999).

Thus an innovative feasibility study was mounted to try to assess how to improve recruitment to a future full scale randomised trial of alternative treatments (given the earlier failure of a Medical Research Council trial). This used qualitative methods to assess men's reasons for participation or non-participation. The design of this study necessitated offering the PSA test to several thousand men in order to identify a small number with the early stage disease who might be randomised to one or other treatment. The design of the study focused particularly on the initial offer of the PSA test. There was an ethical concern that men offered the test should be fully informed about the test and possible treatment side-effects, the favoured approach being that this should be in the context of professional counselling, rather than the conventional information sheet and consent form. The feasibility phase thus examined the

performance of PSA testing in the screening context, men's attitudes to screening, preferences for treatment of early-stage prostate cancer, and their willingness to accept treatment randomisation (the above comments draw upon the author's involvement in the planning of the study in question).

This development, toward a 'counselling' model of men's decision-making about consent to participate in a randomised control trial, is clearly consistent with the direction of both the NHS Centre for Reviews and Dissemination (NHS CRD, 1997) and British Association of Urological Surgeons' recommendations to constrain use of the PSA test, though perhaps elevating the degree of information provision and magnifying the health service incursion into asymptomatic men's appraisal of personal risk and healthcare decisions. This can be contrasted with the prevailing approach in the USA, expressed here by David Kessler, Commissioner of the regulatory body for new medical devices the Food and Drug Administration (FDA) at the time the PSA test was approved for use in early detection. At this time screening programmes were already widely diffused, bolstered by public health campaigns such as 'Prostate Cancer Awareness' weeks: 'This test - used with other procedures - can help detect those men at risk for prostate cancer early on when more treatment options are available..But for the test to help, men must be aware of the importance of early checkups and get them on a regular basis' (FDA, 1994).

The feasibility study mentioned above led to the support of a primary research project in 2001 in the UK, a randomised control trial to compare treatment strategies for screen-detected early prostate cancer (radical prostatectomy, radical radiotherapy or 'watchful waiting'/active monitoring). This is known as the ProtecT study. The trial involves prostate checks and PSA blood tests for 230,000 men aged 50-69, with the expectation that 2,000 would be detected as having signs of early prostate cancer. It is still under way at the time of writing. It is notable in the context of the discussion in this essay about interdisciplinarity, that this study is stated to combine 'the qualitative traditions of sociology and anthropology, epidemiological and statistical disciplines informing randomised trial design, and academic urology and nursing', and the study 'contravened conventional approaches by being driven not by

the randomised trial design but by the qualitative research' (Donovan et al, 2002). Regardless of the interpretation that we might give to possible power-shifts in the inter-disciplinary and inter-methodological relationships of stakeholders in healthcare science, one outcome of the methodological innovations claimed here has been to increase recruitment rates to the large-scale trial of alternate treatments for early stage prostate cancer from an estimated 30-40%, to 70% (Donovan et al, 2002).

The ProtecT study forms part of a wider NHS Prostate Cancer Programme as part of the broad policy initiative known as the NHS Plan launched by the Department of Health in September 2000 (NHS Executive, 2000). This takes the form especially of a risk management programme aimed at providing asymptomatic men with information about testing in order for us to make 'informed choices' about proceeding with the test; and upon speeding up access to diagnosis and treatment. The Programme also included provision to increase the number of urologists in the NHS by nearly 100 by 2005.

In the context of a policy NOT to introduce a public screening programme, it is clear that at the beginning of the 21st century there is considerable ambiguity in the existing policies and practices, and confusion amongst both the medical profession and men concerned about the disease. Men with urinary problems in the UK are likely to be PSA-tested either by general practitioners or by urologists. Self-testing kits are available commercially. Unsurprisingly perhaps, men with suspected or confirmed prostate cancer are generally in favour of testing and a policy of screening for the early-stage disease (Chapple et al, 2002). Thus as Donovan et al note, screening may be creeping in through the back door (Donovan et al, 2001) – exactly what policymakers have been seeking to avoid. One implication of this is that more men than is justified by the existing science will be exposed to further investigation and to radical treatment, in other words surgery or radiotherapy with their associated side-effects, which include relatively high proportions of incontinence and impotence. As the contribution made by the study presented in Paper 5 has shown, it is highly likely that this trend to increased testing is associated with urologists at a relatively early stage in their careers and with clinical centres where there is sub-

specialisation in prostate cancer. This suggests that increased clinical specialisation in the disease by younger urologists at an early stage in their careers is likely to promote pressure for men firstly to undergo the test in the absence of symptoms, and secondly to undergo radical treatment in conditions of uncertainty about disease progression. Thus this case study contributes to describing professional risk-shaping dynamics which are deeply embedded in contemporary social processes of medical professional education, controversy within the frame of evidentiality, and the collective organisation of care in healthcare settings.

The example discussed here of early detection of prostate cancer shares common threads with the previous case study of artificial hips. As with hip prostheses, the study presents evidence of variability in patterns of outcomes of care delivery with clear implications for inequalities in the delivery of care to different populations and patient groups, and thus concomitant inequalities in the social distribution of healthcare risks. Men exposed to the British National Health Service have unequal chances of detailed investigative testing for the presence of early prostate cancer. As with orthopaedic surgeons, specialist urologists display collective patterns of clinical knowledge and beliefs which are expressed in their clinical decision-making and are associated with features of the professional *organisation* of care and local institutional patterns. These institutionalised practices ('institutional signatures' – see case study of outpatient services below) contribute to shaping the context in which risk of prostate cancer may be appraised by men exposed to the healthcare system. Regulatory policy - resistance to mass screening programmes - in this field shows signs of underlying ambivalence related to tensions between a search for robust evidence about the comparative effectiveness of different treatment approaches for the early-stage disease, and the attitudes of men and the medical profession toward the relative benefits and harms of early detection.

4. Variability in outpatient care: collective socio-medical practices

4.1 Background

Hospital outpatient clinics in the United Kingdom are the point of entry to the healthcare system at which the majority of people with a health problem encounter *specialist* clinical expertise. Diagnosis, testing and treatment may all occur in the outpatient setting. In the UK system, typically the person has been referred to the specialist by a general practitioner who in principle acts as gatekeeper for these services. It is in the outpatient setting that the full range of medical and surgical specialisation and sub-specialisation, and a range of diagnostic technologies is on display. It is well known that the styles of behaviour produced by professionals in these settings can be characterised by their routinised nature, and that the interactions in these settings contend with a wide variety of tensions arising partly from the uneasy institutionalisation of medical authority within 'bureaucratic' organisational settings. This has been the topic of one of medical sociology's most well-known works, which detailed the 'ceremonial order' of interactions in outpatient clinics in the UK and the USA (Strong, 1979). Thus it is important to recognise that the techno-practices of outpatient care are expressed through a very complex mixture of expertise, knowledge, organisation, management, routine, interacting participants, interests, material technologies and so forth. It has also become one of the sites in healthcare organisation where broad movements such as contemporary policies in the areas of evidentiality, patient participation and resource use may become activated (e.g. Sanders & Harrison, 2004). But it is also important to recognise that outpatient clinics have functions other than to provide services to patients. They also function as sites for training of staff and for research and development. Outpatient services attracted increasing attention amongst healthcare policy communities in the early 1990s, and this provided the impetus for a broad HSR research agenda which aimed to improve knowledge about the characteristics of the outpatient care system and its underlying dynamics. The case studies discussed here formed part of this agenda.

Outpatient care has always been the poor relation of inpatient care within the British National Health Service. The number of single outpatient attendances in the NHS is massive, on average nearly one per head of population per annum. A National Audit Office study published in 1991 (NAO, 1991) found that outpatient visits varied between English districts in 1988-9 by 352 to 1726 visits per 1,000 resident population. Towards the end of the 1980s and in the early 1990s increasing policy concern began to manifest itself at the general effectiveness of outpatient services. Unlike the longstanding and politically debated 'waiting list problem', which refers to inpatient waiting lists, waiting times for initial outpatient appointments became a political issue only in the early 1990s with the arrival of the Patient's Charter symbolising the growing movement toward standard-setting and consumerisation in healthcare. During the early 1990s the NHS introduced a set of Outpatient Demonstration Sites in an attempt to highlight perceived shortcomings and to introduce innovative improvements (NHS Executive, 1995).

An analysis of the rationality of the operation of outpatient services produced the trenchant conclusion that '..the system seems to be geared to bringing back for consultation patients who have neither the clinical need nor the inclination to attend hospital, whilst blocking the way for new patients to undergo speedy assessment' (Frankel and Robbins, 1993: 93). The particular focus of concern was thus the phenomenon of patients' *re-attendance* at outpatient departments. In 1989-90 only 8.5 million of over 36 million attendances in England were 'new', first-time attendances by patients referred from outside the hospital system (Department of Health, 1990). There was thus a strong perception that re-attendances were a burden on the services available and this was acknowledged by some well-known views within the medical professions that supported notions of a 'merry-go-round' of patients being recycled from one outpatient visit to another because of some form of institutionalised inertia. A structured review of research in outpatient services noted that a number of studies suggested, though not unequivocally, that high re-attendance rates might be especially associated with the subordinate roles and statuses of junior hospital staff in relation to senior, consultant staff in hospital outpatient practice (Faulkner and Frankel, 1993). Junior staff it was suggested may not have had the

confidence or authority to discharge patients to the community or to primary care. 'TCA' (To Come Again) written in medical notes, became synonymous with this alleged lack of authority or initiative of junior doctors.

During the early 1990s, cost pressures and the introduction of GP fundholding responsibilities provided stimulus for general practitioners to question and for specialists to have to justify long term re-attendance at the hospital outpatient clinic. It was recognised that clinical activity delivered by the major acute clinical specialties was more dispersed than popular belief might suggest. Around 10-20% of outpatient attendances in some regions in high volume specialties were acknowledged to take place outside major hospitals, in community hospitals and other settings. This 'outreach' model of specialist care was supported by the government policy of moving toward a 'primary care-led NHS' (Coulter, 1995). Day case surgery developed largely in a research vacuum as regards evidence for population requirements for services, and there was negligible evidence of it substituting directly for equivalent inpatient procedures. Relatively minor procedures which had been provided previously, but not counted, in outpatient clinics were moving toward being provided, counted, and paid for as new episodes of day case care. Contract incentives for general practitioners to undertake minor surgery further challenged the outpatient department as a site for surgical treatment.

Such developments indicated moves toward a blurring of boundaries, in which the once clear lines between general practitioner and hospital specialist were becoming less easy to discern. This fragmentation of roles, settings, expertise and activity around the primary-secondary care interface suggested that the traditional concept of hospital outpatient services was itself becoming outmoded, and that a more flexible model of the complexities of 'ambulatory care' was required.

The organisational phenomenon of outpatient visits thus became increasingly visible to the healthcare policy community. Three published papers (Papers 6-8) are presented which report on three separate empirical studies examining aspects of the

outpatient problem with a primary focus upon aspects of the re-attendance of 'season-ticket holders'.

4.2 Objectives

In summary, given the background outlined above, the three studies encompass a description of the range of variability, especially with regard to re-attendance at outpatient clinics in one English region, an assessment of the scope for improvement in a single department, and an analysis of the obduracy of the socio-medical settings in which health professionals' outpatient care practices are embedded.

More specifically, the objectives of Papers 6 –8 were:
- to describe variation in re-booking rates between a set of consultant firms and 'provider unit groups' in surgical and medical specialties across the south west of England; to assess possible reasons for variation observed; to assess the value of 'routine' NHS data for this type of study (Paper 6);
- to assess the scope for reducing 'unnecessary' clinic attendances in a 'benchmark' department in the UK (Paper 7);
- to explore the experience and views of patients, general practitioners and hospital clinicians which might underlie the observed variations between long-term and short-term outpatient clinic attenders (Paper 8).

Taken together, therefore, the studies describe the extent of variation in re-booking practice in a wide range of hospitals/specialties/ consultant teams; analyse the extent of 'avoidable' re-bookings in an already-efficient setting; and paper 8 offers social and organisational reasons for the types of practices and clinical routines suggested by the analyses in papers 6 and 7.

4.3 Methodology

Two of the studies are based upon primary data collection, an interview study (8) and a clinic-based survey using a custom-designed questionnaire (7), and one uses NHS

'routine' datasets (6). Given this essay's introductory comments about disciplinarity and cross-disciplinarity, it can be noted that Paper 8 draws upon a sociologically-informed frame of reference in investigating the sociomedical and organisational aspects of the networks of relationships in which participants in outpatient care are embedded. The paper reports work that formed part of a larger, cohort study. Thus this can be regarded as exemplifying two of the major types of contributions that 'qualitative research' can make to 'quantitative' studies, namely the capacity to generate hypotheses and the capacity to offer explanations for their findings (Black,1994), through interpretative analysis. In Paper 7 a quantitative analysis of re-attendance rates is presented, and this is supplemented by analysis of open-ended questions which were, again, able to produce valuable insights about the reasons underlying and constituting the clinical behaviour which was expressed in patterns of clinicians' outpatient decision-making. The study on which Paper 6 is based is primarily descriptive in its approach to quantitative data-gathering, and here the possible explanations for the observed patterns of behaviour are limited because data that might have provided explanatory resources were not sought. Nevertheless, it was possible to discount some possible confounding variables, for example sub-specialisation by clinicians.

4.4 Results

As a body of work, the case study of outpatient services explores the extent of variability in rates of outpatient care delivery in different medical specialties, the potential avoidability of re-attendance, and offers accounts of 'good organisational reasons' (Garfinkel, 1967) for these apparently non-rational practices.

Paper 6 shows that there is substantial variation in 'new to old' attendance ratios at different levels of analysis, that this variation cannot be accounted for by epidemiological factors – disease prevalence - alone, and that although there is evidence of individual consultant variations (a 'consultant signature' – suggested in previous research on service rate variations (Wennberg, 1982)), there is also the suggestion that shared cognitive and behavioural factors are operating at the

institutional level (I term this the 'institutional signature'). An important conclusion of the analysis in this paper was confirmation that surgical specialties can be expected to have lower rates of re-attendance than medical specialties. This in turn highlights the conclusion of Paper 7, which shows that there may be significant scope for reduction in 'unnecessary' clinic attendances even in a surgical centre known to operate relatively efficiently in terms of re-attendance rates. Finally, the results of paper 8 suggest that while there are complex social and cultural aspects of both clinical and patients' experiences of outpatient care, a dissonance in 'power perceptions' between general practitioners and hospital clinicians might offer an opportunity for effecting clinically acceptable change to some of the conventional practices of outpatient re-attendance illustrated in the other two studies.

4.5 Conclusions and contribution

Outpatient services have continued to be an object of concern in the healthcare policy community (e.g. CSAG, 2000). These studies taken together show tensions between evidence of potential opportunities and locales for intervention in health professionals' conventional beliefs, practices and organisations, and the resistances to these opportunities represented by deeply embedded social and collective inter-professional networks and meanings. The 'outpatient problem' is, in a sense, a construct that reflexively supports the conventionality of the healthcare system and the professional work that sustains it. Such tension between intervention and convention is evident also in debate about the changes in organisational structure and responsibility implied by UK government policy of moving toward a 'primary care-led NHS'. In the mid-1990s, the NHS R&D programme launched an initiative in 'the primary-secondary care interface', which examined the feasibility of shifting the balance of care in the direction of the primary care sector. However, it was noted from an increasingly consumer-oriented perspective, that the examination of the feasibility of structural change should not be made without recognising the needs, values, perceptions and attitudes of patients (Coulter,1995). It is in furthering the understanding of this aspect of the primary-secondary care interface that Paper 8 presented here has contributed.

Moves toward the dismantling of traditional, conventional patterns of follow-up of outpatients in hospital clinics following surgery can be found in subsequent research which has examined *inter alia* the follow-up of patients after surgery for breast or colorectal cancer. Kievet (2002), for example, performed an extensive meta-analysis of the impact of traditional follow-up regimes on patients' outcomes (mortality rates) after surgery for colorectal cancer, concluding that continuing routine surveillance should focus primarily upon patient support and was best carried out by a general practitioner or nursing personnel. Likewise, Brown et al (2002) reported a small randomised control trial in the United Kingdom that concluded that patient-initiated follow-up is a potential alternative to standard clinic follow-up for women following treatment for Stage 1 breast cancer, with no adverse effects. This case study can also be regarded as contributing to the research literature investigating the modernising trend in contemporary health care toward shifting the balance of services at the primary-secondary care interface, and, indeed, toward a larger measure of self-care (cf. Department of Health, 2001). Papers 6 and 7 presented here can thus be read as contributing to the understanding of conventional clinical practices whose foundations may not meet criteria of clinical rationality nor even produce the best outcomes for particular groups of patients.

Re-structuring of services at the primary-secondary care interface points the way to another way in which the outpatient problem may be tackled, albeit by default. The outpatient problem as conventionally formulated rests largely upon methods and attitudes of professional working that seem, as they are challenged, increasingly archaic. Research and development priorities provide a good indicator of the new directions of 'outpatient' health care. Both the Primary Care Development Fund and the recent priorities for research in the interface between primary and secondary care in the NHS research and development strategy confirm the demise of the traditional model of outpatient services. The themes highlighted are transfer of follow up care to general practitioners, direct access of general practitioners to some services, reorganisation of diagnostic and investigative services, extension of primary care roles for care of people with chronic conditions, treatment in the primary care setting,

specialist outreach, and shared care. Candidates for shared care are being extended from people with diabetes and asthma to those with rheumatological, ophthalmological and orthopaedic problems. Clinical guidelines are reaching further into the taxonomy of illness, for example, to cancers and other endocrine problems. The growth of such innovative services around the primary-secondary care interface can be understood to be re-defining the conventional notion of outpatient services in such a manner as to define the outpatient problem and its attendant mythology out of existence (Frankel and Faulkner, 1994). However, it appears that the notion of 'ambulatory care' mentioned above, which is more associated with the American healthcare system, has not gained significant ground as a shaping concept in UK policymaking.

Thus the case study of outpatient re-attendance shares a number of features with the previous case studies of artificial hips and prostate cancer detection. The papers discussed show evidence of substantial patterned variability in care delivery, associated with conventional clinical practices in secondary care settings. The institutionally patterned nature of these techno-practices I have termed the 'institutional signature' (Paper 6; the significance of this concept is further discussed below: section 5.1). So patients' risk of being re-booked for re-attendances at outpatient clinics is associated with factors other than rationally-defined clinical need. The demythologisation of conventional clinical practices, many associated with specialisation and sub-specialisation in particular clinical fields, in this area contributes to a contextual re-structuring of the sites in which health risks may be managed by patients and health professionals. This is illustrated, in turn, in moves to shift responsibility for monitoring of chronic or recurring conditions to ourselves as citizen-patients in 'the community' and to primary healthcare teams. Regulatory policy here attempts to balance efficiency of care delivery with timely access to appropriate services by patients. Behind regulatory standards such as outpatient waiting times targets (for first referrals from GPs), for example, can be seen a plethora of tensions between clinically defined needs for hospital outpatient attendance, uncertainties about the healthcare science evidence for optimal follow-up

regimes for many medical conditions, political pressures for a 'primary care-led health service', and collective forces of medical professional power.

5. Discussion/Conclusions

Contemporary society is characterised by some sociologists as being in a phase of 'late modernity' in which the application of technologies and technical knowledge has become of the essence. In this review essay I have drawn out the common threads in three fields of healthcare, with a focus upon the delivery of diagnostic and treatment services. I note that these three case studies have as their major focus material *technology* and patterns of clinical *practices*. As discussed above, I take it that technologies and practices are inextricably linked and are 'social'. They are 'socio-technologies' (Brown & Webster, 2004:27 and *passim*). In recognition of this, while I do not use the broad metaphor of 'technologies' as a conceptual tool in this essay (or elsewhere), I have introduced the term 'techno-practice' to denote these practices that can be seen as underlying and indeed partly engendering the patterned and institutionalised structures of the healthcare system. This approach giving high priority to socio-technologies is akin to that adopted in other theorisations of HSR/HTA, for example Lehoux & Blume (2000) and May et al (2001; 2003) who argue that 'the institutional superstructure of health care rests largely on the application of technologies to problems of understanding the nature and distribution of disease, its diagnosis, treatment and management, and the organisation of service provision'.

To re-visit this essay's introductory comments about disciplinarity and inter-disciplinarity, I point out here that this essay can be read, with an eye to the title of my thesis, as re-formulating the disciplinary pedigree of those case study papers originally framed within the HSR/HTA paradigm. During the 1990s, the new practitioners of healthcare science became strongly enrolled in the problem-oriented agendas of health policymakers. A discourse of effectiveness and cost-effectiveness shaped much of the healthcare science produced, and (as noted in Paper 1) the

epistemology of biomedicine and disease epidemiology was extended outside its earlier disciplinary boundaries to be applied to the science of healthcare evaluation, as 'clinical' epidemiology and HSR/HTA. The disciplinary agendas of this early HSR/HTA are embodied in much of the case study work included in this thesis and published during the 1990s. Reformulating this healthcare knowledge in the way I am attempting now in this essay draws attention to the tensions between disciplinarity and problem-oriented, multidisciplinary HSR, and to the shaping of the endeavours of healthcare science by the institutional authority and rich resources of co-ordinated, powerful healthcare policy communities. Thus the multidisciplinary, or a-disciplinary analysis of outpatient service variations, of factors associated with variability in prostate cancer test interpretation, and of artificial hip technologies, originally framed primarily in terms of the 'effectiveness and efficiency' agenda, I re-conceptualise here through the concepts of techno-practice and socio-technology to point to their significance within a new and different agenda for a sociology of the dynamics of healthcare.[2]

The discussion in this concluding section of the thesis recapitulates the nature of the 'problems' which were taken during the early 1990s to provide the rationale for the research agendas of which the case studies form a part. It discusses the work presented in the three substantive areas of techno-practice relevant to patterns of health care delivery and technology, summarising the original contribution to healthcare knowledge which they have made *in toto* (section 5.1). Following from this, I adopt a more explicitly sociological perspective in order to consider the contemporary phenomenon of evidentiality in healthcare of which, again, the submitted studies form part (section 5.2). Finally, in section 5.3, I argue that healthcare evidentiality as represented by the new healthcare sciences, its proponents and its institutional vehicles, should be considered as one of the social forces that must be embraced by a sociopolitics of the dynamics of healthcare innovation and governance.

[2] It is interesting to note that one of the early and foundational texts of medical sociology, 'Medical Sociology' by David Mechanic (1968) includes epidemiology as a small sub-topic.

5. 1 Patterns of health care techno-practice: beliefs, practices and healthcare risk

In this section I discuss the generality of the contribution made by the case study papers which I associate with the Health Services Research model of the multidisciplinary study of health care. From the perspective of the disciplines of clinical sciences the subjects addressed in this thesis appear diverse, but seen through the lens of a sociological perspective their commonalities can be brought into view. Thus this essay has enabled a number of common analytic aspects of contemporary healthcare to be highlighted.

The 'problem-oriented' nature of the multidisciplinary production of healthcare knowledge in the new healthcare sciences was discussed in Section 1 of this essay. In the case of outpatient care the problem was voiced by NHS hospitals and health authorities, and took the form of concerns about, in essence, the efficiency and quality of outpatient care services. In the case of early prostate cancer the issue of screening and diagnostic testing had been identified as a policy issue of national priority by the NHS Executive and the national Health Technology Assessment research programme; concern was also expressed in public forums as the public profile of the disease increased. In the case of total hip replacement, this also had been identified by the national HTA programme, had evoked adverse media comment, and orthopaedic waiting lists were a government and public concern. In summary, therefore, all three fields of healthcare techno-practice affected large proportions of the population, and raised questions of efficiency of healthcare organisation, effectiveness of interventions, and risks associated with exposure to the healthcare system itself.

Turning to the case study papers themselves, each of the three case studies describes variability and patterning in aspects of the collective delivery of healthcare by health professionals. By comparing the framing of the research questions in the different papers, as formulated in their stated objectives, it can be seen that the majority of the papers have addressed themselves to concerns with aspects of health care practices

and technologies whose patterning and variation has obvious implications for the quality and effectiveness of health care delivery. In each case study one implication of the variability shown is that patients receive treatment or diagnostic services based to some extent on vagaries of their access to some health service settings rather than others, which vary for reasons not connected with epidemiologically-defined 'need' or requirements (e.g. Coast et al, 1996).

The case studies have demonstrated the importance, for example, of underlying institutional professional practices that shape variations in patterns of healthcare delivery - 'institutional signatures' in outpatient practice; specialist centres in urological practice for prostate cancer, and centres for hip replacement favouring certain commercial models of prosthesis over others. I am not claiming here that such variations in access related to patterns of service provision are obviously or necessarily harmful for patients. At the level of the healthcare system, the variation shows that some patients receive services or technologies that others do not. The healthcare implication of this is a different question. Often it will be the case that technically 'more advanced' practices will be of greater benefit to patients, but it is clear, for example from the case of early detection of prostate cancer, that access to services which might be regarded as the most 'modern' or specialised cannot always be interpreted as unequivocally in patients' best health-related interests. Underlying the descriptions of healthcare patterns produced by the comparative 'method of variations' are analyses of the institutional beliefs and local medical socialisation practices that result in the outcomes of variability in healthcare and technology choice. In the terms adopted in this review essay, these 'outcomes' can be interpreted as potential risks arising from encounters with healthcare that as citizen/patients we experience as part of our participation in particular local systems of practice.

The concept of 'institutional signatures' represents an attempt to capture the notion of sources of variability in healthcare techno-practice that do not stem from individual practitioner preferences, beliefs and recipes for practice. Some variability in healthcare patterns is doubtless due to 'clinical uncertainty' about optimal

treatments. Numerous studies emanating from the clinical sciences attempt to identify factors that might account for sub-optimal service delivery when assessed against standards. Some of these studies which may combine consideration of patient morbidity, resource factors, geographical area, and individual practitioner beliefs may also assess some aspect of 'practice patterns' such as size of a healthcare team (e.g. Wauters et al, 2004). Variation attributable to individual factors is, of course, important but is not the focus in this thesis. The concept of the individual 'consultant signature' was first introduced, as mentioned above, by John Wennberg whose research has led to a gamut of 'small area variations' studies (Wennberg & Gittelsohn, 1982)[3]. However, the organisational, professional and institutional aspects of healthcare have remained under-researched and under-conceptualised in sociological studies. The recent silver anniversary edition of the journal *Sociology of Health and Illness* published overviews of 'organisational research' in the field, and they note the relative neglect of the area and a failure to build upon existing research (Davies, 2003; Griffiths, 2003). Although related concepts such as distinctive 'ideologies' in different clinical teams have been shown in early work in the field (Strauss, 1964), it is surprising that there appears to be no real equivalent of the clinical science 'practice patterns' concept in the sociological/organisational research literature, and certainly not as a widespread concept. It thus appears that this concept should be used to frame further empirical investigation and should be tested for its utility in relation to more established concepts in the sociological understanding of clinical practice, organisation and change. It is notable that quantitative and comparative research methods were used or re-analysed in the case studies presented here, and it may be that such approaches are required to further develop research in this field.

I now move on from formulating the contributions to healthcare knowledge of the three case studies, to assess the phenomenon of healthcare evidentiality as a key aspect of contemporary healthcare policy processes within contemporary society.

---

[3] Adequate exploration of the complex theoretical and empirical questions of *linkages between* individual and collective behaviour patterns is beyond the scope of this essay. It can be noted here that the variable power of individual consultants in the NHS – to influence clinical teams' practices, to control categorisations of patients (cf. Griffiths, 1997 & 2001), and to access resources for patients will account for a degree of the articulation between 'individual' and 'collective' characteristics.

5.2 Health care evidentiality: credibility and regulation

In this section I discuss the production of healthcare knowledge as 'evidence' in the healthcare science movements represented primarily by HTA and EBM. This is conceptualised as 'evidentiality', the term denoting the rhetorical and actual increased commitment to scientific empirical knowledge in contemporary healthcare policy-making. Firstly the epistemology of healthcare evidence is discussed, and secondly consideration is given to the institutional locales which process evidence and are involved in processes of healthcare policymaking. The role of evidentiality in *regulation* of healthcare innovations is then explored. I argue that the new evidentiality has emerged in parallel with (in response to, but also, paradoxically, engendering) the legitimation problems of healthcare/medicine. I suggest that the evidentiality movement has the twin functions of highlighting tension between on the one hand, credibility claims, and on the other hand evidential uncertainty, and of contributing to societal attempts at control over innovation in the healthcare system via 'regulation'.

The epistemology of healthcare evidence

The studies submitted with this essay are illustrative of the methodological variety in the development of the healthcare sciences during the 1990s. Many different research agendas and designs, within the HSR/HTA paradigms, can be used to produce evaluations of health care delivery issues, but the core methodologies for assessing causal effectiveness of interventions use experimental methods in the form of controlled trials or the systematic review. Knowledge about healthcare is highly diverse but the professional scientific movements of HTA and EBM are quite focused in their research agendas, question formulations, research methodologies and modes of organisation of 'expert' knowledge.

The nature of evidence in these movements has been subjected (as suggested in Paper 1) to a massive amount of methodological self-scrutiny by its proponents and

practitioners. This can be regarded as an attempt to strengthen the scientific foundations upon which the results of investigations into healthcare are based. Paper 1 analyses the core epistemology of healthcare science which is clearly an application to healthcare systems and practices of the experimental bias-elimination principles of biomedical and epidemiological quantitative and comparative research. The terminology of 'outcomes measurement' and HTA's focus upon generic outcomes such as 'quality of life' indicate the clinically-driven framing of research questions in which causality is framed by the primary focus upon medical conditions and 'interventions' which produce effects on the health status of patient populations. The nature of this epistemology has been elaborated by Tanenbaum (1994) and Harrison (1998).[4]

It has been argued that high degrees of emphasis on standardisation and quantification are symptomatic of professions/organisations suffering loss of public trust (Porter, 1995). Thus the movement toward the aim of knowledge-based rationalisation of healthcare systems can be regarded, as suggested at the end of Paper 1 presented here (1997: 203), as an essentially modernist response to the legitimation crisis well-documented as 'challenges to medicine' that have been observed over the last two decades in medicine/healthcare (Gabe et al, 1994). Within this perspective, therefore, it is possible to understand the numerous efforts to standardise the methodologies for the production of healthcare knowledge, for example through handbooks of systematic review methods (NHS CRD, 1996) and guidelines for the conduct of randomised controlled trials. As citizens and patients, therefore, we are invited to renew our trust in medicine/healthcare because of the reinforcement of its scientific underpinnings. As also noted in Paper 1, it is paradoxical that a movement to restore public trust proceeds by a process that involves a heightened questioning of the evidential basis of medical techno-practice. It thus appears that the evidentiality movement attempts to 'manage' both scientific

[4] Harrison in this article, like many other authors subsequently, tends to equate 'evidence-based medicine' (EBM) with all manifestations of the broader growth of what I have termed in this essay the healthcare sciences. While I cannot explore this point in detail here, it is important to distinguish the different strands within the healthcare science movement. HTA as a movement and as a practice, for example, does not share the ideological and professional discourses of EBM focused on 'individual clinical expertise' (cf. Sackett et al, 1996).

uncertainty and public confidence. I suggest that it does so via an institutional positioning between the practitioners of healthcare science, policy processes and patients and publics. Thus it exhibits a form of what might be termed 'bounded uncertainty' in attempting to negotiate between actors that produce uncertainty on the one hand, and actors that seek credible and evidentially robust public services on the other (see sub-section below on the 'organisational sites of evidentiality' for discussion of the related topic of the intermediary role of 'boundary organisations').

Generally, the knowledge-producing approaches of HTA and EBM can be criticised, from a sociological or 'policy analysis' point of view as being, by virtue of their methodology, impervious to various important factors in the policy environment of healthcare, such as manufacturers' interests, the concerns of patient advocacy groups, and existing regulatory policy. This point is taken up in section 5.3 below. A caveat to this assertion is that the HSR/HTA paradigm is *now* increasingly according some recognition to the importance of 'patients' preferences' in its research agendas, though perhaps in ways that could be regarded as constructing 'patients needs' in narrow ways (e.g. Haynes et al, 2002). The Medical Research Council's 'Health Services Research Collaboration' national centre poses 'How can we gain an understanding of people's needs and the population's expectations and experience of health care?' as a key question in its research agenda (MRC HSRC, 2004). These aspects typically require 'qualitative' research methodologies. It can be noted that integration of these with experimental and quasi-experimental methods, and varieties of survey methodology, is seen by the practitioners of the new healthcare sciences as important to the development of HSR/HTA methodology (e.g. Pope & Mays, 1995).

The organisational sites of evidentiality

The relationship between science and policy is a central concern in contemporary regulatory states, as it is in the academic field of science and technology studies. Under the conditions of mode 2 knowledge production/post-normal science (see

section 1), this relationship is changing. It is argued that society in general is moving from an era of technocratic policy-making towards a more inclusive regime (Abels, 2002), though this is not a universal observation across all technology sectors (Abraham & Lewis, 2000: 205). Interaction between science and policy has been conceptualised in terms of 'boundary organisations' (Guston, 1999) which are interpreted as stabilising the boundaries at the interface between the two spheres of activity. Analysis of the structure of interactions between evidence-producing institutions and healthcare policy communities suggests that they have such an intermediary role, though below I note reservations about their stabilising function.

Examples of evidence-producing and processing institutions in healthcare R&D in the UK are the NHS Centre for Reviews and Dissemination, the Cochrane Collaboration, the National Co-ordinating Centre for Health Technology Assessment, and the National Institute for Clinical Excellence (NICE; the latter conducts its own 'technology appraisals'). These newly-designed institutions can usefully be seen as boundary organisations acting to promote the generation and processing of healthcare science in the service of policy but at the same time acting as buffers between the evidence-producers and the evidence-consumers. These hybrid organisations permit healthcare governance processes, drawing upon scientific appraisal of new techno-practices, to design into the policymaking process other actors such as manufacturers, health service managers and patient advocate groups. In other words these organisational sites of evidentiality allow for 'the politics' of the results of healthcare science to be considered. In a similar manner, an intermediary role of research commissioners and producers has been suggested by May et al (2001) who apply perspectives from actor-network theory (e.g. Callon, 1986) to conceptualise funding bodies for evaluation of service development initiatives as 'points of passage' which clinical innovators must pass through in attempts to gain support and legitimacy for their innovations. Thus the production of evaluative evidence under the principles of HSR/HTA methodology mediates the processes of the policy appraisal and subsequent adoption of new healthcare techno-practices.

However, while boundary organisations introduce apparent *structural* stability, the boundaries between science and policy that they embody remain permeable and malleable (cf. Gieryn, 1983). They are better conceptualised, in my view, as boundary zones, rather than clear lines. Such zones become sites for contestation of interests and evidence, suggesting that the uneasy enrolment of HSR/HTA and clinical science into the disciplines and structures of healthcare policymaking should be viewed as a dynamic process that shapes and re-shapes such boundary zones. In the case of the NHS in the UK, a focus on boundary-work draws attention to conflicting role-definitions amongst stakeholders and to their interdependencies and power-relations. Drawing upon my own participant experience in UK national HSR and HTA, I suggest that the activity of making 'policy recommendations' was an issue of contention that highlighted boundary-definitions in the early history of the national HTA movement. At this time, in the case of Paper 3 (systematic review of total hip replacement) and Selley et al (1997; review of diagnostic and screening tools for early prostate cancer) it was unclear the extent to which healthcare policy recommendations might be offered by researchers, in other words by the healthcare scientists including myself. Attempts so to do, in fact, met with the disapproval of the NHS policy community who commissioned the research. A similar phenomenon can be observed in the case of NICE. NICE has the status of a Special Health Authority, and operates in a quasi-legal manner. Decisions about guidance to the NHS on adoption policy for new pharmaceuticals or technologies are made by a select group of individuals. But the evidence they refer to and their interpretation of it can then be challenged by interested parties. Multi-stakeholder 'hearings' for the consideration of the evidence – now in both the 'scientific' *and* quasi-legal senses of the word - about particular new techno-practices are therefore part of the *modus operandi* of NICE. Often, 'appeals' from industry are presented. Such processes as these can be seen, therefore, to be at the same time threatening *and* constitutive of boundary zones between healthcare science and healthcare policy. It is inevitable, and perhaps socially beneficial, that such controversies are generated in and around boundary organisations such as NICE and the national HTA enterprise.

The institutional vehicles of evidentiality thus represent key actors in analysis of the healthcare innovation process for any given techno-practice. This suggests that evidence produced under the epistemological and organisational auspices of the new healthcare sciences has a *regulatory* function with respect to innovation in healthcare. This point is discussed further below.

Evidentiality and regulation

As argued above, and in submitted papers 1 and 4, I suggest that 'Evidence' (as embodied in the HSR/HTA and EBM movements) has a quasi-regulatory function. National HTA constructs distributed laboratories in which new healthcare technologies are tested in comparison to existing technologies. Paper 1 showed that the application of HTA to an innovative healthcare technology could provide a sort of 'safety zone' for the technology – and for the risk-related national policy decision about its deployment – while scientific uncertainty was considered. As I wrote in Paper 1 'Laboratories harbour uncertainty while engaged in projects to reduce it.' (p203). This principle has subsequently been demonstrated even more strongly in the working of the National Institute of Clinical Excellence, which routinely issues guidance that technologies are to be regarded as experimental pending the production of new knowledge in the form of appropriate clinical trial evidence. Meanwhile, patients should only be exposed to the technology in the context of such evaluation. Clearly, therefore, the new healthcare sciences are enrolled by the state into regulatory control over the diffusion of healthcare innovations.

One of the responses to risky healthcare variations has been the development of the movement to produce 'evidence–based guidelines' with the aim of standardising techno-practice (clinical practice patterns). This can be regarded as a modernist, rationalist response which has emerged with the possibility of restoring legitimation to the healthcare system. Evidence-based guidelines are clearly also part of the new regulation of healthcare. Many health-professional representative bodies produce their own guidelines. NICE itself produces a wide-range of authorised guidelines in the UK with its own stamp of credibility as the National Health Service's central

body for the assessment of the science of healthcare interventions. Guidelines represent an explicit attempt to shape frontline medical techno-practice by translating the evidence produced by healthcare science into working rules and routines for decision-making and clinical policy. However, guidelines represent a form of attempted standardisation of clinical behaviour that has to contend with evidential uncertainties as well as the vagaries of individual and institutional working practices. The standardisation associated with guidelines in the context of the Evidence-Based Medicine movement, in a sociopolitical analysis that to some extent exemplifies the approach that I argue for in this essay, at the level of medical practice, has been explored in detail by Timmermans and Berg (2003b).

The importance of the link between governance and evidentiality is increasingly recognised within professional networks of medicine. To give an example from the case studies discussed in this thesis, in the controversial case of policy for detection of early prostate cancer, 'Stronger and braver governance is required to ensure that responsible decisions about risk management emerge for areas such as screening, which have such potentially enormous individual and societal consequences. These decisions must be based on sound research and proper partnerships' (Thornton & Dixon-Woods, 2002).

This analysis is echoed elsewhere in recent research literature on scientific expertise and policy formation. At a transnational level of analysis 'regulation by information' for example has been suggested as characteristic of the (European) regulatory state (Majone, 1997), and Lehoux and Blume (2000) have conceptualised HTA explicitly as a 'regulatory science'. Majone's analysis suggests that information can only change expectations and behaviour if it has credibility. The evidence-producing movement in health care has a number of features that that may be regarded as more or less 'designed' to build and enhance societal credibility. These features are evident in the organisations that have been considered above as the key national organisational sites of healthcare evidentiality in the UK. These institutions have claimed and been accorded a high degree of authority within healthcare policy. Perhaps most obviously they lay claim to credibility through processes of

centralisation. This applies both to the processes by which evidence is brought together and summarised, in particular through the use of systematic review methodology, and to the institutions' organisational forms as symbolic 'national' centres.

I have characterised HSR/HTA in this essay as being healthcare sciences engaged in scientific activity focused primarily upon issues of efficiency, quality and effectiveness of healthcare delivery. These sciences produce knowledge using the epistemology discussed above and organised through some of the institutions identified in this discussion. Regarding these as 'regulatory science' (Jasanoff, 1990; Irwin et al 1997; Lehoux & Blume, 2000) recognises the relationship between knowledge processes and societal control. It draws attention to 'the politics' of the linkages between science and policy: for example, regulatory institutions vary in their appraisal and decisions about techno-practices (cf Abraham & Lewis, 2000); expert scientific knowledge may be challenged by citizen-experts (e.g. Epstein, 1996); and innovations in scientific knowledge may be shaped by the commercial interests of industry (Blume, 1992). These considerations make it clear that the production and deployment of knowledge about healthcare should be regarded as in part a process that concerns the regulatory control of new healthcare techno-practices.

I argued in Paper 1 that HTA shapes the production of knowledge using methodologies that effectively exclude the social, political and ethical agendas that HTA makes programmatic claims for. It is thus left to other disciplinary perspectives, such as sociology of science/science and technology studies, to explore sociological and political analysis of healthcare techno-practices, including the intermediary boundary between the various actors producing healthcare knowledge about innovative practices and those formulating healthcare policy which may or may not promote innovation. This leads us to an explicit consideration of the dimensions of a sociopolitics of the dynamics of healthcare innovation and regulatory policy. This is developed in section 5.3 below.

5.3 The sociopolitics of health care innovation and policy

The previous section considered the production and organisation of healthcare knowledge as a form of 'evidentiality'. The final section of this essay reflects upon the components of a 'sociopolitics' of the dynamics of technology/practice (techno-practice) innovation in healthcare systems, of which the evidentiality of the new healthcare sciences constitutes a significant part. The intermediary knowledge-processing organisations are an increasingly important actor in the sociopolitics of healthcare innovation. The argument builds upon glimpses of such concerns that are to be found in submitted Papers 1 (HTA), 2 and 4 (artificial hips innovation and policy) and 8 (power dynamics at the primary-secondary care interface).

A number of trends mark the development of the contemporary healthcare system in the United Kingdom. These are associated with broader movements in public policy and service delivery, and structural and cultural developments at a societal level. These broader trends were alluded to above (section 1.2) in considering the origins of transdisciplinary knowledge production. Thus societal changes are in train that have been conceptualised as the regulatory state, the knowledge-based economy, network governance, risk society, citizen science and consumerisation, finding expression in part in a re-structuring of healthcare and healthcare policy processes. In healthcare trends are to be seen toward shared medical decision-making; participatory governance; 'patient pathways'; 'knowledge-based' service/evidentiality; technological modernisation; lifelong learning; risk-based regulation; standards and targets-setting; and intensification of specialist services and 'outplacing' i.e. devolution to primary, community or 'self' care. Naturally, these trends are not always acting in harmony.

It has been the aim of this review essay to set the substantive topic areas in their sociomedical and policy contexts. Throughout, the concept of collective techno-practices has been emphasised. Paper 8 was able to show dissonance in the 'power perceptions' of hospital doctors and general practitioners across the sociomedical interface of primary and secondary care. Paper 1 drew on sociological perspectives

to analyse Health Technology Assessment as one strand of the broad movement in the new healthcare sciences - part of 'evidence-based healthcare' - whose growth characterised the 1990s. Further, Paper 4 introduced ideas about the need to extend analysis of clinically-defined policy issues to embrace broader, sociopolitical analysis of the policy environment in which the clinical problems are embedded. This perception is paralleled also in recent analysis of health technology assessment other than my own (Lehoux and Blume, 2000). This extension of the toolkit for analysing clinical/healthcare policy into the realm of the sociopolitical has been echoed by May *et al* who note 'looking at telehealthcare through the lens of clinical studies means that other substantial sectional interests are excluded from analysis' and in terms of the organisation of 'policy-related clinical science' (2001:905), the interests of multiple stakeholders including not only clinical constituencies but also manufacturers and patient groups should be taken in to account (2001:133).

This review essay, therefore, points to tensions between state or professional forms of control over healthcare practices and countervailing forces of innovation. Such forces of innovation come both from within modernising medical professions and from 'outside', for example commercial organisations or societal demands. 'Modernisation' is enshrined in the policy of the NHS, and may run counter to the institutionalised processes of evidentiality discussed above in terms of the intermediary boundary between healthcare science and policy. These two policy directions may conflict with each other in ways which determine the experience at the boundaries of the healthcare system of specific innovative techno-practices. An empirical analysis of such an effect has been produced in the case of telehealthcare (May et al, 2001) and its constituent elements in the practical conduct of healthcare technology evaluation conceptualised in a contingency model (May et al 2003).

The silver anniversary edition of the journal Sociology of Health & Illness contained two chapters on health technologies intended to review the field (Heath et al 2003; Timmermans & Berg 2003a). One might expect that such a professionally high-profile degree of attention to technology would highlight the importance of a politico-economic analysis of healthcare technologies, their production, processes of

adoption and diffusion, interface with healthcare systems and their health and social consequences. However, this is not the case. The chapters that deal with technology do so almost entirely by viewing them in a context of professional-patient interaction. While this site of technology is, of course, important, it is a testimony to the neglect of socio-political analysis of health technology innovation that these key review articles take the focus that they do. In this review essay, therefore, I hope to have argued the case for formulating healthcare science as embodied in HTA/HSR as one important dimension to be considered in a socio-political analysis of contemporary healthcare innovation.

BIBLIOGRAPHY

Abels G. (2002). Experts, citizens, and Eurocrats - towards a policy shift in the governance of biopolitics in the EU. *European Integration Online Papers* http://eiop.or.at/eiop/eiop/pdf/2002-019.pdf. Accessed 9/2004.

Abraham J, Lewis G. (2000). *Regulating medicines in Europe: competition, expertise and public health.* London, Routledge.

Altenstetter C. (1996). Regulating healthcare technologies and medical supplies in the European Economic Area. *Health Policy*, 35, 33-52.

Barry MJ, Fowler FJ Jr, Mulley AG Jr, Henderson JV Jr, Wennberg JE.(1995). Patient reactions to a program designed to facilitate patient participation in treatment decisions for benign prostatic hyperplasia. *Medical Care*, 33, 8, 771-82.

BBC2. (1993). *The High Price Of Hips*. BBC Television programme, February 26th.

Beck U. (1992). *Risk society: towards a new modernity.* London: Sage Publications.

Beck U. (1999). *World risk society*. Cambridge, Polity Press.

Beck U., Giddens A., Lash S. (1994). *Reflexive modernization: politics, tradition, and aesthetics in the modern social order*. Cambridge, Polity Press.

Bero LA, Grilli R, Grimshaw JM, Harvey M, Oxman AD, Thomson MA. (1998).Getting research findings into practice: closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. *British Medical Journal*, 317, 465-468.

Black N. (1994). Why we need qualitative research. *Journal of Epidemiology and Community Health*, 48, 425-6.

Bloor M. (1976). Bishop Berkeley and the adenotonsillectomy enigma: an exploration of the variation in the social construction of medical disposals. *Sociology*, 10, 1, 43-61.

Blume SS. (1992). *Insight and industry: on the dynamics of technological change in medicine*. Cambridge (Mass.), MIT Press.

BOA. British Orthopaedic Association (1999). *Total hip replacement: a guide to best practice*. BOA, London.

Brown L, Payne S, Royle G. (2002). Patient initiated follow up of breast cancer. *Psychooncology*,11, 4, 346-55.

Brown N, Webster A (2004). *New medical technologies and society: re-ordering life*. Cambridge, Polity Press.

Bulstrode CJK, Murray DW, Carr AJ, Pynsent PB, Carter SR. (1993). Designer Hips. *British Medical Journal*, 306, 732-3.

Callon M. (1986). The sociology of an actor network: the case of the electric vehicle. In: Callon M, Law J, Rip A. (Eds.). *Mapping the dynamics of science and technology: sociology of science in the real world.* London, Macmillan, pp19-34.

Cameron E, Bernardes J. (1998). Gender and disadvantage in health: men's health for a change. *Sociology of Health and Illness*, 20, 5, 673-693.

Chamberlain J, Melia J, Moss S, Brown J. (1997). The diagnosis, management, treatment and costs of prostate cancer in England and Wales. *Health Technology Assessment,* 1 (3) (whole volume).

Chapple A, Ziebland S, Shepperd S, Miller R, Herxheimer A, McPherson A. (2002). Why men with prostate cancer want wider access to prostate specific antigen testing: qualitative study. *British Medical Journal*, 325, 737-741.

Coast J, Donovan J, Frankel S. (Eds.). (1996). *Priority-setting: the healthcare debate*. London, Wiley.

Coulter A. (1995). Shifting the balance from secondary to primary care. *British Medical Journal*, 311, 1447-1448.

CSAG. Clinical Standards Advisory Group (2000). *A study of outpatient services.* http://www.dh.gov.uk/assetRoot/04/01/46/45/04014645.pdf. Accessed 09/2004.

Davis HTO, Nutley SM. (1999). The role of evidence in public sector policy and practice. the rise and rise of evidence in health care. *Public Money and Management*, 19(1), 9-16.

Davies C. (2003). Some of our concepts are missing: reflections on the absence of a sociology of organisations in *Sociology of Health and Illness. Sociology of Health and Illness*, 25, 3, 172-190.

Department of Health. (1990). *Health and personal social services statistics for England, 1990.* London:HMSO.

Department of Health (2000). *National joint replacement registry. A consultation document.* London, Department of Health.

Department of Health (2001). *The expert patient: a new approach to chronic disease management for the 21st century.* Department of Health.

Donovan JL, Frankel S, Faulkner A, Gillatt D, Hamdy FC. (1999). Dilemmas in treating early prostate cancer: the evidence and a questionnaire survey of consultant urologists in the UK. *British Medical Journal*, 318, 299–300.

Donovan JL, Frankel SJ, Neal DE, Hamdy FC. (2001). Screening for prostate cancer in the UK. *British Medical Journal*, 323, 763 – 764.

Donovan J, Mills N, Smith M, Brindle L, Jacoby A, Peters T, Frankel S, Neal D, Hamdy F, Little P. (2002). Quality improvement report: Improving design and conduct of randomised trials by embedding them in qualitative research: ProtecT (prostate testing for cancer and treatment) study. *British Medical Journal*, 325, 766 - 770.

Edwards RHT. (1999). Is it time for an Illich Collaboration to make available information on the harms of medical care? *British Medical Journal*, 318, p58. [Letter].

Epstein S. (1996). *Impure science: AIDS, activism and the politics of knowledge.* Berkeley, University of California Press.

Faulkner A, Frankel S. (1993). *Evaluation of delivery and use of outpatient services.* Bristol, University of Bristol Department of Epidemiology and Public Health.

FDA. Food and Drug Administration (1994). Press release 29.8.94. - *FDA approves test for prostate cancer*, Rockville, FDA.

Fitzpatrick D, Shortall E, Sculpher M et al. (1998). Primary total hip replacement surgery: a systematic review of outcomes and modelling of cost-effectiveness associated with different prostheses. *Health Technology Assessment*, 2, i-iv,1-64.

Frankel S, Robbins M. (1993). Entering the lobby: access to outpatient assessment. In: Frankel S, West R (Eds.). *Rationing and rationality in the National Health Service: the persistence of waiting lists*. Basingstoke, Macmillan.

Frankel S, Faulkner A. (1994). The end of the outpatient problem? *British Medical Journal*, 309, 1308.

Funtowicz, S. and Ravetz, J. (1992). Risk management as a post-normal science. *Risk Analysis*, 12(1), 95-97.

Gabe J, Kelleher D, Williams G. (Eds.). (1994). *Challenging medicine*. London, Routledge.

Garfinkel H. (1967). *Studies in ethnomethodology*. Englewood Cliffs, NJ, Prentice Hall.

Gibbons, M., et al. 1994. *The new production of knowledge: the dynamics of science and research in contemporary societies*. Newbury Park, CA.

Giddens A. (1991). *Modernity and self-identity*. Cambridge: Polity Press.

Gieryn TF. (1983). Boundary-work and the demarcation of science from non-science: strains and interests in professional ideologies of scientists. *American Sociological Review*, 48, 781-95.

Greener I. (2003). Patient choice in the NHS: the view from economic sociology. *Social Theory and Health*,1,1,72-89.

Griffiths L. (1997). Accomplishing team: teamwork and categorisation in two community mental health teams. *The Sociological Review*, 45, 59-78.

Griffiths L. (2001). Categorising to exclude: the discursive construction of cases in community mental health teams. *Sociology of Health and Illness*, 23, 5, 678-700.

Griffiths L. (2003). Making connections: studies of the social organisation of healthcare. *Sociology of Health & Illness*, 25, 3, 155-171.

Guston D. (1999). Stabilizing the boundary between politics and science: The role of the Office of Technology Transfer as a boundary organization. *Social Studies of Science*, 29, 1, 87-112.

Harrison S. (1998). The politics of evidence-based medicine. *Policy and Politics*, 26, 1, 15-31.

Haynes RB, Devereaux PJ, Guyatt GH. (2002). Physicians' and patients' choices in evidence based practice. *British Medical Journal*, 324, 1350. [Editorial].

Heath C, Luff P, Svensson MS. (2003). Technology and medical practice. *Sociology of Health & Illness*, 25, 3, 75-96.

Horlick-Jones T, Sime J. (2004). Living on the border: knowledge, risk and transdisciplinarity. *Futures* 36, 441-456. (Special Issue on Transdisciplinarity).

Howson A. (1998). Surveillance, knowledge and risk: the embodied experience of cervical screening. *Health*, 2, 2, 195-215.

Huiskes R. (1993). Failed innovation in total hip replacement. Diagnosis and proposals for a cure. *Acta Orthopaedica Scandinavica*, 64(6), 699-716.

Illich I. (1975). *Limits to medicine: medical nemesis, the expropriation of health.* London, Calder and Boyars.

Irwin A. (1995). *Citizen science: a study of people, expertise and sustainable development.* London, Routledge.

Irwin A, Rothstein H, Yearley S, McCarthy E. (1997). Regulatory science - towards a sociological framework. *Futures*, 29 (1), 17-31.

Kent J, Faulkner A. (2002). Regulating human implant technologies in Europe - understanding the new era in medical device regulation. *Health, Risk & Society*, 4, 2, 190-209.

Kievit J. (2002). Follow-up of patients with colorectal cancer: numbers needed to test and treat. *European Journal of Cancer*, 38(7), 986-99.

Klein, J.T. (1996). *Crossing boundaries: knowledge, disciplinarities and Interdisciplinarities*. Charlottesville, University Press of Virginia.

LeFanu J. (2000). *The rise and fall of modern medicine*. Little Brown.

Lehoux P, Blume SS. (2000). Technology assessment and the sociopolitics of health technologies. *Journal of Health Policy, Politics and Law*, 25 (6), 1083-1120.

Lupton D. (1995). *The imperative of health: public health and the regulated body*. London, Sage.

MacKenzie D. (1999). Theories of technology and the abolition of nuclear weapons. In: MacKenzie D. and Wajcman J. (Eds.). *The social shaping of technology*. Buckingham, Open University Press, 2nd Ed., pp419-442.

Majone G. (1997). The new European agencies: regulation by information. *European Journal of Public Policy*, 4, 2, 262-275.

May C, Mort M, Mair F, Williams T. (2001). Factors affecting the adoption of tele-healthcare in the United Kingdom: the policy context and the problem of evidence. *Health Informatics Journal*, 7, 131-134.

May C, Mort M, Williams T, Mair F, Gask L. (2003). Health technology assessment in its local contexts: studies of telehealthcare. *Social Science & Medicine*, 57, 4, 697-710.

McPherson K. (1988). Variations in hospitalisation rates: why and how to study them. In: Ham C. (Ed.). *Health care variations: assessing the evidence*. London, King's Fund Institute, pp15-20.

Mead N, Bower P. (2000). Patient-centredness: a conceptual framework and review of the empirical literature. *Social Science & Medicine*, 51, 7, 1087-1110.

Mechanic D. (1968). *Medical sociology: a selective view*. New York, Free Press.

Moran M. (2001). The rise of the regulatory state in Britain. *Parliamentary Affairs*, 54, 1, 15-31.

MDA. (1998). *Guidance on the medical devices vigilance system for CE marked joint replacement implants*. Medical Devices Agency, London.

MRC HSRC (2004). *The MRC HSRC research agenda.* http://www.hsrc.ac.uk/Current_research/res_agenda.htm. Accessed October 2004.

Murray DW, Carr AJ, Bulstrode CJ. (1995). Which primary total hip replacement? *Journal of Bone and Joint Surgery*, 77-B, 520-527.

National Audit Office (1991). *NHS outpatient services.* London:HMSO.

National Audit Office. (2000). *Hip replacements: getting it right first time.* Report by the Comptroller and Auditor General. Stationery Office, London. 17 April 2000.

National Audit Office. (2003). *Hip replacements: an update.* Report by the Comptroller and Auditor General. Stationery Office, London. 14 July 2003.

NICE. (2000). *NICE issues guidance to the NHS on artificial hip joints.* National Institute of Clinical Excellence, London.[Press release].

NHS CRD. Centre for Reviews and Dissemination (1996). *Undertaking systematic reviews of research on effectiveness: CRD guidelines for those carrying out or commissioning reviews.* CRD Report 4. York, NHS CRD, York University.

NHS CRD. Centre for Reviews and Dissemination (1997). *Effectiveness Matters. Screening for prostate cancer: the evidence. Information for men considering or asking for PSA tests.* York, University of York, NHS CRD.

NHS Executive (2000). *The NHS Prostate Cancer Programme.* http://www.doh.gov.uk/cancer/prostate.htm. Accessed 17 May 2003.

NHS Executive (1995). *The Patient's Charter. Demonstrating quality: good practice in outpatient departments.* Leeds, NHSE.

NSC. National Screening Committee (1998*). First Report of the National Screening Committee.* Health Departments of the United Kingdom.

Petersen A. (1996). Risk and the regulated self: the discourse of health promotion as politics of uncertainty. *Australian and New Zealand Journal of Sociology*, 32,1, 44-57.

Pope C, Mays N. (1995). Qualitative Research: Reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. *British Medical Journal*, 311, 42-45.

Porter TM. (1995). *Trust in numbers: the pursuit of objectivity in science and public life.* Princeton University Press.

Ravetz J. (2001). Science advice in the knowledge economy. *Science and public policy*, 28, 5, 389-393.

Rhodes R.A.W. (1997). *Understanding governance: policy networks, reflexivity and accountability.* Buckingham, Open University Press.

Robertson A. (2000). Embodying risk, embodying political rationality: women's accounts of risk for breast cancer. *Health, Risk and Society*, 2, 2, 219-235.

Sackett DL, Rosenberg WMC, Muir Gray JA, Haynes RC, Richardson WS. (1996). Evidence based medicine: what it is and what it isn't. *British Medical Journal*, 312, 71-72. [Editorial].

Sanders T, Harrison S. (2004). *Use of scientific evidence in heart failure: assessing the impact on patient participation.* Paper presented to British Sociological Association Medical Sociology Group Annual Conference, University of York, York. Abstract at: http://www.britsoc.co.uk/Library/AbstractBk2Sept.pdf. Accessed 09/2004.

Selley S, Donovan J, Faulkner A, Coast J, Gillatt D. (1997). Diagnosis, management and screening of early localised prostate cancer. *Health Technology Assessment*, 1997; 1 (whole volume).

Sherriff R, Best L, Roderick P. (1998). Population screening in the NHS: a systematic pathway from evidence to policy formulation. *Journal of Public Health Medicine*, 20,1, 58-62.

St Leger A, Schnieden H, Walsworth-Bell JP. (1992). *Evaluating health services' effectiveness.* Milton Keynes, Open University Press.

Stewart-Brown S, Farmer A. (1997). Screening could seriously damage your health. *British Medical Journal*, 314, 533.

Strong PM. (1979). *The ceremonial order of the clinic: doctors, parents and medical bureaucracies*. London, Routledge & Kegan Paul.

Tanenbaum SJ. (1994). Knowing and acting in medical research: the epistemological politics of outcomes research. *Journal of Health Policy, Politics and Law*, 19, 27-44.

Thornton H, Dixon-Woods M. (2002). Prostate specific antigen testing for prostate cancer. *British Medical Journal*, 325, 725-726.

Timmermans S, Berg M. (2003a). The practice of medical technology. *Sociology of Health & Illness*, 25, 3, 97-114.

Timmermans S, Berg M. (2003b). *The gold standard: the challenge of evidence-based medicine and standardization in health care*. Philadelphia, Temple University Press.

Wauters JP, Bosson JL, Forneris G, Turc-Baron C, Golshayan D, Paternoster G, Martina G, Hurot JM, von Albertini B, Foret M, Cordonnier D, Piccoli G. (2004). Patient referral is influenced by dialysis centre structure in the Diamant Alpin Dialysis cohort study. *Nephrology Dialysis Transplantation*, 19, 9, 2341-2346.

Webster A. (2004). Health technology assessment: a sociological commentary on reflexive innovation. *International Journal of Technology Assessment in Health Care*, 20, 1, 1-6.

Wennberg J, Gittelsohn A. (1982). Variations in medical care among small areas. *Scientific American*, 246, 100-112.

Wennberg J. (1988) Practice variations and the need for outcomes research. In: Ham C. (Ed.). *Health care variations: assessing the evidence*. London, King's Fund Institute, pp32-35.

Woolf SH, Henshall C. (2000). Health technology assessment in the United Kingdom. *International Journal of Technology Assessment in Health Care*, 16, 591-625.

Appendix 1.

Full bibliography of Published Work included

The papers are listed in logical (rather than chronological) order, to reflect the discussion in the critical review presented in the thesis.

1. Faulkner A.(1997). 'Strange bedfellows' in the laboratory of the NHS? An analysis of the new science of health technology assessment in the United Kingdom. In: Elston MA (ed) The sociology of medical science and technology. *Sociology of Health and Illness* Monograph No. 3, pp 183-207.

2. Faulkner A. (2001). 'Casing the joint - the material development of artificial hips'. In Ott K, Serlin D & Mihms S (Eds.) Artificial Parts, Practical Lives - modern histories of prosthetics. NY: New York University Press, 2002.

3. Faulkner A, Kennedy LG, Baxter K, Donovan J, Wilkinson M, Bevan G. (1998). 'Effectiveness of hip prostheses in primary total hip replacement: critical review of evidence, and an economic model'. *Health Technology Assessment,* vol 2 no.6 (whole issue).

4. Faulkner A, Kent J. (2001). 'Innovation and regulation in human implant technologies: developing comparative approaches'. *Social Science & Medicine,* 53, 895–913.

5. Faulkner A, Brookes S, Donovan J, Selley S, Gillatt D, Hamdy F.(2000). 'The use of prostate-specific antigen testing in the detection of localised prostate cancer: current opinion and urological practice in the United Kingdom'. *European Journal of Public Health,* 10(4): 289-295.

6. Faulkner A, Harvey I, Peters T, Sharp D, Frankel S. (1997). 'Profiling outpatient workload: practice variations between consultant firms and hospitals in South West England'. *Journal of Epidemiology & Community Health,* 51(3): 310-314.

7. Faulkner A, Saltrese-Taylor A,Williams M, O'Brien J, Collins CD, & Frankel S. (1995). 'Outpatients revisited: subjective views and clinical decisions in the management of general surgical outpatients in South West England'. *Journal of Epidemiology & Community Health,* 49: 599-605.

8. Somerset M, Faulkner A, Shaw A, Dunn E, Sharp D. (1999). 'Obstacles on the path to a primary care-led NHS: complexities of outpatient care'. *Social Science & Medicine,* 48: 213-225

Appendix 2.

List of publications describing the Author's (AF) contribution to co-authored works

1. Faulkner A. (1997). 'Strange bedfellows' in the laboratory of the NHS? An analysis of the new science of health technology assessment in the United Kingdom. In: Elston MA (ed.). The sociology of medical science and technology. Sociology of Health and Illness Monograph No. 3, pp 183-207.

AF Sole author.

2. Faulkner A. (2001). 'Casing the joint - the material development of artificial hips'. In Ott K, Serlin D & Mihms S (Eds.) Artificial Parts, Practical Lives - modern histories of prosthetics. NY: New York University Press, 2002.

AF Sole author.

3. Faulkner A, Kennedy LG, Baxter K, Donovan J, Wilkinson M, Bevan G. (1998). 'Effectiveness of hip prostheses in primary total hip replacement: critical review of evidence, and an economic model'. Health Technology Assessment, vol 2 No.6 (whole issue).

This publication arose from a project led by AF, undertaken for the UK's national Health Technology Assessment programme. AF planned the structure of the document, a 'systematic review', which gives emphasis to studies with a comparative element, planned the detailed analyses, conducted online literature searching in multiple databases, and undertook structured critical appraisal of some 100 published evaluations of the performance of different models of artificial hip (including randomised control trials, other comparative studies, and 'observational series'). These appraisals were done using checklists of methodological criteria developed from previously published guidance. Sections on epidemiology and history of hip prostheses were authored by AF. Kennedy undertook critical appraisal of a sub-set of the other studies reviewed in the document. Donovan interviewed 10 orthopaedic surgeons about clinical choice of prostheses. Wilkinson undertook preliminary literature searching/review. The section presenting the economic model was authored by Baxter and Bevan.

4. Faulkner A, Kent J. (2001). 'Innovation and regulation in human implant technologies: developing comparative approaches'. Social Science & Medicine, 53, 895-913.

AF initiated the concept for the article. The introductory section was authored primarily by AF, and the section on artificial hips' innovation and regulation is wholly authored by AF. Data used included company reports, clinical research articles, newspaper articles, and regulatory agency and professional body publications. The section on breast implants is authored by JK. The structure of the

Conclusion/discussion was devised by AF and the analytic material in the discussion is jointly authored by AF and JK.

5. Faulkner A, Brookes S, Donovan J, Selley S, Gillatt D, Hamdy F.(2000). 'The use of prostate-specific antigen testing in the detection of localised prostate cancer: current opinion and urological practice in the United Kingdom'. European Journal of Public Health, 10(4): 289-295.

The article draws on data from a UK-wide postal survey of urologists, connected to a systematic review of screening/diagnostic tests for early-stage prostate cancer, to which AF contributed as critical reviewer of published studies. AF contributed to formulating the questionnaire for the survey (with Donovan). For the paper AF formulated the research questions and planned the analysis. AF analysed verbatim responses to open-ended questions and performed descriptive statistical analysis of the dataset using STATA software. The multi-variable statistical analysis was undertaken by Brookes supported by a senior statistician. AF led the authorship of the text of the paper in collaboration with Donovan and Hamdy. Gillatt and Hamdy were clinical collaborators. Selley administered the survey.

6. Faulkner A, Harvey I, Peters T, Sharp D, Frankel S. (1997). 'Profiling outpatient workload: practice variations between consultant firms and hospitals in South West England'. Journal of Epidemiology & Community Health, 51(3): 310-314.

AF obtained anonymised summary data on outpatient activity rates from the local regional health authority. AF planned the analysis in collaboration with Harvey and Peters, and carried out the descriptive statistical analysis and graphical presentations in the paper. The components of variance analysis was carried out by Peters (Tables 2 and 3). The main text of the paper was written by AF with the support of Harvey.

7. Faulkner A, Saltrese-Taylor A, Williams M, O'Brien J, Collins CD, & Frankel S. (1995). 'Outpatients revisited: subjective views and clinical decisions in the management of general surgical outpatients in South West England'. Journal of Epidemiology & Community Health, 49:599-605.

The article is based on an analysis of a dataset derived from a self-completion and case-note based survey in a single hospital surgery department, administered by Saltrese-Taylor. AF cleaned the dataset and re-coded where necessary, planned and designed the analysis presented, undertook it using Epi-Info statistical software, and wrote the bulk of the text and produced the tables/graphics. AF coded and analysed verbatim responses to open-ended questions. Advice on quantitative analysis was provided by MW. JO'B and CDC were clinical collaborators involved in arranging practicalities and permissions within the hospital site for the original fieldwork. Frankel provided oversight.

8. Somerset M, Faulkner A, Shaw A, Dunn E, Sharp D. (1999). Obstacles on the path to a primary care-led NHS: complexities of outpatient care. Social Science & Medicine, 48: 213-225

AF formulated the NHS policy-oriented theme of this paper. AF was primary author of the 'Background' section (a referee for the journal commented that this in itself was an original contribution to knowledge), and made a significant contribution to the Discussion section. The paper draws on data provided from project fieldwork carried out by Shaw and Dunn under the management of Somerset. Sharp was 'project leader' of the project (of which AF was a member). AF provided advice on content of the interview schedule used in the fieldwork.

Appendix 3.

Published Work - Papers 1 to 8

The papers that follow are presented in logical order as listed in the Contents page and Appendix 1. Each is prefaced by a page giving the full bibliographic detail and acknowledging the permission of the publisher to reproduce it.

PAPER 1

Originally published 1997

Author: A. Faulkner

'Strange bedfellows' in the laboratory of the NHS? An analysis of the new science of health technology assessment in the United Kingdom.

Published in: Elston MA (ed). The sociology of medical science and technology. Sociology of Health and Illness, Monograph No. 3, pp 183-207.

Acknowledgment: Blackwell Publishing for permission to reproduce this chapter.

7. 'Strange bedfellows' in the laboratory of the NHS? An analysis of the new science of health technology assessment in the United Kingdom

Alex Faulkner

Introduction

A recent disease-focused history of the development of medical research in the United Kingdom, written by a professor of medicine, refers in passing to a new type of health care research. The author notes the multidisciplinary nature of this type of 'medical research' and draws our attention to alleged difficulties of research methodology:

> This new branch of medical research brings together some strange bedfellows, ranging from psychology and the social sciences to biomathematics. It presents many difficulties, not least the uneasy amalgamation of the relatively 'soft' science of interviewing techniques with some fairly sophisticated mathematics (Weatherall 1995: 312).

Current developments in the National Health Service suggest that research knowledge is seen amongst state authorities in the United Kingdom as a means of improving and controlling the development of healthcare services. The new type of health care research known as Health Technology Assessment (HTA) has emerged as both the largest financially and, symbolically, the highest-profile of the research programmes within the new NHS R&D strategy (Department of Health 1993). The NHS Executive has spearheaded the state's co-ordinating action in this area, in spite of the existence of other agencies with kindred health policy interests in the Department of Health. HTA is a multi-faceted movement constituted in the interactions of state policymaking bodies, the medical and healthcare professions, academia, hospital management and healthcare commissioning authorities. Overall the NHS R&D strategy aims to increase the proportion of annual NHS expenditure on research and development from 0.9 per cent to 1.5 per cent. It explicitly advocates the development of an evaluative culture within the NHS, aimed at developing a 'research-based' or 'knowledge-based NHS' (Department of Health 1993). The 'exploitation' of knowledge as a resource generated through research is one of the most important requirements of contemporary industrial capital (Webster 1994). New structures and management in public sector services are accompanied by expansion of research and other knowledge generating practices (Hoggett 1991, Hughes

© Blackwell Publishers Ltd/Editorial Board 1997. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main Street, Malden, MA 02148, USA.

and McGuire 1992). A consequence of this is the formation of new strategic alliances between the various producers and consumers of research knowledge, in which research disciplines participate in the strategic action of healthcare policymaking agencies.

The scope of the national health technology assessment movement brings into view many healthcare issues of interest to a sociological analysis. It raises issues to do with interests, values and inter-organisational relationships. It raises questions about the shaping of research agendas, healthcare policies, regulation of health technologies and, indeed, about the evolving patterns and methods of healthcare which we might have to call upon as health service users. It also raises issues germane to some of the prime concerns of the sociologies of medicine/healthcare and science/technology, including trust and contestability in medical authority, construction of health and healthcare risks, rhetorics of scientific projects and knowledge claims, relationships between the disciplines of medical and healthcare knowledge, and relationships between healthcare experiments, laboratories and technology tests.

In this chapter I am concerned with only a limited set of the possible themes. The focus, therefore, is upon the use of rhetorical discourse in the construction and shaping of a 'need' for health technology assessment around new constellations of institutions and disciplines – 'strange bedfellows' – which are negotiating agendas in healthcare knowledge. I locate this analysis by developing the notion of the NHS as a massive laboratory in and around which healthcare knowledge is produced. Implications for key themes in the sociological study of science/technology and health/illness are discussed. I confine myself largely to considering some of the key activities and developments which are part of the formal national NHS HTA movement in the United Kingdom, with some reference to related activities in the Medical Research Council (MRC), rather than the wider range of healthcare research activity much of which might also follow a broadly health technology assessment model.

Before moving on to consider the main themes of the chapter, I sketch briefly the state-co-ordinated formal structure which has been created for national HTA activity in the United Kingdom, followed by a description of the methods used in producing the analysis of HTA presented here.

Formal organisation and function of national HTA

Health technology assessment is the only one of the many research programmes, set up under the new NHS Research and Development Directorate, to receive support through the formation of a permanent standing group to oversee it, the Standing Group on Health Technology (SGHT). The SGHT identifies priorities for assessment through nation-wide

© Blackwell Publishers Ltd/Editorial Board 1997

PAPER 1

consultation, inviting topic suggestions from health professionals, representative bodies and others. Suggestions are considered by six advisory panels designed to reflect the full range of different sectors within healthcare. Five panels deal with healthcare sectors or types of technologies: the acute sector, primary and community care, pharmaceutical, diagnostics and imaging, and population screening panels. The sixth is concerned with the methodology of HTA. These groups score and rank technologies suggested for assessment which are then passed to the SGHT. SGHT performs a similar exercise to reach recommendations for commissioning assessment projects from research organisations. Membership of these panels and the standing group are of importance when considering the constituencies represented in national HTA. The major criteria for assigning priorities to assessments are stated to be: benefits in terms of improved outcomes for patients; methodological gains; timescale of potential benefits; value for money of assessment; importance of *early* assessment; and factors relating to Health of the Nation policy, prevalence and social/ethical considerations (Department of Health 1995: 46).

Methodology

For data about HTA in the United Kingdom, I draw upon reports published by the major new institutions representing the HTA movement, minutes and papers relating to meetings of advisory groups, comment and debate in the medical and health services press, discussion and communications with some key participants, and upon my own participation in HTA activity, akin to the 'participant comprehension' which Collins (1984) has described. Collins assumed that the researcher adopting this approach would have a single, clear professional identity as sociologist, but nevertheless describes how he and colleagues 'became scientists ourselves' (1984: 60). My professional affiliations are both to sociology and 'health services research'. For one year, in the capacity of a health services researcher, I became a 'scientific secretary' to one of the six advisory groups working with the major group charged with co-ordinating national HTA, the Central Research and Development Committee's Standing Group on Health Technology. I have also been involved in academic research work much of which can be described as health technology assessment, and some of which forms part of the national HTA programme. I thus also reflect upon familiarity with the practices, discourses, networks and institutions of HTA to inform the account presented here. Particular 'data' drawing on this experience are presented in the text in quotation marks and noted parenthetically as being 'author observation'.

© Blackwell Publishers Ltd/Editorial Board 1997

PAPER 1

Laboratories, experiments, technology testing and rhetoric

There have been recent calls for and signs of a rapprochement between the sociologies of medicine/healthcare and science/technology (Bartley 1990, Berg 1995, Casper and Berg 1995). The notion of a 'knowledge-based NHS' is of dual interest because it signals a bringing-together of scientific knowledge, which has been investigated in social studies of science, with healthcare practice, which has been investigated by sociologists of medicine and healthcare. Central to the practice of science are laboratories. Recent developments in the sociology of scientific knowledge and the history of science have regarded laboratories as both empirical (*e.g.* Latour and Woolgar 1979) and metaphorical (*e.g.* Macleod and Rehbock 1994) locations for the study of scientific activity and the production of scientific knowledge. Bartley (1990) suggested that medical sociology, using the tools and concepts of the sociology of science, might investigate medical scientists' empirical laboratories as the locations where medical knowledge is constructed. The notion of the NHS itself as a massive laboratory enables such approaches to be applied to the broader fields of healthcare practice and policy which come under scrutiny in the HTA enterprise.

The primary form of scientific activity which takes place in laboratories is the experiment and its interpretation. In social studies of science the variable relationship between experiments and laboratories is an important subject (Knorr Cetina 1992): ' . . . laboratories and experiments combine differently in different fields' (1992: 114). The sociological focus upon laboratories has permitted a view of science which goes beyond taken-for-granted notions about the ability of experimental methodology to support or negate hypotheses by applying bias-eliminating designs, to allow consideration of experiments in the context of the resources and practices employed in conducting them. Laboratories are not merely the location in which experiments are conducted, but they also involve, for example, the deployment of equipment and measurement instrumentation, the formation of strategic inter-individual and inter-organisational alliances, and the use of persuasive literary techniques in the presentation of 'findings' of experiments in scientific publications.

MacKenzie (1989) has drawn attention to the similarity between scientific experiment and technological testing. The approaches developed for the understanding of scientific knowledge can be applied also to the examination of the testing of (hardware) technologies. While the construction of scientific knowledge typically involves interpretation of particular observations in terms of theory, technology testing involves such interpretation in terms of predicted 'real-world' performance. Knorr Cetina (1992: 116) has described the laboratory as an 'enhanced environment', in which often obscure underlying processes are rendered legible by means of instrumenta-

© Blackwell Publishers Ltd/Editorial Board 1997

PAPER 1

tion and measurement. Thus experimentalists and technology testers work with 'traces' of processes rather than the processes themselves. Technology testing involves 'projection' from the test to observed performance (Pinch 1993). Observations produced 'under laboratory conditions' or under test conditions, are projected by interpretive techniques as predictions of how a technology would perform if applied in the real world. Thus the concepts which have been developed in social studies of science and technological testing can be used in the sociological investigation of the 'assessment', in other words the testing, of medical technologies. Health technology assessments enact various forms of projection, which construct health professionals, healthcare technologies and ourselves (as patients) as participants in the experimental laboratory of the NHS.

The sociology of scientific knowledge (SSK) has also drawn upon social constructionism to show how scientific facts are constructed through the manifold practices of scientists (Latour 1987). Key to this approach to science has been the concept of rhetoric. Having been historically regarded as antithetical to science, rhetorical discourse is seen in SSK as an intrinsic aspect of the methods used by scientists not merely in publicising the results of science but in the constitution of scientific theory and fact (Beer and Martins 1990). Rhetoric in this context can be defined as discourse which implicitly or explicitly persuades or proposes. Here the term has no pejorative connotation. Disciplines have disciplinary rhetorics, and fields of science characterised by the activity of several disciplines, such as health technology assessment, are likely to have multi-disciplinary or inter-disciplinary rhetorics. We can thus speak of the rhetorical constitution of disciplines or scientific fields as an aspect of the production of scientific knowledge.

The deployment of disciplinary rhetorics constitutes 'boundary-work' in Gieryn's sense (Gieryn 1983). In this conception, disciplinary discourse is to be seen as laying claim to professional territory by defining appropriate methods, concepts and agendas for a scientific programme. Such rhetorical work may be especially characteristic of new, changing, weak or interdisciplinary fields of scientific activity, where shared meanings and concepts are lacking (Porter 1995: 228). Tension between disciplinarity and interdisciplinarity is likely (Good 1993). Health technology assessment in the United Kingdom is just such a field of scientific activity.

The relation of disciplinary knowledge to the NHS is a matter which disciplinary activists play a part in constructing. This has been shown, for example, in the work of Ashmore *et al.* (1989) in their discussion of health economics in relation to the medical profession. The authors regard health economists as engaged in an 'educative strategy' (1989: 186) using programmatic discourses, in other words, discourses describing and promoting programmes of desirable change in NHS practices, which might be achieved by the adoption of lessons from the discipline of economics. Similarly, Pinch *et al.* (1992) discerned rhetorical devices at work in the social processes inherent in the introduction of

© Blackwell Publishers Ltd/Editorial Board 1997

PAPER 1

clinical budgeting technology in the NHS. In this chapter I focus primarily upon 'formal' rhetoric produced in policy statements and 'public' professional scientific debate to analyse the project of national HTA.

Constructing and supporting a legitimate 'need' for health technology assessment

In the late 1970s and early 1980s, a repertoire of accounts emerged in the discourse of international healthcare policymaking which might be drawn upon by different actors in constructing rhetorical justifications of a need for HTA. The first strand in this repertoire is a construction of policymakers' 'concerns'. In the late 1970s, a concern to promote cost containment as the key issue was paramount amongst policymakers' discourse on medical technology. For example:

> Technology has been identified as a major cause of increasing health care expenditures, . . . controlling new technology is required to contain health care costs (Committee on Technology and Health Care *et al.* 1979: 14).

Such accounts referred frequently to the uncontrolled proliferation of 'new medical technology', especially novel, high-cost, high-tech diagnostic equipment such as the CT scanner and magnetic resonance imaging (Jennett 1986). As the above account indicates, part of the policymakers' agenda was to control diffusion of this type of equipment because it caused escalating expenditure. In the perspective adopted here, this 'cost-containment' justification of a need for HTA is regarded as a rhetorical device which is constitutive of the boundaries of HTA, enabling certain interests and disciplines to lay claim to a stake in its activities. It does not follow that financial expenditure is the aspect of 'costs' which will necessarily be dominant in the evolving model of HTA practice. While concern for expenditure might have been presented as the primary cause, a *set* of 'concerns' has now come to be portrayed in healthcare policymakers' discourse as underlying the need for assessment of technologies:

> There is growing concern relating to the health benefits and risks of technology, its financial costs, and its social implications (Banta and Gelijns 1987: 255).

The first of these authors is a high-profile international adviser on health technology policy who was involved in meetings such as a seminal and symbolic multi-interest 'Tidal Wave' conference held in England in 1991 (Hoare 1992). The set of concerns Banta cites is notable for its reference not only to health benefits and risks, but also to social implications, a theme which can be traced back to the birth of the original technology assessment movement in the United States, where the concept emerged in the mid-1960s associated with

© Blackwell Publishers Ltd/Editorial Board 1997

PAPER 1

liberal political movements. Prominent in this early concept were notions of the unintended consequences and indirect effects of technologies such as industrial processing (O'Brien and Marchand 1982: 7). This model has been taken up within the programmatic discourse on HTA in the United Kingdom. Another high-profile expert and early proponent of assessment of medical technologies, Barbara Stocking, conducted case studies of expensive healthcare technologies in the United Kingdom. Then Director of the King's Fund Centre for Health Services Development, one of the United Kingdom's major independent research centres for research and consultancy on health services issues, she described technology assessment as including:

> the technical and clinical evaluation of a technology, as well as its economic, social and ethical implications (cited in Hoare, 1991: 1).

Rhetorically deployed concerns about these different dimensions of technology support a 'need' for HTA activities in the NHS. The grouping of 'concerns' of healthcare policymakers also promotes a particular model of the disciplines which might appropriately participate in it. In an article in the *Lancet* aimed at a medical audience, the current (1997) Director of NHS R&D explicitly makes a case for combining specialist disciplines:

> The need to reassemble these fragments and to define their limits is central to the successful incorporation of science into health care (Swales 1997: 1319).

He identifies the major topics of this science as being medical outcomes, cost and quality of life (1997: 1320). The second rhetorical strand is thus a programme to amalgamate diverse disciplines: multi-disciplinarity. Linked to multi-disciplinarity is the 'discovery' of common ground between disciplines via the dusting down and applauding of new 'founding fathers'. This is seen in the canonisation of individuals who according to HTA activists espoused and promoted experimentalism and the monitoring of the effects on patients of healthcare practitioners' work. Individuals who are being acclaimed in this rhetorical discourse include, notably, the epidemiologist Archibald Cochrane who is discussed below in relation to the role of epidemiology and methodology in HTA. For example, Cochrane has been commemorated recently by the publication of a collection of essays in his honour (Maynard and Chalmers 1997), and 'a large photographic portrait of Cochrane has pride of place in the reception area of our (academic) department' (author observation). Two other individuals have been elevated in this way. Ernest Codman, an American who in 1900 instigated a hospital system for monitoring the end results of surgical care, was praised by the Director of the NHS Centre for Reviews and Dissemination (Sheldon and Faulkner 1996). And the 18th-century surgeon John Hunter, credited within the surgical profession as being the first scientific surgeon, received a strikingly titled homage from the chairman of the SGHT (Irving 1993).

© Blackwell Publishers Ltd/Editorial Board 1997

PAPER 1

Appeals to common ground can be seen in the heart of the state-promoted definition of 'health technology' itself. What counts as technology? The meaning of 'technology' is indexical (Pinch *et al.* 1992), that is, it is related to the occasions and contexts of its use. It is also a metaphor. Some metaphors become more dominant than others, and thus can be interpreted as traces of the workings of power across communities, networks or disciplines (Leigh Star 1991: 52). Whereas earlier policymakers' discourse concerned expensive new medical equipment, health technology has been promoted in the programmatic discourse of the state-orchestrated HTA movement in the United Kingdom as a highly abstract, all-embracing metaphor. Far from being confined to high-cost high-technology equipment, it has been promoted by the NHS Executive (and this was a usage already common amongst healthcare policymakers in a number of other countries) as being:

> . . . deliberately defined as broadly as possible. It encompasses all methods used by health professionals to promote health, prevent and treat disease and improve rehabilitation and long term care. It includes the activities of the full range of health care professionals, the use of equipment and procedures, and the administration of pharmaceutical products (Department of Health 1995:8).

Health technology is the experimental matter in the laboratory of the NHS. It forms part of the discourse of the 'knowledge-based NHS'. The boundaries of health technology are being drawn very broadly, by the NHS Executive arm of the state.

It is possible to characterise in a number of ways the range of topics which fall within the purview of health technology assessment. In order to draw attention to the metaphorical and cross-disciplinary deployment of health technology terminology, I present below a typology which characterises technologies from the viewpoint of a healthcare practitioner, with examples mostly drawn from the HTA programme.

A typology of 'health technologies' from a health professional's perspective

Type of Technology	*Examples*
Information systems	Picture Archiving and Communication System (PACS)
Material artefacts	Drugs, bronchodilator, hearing aid
Organisations	Regionalisation (of intensive care), specialised versus local access service (for vascular surgery), primary care-based emergency centre
Interpreted techniques	Diagnostic technologies and screening tests
Technique-assisted interventions	Physiotherapy, laser treatment, coronary artery bypass graft, knee prosthesis
Interpersonal communication	Counselling, psychological treatment, health visitor domiciliary visiting

© Blackwell Publishers Ltd/Editorial Board 1997

PAPER 1

For Pinch *et al.* (1992), health technologies are 'social technologies' in that they are intended to change human behaviour. This notion can be extended by asking what sorts of behaviour might be changed. Here the targets of intended change are health professionals and healthcare policies. Examples would be modes of organisation of healthcare delivery, healthcare techniques, or practitioners' choice between alternative healthcare options. The majority of HTA topics focus upon a condition (such as stroke or low back pain) and its corresponding technologies (such as rehabilitation techniques or spinal surgery) (Department of Health 1995). Some topics include questions of organisational models for the utilisation of technologies, division of labour, job design, and skills. The National Co-ordinating Centre for HTA is charged with developing an inventory of existing health technologies and briefings on their effectiveness and cost-effectiveness (Research and Development Directorate 1995). This re-interpretation of the earlier principles of HTA is commensurate with the radical approach to creating a 'knowledge-based' health service which the state has promoted. The move from new hardware-embodied technologies to a more general notion of interventions having health effects is again evident.

Further rhetorical discourses take the form of 'horror stories'. First, identifiable primarily with the health care purchaser's perspective,[1] it has been a frequently quoted statistic that only some 15–20 per cent of health care interventions had a firm basis in research evidence (Hoare 1992). The editor of the *British Medical Journal* greeted the launch of the first annual report of the Standing Group on Health Technology with the words: 'Few decisions made in health services are made with good evidence' (Smith 1994). So a vast array of healthcare practices are being performed, in this account, without the 'evidence' to 'support' them. This has been countered by proponents of 'evidence-based medicine' amongst the medical professions who uphold the need for individual clinical expertise alongside a commitment to draw on the best available evidence in making care decisions (Sackett *et al.* 1996). This counter claim can be viewed as a version of the conventional medical professional strategy of appeals to a need for clinical freedom.

The second type of horror story takes the following form: unevaluated health technologies can be dangerous and a risk to health. This is illustrated by the construction of the need for assessment of health technologies by reference to previous health technologies which, allegedly unassessed, found their way into routine medical practice only for their deficiencies then to become apparent. The classic example, frequently quoted, is that of gastric freezing (Challah and Mays 1986, Department of Health 1995: Foreword), in retrospect a painful and ineffective treatment for duodenal peptic ulcers.[2]

There are thus several linked strands in the rhetorical repertoire at the disposal of proponents of HTA. A rhetoric of broad rational concerns, employing appeals to medical, social and economic values is linked to appeals to cross-disciplinary common ground; this is supported by a

© Blackwell Publishers Ltd/Editorial Board 1997

rhetoric of appeal to possible danger and risks from healthcare provided in the absence of an adequate evaluative knowledge base. These appeals are not specific to particular disciplines within HTA, and so they promote multi-disciplinary participation in it. Appeals are both to clinical and economic concerns and to a need to produce new healthcare knowledge. Health technology is a 'trans-disciplinary' concept, like other such notions 'metaphorically encompassing the several parts of material handled separately by specialised disciplines' (Good and Roberts 1993: 6). However, as the above interpretations suggest, these rhetorical resources are flexible in their potential use, and are contestable. This draws our attention to power relations at work in the negotiation by the various disciplines for positions in the multidisciplinary world of HTA.

Consideration of these persuasive discourses would be important in a sociological history of the emergence of HTA in the United Kingdom. Alongside them should be set an historical account of the 'official' policy development which identified a need for nationally co-ordinated HTA. This of course would be another, but different, rhetorical discourse. It is thus worth noting the appearance during the 1980s and more recently of committees and advisory groups, and reports from national bodies, which would doubtless form part of such an account. These include: a call from the Council for Science and Society in 1983 for evaluation of all expensive new medical technology; a Medical Research Council (MRC) committee dealing with research on health service delivery; a national Health Technology Assessment committee established in the Office of the Chief Scientist in 1987; a House of Lords Select Committee on Science and Technology which reported on 'Priorities in Medical Research' in 1988, noting a wide disparity between the research needs of the NHS and the activities of the MRC; and a government advisory committee reporting on 'Medical Research and Health' in 1993, explicitly supporting the concept of health technology assessment.

Research alliances and institutional co-ordination

Via HTA, the NHS is gaining access to the means of production of research knowledge, and knowledge-producing disciplines and institutions are gaining access to the laboratory of the NHS. Research alliances are emerging in interactions and negotiations between state orchestration, healthcare organisations and the disciplines of healthcare knowledge production. It is not possible here to analyse the full range of processes which constitute the organisational and disciplinary membership of HTA. In this section, therefore, I give examples of alliance-formation, co-ordination and centralisation which are evident. In doing so, I am aware that I am glossing over many of the tensions which undoubtedly exist around the state-promoted HTA project.

© Blackwell Publishers Ltd/Editorial Board 1997

PAPER 1

The participating constituencies in HTA are broad. The ways in which individuals and institutions become engaged in HTA networks include memberships of the SGHT and its advisory panels, location of co-ordinating bodies such as the National Co-ordinating Centre for Health Technology Assessment within academic institutions, involvement of 'acknowledged experts', often researcher-clinicians, in the process of defining assessment questions prior to funding, consultation within the NHS and more widely amongst representative groups such as medical royal colleges and the Patients' Association, and the advertising and award of assessment project funds. The SGHT and its advisory panels have multi-disciplinary membership, including healthcare purchasers and providers, clinicians and other health professionals, health economists, and individuals seen by members of the R&D Directorate as methodology specialists.

Closer alliances between academic researchers (from the disciplines of epidemiology, medical statistics, public health medicine, human sciences, psychology, sociology, social anthropology and economics), clinician-researchers and purchasing authorities (especially in the form of public health departments) are being developed. Novel forms of collaboration are arising between academic researchers and health service practitioners. Examples are the location of health service research centres in NHS hospitals, and formal collaborations between NHS public health departments and academic research centres.

At inter-institutional level, long-standing co-ordination between the MRC and the Health departments has been extended (Medical Research Council 1991). In 1992, health services research funded by the MRC was seen to need its own managerial Board, the Health Services and Public Health Research Board (HS&PHR) (Medical Research Council 1992, Medical Research Council 1995). The research councils, the Department of Health, the MRC HS&PHR Board and the SGHT are all linked by cross-memberships.

Thus there are multiple signs of new inter-institutional networks, alliances, co-ordination, and centralisation in the broad institutional development of national health technology assessment. The growth of co-ordination of health technology assessment activity attests to the orchestrating role being played by state agencies. However, the term 'health technology' has not been strongly embraced by academic producers of HTA knowledge – 'they generally prefer to describe their organisations, as we do, in different terms, such as "health services research" or "health care evaluation" or "health science" ' (author observation). This terminology can be interpreted as signalling an independent position in preserving a capacity for determining local research strategies without becoming over-dependent upon the NHS-driven agenda of HTA.

Given these organisational developments, the following section considers the relationships between the disciplines and methodologies of HTA.

© Blackwell Publishers Ltd/Editorial Board 1997

PAPER 1

Multidisciplinarity, experimentalism and accuracy of projection

In health technology assessment, projection – the extrapolation of findings from the test environment to the real world – takes a number of forms depending on the methods adopted by the different disciplines of healthcare knowledge. The current incarnation of HTA in the United Kingdom comprises a multi-disciplinary approach drawing upon both quantitative and qualitative methodologies, with a distinctive mixture of disciplines and skills. The research commissioning groups of the HTA programme and the MRC have a preference for research teams made up of certain disciplines. Informal discussion with HTA administrative support staff confirms that projects lacking clinical specialists, elements of economic evaluation or access to expertise in medical statistics are unlikely to be supported. In seeking tenders for a National Co-ordinating Centre to manage and support national HTA activity, the NHS Executive sought a single centre capable of providing: 'Multidisciplinary scientific skills . . . [which] should include clinical (medical, nursing or therapy) and epidemiology, health economics, sociology and other social science disciplines' (Research and Development Directorate 1995).

As Pickstone (1993) argues, until very recently the experimentalism of basic biomedical research has been the primary world of medical research, clinical medicine a 'poor, confused imitation' (1993: 452), and research has been discipline-driven. The impact of scientific research on health service delivery was relatively indirect and long-term (Austoker 1989, Booth 1989, Medical Research Council 1995a). With the arrival of HTA, the organisational changes described above have been implicated in a re-alignment and shifting of power across and between professional disciplines, and a disruption of the previously strong boundary between high status experimental biomedical science and the scientifically lower status clinical research.

Exponents of epidemiology and medical statistics have been the leading disciplinary activists in linking experimental science and clinical practice. The primary form of projection ('generalisability' in the language of HTA), promoted as the vehicle of this linking, is the randomised controlled trial (RCT). In principle the RCT enables valid generalisations to be made because it eliminates sources of bias. In promoting the RCT as the preferred methodology for what he called 'applied medical research', Archibald Cochrane – in what proponents of HTA now regard as a core text in the application of the experimental method to health service research (Cochrane 1971) – aligned clinical research with the experimentalism of the more prestigious biomedical sciences. Skills in experimental research design for causal analysis of disease patterns are strongly claimed within the disciplinary boundaries of epidemiology. Its practitioners have advanced their influence, applying the same methodological concepts to questions of technological

© Blackwell Publishers Ltd/Editorial Board 1997

cause and health effect within health services. Cochrane's name is now used rhetorically to symbolise an international search for bias-free evidence about the effects of health technologies, gained by the collection of results *only* from RCTs. This is known as the Cochrane Collaboration.

Further entrepreneurialism in the discourse of epidemiology can be seen in the 'systematic review', a second methodology aimed at identifying bias-free, generalisable relationships between health technologies and their effects. In strong programmatic statements from policy centres for HTA, this assessment method is also construed as a scientific method owing its principles to epidemiology. The method has been enshrined in the Cochrane Collaboration handbook on 'Preparing and maintaining systematic reviews' (The Cochrane Collaboration 1994). Its introduction, entitled 'The science of reviewing research', begins:

> The scientific principles that apply to epidemiological surveys apply also to systematic reviews: a question must be posed, a target population of information sources identified and accessed, appropriate information obtained from that population in an unbiased fashion, and conclusions derived. Often statistical analysis can help in reaching conclusions (1994: VI-1).

Here the information sources referred to are the results of existing, completed RCTs. The epitome of the method, referred to above as 'statistical analysis', is meta-analysis, the statistical summarisation of the data from several empirical studies deemed to be comparable (Dickersin and Berlin 1992). The formal commitment to elimination of bias has resulted in an attempt to construct the systematic review also as an experimental scientific procedure, along the lines of the randomised controlled trial.

However, a purely quantitative approach to systematic reviews is implicitly criticised by the NHS' national centre for this activity in the formal guidance it has issued (NHS Centre for Reviews and Dissemination 1996). It commends an approach which:

> considers all the results taking into account not only the methodological rigour, and therefore reliability of these studies, but also helping to highlight and explore differences. A qualitative analysis of the evidence is therefore an essential step in the assessment of the effectiveness of a health technology (1996: 46).

The document bemoans an exclusive focus upon the 'narrow' methods of statistical pooling of data. These passages suggest that rather than being a matter – for those with appropriate statistical expertise – of cumulation of results from separate studies, the review requires qualitative judgements to be made about the homogeneity, or otherwise, or studies. Thus what is at issue is the basis on which accurate projections from 'results' of experiment (or combined experiments) might be made: in other words of how bias or subjectivity might best be controlled.

© Blackwell Publishers Ltd/Editorial Board 1997

PAPER 1

Thus the contribution of epidemiological discourse to the construction of the HTA agenda powerfully promotes scientific experimental methodology. However, even around this formal programmatic agenda, conflict is in evidence, for example over the status of the RCT (Black 1996) and of meta-analysis (Eysenck 1994) as methodologies of knowledge production.

Epidemiology also figures in the framing of the national HTA agenda through its focus upon the prevalence of medical conditions. Whereas historically clinical science has been greatly occupied with relatively rare conditions (Frankel 1989), epidemiological analysis of the distribution of remediable conditions has paved the way for a re-distribution of clinical scientific research attention to some of the most common afflictions. And, as noted earlier, prevalence is one of the criteria explicitly used in the formalised priority-setting processes of the SGHT in identifying topics for assessment.

There is increasing interest on the part of medical practitioners in so-called 'qualitative' research methods, coming from within the arena of clinical research. For example, there has been a series of editorials, appearing in the most prestigious medical journals, including the *British Medical Journal*, presenting the case for qualitative methods in healthcare research, and analysing the interface between qualitative methods and clinical experimental research designs. Programmatic statements from a medical viewpoint shape the relationship between qualitative and experimental research. Qualitative research is held to be especially relevant for hypothesis generation, explanation of experimental or quantitative findings and understanding of factors affecting implementation of research results (Jones 1995). Papers written by sociologists have also been published in medical journals, outlining some of the main qualitative research methods which can be applied in healthcare research (*e.g.* Mays and Pope 1995).

The definition of medical symptoms and outcomes of treatment for research purposes has, historically, been largely the preserve of medical specialists performing clinical research. In the new model of health technology assessment, the boundaries of this territory appear less clear. New contributions are being made predominantly by research techniques derived from sociology and psychology. This is embodied in the increasing development of measures of health status and health 'outcomes' measures, in particular measures informed by our own experience, as health service users, of our symptoms and health states. Patient-derived measures of the 'outcomes' of 'interventions' – are being increasingly incorporated into the array of measures used in experimental tests of health technologies. The novelty is that they are constructed from 'qualitative' research, usually interviews, into our experience of symptoms. A typical example is the development of an 'instrument' to assess the outcomes of total hip replacement surgery – one of the highest priority technologies identified in the early HTA programme – from patients' perspectives: 'Many questionnaires . . . intended for general use . . .

© Blackwell Publishers Ltd/Editorial Board 1997

may be . . . insensitive to the specific changes in health produced by a particular intervention . . . Questionnaires are needed therefore which address patients' perception of a single disease entity . . . ' (Dawson *et al.* 1996: 185). These authors' disciplines combine sociology and orthopaedic surgery. Psychometric statistical methods in particular are being used to 'validate' such measures. Validation in this sense is another form of 'projection', akin to the testing of measurement instrumentation technology in laboratory settings. The construction and validation of such measures are presented as processes establishing the scientific accuracy of the representation of patients' subjective experience. This enterprise results in calls for the 'standardisation' of such measures (McDowell and Jenkinson 1996), because the generalisability of results depends partly upon having reliable methods of enabling the effects of the application of health technologies to produce 'traces', to use the SSK term described above, which can be apprehended, quantified and interpreted by statistical analysis.

The proliferation of measures of health status and quality of life extends the definition of what counts as an effect of clinical intervention, and thereby re-shapes the meanings which health and illness might have in the discourse of healthcare policymaking. This process is being strongly supported by the involvement of proponents of health economics in HTA. Health economics has established itself as a core discipline in the health technology assessment repertoire. Economic evaluations are increasingly a part of clinical trials, and it is policy to encourage this development both in the HTA programme and in the MRC. The performance of economic evaluations alongside clinical trials is promoted by professional economists within the discourse of NHS R&D policymaking as a matter for professional health economists rather than as an activity which can be accomplished by 'doing-it-yourself' (Drummond 1994: Foreword). Assessing costs and benefits in health technology assessments enhances the perceived need for health status measurement tools, because changes in health status, measured either as survival or quality of life, are the 'effects' against which differences in costs between alternative technologies are evaluated in economic analysis. Ashmore *et al.* (1989) have drawn attention to the processes by which economists, in interaction with others, construct quality of life measures (QALYs) as representations of generalised public preferences for states of health. Here it can be said that this too, like the discourse of epidemiology and medical statistics, is a methodology for controlling bias and attempting to take a viewpoint which might somehow make it possible for judgements based on social values to be expressed on behalf of a population. This is explicit in the texts of health economists who state that they perform their analysis 'from the perspective of the NHS' or from the 'societal perspective'. Utilitarian philosophies are preferred. But here too, within the expert system of health economics, conflict and debate can be found. It has been questioned, for example, whether the focus upon the *health*

© Blackwell Publishers Ltd/Editorial Board 1997

PAPER 1

benefits of health technology is too narrow. Health economics might fail to take account of other benefits or costs which, as citizens rather than as patients, we might derive from healthcare (Ryan and Shackley 1995).

In interdisciplinary fields, expertise from more than one discipline comes together to seek common goals (Good and Roberts 1993). Within the constellation of HTA disciplines, the clinical sciences and health economics are the ones most clearly delineated. Some interdependencies suggest a blurring of boundaries between the constituent HTA disciplines, reflecting research alliances which have been formed. These include, for example, an interdependence, centred on outcome measurement, between the disciplines promoting expertise in interviewing and questionnaire design and the corresponding clinical specialisms and body systems. A further interdependence may be discerned around cost-effectiveness between the disciplines concerned with health status and quality of life measurement, and health economics. On the other hand, conflicts and tension exist between disciplinary and inter-disciplinary rhetorics of HTA. For example, 'adherents of qualitative methods commonly criticise HTA's emphasis on the randomised controlled trial' (author observation); a public health consultant reports that, in a leading teaching hospital, 'powerful clinical scientists' believe that 'the HTA side is contentious, as there is an obvious link to rationing . . . [though] . . . formalising an HTA process in the NHS Trust would certainly ring the management bell' (Ayres, P. personal communication); and leading voices in the formal programme of health economics criticise the clinical professions' agenda of applying scientific evidence to healthcare practice (Maynard 1997).

The rhetorical claims of health technology assessment constitute the activity of health technology assessment as having the legitimacy of science at a time when novel alignments of professional disciplines are developing, when professional identities and disciplinary authorities exist in tension with each other, and when research communities are being invited to respond to a culture of health service 'needs'. Those involved in the disciplines of HTA perform experiments in and around the laboratory of the NHS. These experiments use measurement instruments to test health technologies. In these tests projections are made from the laboratory world to the 'real-world' of health technology in healthcare practice. The different disciplines of HTA are able to mobilise their technical, methodological and rhetorical resources to different degrees in constructing the HTA agenda.

The assessment agenda

The rhetorics of HTA discussed above might be taken as a guide to the assessment agenda which we would expect to find embodied in the national HTA programme: 'concerns' for benefit and risks to health, for cost and for

© Blackwell Publishers Ltd/Editorial Board 1997

PAPER 1

social implications. We can examine the HTA programme to see what form these concerns might take, shaped by and shaping the organisational and disciplinary processes described above.

Turning to the aspects of health technologies which are typically assessed, it is clear that 'effectiveness' and 'cost-effectiveness' are the two hinges on which the formal NHS HTA world-view swings. Strong statements confirming this abound: for example, 'The aim . . . is to help those conducting and funding trials to ensure that their work tackles the issues of cost-effectiveness as well as effectiveness whenever it is feasible to do so' (Drummond 1994). The MRC approach is essentially the same. Its studies 'measure effectiveness and efficiency in relation to health outcomes' (Medical Research Council 1995b: 107).

Turning first to costs, in what ways are 'cost implications' being constructed in the HTA agenda? A concern with economic implications might suggest that encouragement of potentially income-generating technologies might be high on the agenda, or that high-technology high-cost equipment would be of priority for assessment, but the voice of commercial exploitation is noticeable by its absence from the formal discourse and practices of national HTA, in spite of policymakers' early plans to create collaboration between the NHS R&D strategy and industry (Department of Health 1993), and attempts by commercial interests to stimulate alliances via national 'research foresight' in health-related technologies.[3] It appears that the strength of the clinical and academic economic discourses on benefits and costs of health technologies have resulted in an institutionally bounded, circumscribed agenda which is relatively impervious to commercial networks and interests. This is not to say that practitioners of HTA research, for example, might not transact with both national HTA and commercial organisations in seeking to negotiate favourable conditions for themselves locally, but the clinico-economic discourse of health costs and benefits effectively excludes commercial interests.

Concerns for risks to health were part of the policymakers' early rhetoric of a need for HTA. Examination of the topics of national HTA suggests that *absolute* safety and risks to health are not the major issues in the HTA agenda. Rather, safety and health risks are aspects of outcome measurement in assessments where different technologies are being compared. Of course, in the case of pharmaceuticals and 'medical devices', statutory agencies exist with responsibility to ensure safety: the Committee on the Safety of Medicines, and the Medical Devices Agency in the Department of Health. But the over-riding focus upon comparative effectiveness and cost-effectiveness in the HTA disciplines projects quite different definitions of risk. The major dichotomy evident here might be characterised, if somewhat glibly, as one between risk to the public health and risk to the public purse. Like processes of risk assessment (Carter 1995), health technology assessments construct boundaries between safety and danger. However, unlike

© Blackwell Publishers Ltd/Editorial Board 1997

PAPER 1

discourses of risk to the healthy body, the discourses of health technology assessment amalgamate concerns with transgression of boundaries of health with concerns regarding transgression of boundaries of financial budgets. They thus generalise and extend the notion of risk. The influence of epidemiological and economic perspectives results in the construction of risk at the level of society as a whole.

In HTA health benefits and costs are aspects of both clinical and economic discourses. The agenda of health status measurement and quality of life measurement has been referred to above. 'Costs' and 'benefits', like 'health technology', themselves develop metaphorical meanings. Discourses of clinical research, incorporating assessments of the effectiveness or efficacy of healthcare technologies derived in interactions with patients, refer to effects of technologies such as complications, mortality, and patient 'outcomes'. Health economics constructs costs and benefits in both financial terms and in terms of the health outcomes of quality of life and mortality. Experimental science, as embodied in the epidemiological and statistical methods of HTA, requires standard, reproducible forms of measurement so that stable comparisons between different patients or groups of patients at different times may be engineered. The other disciplines of HTA deploy techniques which enact this requirement. Economists' cost-effectiveness analyses frequently proceed by constructing a model of a typical health district or general practice population in order to examine cost and benefit implications (*e.g.* Bachmann and Nelson 1996). This represents another form of testing of health technologies, in this case projecting from simulations to actual populations.

Turning to the shaping of 'social implications', it has been argued by leading HTA proponent David Banta that: 'The social implications of a new or existing technology can be the most challenging and difficult aspects of evaluation . . . the methods for assessing social implications are relatively undeveloped . . . ' (Banta and Luce 1993: 132). The implication is that when methods are better developed, HTA will be able to address these issues better. However, it could be argued that if HTA practitioners are to focus on a scientific agenda built around generalisability, elimination of bias, and the representation of a form of aggregated public interest, this would preclude examination of substantive social and ethical issues. Economic evaluation is leading to consideration of ethical issues in healthcare provision, but this is focused upon questions of choice between alternative services (the rationing debate), rather than the examination of the substantive social and ethical issues arising with particular technologies such as, for example, animal to human organ transplantation. 'Xenotransplantation was considered by the Acute Sector Advisory Panel of the SGHT, but not accorded priority, in the first year of operation of the national HTA programme' (author observation). It may be that mechanisms outside health technology assessment, such as the investigative activity of the mass media and specially convened

© Blackwell Publishers Ltd/Editorial Board 1997

PAPER 1

groupings of state-identified experts (Advisory Group on the Ethics of Xenotransplantation 1996), will be the major means by which those technologies, with apparently broad social and ethical implications, are addressed within society. This would be despite the HTA claim to this agenda.

The discourses of epidemiology, statistics and economics construct the population's health status as the indirect aggregate of many individual outcomes of health technology. The current HTA agenda allows *directly* for the patient's voice in closely limited ways. Users' representatives such as the College of Health are included in the consultation process on priority topics for assessment. A few health technology assessment topics include lay or patient attitudes or actions as a major strand in the assessment question. One example is the study of the effect on treatment rates of patient participation in care decisions, using interactive video disks containing personal testimonies of former patients and information about the probability of different outcomes following surgical treatments (NHS Executive 1996). As in this case, projects which do incorporate users' voices contain them within the design of the project. They give priority to measuring the outcomes of that participation by the experimental methods of bias-elimination and in terms of aggregated health outcomes. Multidisciplinary HTA defines the healthcare user population as the object of more or less cost-effective healthcare interventions.

Lay incursion into the expert domains of design, conduct and interpretation of assessments has occurred with AIDS activists in the United States (Epstein 1995). Such engagement is not evident in HTA. And, as Elston (1991: 82–3) has noted, experimental treatments and illnesses where curative medical science has little to offer are specific, limited areas perhaps not representative of the overall capacity of healthcare consumers to challenge the authority of medical science.

The embryonic agenda of HTA, considered at this general level, is not about what the sociology of scientific knowledge is most often concerned with, the processes of building particular facts. It is, rather, about preparing the ground which makes it possible for some sorts of facts – rather than others – to grow, what sorts of facts are going to be important. This is perhaps the reason why so much of the early funding in the national HTA programme has been put into study of the methodology of HTA itself. A reflexive emphasis on methodology reflects and facilitates the engagement of the different professional disciplines 'taking part' in HTA. This is also a debate about what is the appropriate instrumentation for the laboratory of the NHS. It again reflects the tension between disciplinarity and interdisciplinarity in HTA. While the major emphasis has certainly been upon the bias-reduction methodologies of the controlled trial and meta-analysis, no exclusive coalition can be said to have emerged. The voices of non-experimental assessment methodologies and non-population oriented

© Blackwell Publishers Ltd/Editorial Board 1997

PAPER 1

disciplines have also negotiated places – at least for the time being – in the emerging agenda. Tensions are evident, for example, in the small number of studies funded on topics such as the role of qualitative methods and action research in HTA (NHS Executive 1996). 'Similar conflicts occur routinely in informal discussions about the respective status of the randomised controlled trial in comparison to non-experimental assessment methods' (author observation). The effect of such internal ambiguities is to enhance the authority of health technology assessment discourse and its participating voices by retaining control over its agenda. This indeterminacy in turn can be seen as a boundary-constructing control strategy (Jamous and Peloille 1977, Gieryn 1983).

Conclusion

'Health technology' is an abstract metaphor shorn of specific disciplinary or methodological signification. As the quotation at the beginning of this chapter suggests, unresolved tensions between disciplines can exist in the interpretative elasticity which this creates. HTA may or may not be interpreted by different interests as an extension of medical research. The trans-disciplinary concept of health technology has been promoted by the state, and the healthcare knowledge disciplines orient themselves to it in different ways. The metaphor enables these diverse actors including HTA research policymakers and research knowledge producers to shape their activities in diverse ways in and through the production of healthcare knowledge. The workings of power are evident within and between the disciplinary discourses of HTA. The health technology metaphor serves the state by enabling the participation of the agencies of knowledge production required by the vision of a knowledge-based health service. The elasticity of HTA's metaphors enables uneasy partnerships to exist between disciplines and methodologies, such as qualitative and experimental methods, and clinical science and sociology.

In health technology assessments, such as clinical trials, the application of technologies in healthcare provision and the testing of them are conflated for a proportion of health care users. Patients are entered into trials of alternative treatments, and models are built to project the effect of health technologies on real patient populations. Thus HTA defines the NHS, its patients and health professionals, as a research laboratory for the field-testing of the effectiveness and cost-effectiveness of healthcare technologies. The inclusion of 'patient preferences' in health technology experiments (*e.g.* Torgerson *et al.* 1996) suggests that, as consumers, our intimate feelings about treatment or other healthcare choices can be incorporated into health technology. If the exercise of choice itself might affect health, then personal healthcare choices become part of the material from which some health technologies are manufactured.

© Blackwell Publishers Ltd/Editorial Board 1997

PAPER 1

The different combinations of the variable relations between experiments and laboratories in health technology assessment, such as the multi-centre randomised controlled trial combined with cost-effectiveness analysis, pose methodological and analytic challenges for the sociology of healthcare knowledge. For example, HTA projects construct *ad hoc* laboratories in the healthcare environment, which are not marked by separate physical space, and resources for experimental work are widely dispersed between organisations and disciplines. Analytically, a challenge is presented by the perception that, in HTA, health professionals and patients themselves constitute part of health technologies and the measurement instrumentation employed in experiments.

Health technology assessment provides a locus in which relationships between rational science, scientific uncertainty, trust and authority may be examined. Its processes constitute and define areas of expertise characterised by scientific uncertainty. The negotiable 'boundaries' between acceptable and non-acceptable risk, conventionally if implicitly conceived of as lines of demarcation, might better be re-conceptualised as broad bands where questions of comparative risk to health and budgets are held, pending expert, political and public actions related to scientific evidence. Laboratories harbour uncertainty while engaged in projects to reduce it. National health technology assessment policy exemplifies a rational-planning approach to the dynamics of science and health care. From the perspective of the sociology of scientific knowledge, it is an activity which is open to analysis as a scientific practice, with associated contestable knowledge claims. Contestability of expertise is frequently taken to be one of the defining features of health-related issues in contemporary societies, and it is associated with a decline in trust of medical authorities (Gabe and Bury 1996). However, it is not obvious that this is a clear, unopposed trend across healthcare. The HTA movement, in its engagement of expert disciplines, its research alliances, its encompassing metaphors, its containment of uncertainty, its consultation of public voices, its bias-elimination methodologies, and its use of a rational aggregated voice speaking for the good of the public health, might be seen as part of a dialectical process involving decline in traditional medical authority and a reconstruction of a new framework of scientific authority. We are invited to place trust again in this new form of scientific expertise. It is paradoxical that this invitation to place our trust in science is associated with the radical questioning of more established forms of healthcare knowledge which the health technology assessment movement promotes.

Acknowledgements

I should like to thank the editor and the anonymous referees for comments on previous versions of this chapter, and Caroline Brown for comments and advice.

© Blackwell Publishers Ltd/Editorial Board 1997

PAPER 1

Notes

1 Since 1991, the NHS has been organised by a division between health authorities – 'purchasers' or 'commissioners' – which contract for health services with 'providers' (called NHS Trusts) in the form of hospitals and community health service agencies.
2 The use of the example of gastric freezing to support the case for assessment is something of a selective reading of the assessment history of this particular operation. It was invented and unevenly diffused primarily in the USA, after testing with apparent success on animals. At least thirty-six trials of the procedure were published in American medical scientific journals during the 1960s, the early studies being at least 'qualified favourable'. With hindsight, many of these studies are regarded as having methodological flaws. The negative results of the only multicentre randomised controlled trial were not published until the procedure had fallen out of favour (Fineberg 1979). Gastric freezing spread in spite of assessments, rather than because of lack of assessment.
3 For example, a multi-interest conference 'Planning national research priorities: foresight and the science base in wealth and health creation', held in Cambridge in 1994, organised and supported by SmithKline Beecham Pharmaceuticals and attended by the NHS Director of Research and Development.

References

Advisory Group on the Ethics of Xenotransplantation (1996) *Animal Tissue into Humans*. London: Department of Health.

Ashmore, A., Mulkay, M. and Pinch, T. (1989) *Health and Efficiency: a Sociology of Health Economics*. Milton Keynes: Open University Press.

Austoker, J. (1989) Walter Morley Fletcher and the origins of a basic biomedical research policy. In Austoker, J. and Bryder, L. (eds) *Historical Perspectives on the Role of the MRC*. Oxford: Oxford University Press.

Bachmann, M. and Nelson, S. (1996) *Screening for Diabetic Retinopathy: a Quantitative Overview of the Evidence, Applied to the Populations of Health Authorities and Boards*. Bristol: Health Care Evaluation Unit, Department of Social Medicine, University of Bristol.

Banta, H.D. and Gelijns, A. (1987) Health care costs: technology and policy. In Schramm, C.J. (ed) *Health Care and its Costs*. New York: W.W. Norton and Co.

Banta, H.D. and Luce, B.R. (1993) *Health Care Technology and its Assessment: an International Perspective*. New York: Oxford University Press.

Bartley, M. (1990) Do we need a strong programme in medical sociology?, *Sociology of Health and Illness*, 12, 371–89.

Beer, G. and Martins, H. (1990) Introduction, *History of the Human Sciences*, 3, 163–75.

Berg, M. (1995) Turning a practice into a science: reconceptualising postwar medical practice, *Social Studies of Science*, 25, 437–76.

Black, N. (1996) Why we need observational studies to evaluate the effectiveness of health care, *British Medical Journal*, 312, 1215–18.

© Blackwell Publishers Ltd/Editorial Board 1997

Booth, C.C. (1989) Clinical research. In Austoker, J. and Bryder, L. (eds) *Historical Perspectives on the Role of the MRC*. Oxford: Oxford University Press.

Carter, S. (1995) Boundaries of danger and uncertainty: an analysis of the technological culture of risk assessment. In Gabe, J. (ed) *Medicine, Health and Risk. Sociology of Health and Illness Monograph*. Oxford: Blackwell.

Casper, M.J. and Berg, M. (1995) Constructivist perspectives on medical work: medical practices and science and technology studies, *Science, Technology and Human Values*, 20, 395–407.

Challah, S. and Mays, N.B. (1986) The randomised controlled trial in the evaluation of new technology: a case study, *British Medical Journal*, 292, 877–9.

Cochrane, A.L. (1971) *Effectiveness and Efficiency: Random Reflections on Health Services*. London: Nuffield Provincial Hospitals Trust.

Collins, H. (1984) Researching spoonbending: concepts and practice of participatory fieldwork. In Bell, C. and Roberts, H. (eds) *Social Researching: Politics, Problems, Practice*. London: Routledge and Kegan Paul.

Committee on Technology and Health Care, Assembly of Engineering, National Research Council and Institute of Medicine (1979) *Medical Technology and the Health Care System: a Study of the Diffusion of Equipment-embodied Technology*. Washington, DC: National Academy of Sciences.

Dawson, J., Fitzpatrick, R., Carr, A. and Murray, D. (1996) Questionnaire on the perceptions of patients about total hip replacement, *Journal of Bone and Joint Surgery*, 78-B, 185–90.

Department of Health (1993) *Research for Health*. Leeds: Department of Health.

Department of Health (1995) *Report of the NHS Health Technology Assessment Programme 1995*. London: Department of Health.

Dickersin, K. and Berlin, J.A. (1992) Meta-analysis: state-of-the-science, *Epidemiologic Reviews*, 14, 154–76.

Drummond, M. (1994) *Economic Analysis alongside Controlled Trials*. London: Department of Health.

Elston, M.A. (1991) The politics of professional power. In Gabe, J., Calnan, M. and Bury, M. (eds) *The Sociology of the Health Service*. London: Routledge.

Epstein, S. (1995) The construction of lay expertise: AIDS activism and the forging of credibility in the reform of clinical trials, *Science, Technology and Human Values*, 20, 408–37.

Eysenck, H. (1994) Meta-analysis and its problems, *British Medical Journal*, 309, 789–92.

Fineberg, H.V. (1979) Gastric freezing – a study of diffusion of a medical innovation. In Committee on Technology and Health Care *Medical Technology and the Health Care System: a Study of the Diffusion of Equipment-embodied Technology*. Washington, DC: National Academy of Sciences.

Frankel, S. (1989) The natural history of waiting lists – some wider explanations for an unnecessary problem, *Health Trends*, 21, 56–8.

Gabe, J. and Bury, M. (1996) Halcion nights: a sociological account of a medical controversy, *Sociology*, 30, 447–69.

Gieryn, T.F. (1983) Boundary-work and the demarcation of science from non-science: strains and interests in professional ideologies of scientists, *American Sociological Review*, 48, 781–95.

Good, J.M.M. (1993) Quests for interdisciplinarity: the rhetorical constitution of

© Blackwell Publishers Ltd/Editorial Board 1997

PAPER 1

social psychology. In Roberts, R.H. and Good, J.M.M. (eds) *The Recovery of Rhetoric: Persuasive Discourse and Disciplinarity in the Human Sciences*. London: Bristol Classical Press.

Good, J.M.M. and Roberts, R.H. (1993) Introduction: persuasive discourse in and between disciplines in the human sciences. In Roberts, R.H. and Good, J.M.M. (eds) *The Recovery of Rhetoric: Persuasive Discourse and Disciplinarity in the Human Sciences*. London: Bristol Classical Press.

Hoare, J. (1992) *Tidal Wave: New Technology, Medicine and the NHS*. London: King's Fund Centre.

Hoggett, P. (1991) A new management in the public sector, *Policy and Politics*, 19, 243–56.

Hughes, D. and McGuire, A. (1992) Legislating for health: the changing nature of regulation in the NHS. In Dingwall, R. and Fenn, P. (eds) *Quality and Regulation in Health Care: International Experiences*. London, New York: Routledge.

Irving, M. (1993) *Hunter's Baton*. Oration delivered on the Bicentenary of the death of John Hunter. London: Royal College of Surgeons of England. 14 September 1993.

Jamous, H. and Peloille, B. (1977) Professions or self-perpetuating systems? Changes in the French university hospital system. In Jackson, J. (ed) *Professions and Professionalisation*. Cambridge: Cambridge University Press.

Jennett, B. (1986) *High Technology Medicine: Benefit or Burden?* Oxford: Oxford University Press.

Jones, R. (1995) Why do qualitative research?, *British Medical Journal*, 311, 2.

Knorr Cetina, K. (1992) The couch, the cathedral, and the laboratory: on the relationship between experiment and laboratory in science. In Pickering, A. (ed) *Science as Practice and Culture*. Chicago: University of Chicago Press.

Latour, B. (1987) *Science in Action*. Milton Keynes: Open University Press.

Latour, B. and Woolgar, S. (1979) *Laboratory Life*, Princeton, NJ: Princeton University Press.

Leigh Star, S. (1991) Power, technology and the phenomenology of conventions: on being allergic to onions. In Law, J. (ed) *A Sociology of Monsters: Essays on Power, Technology and Domination*. London: Routledge.

MacKenzie, D. (1989) From Kwajalein to Armageddon? Testing and the social construction of missile accuracy. In Gooding, D., Pinch, T. and Schaffer, S. (eds) *The Uses of Experiment: Studies in the Natural Sciences*. Cambridge: Cambridge University Press.

Macleod, R. and Rehbock, P.F. (1994) *Darwin's Laboratory: Evolutionary Theory and the Natural History of the Pacific*. Honolulu: University of Hawaii Press.

Mays, N. and Pope, C. (1995) Rigour and qualitative research, *British Medical Journal*, 311, 109–12.

Maynard, A. (1997) Evidence-based medicine: an incomplete method for informing treatment choices, *Lancet*, 349, 126–8.

Maynard, A. and Chalmers, I. (1997) *Non-random Reflections on Health Services Research*. London: BMJ Publishing Group.

McDowell, I. and Jenkinson, C. (1996) Development standards for health measures, *Journal of Health Services Research and Policy*, 1, 238–46.

Medical Research Council (1991) *Annual Report 1990–91*. London: Medical Research Council.

© Blackwell Publishers Ltd/Editorial Board 1997

PAPER 1

Medical Research Council (1992) *Annual Report 1991–92*. London: Medical Research Council.

Medical Research Council (1995a) *Research Developments Relevant to NHS Practice, Public Health and Health Departments Policy*. London: Medical Research Council.

Medical Research Council (1995b) *Scientific Strategy*. London: Medical Research Council.

NHS Centre for Reviews and Dissemination (1996) *Undertaking Systematic Reviews of Research on Effectiveness: CRD Guidelines for Those Carrying Out or Commissioning Reviews*. CRD Report 4. York: NHS Centre for Reviews and Dissemination, University of York.

NHS Executive (1996) *Report of the NHS Health Technology Assessment Programme 1996*. Leeds: NHS Executive.

O'Brien, D.M. and Marchand, D.A. (1982) Politics, technology, and technology assessment. In O'Brien, D.M. and Marchand, D.A. (eds) *The Politics of Technology Assessment: Institutions, Processes, and Policy Disputes*. Massachusetts, Toronto: D.C. Heath.

Pickstone, J.V. (1993) Ways of knowing – towards a historical sociology of science, technology and medicine, *British Journal for the History of Science*, 26, 433–58.

Pinch, T. (1993) 'Testing – one, two three . . . testing!': Toward a sociology of testing, *Science, Technology and Human Values*, 18, 25–41.

Pinch, T., Ashmore, M. and Mulkay, M. (1992) Technology, testing, text: clinical budgeting in the UK National Health Service. In Bijker, W.E. and Law, J. (eds) *Shaping Technology/Building Society*. Cambridge, MA, MIT Press.

Porter, T.M. (1995) *Trust in Numbers: the Pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press.

Research and Development Directorate, NHSE (1995) *Management and Support for the NHS Health Technology Assessment (HTA) Programme*. Unpublished tender specification document.

Ryan, M. and Shackley, P. (1995) Assessing the benefits of health care: how far should we go?, *Quality in Health Care*, 4, 207–13.

Sackett, D.L., Rosenberg, W.M.C., Muir Gray, J.A., Haynes, R.B. and Richardson, W.S. (1996) Evidence based medicine: what it is and what it isn't, *British Medical Journal*, 312, 71–2.

Sheldon, T. and Faulkner, A. (1996) Vetting new technologies, *British Medical Journal*, 313, 508.

Smith, R. (1994) Towards a knowledge based health service, *British Medical Journal*, 309, 217-18.

Swales, J.D. (1997) Science in a health service, *Lancet*, 349, 1319–21.

The Cochrane Collaboration (1994) *Handbook Section VI: Preparing and Maintaining Systematic Reviews*. Oxford: Cochrane Collaboration.

Torgerson, D.J., Klaber-Moffet, J. and Russell, I.T. (1996) Patient preferences in randomised trials: threat or opportunity? *Journal of Health Services Research and Policy*, 1, 194–7.

Weatherall, D. (1995) *Science and the Quiet Art: Medical Research and Patient Care*. Oxford: Oxford University Press.

Webster, A. (1994) University-corporate ties and the construction of research agendas, *Sociology*, 28, 123–42.

© Blackwell Publishers Ltd/Editorial Board 1997

PAPER 1

PAPER 2

Originally published 2002

Author: A. Faulkner

Casing the joint - the material development of artificial hips.

Published in: Ott K, Serlin D & Mihms S (eds.) Artificial Parts, Practical Lives - modern histories of prosthetics. pp199-226.

Acknowledgment: New York University Press for permission to reproduce this chapter.

7

Casing the Joint

The Material Development of Artificial Hips

Alex Faulkner

INTRODUCTION

Hundreds of thousands of people worldwide walk with the hidden aid of entirely artificial hips. Many have had artificial replacements installed for both of their hip joints. These devices, implanted by orthopedic surgeons in risky and expensive operations, restore locomotor function and reduce pain from arthritic or otherwise damaged joints. Indeed, artificial hips are perceived to be one of the success stories of modern technological surgery. Most people who function physically with these invisible aids, however, are unaware of their material composition, the process by which they have been manufactured, or even their brand name. For the user, their success is defined by their functionality: as long as it works and it reduces pain.

Today, total hip prostheses, as orthopedic surgeons and manufacturers call them, are one of the major products of multinational companies that specialize in medical technology. In contemporary medicine, designs and materials frequently change as innovating surgeons and manufacturers seek improved performance, a broader range of potential implantees, and improved profitability for their devices. Innovation in materials and designs should thus be understood in the context of a commercial environment as well as one in which medical practitioners and patients seek technologies to alleviate pain and improve physical functioning. This essay discusses the various materials used in modern devices that replace the hip joint completely. By tracing the range of ma-

terials used as the prosthetic technology has evolved, the essay will outline the materials used in the different types of designs produced today.

In medical terminology, the formation of an artificial joint between bones is known as arthroplasty. The hip joint is essentially a ball-and-socket joint, the "ball" being the rounded head of the thigh bone (femur), the socket being a cavity in the hip bone (acetabulum) itself. Human implant materials, to be successful, must exist in the tissue of the human body without causing adverse reaction, either to the tissue or to the materials. This means that the search for suitable materials is in part a search for "inert," biologically compatible materials as well as materials that will withstand the forces exerted on joints by physical activity. As this essay will show, surgeons and engineers have experimented with a wide variety of materials—including metals, plastics, and ceramics—during the evolution of this type of implant. Some of the major developments in industrial engineering and materials science in the twentieth century, as one might expect, have contributed to the history of artificial hips.

This essay divides its account of the development of artificial hips into four broad chronological stages: early history from the nineteenth century to the early 1940s; a "premodern" phase of accelerated activity from the mid-1940s to the mid-1950s; a "modern" phase from the mid-1950s through the 1970s; and recent developments over the last twenty years. Since clinical developments in artificial hip replacement were taking place around the world simultaneously, this historical account is organized around developments that involve similar materials and similar surgical techniques. It is worth saying that any study of the development of artificial hip implants must contend with the fact that whatever historical record already exists has been compiled primarily by members of the profession responsible for developing and implanting these devices. Thus, largely orthopedic surgeons and designers have determined the existing literature. By contrast, only now is a view of these devices emerging from the perspective of those who have received implants, especially as a result of multidisciplinary research into the efficacy of health care technologies in the West.

A leading commentator on the state of design and materials in hip implant technology at the end of the twentieth century has described the process of orthopedic innovation in hip prostheses as a "trial and error culture" (Huiskes, 1993). Thus, this chapter is concerned with the products of that culture, which is characterized by continual innovation and experimentation. It is easy to give the misleading impression that a technology has a clear, unilinear, predominantly technical and, with hindsight, predictable trajectory, in which one can discern the rational march of technological progress. And it is certainly the case, as noted above, that many perceive the artificial hip in its various forms to be a highly successful technology, the development of which is punctuated by heroic individuals and breakthrough moments. The literature on orthopedic research contains many examples of this type of history. From a sociological perspective, however, technological artifacts represent the product of networks of relations between groups of people who have been able to establish their stake in that technology. The "success" or "failure" of technological developments is of equal importance in seeking explanations for why some become dominant and accepted for routine use, and why others do not. The history of artificial hips has many examples of materials that did not "work" well, but also some where satisfactory explanations for changing material or design are less obvious.

The history presented here, then, is an account of the activities of a range of actors who have produced and used artificial hip replacement technology. Technology, in effect, is society made material, and the production of technologies is accompanied by the "production" of its markets and its users. The radical nature of total hip replacement and its general success in contemporary health care practice have created a high demand and indeed an increasingly large market for this type of prosthesis as both younger and older age groups with degenerative joint disease and other forms of joint damage are seen, or see themselves, as potential implantees.

In order to illustrate the perhaps unexpected variety in the material history produced by this trial and error culture, I chose the visual images presented in this essay in part simply because they show something of the range of the material technologies used over time. In addition to this, however, they draw attention to the different media in which these prostheses are represented within society. Normally, only the inhabitants of the confined worlds of biomechanical laboratories, manufacturing plants, or hospitals see artificial hips. But after looking at these images, one is reminded that these devices have been, or will be, inserted inside the human body in an attempt to mimic the function of human skeletal tissues. There is, surely, something rather unsettling about looking at body parts that are usually hidden from view. This

raises questions about how we discuss the inner material and mechanics of the human body. Society's concern with technological medicine ensures that such images sometimes reach the public domain. But these images also challenge the conventional perception of the external fabric of human bodies, and thus challenge the material foundation of everyday life.

FIRST EXPERIMENTS WITH SURGICAL HIP IMPLANTS

The development of artificial hips is shared between Europe and the United States. In the late nineteenth and early twentieth centuries, a wide variety of materials was used in implant surgery generally, including metals such as zinc, copper, lead, and aluminum. However, none appeared especially suitable and it was customary to regard them as temporary measures, sometimes inserted into the body with the hope of stimulating self-repair by the body tissues. The first recorded successful arthroplasty of the human hip was carried out in 1822 in Westminster Hospital in London, but this did not involve any prosthetic components. The first known case of insertion of foreign material between bone ends occurred in 1840, when wood was used in an attempt to mobilize a stiff jawbone. In the late 1880s, experiments with hip joints were being conducted in France with techniques using muscle tissue as the implantation material, and in Germany with various natural and fabricated materials. Thomas Gluck in 1890 described ball-and-socket joints made from a luxurious, hard, natural material—ivory—and fixed internally with nickel-plated steel screws. He had also developed a glue made with colophony (a derivative of pine resin extract), pumice powder, and plaster to achieve fixation within the body tissue. Later, when this method led to extrusion of the joints, one surgeon employed gold foil.

In spite of Gluck's prophetic experiments, thirty years elapsed before the next significant development in prosthetic hip surgery. Sustained experimentation took place during the 1920s and 1930s, when surgeons used a variety of materials in an attempt to design a durable inverted cup—called a "floating cup" or mold—to fit over the shaped head of the thigh bone and between it and the acetabulum. This was the first type of hip implant to be used in any significant numbers. One surgeon, Otto Aufranc, was to report in 1956 that he had conducted over a

thousand hip arthroplasties over a fifteen-year period. Aufranc asserted that the mold or cup method was the best solution to conditions of the hip resulting from rheumatoid arthritis and traumatic degenerative arthritis (Aufranc, 1957).

The floating cup was pioneered by Marius N. Smith-Petersen, an eminent surgeon from Boston, Massachusetts, whom Scales dubbed the "doyen of arthroplasty of the hip."[1] Smith-Petersen, an emigrant to the United States from Norway in his teens, made many innovations in orthopedic surgery and became a professor of orthopedic surgery at Harvard Medical School and chief of orthopedics at Massachusetts General Hospital. He was acclaimed in the United Kingdom, where he was elected an honorary member of the Royal Society of Medicine in 1952.[2] The principle of Smith-Petersen's mold was to restore function by introducing an artificial cup into the joint, rather than actually replacing a removed part of the joint. Experimenting with diverse materials, Smith-Petersen made successive use of glass, viscaloid, Pyrex, Bakelite, and Vitallium.[3]

Vitallium, a metal alloy of cobalt, chromium, and tungsten, appeared to be the most biocompatible material. By 1940 over 1,200 Vitallium floating mold operations by over sixty surgeons were presented to the Texas Surgical Society (Howmedica Inc., 1998). In fact, for the first—but not the last—time in the history of hip prostheses, this material was transferred to orthopedics from applications in the cognate fields of dentistry and dental implantation. As early as the 1870s, one of celluloid's applications also had been in the production of dental plates; indeed, Smith-Petersen borrowed Vitallium from his own dentist. While many of these materials appear with hindsight unlikely candidates for their intended function, they are extremely interesting for what they emphasize about the range of "man-made" materials, drawn from various embryonic or more established manufacturing industries, with which it was possible to experiment with the boundaries of the human body during the first decades of the twentieth century. Each of these materials, in fact, did have at least one attractive property to justify testing it in this environment. For example, after observing that glass stimulated tissue growth, scientists assumed that other related materials would be both inert and malleable under the appropriate manufacturing treatment.

The first recorded artificial *replacement* in a hip joint by a manufactured part occurred in 1922, when a forty-one-year-old man in

Britain suffered a fracture to the neck of the femur. A graft was taken from his other femur bone, and a head fashioned from ivory was attached to it before surgeons inserted it into the host hip socket.[4] More than a decade passed, however, before doctors implanted the first known, and soon the first successful, metallic replacement of the head and part of the neck of the femur.[5] Possibly the first artificial metal hip socket fixed into a human acetabulum was introduced by a German, E. Rehn, who inserted a steel cup with spikes on its outer surface to fasten the device into the refashioned host bone in cases of congenital dislocation. This was conceived as a temporary device that would enable new bone to form in the hip so that a secure joint would exist when the metal cup was removed. At about the same time perhaps the first metal acetabular cups, made of Vitallium, were introduced by Preston and Albee in the United States around 1940, which were known as the Albee and Albee-Preston cups (Howmedica Inc., 1998).

Meanwhile, in 1938 P. Wiles, a surgeon at Middlesex Hospital in England, had been experimenting for the first time with ground components of a stainless steel to replace both cup and ball elements of the joint, precisely engineered to fit each other, the first attempt at *total* hip joint replacement. Wiles inserted six such devices, bolt- and screw-fixed to the femoral and acetabular bone. At the same time, in the United States, E. J. Haboush in New York developed the concept of the mold arthroplasty further and introduced a hollow ball with a "skirt" that would fit around the shaped head and neck of the femur, made from Vitallium alloy. In the case of the Albee cup, fixation of the head to the prosthesis proved problematic; Haboush concluded that an artificial head and neck of the femur and an artificial socket were necessary. He organized tests in his New York laboratory using a Vitallium head and an acetabular socket made of acrylic. Unfortunately, excessive wear and abrading resulted from these tests.

The development of hip prostheses, therefore, through the early 1940s saw the birth of the trial and error culture in orthopedic implant surgery, in which a wide range of materials drawn from a variety of sources had been tested in different designs, producing a large number of "errors." Neither the mold nor partial joint bone replacement showed great success. The late 1940s and early 1950s, by contrast, were an especially fertile and productive period for the evolution of hip implant technologies—though not without error, of course—as new materials

and modifications to known materials were brought into the test-bed of human hip implants.

EARLY MODERN DEVELOPMENTS

In the mid-1940s a new design signaled the first use of plastics in joint replacement technology, and brought with it the first biomechanically designed hip implant. In 1946 two Parisian brothers named Judet introduced this very influential innovation. The Judets believed that osteoarthritis was primarily a disease that attacked the femoral head; thus, because it was fitted to a diseased foundation, the "mold" arthroplasty was not a viable design. The brothers designed their prosthesis in methylmethacrylate polymer, a glassy thermoplastic that can be cast or molded, most commonly known as acrylic, but also produced under other proprietary names such as Perspex, Plexiglas, and Lucite. The material was first produced in Germany early in the century, but industrial production developed from 1932 onward when acrylic sheet and molding techniques became available. The acrylic industry progressed rapidly during World War II, when the material was used for glazing on aircraft bodies. Like Vitallium, acrylic was first used in medical treatment as a dental implant. Haboush had studied its wear-resistance properties in 1940 when considering his arthroplasty work, but had not used it in his surgical experiments. The Judet design in acrylic consisted of an enlarged hemispherical head and a short thick stem external to the neck of the femur, the design sometimes being referred to as a "mushroom." To insert it into the body, the surgeon needed to ream the socket into the recipient's hipbone.

The Judet acrylic hip joint was the first mass-produced surgical implant of any sort to use a thermoplastic. By 1952, surgeons had inserted more than four hundred Judet implants and the design became widely used throughout Europe. In spite of good biocompatibility of the material, however, the rate of subsequent fractures was high. To combat this, surgeons often embedded a rod of stainless steel or chromium-plated brass in the stem to reinforce the reconstructed hip. These modifications resulted in problems associated with the brittleness of acrylic resin; consequently, the device was produced in "block" or solid nylon. Although this reduced the occurrence of stem fractures, the particles of

nylon material proved to be more destructive to the bony tissue of the hip than the acrylic had been. Some surgeons had similar designs made in cobalt-chrome alloys and stainless steels, but while again reducing fractures, and having improved wear characteristics at the articulating surfaces of ball and socket, it generally suffered from excessive loosening. Nevertheless, interest in this type of design still exists. In the early 1980s German surgeons implanted a version of the Judet-style endoprosthesis (i.e., one that is used for repairing the head of the thighbone), the fair results of which were published in the orthopedic press in the 1990s (Bettin et al., 1993).

A well-known American surgeon, M. d'Aubigne, who produced his own design using acrylic resin material, also used the Judet prosthesis. This featured a very long stem for fitting inside the femur and—in a highly unusual departure—a flat plate as the weight-bearing part, which looks rather like a spatula. Solid resins were used in some later designs: in the early 1960s, for example, the socket component of a model produced in Italy used the polyacetal resin Delrin, which was in fact manufactured in the United States by DuPont. In general, however, despite the popularity of the acrylic resin Judet device, the general tide of opinion among surgeons turned against this material in the 1950s because of high levels of failure due to loosening caused by excessive wear. With hindsight, therefore, the Judet design is frequently regarded as a disaster in the orthopedic community, but it can be regarded as the first artificial hip to be implanted in numbers of patients sufficient to suggest that a totally artificial hip joint of some design could become a regular and accepted part of surgical treatment of joint disease and fracture in populations on a large scale.

The post–World War II years witnessed such an expansion in the development of polymer plastics for consumer goods that it is sometimes known as the "Age of Plastics." Orthopedic implants utilized solid nylon as a material in cup implants as well as in the Judet ball and stem device. The material was used for dry bearings in engineering applications, and its relative resilience was appealing as a possible means to spread the weight load through the femur. A small number of prostheses were implanted with this material; one design followed the Smith-Petersen model while another design was fixed to the head of the femur by three splines. Neither design was successful: like the Judet hip, both designs induced severe wear resulting in tissue reaction with abraded particles of nylon. In the late 1940s, polythene was used in a

non–weight-bearing prosthesis and in a socket cup similar in design to the Smith-Petersen floating cup. The latter, which was weight-bearing, soon showed signs of deterioration and wear with the polythene, like nylon, reacting adversely and producing fine particles that were found embedded in body tissue.

At approximately the same time in 1950, there were major developments in the United States in the use of metals for the femoral component of the hip. Austin Moore, for example, a surgeon at the University of South Carolina, had been involved with the first known metallic replacement of the head and part of the neck of the femur in a human hip. Moore's name is closely linked with that of the surgeon F. R. Thompson in the United States. Both surgeons considered that a suitably shaped stem component could be inserted inside the upper section of the bone of the femur, in the relatively soft honeycomb-like bone of the "intramedullary canal" inside it. Moore's initial concept was of a "self-locking," straight device that achieved stability inside the slightly curved femur by the pressure of jamming it firmly in (later, such a method was to be termed "press-fit") (Moore, 1957). Moore, aided by a professor of engineering at the Austenal Laboratories in New York, produced the prototype of this design in cobalt-chrome alloy, probably Vitallium, and the design included a ball component that was offset from the top of the femoral stem. Figure 7.1 illustrates the design and its development. This type of collaboration between surgeons, engineers, and testing laboratories was to become increasingly crucial to future innovations in orthopedic implants.

The Moore device was technically successful in many ways, but implantees suffered from unstable movement at the hip and, although it did not completely loosen, in most cases "migration" of the device occurred. In closely related work, F. R. Thompson believed in the same concept of intra-femoral implantation. His 1951 design was quite similar in appearance to the Moore device, though Thompson feared breakage and difficulties of removing the stem if it was ingrown with bone tissue. Thompson also preferred cobalt-chrome to steel and used Vitallium for his devices. Haboush experimented further as well, and in 1951 he inserted a cast Vitallium ball and hip socket, fixed with dental acrylic cement, into a patient at the Hospital for Bone and Joint Diseases in New York.

While Vitallium had been the most promising metal alloy for hip implants during the 1940s, between the two world wars a large number

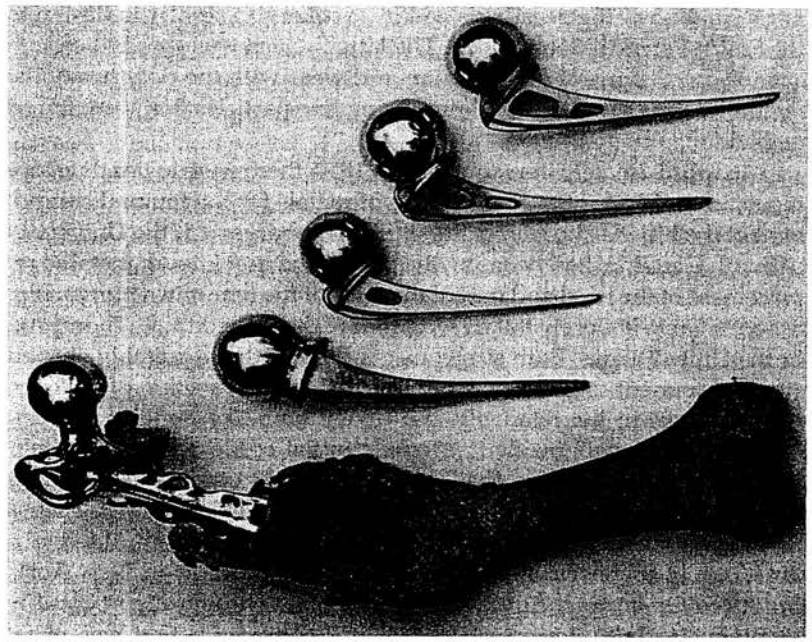

FIG. 7.1. Sequence of Austin Moore designs. This design was one of the first to use embedding of a tapered stem inside the bone of the femur. The first in the sequence shows the prosthesis emerging from the top of a retrieved, decayed thigh bone. The loops in the metal were intended for the attachment of muscle tissues. The material is cobalt-chrome alloy. The initial straight stem was soon modified. The windows in the stem—"fenestration"—were designed to enable the cancellous bone of the femur to grow through, thus fixing the device in position. From Moore, 1957. Reprinted by permission of the *Journal of Bone and Joint Surgery* (American).

of stainless steels and steel alloys were created and used in a wide range of engineering applications in the industrialized nations: "18-8" stainless steel (18 percent chromium, 8 percent nickel) had been first introduced into surgery in 1926 as an alternative to vanadium steel, while the addition of molybdenum produced an even more corrosion-resistant steel. Thus, the tensile strength and inertness of stainless steels helped them become, after World War II, the main alternative to cobalt-

chrome alloys as the material for the load-bearing femoral component of artificial hips. In the United States in the early 1950s, stainless steel was used both in stem designs and in new designs of the mold or floating cup concept. For example, J. C. Adams produced first a hollow stainless steel cup, followed by a solid cup with a small aperture for tight containment of a reshaped femoral head and neck. The design proved to be unstable and bone was reabsorbed around the device. A Judet-style stem prosthesis also was conceived in stainless steel around this time, but failed to gain popularity, and Eicher produced a head-neck replacement, first in a stainless steel and then in cast cobalt-chrome alloy (Scales, 1967).

Stainless steels were also being used in England by G. K. McKee in Norwich in his first attempts to produce a design for a total ball and socket hip implant. McKee had designed and fabricated models of total hip prostheses as early as 1940. Like his counterparts in the United States, he found that his stainless steel devices quickly became loose in vivo, while a cobalt-chrome alloy version survived well for a period. Following a visit to the United States, McKee adopted the design form of the Thompson solid stem prosthesis. To this he added a three-clawed cup that was screwed into the acetabulum, all parts now made from a cobalt-chromium alloy called vinertia.[6] In spite of the developments in stainless steel alloys, McKee had come to the belief that this metal and its alloys were insufficiently inert for human implantation. He also believed that titanium had a tendency to "self-weld" (McKee, 1971: 50). Thus he used cobalt-chrome for cup and socket and femoral components that were produced as a pair for implantation in a single hip replacement, and he marked each component with a number for matching purposes.

Despite the fact that a long-lasting stable artificial hip remained on the horizon, knowledge of suitable materials—in terms of both biocompatibility and strength—had increased significantly by the 1950s. Vitallium and some forms of stainless steel were well established, and many of the trial and error designs described above were conceived in an attempt to preserve and utilize undamaged or reshaped hip sockets, where surgeons inserted them into either a replacement femoral neck and head or a head with a cup/mold positioned over it. On the other hand, the lack of successful hip replacement design was starting to lend support to the idea, already suggested by some surgeons, that for a hip implant to be functional without causing damage either to the human

bone tissue or to the prosthetic materials themselves, surgeons would have to produce a design of highly wear-resistant components in which the articulating movement between ball and socket itself was achieved through the use of totally artificial materials. With benefit of hindsight, in the designs of Moore, Thompson, and McKee, the future of total hip replacement, the concept on which current artificial hip joints is based, can be discerned.

In reviewing these early hip implants we can see that they were pioneered typically by a single surgeon on a small number of patients using custom-made components. It is interesting to note the trend in naming devices after their surgeon-inventor, a trend that continued into the "modern" phase of development. The relative failure of some designs when implanted more widely, which may have seemed promising initially when implanted by the surgeon-inventor, might be explained to some extent by poorer manufacturing techniques used when the device was subjected to mass production techniques (Williams and Roaf, 1973). However, progress was being made in the understanding and production of orthopedic materials in spite of clinical failures.

"MODERN" HIP IMPLANTS

While stainless steels and Vitallium were established as leading contenders in artificial hip fabrication by the early 1950s, one metal—titanium—had yet to make an appearance. Although titanium is now a commonly used material in hip implants, its first recorded use in vivo was by Leventhal in 1957, six years after he had originally presented a case for its use in the *Journal of Bone and Joint Surgery*. Titanium has an oxide layer or film on its surface that reduces tissue reaction to a minimum, making it highly inert, like a ceramic. Titanium is also resistant to saline environments, which enhances its suitability for implantation inside the body. Cobalt-chrome and "316L" stainless steel are corrosion-resistant metal alloys with similar properties in this respect. Materials such as gold, silver, and platinum also have high corrosion-resistance but have poor mechanical properties, rendering them of little interest for heavy-duty implants. Titanium has less elasticity than cobalt-chrome alloys. In fact, titanium (and its alloys) is the only metallic material of significance to be introduced in implant surgery between 1950 and 1970. By contrast, many different plastics, rubbers, fibers, and fabrics were introduced during the same period. The U.S. Department of Defense supported the development of titanium technology, which required new melting and fabrication techniques, at its titanium metallurgy laboratory.[7]

Given its emergence at this time, it is striking that titanium alloys were not considered by the artificial hip's most successful surgeon-innovator, John Charnley, from Lancashire, England. Charnley is best known for integrating the scientific study of joint lubrication and biomechanics with the problems associated with prosthetic hip joint surgery. Indeed, in reviewing the "first 32 years" of total hip replacement in 1991, William Harris, a leading contemporary U.S. surgeon-designer, singled out Charnley in his appraisal of the progress of hip joint replacement technologies. Charnley had in fact developed a preliminary arthroplasty as early as 1946, but had abandoned the idea and throughout most of the 1950s had remained pessimistic about the prospects of a successful design. The main reason for his skepticism at this time lay in his view that the frictional properties of acrylic or metal in the prosthesis and bone and cartilage in the human body were incompatible. Cartilage lubricated with synovial fluid in human joints has an extremely low coefficient of friction, more slippery than a skate on ice (as his biographer notes), which artificial materials cannot match. With collaborators in engineering Charnley built rigs to gauge the friction in joints of bone and the artificial materials then mainly in use: stainless steel, cobalt-chrome, and acrylic (and perspex) (Waugh, 1990: 104). Friction was shown to be much greater with these artificial materials even when animal synovial fluid was used as a lubricant.

Charnley's aim thus became to find two different artificial materials for the ball and socket that would slide freely in contact with each other, but not require artificial lubrication. Charnley turned to manufacturers of synthetic plastics in a search for suitable socket material. He found a company in Bolton, Lancashire, that knew how to apply new polymers in engineering applications. Charnley believed that the best material for his purposes was polytetrafluorethylene (PTFE)—best known under the brand name Teflon, but sometimes marketed under the proprietary name Fluon—the same material now famous for its use in nonstick cooking utensils. Not only did it not stick, but Charnley believed it to be the most inert plastic then known; indeed, he tested this by inserting small pieces into his own leg. He experimented with a floating cup design combined with a socket cup, both in Teflon, but

quickly moved on to a more radical approach: surgical removal of the femoral head, replacing it with a Moore-style femoral component, which was then fixed with the addition of "cement" inside the femur. Today this cementing principle is the mainstay of orthopedic hip replacement practice, especially for implants in elderly people. The Moore stem was the first to use the Teflon socket. McKee in Norwich and others, as we have seen, were using primarily metal, cobalt-chrome or stainless steel, for the femoral head, but this failed in patients with arthritic socket bones that could not withstand the abrasion of this material and design. In 1962, however, the PTFE sockets, made by Charnley himself and inserted into some three hundred patients, began to fail because bone tissue began to deteriorate in reaction to worn particles of the material. Charnley noted that the extreme gratitude expressed by his patients in the early stages following the operation delayed the recognition of the failure. Charnley sought more mechanically robust forms of polyethylene with high density, and the first high molecular weight polyethylene (HMWP) for the artificial hip socket was introduced at the end of 1962. Since that time, the development of the socket component forming the bearing surface with the ball has been dominated by the use of successive refinements of plastic materials.

Charnley's early total joint device appeared promising, and he accepted advice from mechanical engineers that a smaller femoral head would be an improvement, partly because it would enable a thicker, more robust socket cup to be used. This was the origin of his so-called low friction arthroplasty (Charnley, 1979). Charnley redesigned the Moore-style femoral component and had it manufactured in stainless steel by Thackrays, an engineering company from Leeds, Yorkshire, with experience in making surgical instruments. Although he believed that cobalt-chrome would prove to be the preferable metal, he never adopted it in his own surgical practice. Charnley's low friction design is illustrated in figure 7.2. It is worth noting that the two metals as produced at the time had different working characteristics: cobalt-chrome artifacts were cast, while stainless steel was wrought. This difference reflected different modes of production. Whereas cobalt-chrome required large numbers of relatively unskilled female labor, stainless steel was produced by "men who [were] true engineering craftsmen." The masculine image of stainless steel manufacture, a skilled engineering craft strongly associated with northern England, may have held some appeal for Charnley, and may partly explain why he continued to use artificial

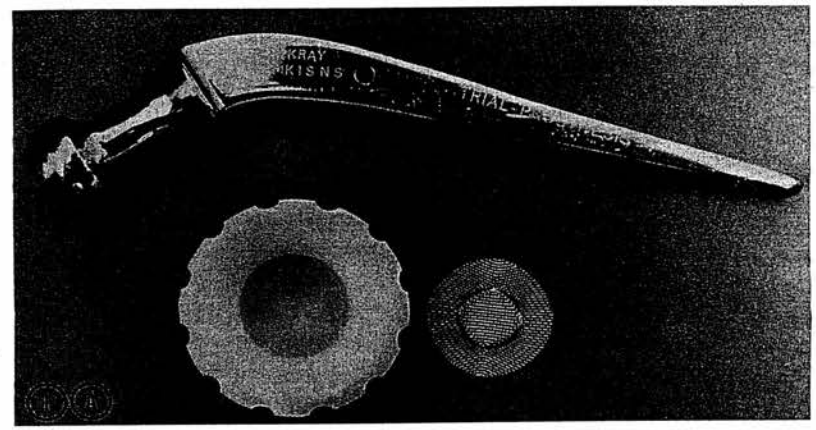

FIG. 7.2. Early Charnley prosthesis. The femoral component is in stainless steel, the cup in high-density polyethylene. The words "Trial prosthesis" are inscribed on the stem; the name of the manufacturers (Thackrays, the medical instrument engineering company in Leeds, England, bought in the early 1990s by DePuy International, one of the major manufacturers of orthopedic implants worldwide) is also legible. The "trial" prosthesis was in fact a manufacturer's model made to test the fabrication process, and would not have been implanted. As can be seen, the small femoral head was absolutely spherical and produced with a high degree of shine, which Charnley believed essential to low friction performance. From Owen, 1971, 69. Reprinted by permission of Professor Robert Owen.

hips produced in stainless steel, in addition to the fact that it was also cheaper. By the late 1960s, Thackrays was producing nine to ten thousand stainless steel artificial hips per year. In fact, the company did produce a cobalt-chrome version of the hip in the early to mid-1970s for the U.S. market, but the plan was not financially successful and was abandoned.

At the same time that Charnley was developing his designs, other surgeons in England were designing prostheses that were to be implanted very widely. For example, in the late 1950s and early 1960s the widely used "Stanmore" devices, named for a town in Essex where the country's National Orthopedic Hospital is located, used cobalt-chromium alloy as the material chosen for all related parts. Another

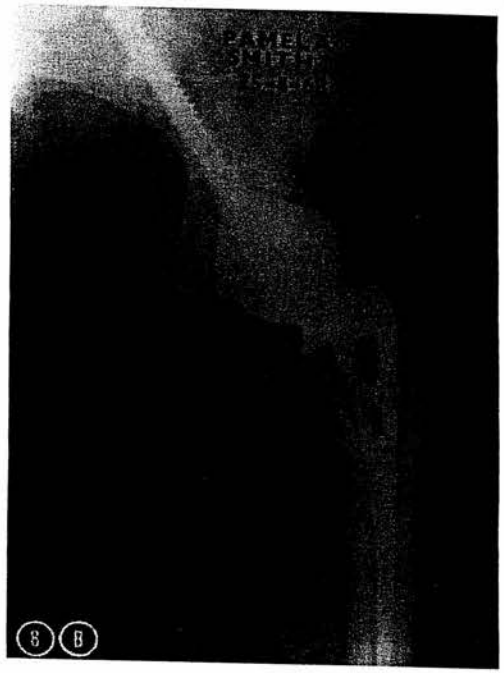

FIG. 7.3. The Ring hip prosthesis. Both components, including bearing surfaces, in cobalt-chrome. Note the very long socket screw fixed into the acetabulum. The name of the patient on the x-ray reminds us that these devices are implanted in the interior of a person's body, but the image is ambiguous in this respect, because, although the person is named, only part of the person is shown and the x-ray image is one generally available only to the specialist orthopedic surgeon and other hospital staff. While a large number of x-ray images of hip prostheses have been published in academic journals by orthopedic surgeons, it is relatively unusual to see the patient's name inscribed on them. From Ring, 1971. Reprinted by permission of Dr. P. A. Ring.

individual orthopedic surgeon, P. A. Ring, developed a cement-free prosthesis with a novel means of attaching the socket component. Again, in the Ring prosthesis, cobalt-chromium alloy was the material used. As can be seen from figure 7.3, the socket design featured a long screw for fixing the prosthesis up into the hip bone. As with the early Charnley prostheses, clinical results of these implants are still being published today.

RECENT DEVELOPMENTS

Over the past ten to fifteen years, enormous institutional and financial resources have been marshaled for the purpose of exploring new materials and designs for artificial hips. In spite of the many apparent advances, the current state of implant technology has been described by one leading commentator as an "innovation impasse" and by another as having hit a "glass ceiling." The proliferation of new models, combinations of materials, and design variations, many of which are considerably more expensive than earlier models, has led some commentators to express concern about a trend toward "designer hips" (Bulstrode et al., 1993). In the 1960s most patients whom surgeons saw for replacements were relatively elderly people who suffered from severe joint pain due to arthritis. But this characterization of the end-user has been expanded as the technology, and demand for, hip prostheses has changed. The reports of good clinical results for prostheses fifteen years or longer after implantation has encouraged an expansion of the market to older and especially younger implantees, such as those who have suffered joint damage from sports-related injury or deterioration. Thus if artificial hips have reached a plateau in their current technological development, this may be because they fell victim to their own success during the 1970s and 1980s.

Perhaps the central issue in the contemporary development of hip implants is the division between those that are implanted with acrylic "cement" and those that are not. In spite of the success of cemented models, the development of cementless models has been spurred by two main factors. The first is the loosening of the components, which in the 1970s was attributed to "cement disease," a condition whose nature and existence are still disputed. It was believed that adverse reactions occurred between tissue and cement, causing particles of cement to aggravate the surrounding tissue, which in turn loosened the implant. Regardless of the veracity of this theory, it certainly is one of the forces behind the development of cementless models that, in turn, has encouraged a search for alternative materials with which to create adequate fixation to the host bone. The other main reason is the continued poorer

performance of the socket components of cemented implants. Moreover, there has been a trend toward dividing the prosthesis into separate functional parts. This is known as modularity, the prime example of which is the production of the head of the femoral component separately from its stem, which gives surgeons greater choice in seeking better anatomical fit of the device.

Surgeons and engineers have thus sought materials and designs that might improve the attachment of the implant to the host bone and tissue. A variety of techniques of encouraging the surrounding bone tissue to actually grow into the prosthesis, thus achieving what is known as "biological" fixation, have emerged. Surgeons did not introduce these uncemented "porous-coated" designs until the early 1980s, although the concept had been investigated as early as the 1960s. There are now more than twenty different models of this type of design on the market (Griffiths, Priest, and Kushner, 1995). The design usually consists of "micro-pores" of the basic implant metal added as a very thin layer to the basic device. The additional layer is formed of minute spherical beads in cobalt-chrome, or wire mesh or honeycomb-like lattice in titanium, requiring advanced production technologies to fix to the basic surface. The stem/head and the cup may be treated, usually all or part of the "shoulder" of the femoral component, and all of the outer surface of the socket. Such designs are increasingly widely marketed but remain somewhat controversial clinically, partly because of concerns about their longevity in the body and partly because pain in the thigh appears to be fairly common with this method of fixation of the femoral component (Faulkner et al., 1998).

The concept of biological fixation was tested in the laboratory as early as the late 1960s, as noted above. This has now been extended to the concept of a bond where, rather than simple ingrowth of bone into the inert prosthesis, a biological interaction between bone and prosthesis is sought. This represents the ultimate form of fixation within the body short of human cell-based tissue engineering. The bioactive material is applied as a further coating, usually in a ceramic material sprayed on to the surface of the porous coat. Hydroxyapatite (HA), a calcium phosphate ceramic, is of particular importance. This can be derived from natural bone, but can also be produced synthetically. Its chemical and crystalline structure is the same as the major mineral constituent of human bone. Dentistry again plays a part in the pedigree of the material, since it was originally used in oral and facial surgery, for example,

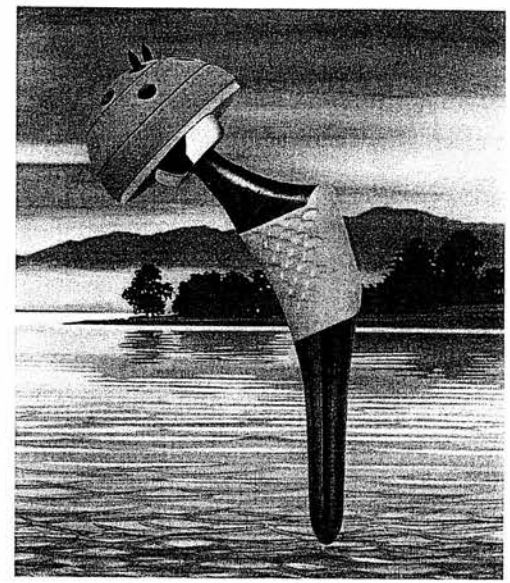

FIG. 7.4. ABG II hip with hydroxyapatite bioceramic coating. The stem is made from Vitallium cobalt-chrome-molybdenum alloy, and the socket and stem are coated with the ceramic hydroxyapatite (HA), characteristically white, for cementless fixation. Note the "macro-interlock" bobbling like fish gills on the upper part of the stem to enhance fixation. The image is from a 1998 advertisement. The backdrop of a calm scene of "Nature" emphasizes the appeal being made to smooth, organic, natural function. Reproduced by permission of Howmedica Osteonics.

as synthetic bone for artificial teeth, but it is unsuitable for weight-bearing because of its brittleness. Advanced forms of industrial technology are required for spraying HA, the coating being built up as a series of layers blasted on to the host metal by a computer-controlled robotic spray-nozzle. An example of an HA-coated device is in figure 7.4.

Long-term results of the performance of HA in the human hip are still awaited, but it is becoming increasingly widely used. Early studies suggest that it may be less associated with thigh pain than many of the earlier porous-coated devices (Faulkner et al., 1998). There is some technical controversy about the optimal metal to use with HA coating.

Titanium has higher elasticity than most other metals but is sensitive to the surface disturbance introduced with coating. Chromium-cobalt-molybdenum alloys remain stronger when coated (Learmonth and Spirakis, 1989), but on the other hand titanium may be preferable because it forms a chemical as well as a mechanical bond with HA coating (Geesink, 1990).

The materials used for the femoral stem and head in cementless implants in the recent period have been mainly stainless steel, cobalt-chrome-molybdenum alloy, and titanium aluminum vanadium alloy (Head, Bauk, and Emerson, 1995). However, stainless steels are much less used now for the stem. The orthopedic research literature disputes the relative advantages of the other two alloys. Again there is support for titanium alloy because of its biocompatibility, the degree of ingrowth of bone into the implant when porous-coated, and its elasticity, which is important for use in smaller stem sizes. Titanium was supported extensively in European centers of orthopedics during the 1980s (Stern et al., 1992). On the other hand, it is said to be an inferior material for the load-bearing surface of the head of the femur, is more prone to abrasive wear, and conveys greater stresses, making it less suitable for use with the particle-structure of acrylic cement (Head, Bauk, and Emerson, 1995). Ceramics and cobalt-chrome alloys are both superior in their wear properties. In fact, the most widely used cementless artificial hip in the United States—and probably worldwide—is the AML (anatomic medullary locking) device, produced by DePuy and manufactured in cobalt-chrome-molybdenum alloy. Where titanium stems are used they are generally combined now in modular devices with stainless steel or cobalt-chrome alloy heads, which are capable of taking a higher degree of polish and are less prone to scratching (Learmonth and Spirakis, 1989). The head component in modular designs may be coated in an attempt to achieve the smoothness of ceramic, for example by the use of titanium nitride.

Metal materials and surface finishes, therefore, especially for the femoral component of the artificial hip, are still being developed. "Superalloys" and composite materials are being tested in an attempt to achieve better biological and biomechanical matching to the flexibility of the femur. Polyacetal, polyethylene, carbon, and acrylic are among the materials that have been employed for this purpose in metal composites (Learmonth and Spirakis, 1994). In another technique to enhance fixation, the stems of devices for cemented implantation may be precoated with acrylic cement, added after either grit blasting to achieve a relatively rough surface or bead-blasting for a smooth surface.

A further major area of current debate and development in hip implant technology is the material used for the bearing surfaces between artificial ball and socket. Different combinations are in use and being tested, the major contenders being ceramics, metal, and polyethylene. Of the ceramics, alumina (the oxide of aluminum), titanium oxides, and zirconia (oxide of the rare crystalline element zirconium) are used and have good biocompatibility (Learmonth and Spirakis, 1994). Zirconia is a composite material, intermediate between metal and nonmetal, marketed as combining the advantages in wear and mechanics of both ceramics and the superalloys. The first all-ceramic (alumina-alumina) bearing in a total hip implant in a human had been implanted in 1970 in France (Nizard et al., 1992). Improved manufacturing after the mid-1970s enabled a smoother surface finish to be achieved, utilizing a dense alumina. In modular designs of implant, ceramic heads may be combined with high-density polyethylene, or alternatively modern cobalt-chrome for both surfaces may be used. McKee and others had used all-metal designs in the 1960s, but in general this was superseded by metal-polyethylene combinations. However, currently the metal-polyethylene bearing surface is seen as one of the weak points in the prosthesis design, and new metal-to-metal units have been developed again since the mid-1980s, for example by Müller in Switzerland, as illustrated in figure 7.5 (Head, Bauk, and Emerson, 1995). Müller's socket devices have a titanium outer shell inside that sits a polyethylene component that is lined with a cup in cobalt-chrome for the bearing surface of the joint. The femoral head is also of cobalt-chrome alloy. Maurice Müller is an extremely high-profile surgeon and orthopedic entrepreneur in Switzerland.

The advancing developments in materials and design appear to have brought their own problems in spite of the great clinical success of total hip implants. The earlier modern prostheses (1960s and 1970s) suffered from problems such as metal fractures and breakages, typically occurring five or more years after implantation. The arrival of superalloys and composite materials has almost eliminated mechanical breakage in normal usage of the artificial hip. The problems that these bring may appear only two or three years after implantation. Describing the effects of the new materials such as ultra-high molecular weight polyethylene, one surgical commentator has stated that "We have unleashed

PAPER 2

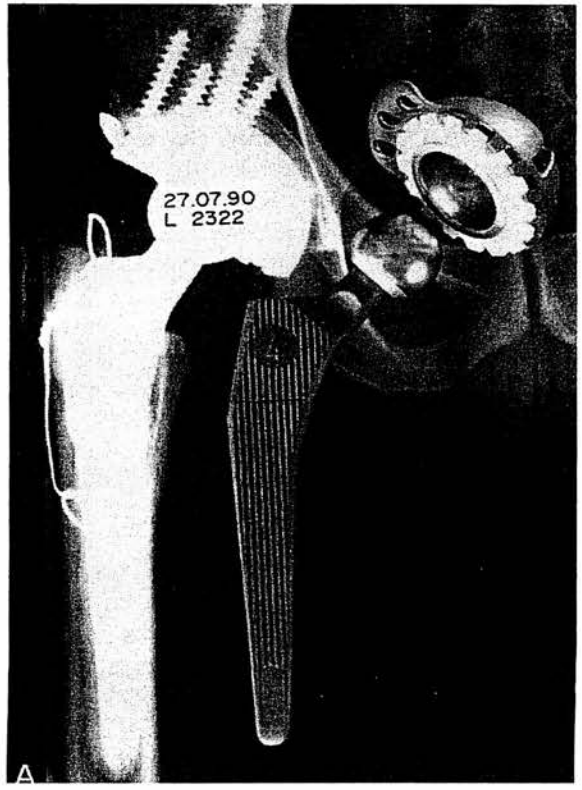

FIG. 7.5. The Müller metal-metal total hip replacement. The stem has vertical grooving for "macro-interlock" inside the thigh bone. The head of the Müller metal-metal model is much larger than the Charnley metal-polyethylene concept. The socket is comprised of three parts, the outer shell of titanium, the inner liner of polyethylene, and the cup, into which the head of the femoral component fits, of cobalt-chrome. In a previous design, now abandoned, the cup surface was treated with titanium nitrite for decreased friction, and the cobalt-chrome head of the femoral component was treated with a thin layer of titanium carbide (Head, Bauk, and Emerson, 1995). As can be seen, the socket is screwed into the hip bone, without cement. The design was developed in 1989 by Müller in collaboration with Sulzer Medical Technology. This distinctive image is unusual in depicting a photograph of the device superimposed upon an x-ray image of the same device in vivo. From Müller, 1995, 57. Reprinted by permission of *Clinical Orthopedics and Related Research.*

a torrent of particles into our joints, producing a devastation far exceeding simple prosthetic failure or fragmentation of parts" (Booth, 1994). While this may be a somewhat exaggerated account, it does appear to be the case that introduction of new materials within and adjacent to the tissue of the human hip and thighbone is a continuing cause for concern.

CONCLUSION

The early history of artificial hip technology was concerned with repair of conditions including fractured bones, congenital abnormalities, and ankylosis (joint stiffness)—conditions where disability was outwardly manifest. Even partial replacement technologies, especially those designed to position a bearing on the top of the femur, were designed primarily in these contexts. The concept of the total hip implant, however, allowed eventually for a relatively routine joint replacement operation, appropriate for people suffering from painful arthritic deterioration of the hip joint. This has turned out to be an enormous market, especially in an aging society, and the technology is now being extended to younger patients. Some of the newer designs, such as those featuring bioactive coatings, are being actively promoted particularly for this younger client group.

Technological medicine provides devices that can replace damaged or worn internal components of the human locomotor system. But in examining the design and material composition of artificial hips, we should not forget that we are peering into the normally hidden world of the human anatomy. Unlike external prostheses, these devices, once implanted, can be visualized only through x-ray and other imaging technologies. Since direct observation is not possible, the material properties of the devices are not part of the end-users' everyday social world. These technologies, then, are used and controlled by health professionals in the context of doctor-patient relationships. Implantees "consume" them only via their effect upon physical functions and experience of sensation such as pain or discomfort. Given their invisibility in the everyday life of implantees, the forms in which artificial hips are represented are of particular interest. The two most widely available sources of images of artificial hips are photographs and x-ray reproductions in clinical research articles in the academic orthopedic press,

and in the advertising images of the orthopedic manufacturing companies. This chapter illustrates examples of these types of image. Orthopedic research articles focus primarily on performance of the prosthesis, defined in terms of technological and clinical criteria such as implantees' pain and physical function in the years following implantation, and the durability of implants. On the other hand, the manufacturers' portrayals focus upon two other key aspects of the prosthesis, its material properties and "operability" from the surgeon's point of view. As one might expect, some modern advertising representations of the devices attempt to incorporate images that express the strength, naturalness, and biocompatibility of modern materials.

Of course, as one would expect with devices that are usually hidden from view, hip prosthesis components in general feature little wording or iconography. Most, in fact, do have serial numbers and some, as seen in the Charnley example above, have the manufacturer's name(s) or lettering or other functional inscriptions. The representation of prosthetic devices within the orthopedic surgical and manufacturing communities has also been referred to in relation to the name, either conventional or legally trademarked, of artificial hips over the course of their evolution. In the 1940s and 1950s there was a strong tendency for hip implants to be known and named for the surgeon-designer responsible (e.g., the Judet, the Charnley) or, in some cases, the geographical origin (e.g., the Exeter, the Minneapolis). This tradition continues, to some extent, but has been joined now by designs that reflect a culture responding less to surgical ingenuity and more to global business and consumerism. This reflects the larger shift from surgeon-driven innovation to alliance with a competitive corporate environment of profit-related innovation where orthopedic implant manufacturers play a powerful role. Recent hip designs, for example, carry names that emphasize functionality and anatomical compatibility, such as the "Natural-Hip System," the "BioGroove Hip System," and the "Taper-Fit Total Hip System."

One can only speculate what these names convey in the marketplace of contemporary technological medicine. The term "system" perhaps suggests a product that has the capacity to fulfill a wide variety of consumers' needs and provides a variety of possible options from which the user can select. The images conjured by the words "Natural" and "Taper-Fit" clearly appeal to ideas about restoring natural and normal function. These appeals are as likely to be directed at surgeons and managed-care organizations as at the implantee-users for whom reduction in pain and restoration of function are generally the most important goals of prosthetic hip surgery. Thus by looking inside the body at these materials, we find ourselves looking also at the intersecting worlds of surgeons, engineers, and manufacturers. Decisions regarding choice of implant for different users rest essentially with surgeons, but these surgeons operate in an environment where they are exposed to manufacturing organizations that shape the prostheses that they will make available. Indeed, some of these practicing surgeons are surgeon-designers or surgeon-design consultants: the design of orthopedic implants now is the outcome of complex relationships between the clinical and bioengineering worlds.

When one contemplates a world in which such large numbers of people participate in "normal activities" with the invisible aid of such prostheses, one of the most striking features is the variety of materials and designs with which people have been provided. While this no doubt reflects the variety of shapes and sizes of the human body, it also reflects to some extent the inventiveness of designers and the competitiveness of manufacturing institutions. It is perhaps ironic to reflect that fashion is one of the driving forces behind an essentially invisible technology; the apparently late uptake of titanium alloys in Europe may be a case in point. The progression of materials and designs in artificial hip technology might be seen simply as the progressive, linear refinement of increasingly better-functioning devices—examples, in other words, of technical rationality in action. The shape and composition, however, of particular technologies might always have been different. Indeed, as this account demonstrates, the material development of artificial hips continues to be shaped by different forces of innovation in the context of contemporary orthopedics. Although some aspects of the design, such as the use of a stem sited within the femur, have achieved consensus among surgeons and engineers, there is no one single set of material technologies that can be said to dominate all areas of the design.

NOTES

1. In Scales, 1967, a survey of the early development of materials used in artificial hips.

2. An obituary of Smith-Petersen appeared in the *Harvard Medical Alumni Bulletin* 28 (October 1953): 36–37.

3. Viscaloid was a celluloid composed essentially of gun cotton and camphor; Pyrex is the borosilicate glass later used widely in heat-resistant cooking ware; Bakelite, a "thermo-setting" plastic, related to celluloid, was first patented in 1907 and used primarily in the development of electrical insulation and automobile parts as well as hundreds of household goods; Vitallium was a metal alloy of cobalt, chromium, and tungsten (tungsten was later replaced by the more easily obtainable molybdenum, which was available in Colorado from about 1918).

4. This was reported in the *British Journal of Surgery* by E. W. Hey-Groves in its 1926–27 volume, with an illustrative diagram (Hey-Groves, 1926–27, cited in Scales, 1967).

5. Implanted by the surgeon Dr. Bohlman of Maryland. Two such replacements failed, but the *American Journal of Surgery* later reported a successful, long-lasting result in one patient after ten years of implantation (Bohlman, 1952; cited in Scales, 1967).

6. See Waugh's account in his biography of John Charnley (1990).

7. See *Encyclopedia Britannica*, 1963.

REFERENCES

Aufranc, O. E. Constructive hip surgery with the Vitallium mould: A report on 1000 cases of arthroplasty of the hip over a fifteen-year period. *Journal of Bone and Joint Surgery* 39A, no. 2: 237.

Bettin, D., B. Greitemann, J. Polster, and S. Schulte Eistrup. 1993. Long term results of Judet's cementless total endoprosthesis of the hip joint. *Zeitschrift für Orthopädie und Ihre Grenzgebiete* 131, no. 6 496–502.

Bijker, W. E., T. P. Hughes, and T. J. Pinch, eds. 1987. The social construction of technological systems: New directions in the sociology and history of technology. Cambridge: MIT Press.

Bohlman, H. R. 1952. Replacement reconstruction of hip. *American Journal of Surgery* 84, no. 3: 1497.

Booth, R. E. 1994. Current concepts in joint replacement: The closing circle: Limitations of total joint arthroplasty. *Orthopedics* 17, no. 9: 757–59.

Bulstrode, C. J. K., D. W. Murray, A. J. Carr, et al. 1993. Designer hips. *British Medical Journal* 306: 732–33.

Charnley, J. 1979. Low friction arthroplasty of the hip: Theory and practice. Berlin: Springer Verlag.

Faulkner A., L. G. Kennedy, K. Baxter, J. Donovan, M. Wilkinson, and G. Bevan. 1998. Effectiveness of hip prostheses in primary total hip replacement: A critical review of evidence and an economic model. *Health Technology Assessment* 2, no. 6.

Geesink, R. G. T. 1990. Hydroxyapatite-coated total hip prostheses: Two-year clinical and roentgenographic results of 100 cases. *Clinical Orthopaedics* 261: 39–58.

Griffiths, H. J., D. R. Priest, and D. Kushner. 1995. Total hip replacement and other orthopedic procedures. *Radiologic Clinics of North America* 33, no. 2: 267–87.

Hahn, H., and W. Palich. 1970. W. Preliminary evaluation of porous metal surfaced titanium for orthopaedic implants. *Journal of Biomedical Materials Research* 4: 571.

Harris, W. H. 1992. The first 32 years of total hip arthroplasty: One surgeon's perspective. *Clinical Orthopaedics* 274: 6–11.

Head, W. C., D. J. Bauk, and R. H. Emerson. 1995. Titanium as the material of choice for cementless femoral components in total hip arthroplasty. *Clinical Orthopaedics* 311: 85–90.

Heck, C. V., and F. A. Chandler. 1954. Material failures in hip prostheses. *Journal of Bone and Joint Surgery* 36A: 1059.

Hey-Groves, E. W. 1926–27. Some contributions to the reconstructive surgery of the hip. *British Journal of Surgery* 14: 486.

Howmedica Inc. 1998. The Vitallium alloy story. <http://www.howmedica.com/book/vitall.htm>.

Huiskes, R. 1993. Failed innovation in total hip replacement: Diagnosis and proposals for a cure. *Acta. Orthop. Scand.* 64, no. 6: 699–716.

Learmonth, I. D., and A. Spirakis. 1989. Current status of total hip replacement: A review of biological and biomechanical factors. *South African Journal of Surgery* 27: 84–88.

———. 1994. Current thoughts on total hip replacement. *South African Medical Journal* 84: 88–90.

Leventhal, G. S. 1951. Titanium, a metal for surgery. *Journal of Bone and Joint Surgery* 33A: 473.

MacKenzie, D., and J. Wajcman, eds. 1985. The social shaping of technology: How the refrigerator got its hum. Milton Keynes: Open University Press.

McKee, G. K. 1971. McKee-Farrar total prosthetic replacement of the hip. In M. Jayson, ed., Total hip replacement. London: Sector Publishing, 47–67.

Moore, A. 1957. The Self-Locking Metal Hip Prosthesis. *Journal of Bone and Joint Surgery* 39(A): 811–27.

Müller, M. 1995. From *Clinical Orthopaedics and Related Research* 57.

Nizard, R. S., L. Sedel, P. Christel, et al. 1992. Ten year survivorship of cemented ceramic-ceramic total hip prosthesis. *Clinical Orthopaedics* 282: 53–63.

Owen, R. 1971. The Charnley total hip replacement. In M. Jayson, ed., Total hip replacement. London: Sector Publishing, 68–85.

Park, J. B. 1984. Biomaterials science and engineering. New York: Plenum.

Rang, M. 1966. Anthology of orthopaedics. Edinburgh: E. and S. Livingston.
Ring, P. A. 1971. Ring total hip replacement. In M. Jayson, ed., Total hip replacement. London: Sector Publishing, London 26–46.
Scales, J. T. 1967. Arthroplasty of the hip using foreign materials: A history. *Proceedings of the Institution of Mechanical Engineers* 181, no. 3: 63–84.
Smith-Petersen, M. N. 1939. Arthroplasty of the hip: A new method. *Journal of Bone and Joint Surgery* 21, no. 2: 269.
Stern, L. L. D., M. I. Cabadas, J. M. F. Fdez-Arroyo, J.M.F., R. Zarzoso, J. C. Sanchez-Barbero, and J. Granado. 1992. LD Hip arthroplasty: Design concepts and first clinical trials of a new modular system. *Clinical Orthopaedics* 283: 39–48.
Venable, C. S., and W. G. Stuck. 1947. The internal fixation of fractures. Springfield, IL: C. C. Thomas.
Waugh, W. 1990. John Charnley: The man and the hip. Berlin: Springer-Verlag.
Williams, D. F., and R. Roaf. 1973. Implants in surgery. London: Saunders.

PAPER 3

Originally published 1998

Author: Faulkner A, Kennedy LG, Baxter K, Donovan J, Wilkinson M, Bevan G

Effectiveness of hip prostheses in primary total hip replacement: critical review of evidence, and an economic model.

Published in: Health Technology Assessment, volume 2 No.6 (whole issue).

Acknowledgment: National Co-ordinating Centre for Health Technology Assessment for permission to reproduce this report.

Effectiveness of hip prostheses in primary total hip replacement: a critical review of evidence and an economic model

A Faulkner
LG Kennedy
K Baxter
J Donovan
M Wilkinson
G Bevan

Department of Social Medicine
University of Bristol

Published July 1998

This report should be referenced as follows:

Faulkner A, Kennedy LG, Baxter K, Donovan J, Wilkinson M, Bevan G. Effectiveness of hip prostheses in primary total hip replacement: a critical review of evidence and an economic model. *Health Technol Assessment* 1998;**2**(6).

Health Technology Assessment is indexed in *Index Medicus*/MEDLINE and *Excerpta Medica*/EMBASE. Copies of the Executive Summaries are available from the NCCHTA web site (see overleaf).

PAPER 3

Contents

PAPER 3

PAPER 3

List of abbreviations

CI	confidence interval*
GP	general practitioner
HA	hydroxyapatite
HBF	heterotopic bone formation*
HMHDPE	high molecular weight, high density polyethylene*
MACTAR	McMaster–Toronto Arthritis Patient Function Preference Questionnaire*
RCT	randomised controlled trial
RSA	roentgen stereophotogrammetry*
SEM	standard error of the mean*
SIFT	Service Increment for Teaching
SSTRR	Standard System of Terminology for Reporting Results*
THR	total hip replacement
WOMAC	Western Ontario and McMaster Universities Arthritis Index*

* Used only in appendix

NB: Abbreviations of the names of specific prostheses are given in the list of prostheses designs

PAPER 3

Prostheses designs

The following prostheses or prosthesis components are discussed in the publications covered by this review and, where possible, the name of the supplier/manufacturer of the prosthesis is given in parentheses. However, in some cases, the supplier/manufacturer was not identified in the original published document and the names of these prostheses are marked with an asterisk (*).

ABG (Howmedica)
Accu-Path*
ACS (DePuy)
AlloPro (Intermedics Orthopedics)
AML (DePuy)
AML Porcoat (DePuy)
Anatomic Medullary Locking – see AML
Anatomic Porous Replacement – see APR
APR (Intermedics Orthopedics)
ARC (Howmedica)
Arthrophor (Joint Medical Products)
ATS (Howmedica)
Aufranc-Turner (Howmedica/Zimmer)
Autophor (Smith and Nephew)
Balgrist (Germany)*
BIAS (Zimmer)
Bichat (Howmedica)
Bimetric*
Biofit (Smith & Nephew)
Biological Ingrowth Anatomic System – see BIAS
Boneloc® (cement)
Brunswick*
Buck 32*
Butel® (Smith & Nephew/Richards)
CAD (Howmedica)
Ceraver Osteal – see Osteal (Aluminia)
Charnley (DePuy)
Charnley Low Friction Arthroplasty – the name given to the original Charnley design
Charnley-Müller (DePuy)
CDH (Howmedica)
Christiansen (Delrin, Dupont)
CLS (Protek)
DF 80*
Dual Lock (DePuy, Zimmer or Protek)
Duraloc (DePuy)
Elite (DePuy)
Engh–Anderson (DePuy)
Exeter Polished (Howmedica)
Femora*
Freeman (Corin Medical)
Furlong (JRI)
Griss (Sulzer AG)
Harris (Howmedica)
Harris Design 2 (Howmedica)
Harris-Galante (Zimmer)
Harris Precoat (Zimmer)
HD-2 – see Harris Design 2
Honnart Patel-Garches*
Howse*
HP-Garches – see Honnart Patel-Garches
HS2P*
ICLH*
Indiana Conservative (DePuy)
Intermedic (Intermedics)
IOWA (Zimmer)
Kirschner Anatomic (Kirschner)
Kirschner Murray Welch (Kirschner)
Link V (Link America)
Lubinus SP (Waldemar Link)
LD*
LMT*
Lord Madreporic*
Mallory Head (Biomet)
Marburg*
McKee-Farrar (Howmedica)
Mecring (Mecron Medical)
Mecron (Mecron Medical)
MHP – see Mallory Head
Miami Orthopedic Surgical Club – see MOSC
Morscher*
MOSC (Biomet)
Müller (Straight Stem) (Protek AG)
Omnifit (Osteonics)
Optifix*
Osteal aluminia (Ceraver)
P-2 – see Protasul 2
P-10 – see Protasul 10
PCA® (Howmedica)
Pennsylvania Total Hip
Porous Coated Anatomic – see PCA®
Precoat – see Harris Precoat
Profile (DePuy)
Protasul 2 (Sulzer Brothers)
Protasul 10 (Sulzer Brothers)
Richard (Richards)
Ring (Zimmer)
RM (R Mathys Company)
Romanus (Biomet)
Scanhip (Mitab AB)
SixTi/28 (Zimmer)
SLF (Corin Medical)
Spectron/Biofit (Richards)
Spectron/ITH (Richards)
Spectron Lubinus – see SP Lubinus
Spectron EF (Smith & Nephew Richards)
SP Lubinus (Waldemar Link)
SRN-REV – see S-ROM Anderson
S-ROM Anderson (Joint Medical Products)
S-ROM Super (Joint Medical Products)
Stackhouse*
Stanmore (Biomet)
STH-2 (Zimmer)
T-28 (Zimmer)
Taperloc (Biomet)
TARA (DePuy)
Tharies*
TiBac (Zimmer)
Ti-Fit (Smith & Nephew Richards)
Titan (Landos)
Total Articular Resurfacing Arthroplasty – see TARA
TR-28*
Trapezoidal-28 – see T-28
Triad (Johnson & Johnson)
Trilock (DePuy)
T-TAP (Biomet)
Wagner*
Weber (Sulzermedica)
Whitesides (Dow Corning Wright)
Wrightington (Howmedica)
Zweymuller*

PAPER 3

Executive summary

Objectives

- To review available evidence on the comparative effectiveness of different prostheses types in total hip replacement (THR) for adults suffering primarily from osteoarthritis.
- To develop an economic model, using cost data from two NHS orthopaedic centres, to model the cost-effectiveness of alternative prostheses under varying resource input assumptions.

Methods

The reviewers had the benefit of a large in-house database. Additional searches were conducted in Medline, 1980–95, using a modified Cochrane strategy for identifying randomised controlled trials (RCTs). Separate searches were conducted in Embase, 1990–96, to identify studies with comparison or control aspects. Further details are given in the full report.

For inclusion, studies had to provide clinical outcome data for specified prosthesis designs, comprising functional assessment, radiographic data or time to failure. There were very few RCTs. Priority was given to studies with an element of comparison. Checklists and simple rating scales were used.

NHS price data and data relevant to costs were obtained directly from two NHS Trusts and their associated orthopaedic centres. The total expected costs of THR included an element for revision of the primary operation.

Results

Appraisal of studies

Most of the studies came from specialist orthopaedic centres; this has a bearing on the generalisability of the results of individual studies. The methodological quality of the studies was generally poor, for example, lack of sample size calculations.

Comparison of prosthesis types

The following tentative conclusions can be drawn about the performance of different types of prostheses. The various designs are described in the full report.

Cemented designs in general show good survival results at 10–15 years plus. Models with good, published, comparable results (at 10 years or more) include the Stanmore, Howse, Lubinus, Exeter and Charnley. The rate of acetabular revision in cemented implants remains problematic. Newer ('second-generation') cementation techniques usually give better results than more traditional techniques.

In comparing short- to medium-term longevity between **non-cemented porous-coated** and **cemented** prostheses designs, there is no clear advantage for either type. Thigh pain is a problem associated with non-cemented porous-coated implants to which cemented designs are not prone.

The small number of studies of **cementless hydroxyapatite (HA)-coated** models report mild to moderate thigh pain in between 0% and about 5% of patients at 2–5 years' follow-up, a good result compared with porous-coated implants.

Hybrid designs are comparable with the best cemented designs for early survival (6–7 years), superior both in terms of survival and thigh pain to porous-coated implants.

The **uncoated press-fit** and **resurfacing** types of hip prosthesis have survival results that are notably inferior to those of other types. Little evidence is available on **fully modular** prostheses.

Economic modelling

Using the economic model developed in this study, the general conclusions under our assumptions are summarised below.

Prosthesis cost, hospital costs and revision rate are the components of the model with the greatest impact in terms of changing total expected costs for THR procedures.

Very high and very low estimates of hospital costs change the total expected costs for individual prostheses but have little effect on their relative cost-effectiveness compared with each other.

PAPER 3

Compared with survival data for the Charnley cemented prosthesis from 'centres of excellence', and assuming a prosthesis cost of £353 including cement, even a 'no revisions' prosthesis should not cost more than about £650 (at 1997 prices) to have equivalent total expected costs over 20 years. Only cemented prostheses are currently available at this price.

In 70-year-old men, for example, a low price prosthesis is generally more cost-effective than a high price prosthesis, even with a very low revision rate. In 40-year-old men, prostheses with high prices and low revision rates can be more cost-effective than low priced prostheses with higher failure rates.

Conclusions

Policy implications

- The major concern is the proliferation of novel designs of prostheses whose effectiveness is unknown. Mechanisms for improving use of appropriate prostheses could be examined. Aspects to consider are suggested in the full report.
- Healthcare commissioners could model costs of alternative prostheses, using their local input resource assumptions and outcome data, along the lines of the model described.
- Commissioners and providers could also:
  - ascertain the range and extent of use of routinely used prostheses known to have results poorer than the best cemented designs, distinguishing different design types and taking account of age-groups, and seek audit of outcomes, including revision rates
  - in the case of significantly new designs, satisfy themselves that appropriate monitoring and evaluation is carried out.

Research recommendations

Some of the key recommendations from the main report are as follows.

General

- Improvements are needed in the design and reporting of research studies in this area.
- Further inclusion of patient-derived quality-of-life measures in studies of hip prosthesis performance is essential, as clinical hip-scoring systems do not take the patient's views into account when assessing outcomes.
- Patients' values and choices regarding quality of life in relation to THR should be investigated.

Prosthesis types

- Reporting of longer follow-up studies, especially of hybrid and cementless HA-coated models, is required in order to assess further their early promising outcomes. Follow-up of the coated acetabular component of hybrid implants is required to ascertain the medium- and long-term performance of this prosthesis design.
- Results for thigh pain and longevity in HA-coated models require longer follow-up periods. The extent and significance to patients of thigh pain associated with porous and HA-coated implants should be assessed. Longer follow-up assessments are also required for porous-coated cementless and fully modular designs.
- Further exploration is required of the associations between radiographic signs of loosening/migration and later mechanical failure.
- More up-to-date information is needed on the use of new cementation techniques, so that their use can be encouraged.

PAPER 3

Chapter 1

Introduction

Total hip replacement (THR) has, since the 1960s, become one of the most frequent orthopaedic procedures undertaken in the NHS; it is, in general, extremely effective in pain relief and improved physical function in, typically, patients aged 60 years or more who are suffering from osteoarthritis. It is an expensive procedure and substantial resources are devoted to it. In the UK, in the year 1994/95, some 32,500 primary replacements were performed within the NHS (according to Hospital Episode Statistics). THR performs very favourably in cost–utility studies that compare it with other surgical procedures.

In assessing prosthetic technology, it is easy to believe that if an optimal design for the implant could be created in bioengineering laboratories, then a standard optimal effectiveness could be defined and implemented across health services. However, THR is a clinical service and such a technical solution, even if it could be engineered, would not have this effect because of the wide range of other factors which necessarily contribute to overall outcomes of this intervention. These factors include:

- surgical technique
- surgical approach
- surgeons' experience
- operating theatre environment
- effects of prophylaxis for thrombosis and pulmonary embolism
- rehabilitation procedures
- patient factors such as bone quality and severity of disease.

Interpreting the evidence on the performance of different prosthesis designs is thus difficult.

Hip prostheses technology is continually changing and many new designs and methods of fixation have been experimented with since the original Charnley Low Friction Arthroplasty cemented concept of the 1960s. Some prosthetic designs have identifiably better outcomes than others, and some fail early and spectacularly. THR technology has been, to some extent, a victim of its own success as its use has been extended to include younger age groups and as increases in the longevity of implants are sought. Repeat THRs (revisions) perform notably less well than primary replacements and, clinically, revisions are regarded as something to be avoided if possible.

The rationale of supply and demand underlying the proliferation of alternative designs, fixation methods and surgical instrumentation is difficult to interpret. There is no statutory or nationally coordinated professional monitoring of processes of innovation and diffusion in the UK. Factors contributing to the difficulty of interpretation include:

- the commercial interest of manufacturing companies active in supplying the orthopaedic profession
- orthopaedic surgeons' creativity and ingenuity
- difficulties in interpreting the comparative results – especially short-term results – of different hip technologies.

In a somewhat critical discussion, orthopaedic innovation in THR technology internationally has been referred to as a 'trial-and-error culture'.[1] Interviews conducted by our research team have indicated that manufacturers exert, in various ways, a degree of influence over the prosthesis models which surgeons might prefer and the choice available to them. Some orthopaedic surgery departments, for example, are supplied by a single manufacturer.

It is known that over sixty different models of THR prostheses are used to at least some extent in the UK. Recent trends in new prosthesis technology have been towards new methods of cemented fixation and various designs of uncemented component. In most cases, uncemented components are significantly more expensive per unit than cemented components (in part, at least, reflecting more complex production processes), the price range in the UK being of the order of £300–400 to £1500–1600 in 1996/97. Such differences have major implications for the comparative cost-effectiveness of alternate technologies and, hence, for the total hospital and other costs associated with THR procedures.

PAPER 3

The major interrelated issues in the use of THR technology in the NHS are:

- the proliferation of new models in the market for prostheses
- the comparative performance following implantation of different types with different costs (performance includes primarily the longevity of implants and their effectiveness in pain relief and functional improvement).

More detailed current issues are:

- cemented versus cementless designs
- indications for different patient groups such as age groups
- optimal alloys for components in terms of elasticity,[2–4] biocompatibility[5] and abrasive wear[3]
- bearing surface materials at the interface of the head and cup components
- pain implicated with uncemented stem components
- the relative merits of different types of coating on uncemented components.

Some authors have suggested that design goals for the hip prosthesis are actually incompatible[1,6] and that this is seldom acknowledged.[1] There are a number of examples:

- strengthening of the cement–prosthesis interface may weaken the cement–bone interface and vice versa
- modular components must try to allow for optimal fit and for maximum initial stability at the same time
- stems must try to be flexible in order to avoid 'stress shielding' (leading to atrophy of the surrounding bone) but be stiff enough for initial stability to promote ingrowth of bone and avoid damage at the bone-prosthesis interface.

Aim of this review

The major focus of this review is on different types of prosthesis technology in terms of methods of fixation. Somewhat less attention is given to related issues such as cementation technique and the mechanics of component loosening associated with different metal alloys or other materials.

This report forms part of an extended systematic review of the effectiveness and cost-effectiveness of total hip prostheses. It includes a critical overview of published research literature on the performance of prosthetic technology in THR and an economic appraisal of the implications of different costs of prostheses in the light of evidence about survival of different models. This takes the form of an economic model which can be used – by healthcare purchasers or providers – to estimate the effects on total THR costs of varying the cost of resource inputs to the model, including the price of prostheses.

PAPER 3

Chapter 2

Indications for primary total hip replacement

THR is undertaken for severe degenerative joint disease, especially arthritis. The two main conditions treated by this approach are osteoarthritis and rheumatoid arthritis. Osteoarthritis is associated with advancing age while rheumatoid arthritis is more likely to occur in young adults. Other diseases treated by the procedure include avascular necrosis, congenital dislocation, Paget's disease, ankylosing spondylitis and traumatic arthritis.

There is uncertainty regarding both the definition of exact criteria for hip replacement surgery and the symptoms which might be associated with maximal benefit from surgery. Variation in the rates of THR surgery across regions in England[7] raise questions about the consistency of decision-making in general practitioner (GP) referrals and in choosing THR treatment. Orthopaedic surgeons have been "making-do without randomised trials"[8] of case selection for hip joint replacement.

Few studies have examined the question of optimal indications for THR in detail. Trials and other studies tend to take surgery as indicated and randomise patients by type of prosthesis or surgical method. Some consensus statements, aimed at distilling opinions on good practice in this area, have been made but the shortcomings of such approaches should be recognised. The US National Institutes of Health Consensus Development states that "candidates for THR should have radiographic evidence of joint damage and moderate to severe persistent pain or disability or both that is not substantially relieved by an extended course of non-surgical management".[9] A recent study employing the Delphi technique specified criteria for identifying appropriate patients for referral to a surgeon for consideration for arthroplasty.[10] Pain and functional status were the key criteria but age, ability to work and other important factors were also considered. Orthopaedic surgeons in the UK who were interviewed as part of this study also supported the primacy of pain and function, with pain being seen as the most important factor. An initiative in New Zealand, which aims at ensuring that those most in need are offered surgery, has involved the development of a scoring system, again using consensus methods, for determining priority for major joint replacement. Pain is the most important component of this score, with functional activity, movement ability, deformity, multiple joint involvement and ability to live independently all contributing to a lesser extent.[11]

It may be concluded that the principal indications for THR are pain and functional limitation; however, this conclusion is the result of consensus rather than primary research. Disease-specific pain is, of course, difficult to define both clinically and as an outcome measure in hip prosthesis follow-up studies.

While there is a basic consensus on the primary indications for total hip replacement *per se*, more detailed indications for the procedure are less clear. During the 1970s and 1980s, patients aged between 60 and 75 years were considered to be most suitable for the procedure; however, more recently, this age range has been extending in both directions. In younger age groups, procedures such as osteotomy and fusion may be considered as alternatives but there is no evidence to suggest that these are preferable. Data on potential risk factors such as age, weight and medication are insufficient guides to treatment for individual patients and there are no clear indications for different surgical approaches and techniques. Choices of different types of implant for different patient groups are surrounded by uncertainty and variations in surgical practice. The extent to which surgeons exercise choice of type of prosthetic component in relation to patient criteria such as age (as a proxy for activity level), diagnosis, bone stock quality and weight (body mass index) is unknown.

Chapter 3

Evolution of different types of prosthetic technology for total hip replacement

The periods of major development in THR technology from the initial cemented procedure to the more recent major design innovations are shown in *Table 1*.

The theoretical advantages and disadvantages of the different designs and fixation methods of hip prostheses are not discussed in detail here. Basic differences between the different types are described below.

In the 1970s, high failure rates of the early **cemented** THRs were found, which were characterised by bone loss (osteolysis) and mechanical loosening of prostheses. The cause was considered by many to be 'cement disease', that is, a direct reaction between the 'cement' (i.e. polymethyl methacrylate) and the body tissue. This belief was a major stimulus in the search for alternative solutions to the problem of long-term fixation and led to the concept of cement-free fixation. Methods of cementation have themselves evolved and are conventionally classified into the three 'generations' with the characteristics noted in *Table 1*.

Various **cement-free** methods have been developed which can be summarised broadly as:

- **press-fit** methods, in which fixation is sought by closeness of fit between prosthesis and bone, often assisted mechanically by techniques such as threading and augmentation by screws, nails or pegs, and 'macro-interlock' design features such as ribbed stems designed to improve fixation by wedging
- **porous-coated**, in which cementless technology is treated at surfaces adjacent to bone with an inert microporous coating in the form of mesh or beads with the aim of encouraging ingrowth of bone into the prosthesis surface
- **hydroxyapatite- (HA-) coated**, which is similar to porous coating in concept but the surfaces adjacent to bone are coated with HA, a form of calcium phosphate ceramic considered to be biologically active and capable of direct chemical bonding to bone.

TABLE 1 *Major developments in THR technology*

Prosthesis type	1960s	1970s	1980s	1990s
Cemented				
1st generation Finger packing	1960s			
2nd generation Intramedullary femoral plug, cement gun, superalloys for stems		mid-1970s		
3rd generation (some still regarded as experimental) Pressurisation, porosity reduction, precoating, rough surface, centrisation			mid- to late 1980s	
Ceramic (heads/cups)		late 1970s		
Uncoated press-fit cementless		late 1970s		
Porous-coated cementless			early 1980s	
Hybrid (cemented stem/ uncemented cup)			early 1980s	
HA-coated cementless			late 1980s	
Fully modular			late 1980s/	early 1990s

In **hybrid** models, a cemented stem is combined with an uncemented cup, which retains the relatively good performance of cemented stems but substitutes possibly superior cement-free cups; this allows immediate weight-bearing and, hence, may be seen clinically as suitable for older patients unable to use crutches.

In the **fully modular** type of prosthesis, the problem of achieving close anatomical fit is tackled by making available a range of sizes of separate subcomponents of the total prosthesis, including the acetabular cup, the femoral

stem, and the separate sleeve and head of the femoral component. Manufacturers are developing increasing modularity, and an increase in modular connections in a prosthesis leads to increased production costs.

Ceramic heads and cups (among combinations of materials) have b in an attempt to lessen wear and t production of damaging particles surfaces of the prosthesis.

Chapter 4

Types of outcome measure

There are, broadly, three types of outcome measure available for review in the orthopaedic literature on total hip prostheses:

- the lifespan of the prosthesis, which is usually referred to as its 'survival' and is typically represented by survivorship analysis or revision rates (i.e. rates of replacement of prostheses)
- prosthesis function *in situ,* which is measured typically by one of several standardised clinical hip scoring systems (outlined below) and which includes symptomatic loosening
- radiographic definitions of possible failure, including bone loss (osteolysis), subsidence of the stem component, migration of the cup component and wear of materials.

In practice, revision rate/survivorship is reported differently in different studies. Many studies use 'survival analysis' to assess the longevity of implants. This calculates, in a given cohort, the number of implants surviving unrevised each year as a proportion of those still *in situ.* In terms of the performance of the prosthesis itself, the key criterion for failure is revision for **aseptic loosening**, that is, as far as possible the defined outcome is caused by characteristics of the prosthesis rather than by confounding factors such as infection or dislocation (which may be due to accidental falls). In practice, reporting of the causes of revision is variable and sometimes just the revision rate is given without any qualification regarding its interpretation. In such cases, it is impossible to know whether causes of revision normally irrelevant to the prosthesis technology, such as infection, have been included. Other dimensions used in prosthesis survival analysis, which may or may not be included in different authors' definitions of failure, include radiological evidence of loosening and patient tolerance of symptoms. Some studies include 'pending revisions', others do not.

The total mechanical failure rate is also frequently reported. This refers (usually) to revision rates caused by aseptic loosening combined with radiographic evidence of loosening, fracture or other mechanical failure of components.

The main clinical hip scoring systems are briefly described in *Table 2.* Scores are allocated by a clinician. The Harris and the Merle d'Aubigne systems are the ones most frequently reported in the studies reviewed here.

Pain is conventionally graded from 'mild' through 'modest' to 'severe'. Some studies report the proportion of patients found to be 'pain-free'. A problem with these grades is that studies rarely indicate whether pain is related to context; for example, whether it is related to a particular

TABLE 2 *Most frequently cited clinical hip scoring systems*

Merle d'Aubigne (also in a version revised by Charnley)	3 dimensions: pain, mobility, walk: each scored 0–6	Hip function graded as 'good; fair; medium; poor'
Harris	4 dimensions (score): pain (0–40), function (0–47), range of motion (0–5), absence of deformity (0–8)	Combined score = 100; < 70 poor; 70–79 fair; 80–90 good; 90–100 excellent (In studies, 'good' and 'excellent' results frequently classified together, and a 'mean Harris' combined score given for a cohort.)
Johnston	5 dimensions: pain, activity, limp/walk, walking aid, ambulation time	
HSS (Hospital for Special Surgery)	4 dimensions: pain, walk, range of motion, function: each scored 10 points	Combined: 32+ excellent; 24–31 good; 16–23 fair; < 16 poor

level of physical activity. Grading is undertaken by the clinician.

Thigh pain, which is fairly frequently reported in studies of uncemented implants because it appears to be a problem in at least some of these, has rarely been reported for cemented implants and cannot be inferred from the pain dimensions included in the common hip scales. This makes comparison of this aspect between the two broad technologies difficult. Some studies in which this outcome is compared for different types of prosthesis are reviewed later in the chapters in which the key results of clinical studies are presented.

Of the outcome measures commonly used in the studies reviewed, reporting of radiographic measurements is the most diverse and difficult to interpret. It is frequently stated that radiographic results do not correlate with clinical findings and prosthetic survival; however, this is a controversial subject in which the evidence from one study to another is conflicting. The possibility of predicting later failure from early radiographic measures is an important issue, especially in the context of the proliferation of unproven novel designs and technologies. In one study it was suggested that failure of the femoral component due to loosening can be predicted with 86% specificity and 78% sensitivity using standard X-ray techniques.[12] Migration/subsidence (of the stem component) of greater than 1.2 mm per year[12] or at 2 years post-implant[13] have been suggested as the best threshold for prediction of failure.

Radiographic evidence is treated as a standard outcome measure in many studies of hip prosthesis technology, regardless of its potentially predictive role. The most common dimensions analysed are loosening, migration (especially of the acetabular component) and subsidence (of the stem). Again, there is variation in the threshholds and criteria used by different authors in defining these measures. Other measures used are stability, presence of continuous radiolucent lines (indicative of possible loosening) around components, change in orientation of components, signs of wear or abrasion of prosthetic surfaces and presence of particulate debris associated with wear.

A number of negative outcome measures are not included in the summaries of studies critically appraised in this review because they do not aid comparison of performance of the prosthetic technology *per se*. These include:

- infection (for example, Ahnfelt and colleagues,[14] in their analysis of 15 different implant models, report no difference between implants when failure leading to revision for infection is the end-point)
- dislocation
- postoperative fracture (often associated with accidental falls)
- mortality
- intra-operative complications (e.g. blood loss)
- deep vein thrombosis and pulmonary embolism.

Measurement of quality of life, pain, activities of daily living and satisfaction, using patient-derived measures, are notably absent from the literature reviewed. Issues in the measurement of outcomes of THR have been discussed in more detail elsewhere (for example, by Heaton *et al.*[15]) and form the focus of a separate report commissioned by the NHS R&D Health Technology Assessment programme.

Chapter 5

Review methodology: search strategies, selection criteria and critical appraisal methods

Literature search

The reviewers had the benefit of a large bibliographic database compiled within their department to support a number of research projects on epidemiology and service provision for THR. Additional searches were conducted employing a modified Cochrane strategy for identifying randomised controlled trials (RCTs) on Medline, 1980–95, and broad criteria for THR/arthroplasty for 1995 on Medline and Embase. As the review project progressed, *ad hoc* searching identified a number of important studies published in 1996. Separate searches using a variety of terms (such as control*, versus, compar*, match*) were conducted on Embase, 1990–96, to identify studies with comparison or control aspects in the study design. The searching was limited historically because prosthesis models change continually; hence, collecting evidence on superseded models was considered to be unproductive. A number of individuals and organisations, for example, the Medical Devices Agency, were contacted directly.

Selection criteria

The following criteria were applied:

- identifiable type of total hip prosthesis, including named models not currently used in the UK
- clinical data given (excluding, for example, laboratory-only studies)
- patient group: adult with a primary diagnosis of hip arthropathies/congenital deterioration, excluding hip fracture
- follow-up period specified
- outcome definition for prosthesis failure to include survivorship and/or revision rate and/or radiographic criteria
- type of evidence: observational or experimental design
- stage of study (end/interim) reported
- only English language articles or abstracts included.

In addition, a large number of bioengineering and prosthesis retrieval studies of the mechanics of loosening, migration, subsidence, and laboratory studies of wear of material components have been collected; however, these are not reviewed in this report.

Excluded studies

Many studies were excluded from the review following inspection of the full text of retrieved articles. The reasons for exclusion at this stage included:

- unusual diagnostic profile of patient group
- lack of primary data included in the article
- rare and obsolete prosthesis design
- high proportions of revision operations in the study group.

The exclusions included a small group of studies of uncemented porous-coated designs with very short (2–3 years) follow-up periods.

Critical appraisal methods

A total of 233 studies giving primary data were included in the review.

It is difficult to isolate the outcomes associated with the prosthetic technology and design in THR from potentially confounding variables, especially surgical techniques, surgeon-specific factors and patient characteristics. Priority in the appraisal of studies has thus been given to studies with an element of comparison or control of these variables. There are very few RCTs to draw upon.

Checklists were used to control the appraisal process. Separate checklists were used for RCTs, non-controlled comparative studies and observational/cohort studies without comparative features. Each study was reviewed by one of the research team (either AF or GK). Blinding to author or affiliation was not employed. The checklist criteria are presented in chapter 6. These were adapted from the similar approach

PAPER 3

used by Cowley in her Medline-based review of the same subject for the Australian Institute of Health and Welfare.[16] Studies were given a rating A, A/B, B, B/C or C based on the extent to which the appraisal criteria were met. In the presentation of results, ratings A, B and C only have been used for the sake of simplicity; A/B ratings were deemed to be A-rated and B/C ratings were deemed to be B-rated. Definite failure to meet one of the key criteria resulted in a C rating. Studies which met all key criteria were rated A or B, depending upon performance against the other criteria. Criteria were not given explicit weights and were not regarded as of equivalent weight in these decisions. They thus have a subjective component. The ratings provide a simple method of summarising the quality of the studies reviewed using the checklist criteria and, hence, should be regarded as shorthand summaries of the detailed appraisals carried out for each study.

The main classifications used in structuring the presentation of studies in this report are the type of research design and types of prosthesis. The criterion-by-criterion appraisal of each study is presented in the appraisal tables in the appendix to this report. Key data were extracted for each study and these are presented in the data tables in the appendix. Individual written summaries of each RCT reviewed are presented at the beginning of the appendix.

Publication bias

The major focus of this review is on the comparative effectiveness of different prostheses rather than the effect size associated with prostheses compared to another type of intervention. Formal methods of assessment of publication bias cannot be applied to the small number of RCTs available for this review. It is possible that more reports of studies giving poor results associated with particular models may be published in non-English language journals, which were not included in this review, but the proportion of the reviewed studies reporting failures and 'poor' results is relatively high, suggesting fairly open editorial policies in which publication bias toward positive or 'good' results is not a major concern. A very small number of English-language abstracts of non-English articles have been included where sufficient information was available. These are noted in the appraisal tables in the appendix. Further abstracts were scrutinised but it was judged that, for the limited additional information likely to be obtained, translation was not warranted. No RCTs were covered by these abstracts. A more important factor affecting interpretation of results is the institutional origin of studies, the majority coming from specialist and teaching centres. Also noteworthy is the small proportion of studies in which some of the authors may be seen to have a direct vested interest in the commercial prospects of the particular component reported upon.

used by Cowley in her Medline-based review of the same subject for the Australian Institute of Health and Welfare.[16] Studies were given a rating A, A/B, B, B/C or C based on the extent to which the appraisal criteria were met. In the presentation of results, ratings A, B and C only have been used for the sake of simplicity; A/B ratings were deemed to be A-rated and B/C ratings were deemed to be B-rated. Definite failure to meet one of the key criteria resulted in a C rating. Studies which met all key criteria were rated A or B, depending upon performance against the other criteria. Criteria were not given explicit weights and were not regarded as of equivalent weight in these decisions. They thus have a subjective component. The ratings provide a simple method of summarising the quality of the studies reviewed using the checklist criteria and, hence, should be regarded as shorthand summaries of the detailed appraisals carried out for each study.

The main classifications used in structuring the presentation of studies in this report are the type of research design and types of prosthesis. The criterion-by-criterion appraisal of each study is presented in the appraisal tables in the appendix to this report. Key data were extracted for each study and these are presented in the data tables in the appendix. Individual written summaries of each RCT reviewed are presented at the beginning of the appendix.

Publication bias

The major focus of this review is on the comparative effectiveness of different prostheses rather than the effect size associated with prostheses compared to another type of intervention. Formal methods of assessment of publication bias cannot be applied to the small number of RCTs available for this review. It is possible that more reports of studies giving poor results associated with particular models may be published in non-English language journals, which were not included in this review, but the proportion of the reviewed studies reporting failures and 'poor' results is relatively high, suggesting fairly open editorial policies in which publication bias toward positive or 'good' results is not a major concern. A very small number of English-language abstracts of non-English articles have been included where sufficient information was available. These are noted in the appraisal tables in the appendix. Further abstracts were scrutinised but it was judged that, for the limited additional information likely to be obtained, translation was not warranted. No RCTs were covered by these abstracts. A more important factor affecting interpretation of results is the institutional origin of studies, the majority coming from specialist and teaching centres. Also noteworthy is the small proportion of studies in which some of the authors may be seen to have a direct vested interest in the commercial prospects of the particular component reported upon.

Chapter 6

Summaries of effectiveness studies and results of critical appraisal

The majority of studies of the outcomes of hip prostheses in primary THR are observational in design. Few RCTs have been published. This review has tried to maximise the use of studies with an element of comparison between prosthesis types. The most studied single model is the cemented Charnley. The great majority of studies have appeared in a small number of specialist orthopaedic journals and emanate from specialist orthopaedic centres and departments, mainly in teaching hospitals. About 12% of the reviewed studies originate in the UK. Length of follow-up is inadequate for the full evaluation of the longevity of more recently introduced types of prosthesis. The methodological quality of studies is, in general, low, especially notable being the lack of sample size calculation in any of the reviewed studies. In most studies, the sample sizes actually reported appear to be notably smaller than would be ideally recommended to achieve valid generalisable results.

Summary of study characteristics

Numbers of reviewed studies for different comparisons and prosthesis types

Meta-analyses

A single meta-analysis exists in the orthopaedic literature on total hip prostheses[17] and is commented on later in this chapter.

RCTs and comparative studies

A total of 78 RCTs and comparative studies are included in this review (17 RCTs, 61 other comparative studies).

The prosthesis (or prosthesis type) comparisons which these studies make possible overall are summarised in *Table 3*.

It can be seen that the most numerous comparisons in the literature are between alternative cemented designs, followed by

TABLE 3 *Numbers of trials and comparative studies included*

Type of prosthesis	Cemented – Charnley	Other cemented	Ceramic	HA-coated	Hybrid	Modular	Press-fit	Porous	Re-surfacing
Cemented – Charnley	1								
Other cemented	12	15							
Ceramic	1	3	1						
HA-coated	1	2							
Hybrid	2	2							
Modular		1				1			
Press-fit	2	5		5			4		
Porous	3	10	1	4	6		2	2	
Resurfacing	2	5	1					2	

NB: This table excludes six studies which did not fit into the above divisions, one being the meta-analysis referred to above. A blank cell indicates that no study for this comparison was found in the review.

PAPER 3

cementless porous-coated versus cemented, HA-coated and hybrid designs of prothesis.

Other RCT/comparative studies

There are 11 other studies which are not straightforward prosthesis versus prosthesis comparisons. These are dealt with separately from the above studies in chapter 10. They include comparisons of patient variables and fixation types and techniques, together with reports assessing why outcomes such as dislocation and fracture may occur.

Observational studies

A total of 145 observational studies were reviewed (see *Table 4*).

TABLE 4 *Numbers of observational studies reviewed*

Type of prosthesis	Number of studies
Cemented – Charnley	45
Other cemented	29
Uncemented – press-fit	13
Uncemented porous-coated	34
Uncemented HA-coated	9
Hybrid	6
Others	5
Total	**145**

A further six observational studies of uncemented threaded press-fit acetabular components were reviewed for data extraction purposes but not appraised, because this design has generally been abandoned.

Journals of publication

The journals in which the reviewed studies of hip prostheses mainly appear and the number of articles are presented in *Table 5* (RCTs and comparative studies) and *Table 6* (observational studies).

In *Table 5*, two studies which were published in separate journals *(Clinical Orthopaedics and Related Research* and *Canadian Journal of Surgery)* reported on the same cohort of patients and are jointly appraised in this report, although they are recorded separately here. The single meta-analysis study was not included.

Only four of the RCTs and comparative studies listed in *Table 5* appeared in non-orthopaedic journals; one each in: *Canadian Journal of Surgery, Australian and New Zealand Journal of Surgery, Investigative Radiology* and *Keio Journal of Medicine.*

TABLE 5 *Journals in which RCTs and comparative studies appear*

Journal name	Number of studies	Percentage of total
Clinical Orthopaedics and Related Research	20	25
Journal of Bone and Joint Surgery [American]	10	13
Journal of Bone and Joint Surgery [British]	12	15
Combined	22	28
Acta Orthopaedica Scandinavica	11	14
Journal of Arthroplasty	8	10
Archives of Orthopaedic and Trauma Surgery	6	8
Others	12	15

TABLE 6 *Journals in which observational studies appear*

Journal name	Number of studies	Percentage of total
Clinical Orthopedics and Related Research	41	28
Journal of Bone and Joint Surgery (American + British combined)	46	32
Journal of Arthroplasty	10	7
Acta Orthopaedica Scandinavica + Acta Orthopaedica Belgica	9	6
Others (e.g. *Orthopedics, Journal of Orthopaedics and Rheumatism, Orthopedic Clinics of North America*)	39	27

Only two of the observational studies *(Table 6)* appeared in non-orthopaedic journals or other specialist sources: *Canadian Journal of Surgery,* and *Australian and New Zealand Journal of Surgery.*

Overall, it is clear that a small number of specialist orthopaedic journals account for the vast majority of publications of primary research studies on total hip prostheses. Very few studies are published in generalist medical journals.

Country/area in which studies are conducted

RCTs and comparative studies

The countries of origin of the RCTs and comparative studies are shown in *Table 7*, both in terms of the number of studies and as percentages of the total. The one joint study, between the UK and Switzerland, is recorded for both countries. The single meta-analysis study is not included.

TABLE 7 *Country of origin of RCTs and comparative studies*

Journal name	Number of studies	Percentage of total
USA	34	43
Sweden	12	15
UK	9	11
Norway	4	5
Switzerland	4	5
Austria	3	4
Denmark	3	4
Finland	2	3
Canada	2	3
Others	6	7

NB: List includes one joint study between the UK and Switzerland, which is recorded for both countries. The single meta-analysis study is not included.

Notably, 47% of RCTs came from Sweden and 18% from the UK; 41% of the comparative studies came from the USA and 10% from the UK.

Observational studies

The approximate percentages of reviewed observational studies by country/area are presented in *Table 8*.

TABLE 8 *Country of origin of observational studies*

Country of origin	Percentage of total
UK	13
Other European	30
USA	41
Others	16

Types of hospital from which published research originates

The great majority of studies came from specialist orthopaedic centres or teaching hospitals. A small number of district general hospitals can be identified, and a few studies report multicentre results in which all types of hospital have been included. The preponderance of specialist centres must be borne in mind when interpreting the studies' results.

Diagnostic profiles of patient groups

The major primary diagnosis in the great majority of published studies is osteoarthritis of the hip. In most series the proportion of patients with osteoarthritis varies between 50% and 80%. The next most prevalent condition is rheumatoid arthritis. In studies of patients in younger age groups the proportion with rheumatoid arthritis tends to be higher. Very few of the cohorts studied have been confined to patients with a single disease entity. Few studies perform subgroup analysis by diagnostic group.

Maximum follow-up periods

The period of follow-up is obviously important in the evaluation of prosthesis technology. Where possible, the maximum stated follow-up period reported in each of the studies reviewed are summarised in *Table 9* (RCTs and comparative studies) and *Table 10* (observational studies). In those studies in which mean or median follow-up period only was stated, an informed guess of the likely maximum period has been made, based on other studies and publication dates. Comparison of these results gives an indication of the typical length of follow-up available from published studies for the different types of prostheses. Losses to follow-up and death mean that numbers of hips reviewed at the maximum period are frequently low.

TABLE 9 *RCTs and comparative studies: maximum follow-up periods*

Type of prosthesis	Number of studies	Approximate average maximum follow-up period (years)
Cemented – Charnley	12	13.5
Other cemented	52	9
Ceramic	4	9
Uncemented press-fit	16	4.5
Uncemented porous-coated	23	5
Uncemented HA-coated	11	3
Hybrid	6	5
Modular	1	6
Resurfacing	4	6.5

PAPER 3

TABLE 10 *Observational studies: maximum follow-up periods*

Type of prosthesis	Number of studies	Approximate average maximum follow-up period (years)
Cemented – Charnley	45	15.5
Other cemented	29	10
Uncemented press-fit	13	7–8
Uncemented porous-coated	33	7
Uncemented HA-coated	10	4
Hybrid	6	6

It can be seen clearly from the tables that follow-up periods for non-cemented types are on average only short term. The longest follow-up period of studies of non-cemented prostheses included in this review was about 10 years.

Sample size

None of the reviewed studies of any type reported a prospective calculation to estimate required sample size.

Sample size as a criterion was not included in the checklist of criteria because the range of different outcome measures used in studies have different implications for sample size. A large difference in clinical hip scores, for example, may be detected with a relatively small sample but small differences in relative survival of prostheses, when revision of the hip is the definition of survival, requires relatively large sample sizes.

An analysis of the reviewed studies, according to our own sample size calculations under the assumptions for each type of study design, is presented below.

RCTs and comparative studies

Two sample size calculations have been made, one stringent and the other more relaxed. In the first, in order to detect a difference of 4% in the survival rates between two prosthesis designs, assuming an expected survival rate of 90% at 20 years (95% confidence interval, 80% power) would require an achieved sample size of 1085 hips per arm. Allowing for death during follow-up, and assuming mean age at operation of 65 years, an initial sample size of 3600 hips per arm might be required at this level of stringency (fewer for younger age groups). Alternatively, assuming a survival rate of 80%, in order to detect a relatively large difference of 10% in prosthesis survival would require an achieved sample size of some 313 hips per arm, or about 480–500 hips per arm if a total follow-up of 300 hips per arm was required for 10 years.

The numbers of reviewed RCTs and comparative studies of prosthesis versus prosthesis meeting these sample size criteria are as follows:

- > 3600 per arm – none
- > 1000 per arm – one (Havelin, *et al.*, 1994[18])
- > 480 per arm – none except the above[18]
- > 300 per arm – the above[18] plus two (Ebramzadeh, *et al.*, 1994;[19] Schreiber, *et al.*, 1993[20]).

In summary, only three of the RCTs and comparative studies reviewed have sample sizes enabling statistically valid comparisons for hip survival between prosthesis groups over the defined minimum of 300 hips.

Observational/cohort studies

A sample size of 600 hips is required in order to have a 95% confidence interval with an estimated precision of ± 0.04 (i.e. confidence interval with width of 8%), for an assumed 60% prosthesis survival rate at 20 years. Larger samples would be required in order to have the same precision for higher percentage survival assumptions. In comparison to studies of the survival of cemented prostheses with long follow-up, a 60% survival rate is a conservative assumption. This sample size estimate assumes that total follow-up is possible which, in practice, it is not because of the death of a proportion of the initial recipients of a prosthesis (this is likely to be about 70% if the mean age at operation is 65 years).[21] Thus, to achieve total follow-up of about 600 patients would require an initial cohort of some 2000 patients (this figure does not take account of the fact that most studies include a proportion of surviving patients who are lost to follow-up).

Only three of the 145 observational studies reviewed include cohorts fulfilling this sample size criterion, those by Ahnfelt and colleagues,[14] based on the multicentre Swedish registry, Dall and colleagues,[22] both in cohorts given the Charnley prosthesis, and Mohler and colleagues,[23] following-up the Iowa Hip cemented prosthesis.

The great majority of observational studies have cohort sizes of between 100 and 500 hips.

Appraisal results

Full results of the appraisal of individual studies are given in the appraisal tables in the appendix. The numbers of A-rated studies and the percentages of

studies meeting the individual appraisal criteria are summarised below. The rating procedure used is that described in chapter 5.

Study using meta-analysis

The single meta-analysis study[17] combines the data from observational studies internationally of uncemented press-fit threaded acetabular cups and compares the results to cemented and porous-coated control groups, also using combined data. Extracted data from the study are included in the data tables in the appendix but the study has not been included in the checklist appraisal. The practice of combining data from observational studies is, in principle, methodologically weak because of the effects of confounding and selection and other biases in single, observational, non-randomised studies. Even given the very careful selection of studies which the authors achieved, it is difficult to account fully for possible sources of heterogeneity in combined studies.[24] However, it should be noted that the results of the major comparison of relative effectiveness in this meta-analysis are supported by orthopaedic surgical practice internationally, in which use of the threaded cup design has been largely abandoned.

A-rated RCTs and comparative studies

The numbers of A-rated RCTs and comparative studies are presented in *Tables 11* and *12*.

The number of A-rated observational studies are presented in *Table 13*.

Proportions of studies meeting appraisal criteria
The overall appraisal results for each criterion

TABLE 11 *RCTs and comparative studies: number of A-rated studies for prosthesis versus prosthesis comparisons*

Type of study	Number of studies	Number of A-rated studies
RCT	15	1[25]
Comparative	51	0

TABLE 12 *Other RCTs and comparative studies: number of A-rated studies*

Type of study	Number of studies	Number of A-rated studies
RCT	2	0
Comparative	10	2[26,27]
NB: Excludes the meta-analysis study.		

TABLE 13 *Observational studies: numbers of A-rated studies for each prosthesis type*

Type of prostheses	Number of studies	Number of A-rated studies
Cemented – Charnley	45	7
Other cemented	29	12
Uncemented press-fit	13	1
Uncemented porous-coated	33	9
Uncemented HA-coated	10	6
Hybrid	6	1
(See appraisal tables in appendix for details)		

for each type of study design are summarised in *Tables 14–17*.

In addition to the appraisal results for RCTs presented in *Table 14*, two other RCTs which do not make prosthesis versus prosthesis comparisons have been reviewed. They meet the appraisal criteria as follows:

criterion 1 – neither
criterion 2 – 1
criterion 3 – neither
criterion 4 – neither
criterion 5 – 1
criterion 6 – 2
criterion 7 – 2
criterion 8 – 2
criterion 9 – 2
criterion 10 – 2
criterion 11 – 2
criterion 12 – 1
criterion 13 – neither
criterion 14 – 2
criterion 15 – neither
criterion 16 – 1
criterion 17 – neither.

A number of general comments on the design of studies of hip prosthesis outcomes in the orthopaedic literature can be made on the basis of the results presented in *Tables 14–17*.

Methodological weaknesses in the studies are of particular concern in the description of study group characteristics from which representativeness might be assessed. Only one in three of the RCTs identified the method of randomisation. In the comparative studies, the major weaknesses are in descriptions of the process of assignment of patients to prosthesis groups, and in establishing the comparability of patients in comparison groups (either through matching or statistical analysis). Fewer than half of the observational studies give an account of the selection of patients included in the study (in addition, there are very few descriptions of clinical indications for prosthesis choice). Review of the observational series suggests that prosthesis selection practice is based largely on the preferences of clinicians and surgical centres.

Prospective sample size calculations have not been evident in the scientific orthopaedic literature on the clinical results of hip prostheses.

Most studies have made neither clinical nor radiological evaluations independently of the operating surgeon nor has blinding to the intervention been employed where appropriate.

TABLE 14 *Appraisal summary of prosthesis versus prosthesis RCTs (n = 15)*

Key criteria	Number meeting criterion (%)
1 Method of randomisation identified and appropriate	5 (33)
2 Patient groups balanced or effect of any difference evaluated in valid statistical analysis	11 (73)
3 Patients blind to prosthesis type	2 (13)
4 Assessments of clinical/radiological outcome blind to prosthesis type if possible	2 (13)
5 Appropriate statistical analysis	10 (67)
6 Number of patients deceased or lost to follow-up reported or included in statistical analysis	13 (87)
7 Follow-up period – mean and range	15 (100)
8 Prosthesis model specified	15 (100)
9 Clearly defined criteria for measuring outcomes	14 (93)
10 Age – mean and range	14 (93)
Other criteria	
11 Quantification of outcomes	13 (87)
12 Follow-up data compared with preoperative data (preferably mean **and** range)	8 (53)
13 Independence of investigators (declared or no vested interest)	6 (40)
14 Numbers of men and women given	11 (73)
15 Weight – mean and range	9 (60)
16 Preoperative diagnoses with percentages/ numbers of patients given	13 (87)
17 Clinical evaluation independent of operating surgeon	4 (27)

TABLE 15 *Appraisal summary of prosthesis versus prosthesis comparative studies (n = 51)*

Key criteria	Number meeting criterion (%)
1 Method of assignment of patients to different prostheses described, and appropriate	17 (33)
2 Patients matched or differences evaluated in valid statistical analysis	14 (28)
3 Appropriate statistical analysis undertaken	32 (64)
4 Number of patients deceased or lost to follow-up reported or included in analysis	29 (57)
5 Follow-up period – mean and range	39 (76)
6 Prosthesis models specified	46 (90)
7 Clearly defined criteria for measuring outcomes	46 (90)
8 Age – mean and range	37 (73)
Other criteria	
9 If retrospective, patients selected without knowledge of outcomes	27 (71)
10 In prospective studies, follow-up assessments blind to prosthesis type, if possible	2 (15)
11 Results given for specific models (and sizes)	38 (75)
12 Quantification of outcome criteria	41 (80)
13 Follow-up data compared to preoperative data (mean and range)	10 (20)
14 Independence of investigators (declared or no vested interest)	12 (24)
15 Numbers of men and women given	41 (80)
16 Weight – mean and range	14 (27)
17 Preoperative diagnoses with percentages/ numbers of patients given	34 (67)
18 Clinical evaluation independent of operating surgeon	4 (10)
19 Radiological evaluation independent and blinded to clinical results	3 (6)

PAPER 3

TABLE 16 *Appraisal summary of other comparative studies (n = 10)*

Key criteria	Number meeting criterion (%)
1 Method of assignment of patients to different prostheses described and appropriate	6 (67)
2 Patients matched or differences evaluated in valid statistical analysis	7 (78)
3 Appropriate statistical analysis undertaken	8 (80)
4 Number of patients deceased or lost to follow-up reported or included in analysis	9 (90)
5 Follow-up period – mean and range	7 (70)
6 Prosthesis models specified	9 (90)
7 Clearly defined criteria for measuring outcomes	9 (90)
8 Age – mean and range	9 (90)
Other criteria	
9 If retrospective, patients selected without knowledge of outcomes	5 (56)
10 In prospective studies, follow-up assessments blind to prosthesis type, if possible	NA
11 Results given for specific models (and sizes)	6 (60)
12 Quantification of outcome criteria	9 (90)
13 Follow-up data compared to preoperative data (mean and range)	3 (30)
14 Independence of investigators (declared or no vested interest)	4 (40)
15 Numbers of men and women given	8 (80)
16 Weight – mean and range	4 (40)
17 Preoperative diagnoses with percentages/ numbers of patients given	9 (90)
18 Clinical evaluation independent of operating surgeon	0 (0)
19 Radiological evaluation independent and blinded to clinical results	2 (22)

NA, not applicable

TABLE 17 *Appraisal summary of observational studies (n = 145)*

Key criteria	Percentage meeting criterion	Number meeting criterion/ number applicable*
1 Method of selection of patients identified	43	58/135
2 Prosthesis models specified	100	145/145
3 Results given for specific models	94	127/135
4 Follow-up period – mean and range	84	113/135
5 Number of patients lost to follow-up or deceased – reported or included in analysis	78	97/125
6 Age – mean and range	79	107/135
7 Preoperative diagnoses of reviewed patients stated with percentages/numbers	83	112/135
8 Clearly defined criteria for measuring outcomes/ quantification of outcomes	91	112/121
Other criteria		
9 Valid statistical analysis	50	63/125
10 Outcome data compared to preoperative data	40	44/111
11 Data given for deceased patients	25	23/93
12 Clinical evaluation independent of operating surgeon	8	9/112
13 Radiological evaluation independent and blinded to clinical results	14	16/112
14 Numbers of men/women stated	89	102/115
15 Weight – mean and range	23	27/115
16 Surgical technique/approach stated	77	89/115
17 Grade/experience and number of surgeons stated	47	54/115
18 Type of hospital/centre (general/ specialist/teaching) stated	81	93/115
19 Unilateral/bilateral results separate	0	0/145
20 Independence of investigators (vested interest) stated	28	32/115

* *The 'number applicable' for each criterion varies for several reasons: a number of studies which clearly failed by one or more of the key criteria were not appraised for the remaining criteria; some criteria (e.g. data given for deceased patients) were not applicable in some studies; in a very small number of studies from which data have been extracted, only an abstract was available. The denominator in calculating percentages has been adjusted accordingly.*

Chapter 7

Effectiveness of hip prostheses: a summary of key results from clinical studies

Given the methodological quality of the reviewed studies, results for different types of prostheses should be treated as estimates with wide confidence intervals. The majority of studies come from specialist centres and this is likely to have a bearing on the generalisability of the results. Clinical outcomes are measured by prosthesis survival, radiographic measurement and hip scoring. Clinical hip scoring is likely to underestimate the qualitative significance for recipients of hip implants of pain and function.[28] Taking these points into account, the following conclusions can be drawn about the performance of different types of prosthetic hip technology on the basis of the evidence summarised in this chapter.

Cemented designs in general show good survival results at 10–15 years plus. Models with good, published, comparable results (at about 10 years or more) include the Stanmore, Howse, Lubinus, Exeter and Charnley designs. The rate of acetabular revision in cemented implants remains problematic. There is some evidence that all-polyethylene acetabular components are preferable to metal-backed designs in terms of longevity of the implant.

Evidence of short-term comparisons between non-cemented porous-coated designs and cemented designs is equivocal. One comparative radiographic study suggests that cemented acetabular components performed better than porous-coated designs but that porous-coated stems performed better than cemented models. The first 10-year survival results for porous-coated models appear to bear comparison with the cemented models for the same follow-up period, especially when the relatively lower average age of the patient groups implanted with the porous-coated models is taken into account.

The comparative evidence strongly suggests that thigh pain is a problem associated with porous-coated (and other cementless) implants, to which cemented designs are not prone. In the observational studies of porous-coated implants, reports of thigh pain prevalence ranged between about 2% and about 25% at 2–7 years follow-up, with several studies reporting prevalence values at about the higher 25% level, including in non-loose stems. In contrast, in the small number of studies of HA-coated models, mild to moderate thigh pain was found in between 0% and about 5% of patients at 2–5 years follow-up. This is a relatively good result in comparison to reports of porous-coated implants and requires further substantiation.

Radiographic studies of cemented versus HA-coated designs suggest that HA-coated models have better early fixation and less migration than cemented models. The lesser migration of HA-coated models may be associated with less early postoperative pain, according to one comparative study. With maximum follow-up periods of only 3–4 years for this form of fixation, longer-term studies of survival and clinical results are required.

Hybrid prostheses appear to do well in the short term but the available studies cannot give any indications for their mid- or long-term results. Given wide confidence intervals, for early (6–7 years) survival this type of design can be regarded as comparable with the best cemented designs. Early survival is superior to that for uncemented porous-coated implants, and early thigh pain in cemented stem components in hybrid implants is minimal or absent compared with porous-coated designs. Longer follow-up, especially of the coated acetabular component of hybrid implants, is required to ascertain the medium- and long-term performance of this design.

Little evidence is available on fully modular prostheses. Theoretically, modularity permits greater intra-operative flexibility for the surgeon and potentially better component fit but further evidence, especially in comparison to cemented implants, is required. One comparative study suggested that a fully modular stem performed less well than cemented stems. Laboratory analysis of retrieved components suggests that mixed-alloy components are more prone to corrosion than single alloy devices.[29]

Evidence for the performance of ceramic hips is equivocal. Wear rates are less than for other materials at the articulating surface of the joint.

Comparative studies have suggested either lower or equivalent revision rates for ceramic versus cemented implants at medium-term follow-up. The implications of laboratory studies of alternative bearing surface materials require further investigation.

The uncoated press-fit and resurfacing types of hip prosthesis generally have survival results notably inferior to the other types of design available.

In the following three chapters the results of the review are presented, as follows:

- results of each comparison of different types of prosthesis taken from RCTs and comparative studies (chapter 8)
- results from selected observational studies for different types of prosthesis (chapter 9)
- results from comparative studies on selected key issues, including thigh pain, bearing-surface materials, inter-surgeon and inter-hospital comparisons (chapter 10).

A brief summary of each RCT included in this review is presented in the appendix (see page 83).

PAPER 3

Chapter 8

Results of studies comparing different types of prosthesis

Each type of prosthesis was compared, in turn, to all other types of THR where permitted by the available studies. For some comparisons there were more than two or three relevant studies which could be summarised (see *Table 3*) and, in this situation, only those studies with the highest rating, longest follow-up and/or survivorship results are included. One comparison noted in *Table 3* (i.e. ceramic versus resurfacing) is not included, because of poor reporting of data in the paper.

Charnley versus Charnley

Comment

Only one paper compared one form of Charnley with another. Flanges on the prosthesis may reduce the incidence of loosening but this was a small study with a C-rating and so no definite conclusions can be drawn.

Hodgkinson and colleagues, 1993[30]

The Charnley hips were implanted flanged (n = 168) or unflanged (n = 182). The patients were well matched statistically, with an approximate mean age of 58–60 years, and 70% had osteoarthritis. All the prostheses were inserted by the same surgeons. Accurate data at 9–11 year follow-up were available for 302 prostheses (152 unflanged, 150 flanged). In total, there were 15/350 (4.3%) revisions (which included 11 patients who were excluded from further analysis because the revisions occurred before the 10-year follow-up), nine of which had radiological evidence of loosening. Analysis of demarcation lines (or the extent and width of any radiolucent lines) showed the flanged prostheses in a better light than the unflanged. No demarcation was seen at the 10-year follow-up in 43% of flanged hips compared with 30% in the unflanged group. Similarly, 19% of flanged prostheses had demarcations grades of 2 or more (indicating radiographic loosening) compared with 25% in the unflanged group. (These differences were significant but the statistical level was not indicated.)
C-rated.

Charnley versus other cemented prostheses

Comment

Three papers were selected from others in the same category on the basis of length of follow-up and survival analyses. All of the papers making this comparison were C-rated and so leave doubts about the validity of the results. In the first two papers summarised here, Charnley is the superior hip, with Lubinus and T-28 giving closely comparable results. However, the third paper suggests the opposite – there is still an 84% survival rate for the Charnley at 10 years but this is neither as high as in other reports nor as high as the Stanmore with which it is compared (however, as no patient details were given no comment can be made on any effect this may have had). Overall, the Charnley prosthesis gives consistently good mid- to long-term results. It cannot be concluded with certainty that any of the other cemented prostheses is consistently equal to the Charnley.

Ahnfelt and colleagues, 1990[14]

Results from this retrospective study were taken from a Swedish multicentre registry. The patients had an approximate median age of 64 years for women and 66 years for men, with a main diagnosis of osteoarthritis. In all cases, the hips had been implanted originally between 1979 and 1986 and had required revision. Survival without revision for loosening of the original THR was observed for various prostheses. The Charnley hips had a 92% survival at 10 years (n = ?). The observed survival in eight out of ten other cemented prostheses for which results were quoted ranged from 63% (Christiansen) to 89% (Stanmore) at 10 years and from 95% (Exeter) at 5 years to 93% (Lubinus) at 9 years. **C-rated.**

Ritter, 1995[31]

The Charnley prosthesis (n = 260) was compared with four other cemented hips: Müller (n = 163), T-28 (n = 642) and 319 MOSC hips, the latter with either an all-polyethylene cup or a metal-backed cup. The average follow-up time ranged from 8.9 to 12.7 years. One surgeon performed all the operations. In all, 66% of the patients had

PAPER 3

osteoarthritis and their mean ages ranged from 59 years (Charnley) to 76 years (Müller). Over 200 patients were lost to follow-up. Of the Müller and MOSC metal-backed, 20% failed within 1 year of the operation, while only 9% of the MOSC all-polyethylene failed (compared with 14% Charnley and 10% T-28) within the same period. An analysis of survival at 10 years showed the Charnley and T-28 to be superior: Charnley, 93% (20 years = 76%), and T-28, 93% (17 years = 75%), MOSC all-polyethylene, 90% (12 years = 87%), Müller, 81% (17 years = 56%), and MOSC metal-backed, 60%. **C-rated.**

Britton and colleagues, 1996[32]

In this prospective study, 205 Charnley and 982 Stanmore prostheses, implanted or implant supervised by one surgeon, were compared after a median follow-up period of 8 years. No patient details were given. In all, 7% of the hips were revised, 38/81 because of loosening. When 'revision' was the end-point, Charnley hips were reported to have an 84% survival rate at 10 years compared with 93% for Stanmore hips and, when the end-point was 'onset of slight pain', a 44% survival rate at 10 years compared with 48% for Stanmore. The survivorship curves were reported to be similar for both prostheses for the first 8 years. Beyond this, Charnley hips became significantly worse irrespective of the end-point chosen, that is, need for revision or the onset of different levels of pain. **C-rated.**

Charnley versus ceramic

Comment

The number of ceramic hips was too small in this study to draw conclusions about them or their performance in relation to the Charnley.

Hoffman and colleagues, 1994[33]

This retrospective study involved a total of 1166 hips in 974 patients. Six surgeons (grade not specified) performed the operations in either a publicly funded hospital (66%) or at a private clinic (34%). The patients involved had an average age of 66 years and 89% had a diagnosis of osteoarthritis; there were slightly more males (54%) than females and 55% of operations were to the right side of the body. This section of the report refers to only part of the overall study. The Charnley prostheses (n = 867) had 72 failures recorded, six caused by loosening; with an annual failure rate of 1.78% and a survival rate of 73% at 15 years. The ceramic prosthesis (Autophor, n = 35) had a failure rate of 15% in 3 years, giving an annual rate of 5%. **C-rated.**

Charnley versus HA-coated

Comment

The numbers involved in this study were too small and the length of follow-up too short for any differences between the two prostheses types to be demonstrated unless one performed extremely badly. Further studies are needed in this area as it is an important comparison which has been neglected.

Bradley & Lee, 1992[34]

In this RCT, the Furlong (HA-coated prosthesis, n = 97) was compared with the Charnley (n = 73). The patients (with primary osteoarthritis, average age 68 years, range 45–75 years) were randomly allocated by year of birth. A total of 139 patients were reported on at 1-year follow-up and 74 at 2-year follow-up. The results at 1 year and 2 years, based on the Harris Hip Scores, were very similar in both groups (no pain: Furlong 98%, Charnley 96%, $p = ?$; number without a limp, walking distance and use of walking aids, ability to climb stairs and put on shoes/socks, all had "uniformly good average results"). There were no revisions or evidence of loosening. **C-rated.**

Charnley versus hybrid

Comment

Hybrid prostheses appear to do well in the short term but the available studies cannot give any indications for their mid- or long-term results. Both studies here had a C-rating. The first had very small numbers and, although a statistical difference regarding absolute rotation was noted, it is not clear if this would translate into greater problems for the Harris-Galante prosthesis in later years and so it is difficult to comment on this result.

Onsten and colleagues, 1994[35]

Charnley stems were implanted in both bilateral hips under the same anaesthesia in 29 patients in this RCT. One hip had a Harris-Galante type 1 cup and the other a Charnley cup. At 27 month follow-up, 21 patients (diagnosis, osteoarthritis; age range, 41–76 years) were assessed. Five Charnley and three Harris-Galante prostheses did not migrate or rotate. Mean values of absolute migration between the groups in any direction did not differ ($p = 0.06–0.98$) but the mean values of absolute rotation did ($p = 0.08–0.008$) – the Harris-Galante hips rotating the most. **C-rated.**

***Callaghan and colleagues, 1995*[36]**
As part of a larger study, 330 Charnley hips (follow-up minimum 20 years) and 89 Charnley hips in patients less than 50 years old (follow-up, 16–22 years) were compared to 130 hybrid hips (Harris-Galante type 1 cup plus IOWA cemented stem; follow-up, 5 years) and 61 similar hybrids used as revisions (follow-up, minimum 5 years). No patient details were given. In the Charnley groups the cup revision incidence was 10.6% in the 20-year follow-up and 13% in the younger age group; loosening was 12.8% and 37%, respectively. The stem revision incidences in the two Charnley groups were 3.2% and 2.2% for the 20-year follow-up and the young age group, respectively, while loosening was reported in 4.3% and 6.1%, respectively. This is in contrast to the hybrid groups where no revisions (or re-revisions) were reported and only one migration in the revision group occurred; however, follow-up was only for 5 years or more. Wear rates were also assessed in this study, with the Harris-Galante cups (28 mm head) reported as having less wear than other groups, but unfortunately no details were given. **C-rated.**

Charnley versus press-fit

Comment
Even at 1-year follow-up, the Charnley showed better results than the press-fit design. A reduced probability of survival, higher revision rates caused by loosening, increased risk of subsidence and general lack of performance does not inspire confidence in this type of prosthesis in comparison to the Charnley.

***Wykman and colleagues, 1991*[37]**
Charnley and Honnart Patel-Garches prostheses were each inserted into 75 patients during an RCT; 15 patients in each group had bilateral arthroplasties (age range, 29–82 years; osteoarthritis, 77%). The two prostheses had a similar probability of survival at 5–6 years approximately (Charnley, 88%; Honnart Patel-Garches, 82%; *p*, not significant). More revisions were required in the Honnart Patel-Garches group over 5 years (18.7%, all for loosening, all but one causing mid-thigh pain) compared with the Charnley group (11%, five for loosening, no mid-thigh pain). A further five in the Honnart Patel-Garches group had a possible need for revision caused by mid-thigh pain (increasing the revision rate to 25%). Subsidence of more than 4 mm occurred in 5% (Charnley) and 33% (Honnart Patel-Garches). **C-rated.**

***Olsson and colleagues, 1985*[38]**
A total of 119 patients had either a cemented (Charnley, n = 61, mean age 67 years) or non-cemented (Honnart Patel-Garches, n = 59, mean age 64 years) prosthesis implanted; 82% of patients had osteoarthritis. Clinical evaluation showed similar preoperative results but the Charnley prosthesis performed better at the 1-year assessment – Harris Hip Score and Limp, Charnley versus Honnart Patel-Garches, $p < 0.001$; maximal walking speed, Charnley versus Honnart Patel-Garches, $p < 0.05$ (twice as many patients with the Honnart Patel-Garches prosthesis required a device to assist them). A quantitative analysis of gait showed the latter group to have slightly better preoperative results but 1 year after surgery the Charnley group had greater improvement. No revisions were reported. **C-rated.**

Charnley versus porous-coated

Comment
In both of the studies summarised below relatively few porous-coated prostheses were assessed; the porous hips had only short- to mid-term follow-up compared with the Charnley hips. It is therefore very difficult to make conclusive judgements about relative performance. However, it would appear that porous hips have good short-term survival and, if this were to continue in the longer term, may be comparable to the Charnley.

***Callaghan and colleagues, 1995*[36]**
As part of a larger study, 330 Charnley hips (minimum follow-up of 20 years) and 89 Charnley hips in patients under 50 years of age (16–22 years follow-up) were compared with 100 PCA prostheses (minimum follow-up of 7 years). No patient details were given. In the Charnley groups, cup revision incidence was 10.6% in the 20-year follow-up group and 13% in the younger age group; loosening was 12.8% and 37%, respectively. Stem revision incidences in the two Charnley groups were 3.2% and 2.2%, respectively, with loosening reported in 4.3% and 6.1%, respectively. The porous-coated group had a cup revision incidence of 4% and a migration incidence of 5%, which included two revised cases. **C-rated.**

***Hoffman and colleagues, 1994*[33]**
This retrospective study involved a total of 1166 hips in 974 patients. Six surgeons (grade not specified) performed the operations in either a publicly-funded hospital (66%) or at a private clinic (34%). The average age of patients was

66 years, 89% had a diagnosis of osteoarthritis, there were slightly more men (54%) than women and 55% of operations were to the right side of the body. This section of the report refers to only a part of the overall study. There were 72 failures of the Charnley prostheses (n = 867) recorded, six caused by loosening; with an annual failure rate of 1.78% and a survival rate of 73% at 15 years. Neither the Harris-Galante (n = 105) nor the PCA prostheses (n = 38) had any revisions in approximately 3–4 years of follow-up. **C-rated.**

Charnley versus resurfacing

Comment

The results do not encourage the use of resurfacing hip prostheses.

Ahnfelt and colleagues, 1990[14]

The results of this retrospective study were taken from a Swedish multicentre registry. The patients' median age was approximately 64 years for women and 66 years for men; the main diagnosis was osteoarthritis. In all cases the hips had been implanted originally between 1979 and 1986 and had required revision. Survival without revision for loosening of the original THR was observed for various prostheses. Charnley hips had a 92% survival at 10 years (n = ?) whereas the Wagner resurfacing hip prosthesis had only 28% survival at 10 years – the worst result in the study. **C-rated.**

Cemented versus cemented (non-Charnley)

Comment

There are a great number of cemented prostheses available and, from the selection of studies below, it can be seen that results can vary considerably. There have been many design modifications over time, some apparently beneficial and others not. Some modifications, such as in the well-known case of the Christiansen prosthesis (see, for example, Ahnfelt *et al.*, 1990[14]), have had disastrous results. From these studies, prostheses such as Stanmore and T-28 appear to give good results over mid-term follow-up as does the Spectron/ITH combination in the short-term. The Spectron/Biofit and Lubinus SP hips may prove to have good outcomes but the numbers were too small to draw any conclusions. The Ritter study[31] suggests that cemented all-polyethylene acetabular components perform better than cemented metal-backed components.

Ahnfelt and colleagues, 1990[14]

The results of this retrospective study were taken from a Swedish multicentre registry. The patients' approximate median age was 64 years for women and 66 years for men; the main diagnosis was osteoarthritis. In all cases the hips had been implanted originally between 1979 and 1986 and had required revision. Survival without revision for loosening of the original THR was observed for various prostheses. Eight out of ten cemented prostheses had results quoted for them. The observed survival ranged from 63% (Christiansen) to 89% (Stanmore) at 10 years and from 95% (Exeter) at 5 years to 93% (Lubinus) at 9 years. The Christiansen prosthesis gave very poor results compared with the other cemented types. (This 'trunnion-bearing device' was popular in the late 1970s in Sweden and in 5 years more than 5000 were implanted. By 1986, 1524 of them had been revised and survival analysis predicted that 200 more would require revision in the following 4 years). **C-rated.**

Ritter, 1995[31]

Four cemented hips were compared – 163 Müller, 642 T-28, and 319 MOSC hips with either an all-polyethylene cup or a metal-backed cup. The average follow-up time ranged from 8.9 years to 10.1 years. One surgeon performed all the operations. The mean ages of the patients, 66% of whom had osteoarthritis, ranged from 62 years (T-28) to 76 years (Müller); 13% of patients were lost to follow-up. Within 1 year of the operation, 20% of both the Müller and MOSC (metal-backed cup) failed, while only 9% of the MOSC (all-polyethylene) and 10% of the T-28 failed within the same time. Apart from metal-backing, the larger femoral head size of the Müller may also be implicated. An analysis of survival at 10 years showed the T-28 and MOSC (all-polyethylene) to be superior to the others: T-28 93% (75% at 17 years), MOSC (all-polyethylene) 90% (87% at 12 years), Müller 81% (56% at 17 years) and MOSC (metal-backed) 60%. **C-rated.**

Espehaug et al., *1995*[39]

In a Norwegian multicentre survey, 12,179 hips were followed-up for a mean of 3.2 years (maximum 6.4 years) at various hospitals. Approximately 50% of the patients were in the age range 65–74 years (diagnosis was not given). The 5-year failure rates for the prostheses ranged from 7.33% for Müller type hips (n = 116) and 4.96% for Spectron/Lubinus combinations (n = 302) to 0.85% for Spectron/ITH combination (n = 1034) and Spectron/Biofit (n = 152); Lubinus SP hips (n = 129) required no revisions. **C-rated.**

Non-Charnley cemented versus ceramic

Comment

There were too few results (all from C-rated studies with small numbers of patients) to make convincing statements regarding the relative benefits of ceramic and cemented. Ceramic prostheses were originally designed to reduce wear and one comparative study (Zichner & Willert, 1992,[40] in which the same surgeons performed all operations) showed this to have been achieved in the hips assessed, together with a lower revision rate than for cemented prostheses. The short-term results compared with cemented prostheses appear to be worth investigating further.

Zichner and Willert, 1992[40]

A Müller-type endoprosthesis was inserted into 354 hips in 313 patients between 1970 and 1980: 149 with a Protasul-2 ball (average follow-up 66 months); 105 with Protasul-10 ball (average follow-up 46 months); 100 with a ceramic ball (average follow-up 73 months). All the prostheses were implanted at the same clinic by the same surgeons using the same technique. As a result of loosening, 10% Protasul-2 and 4.8% Protasul-10 prostheses were revised compared with 2% of the ceramic types. Displacement rates were also assessed: 30% of the non-revised Protasul-2 ball hips had a displacement rate of > 0.2 mm/year, 8% with a displacement rate of > 0.3 mm/year; 20% of the non-revised Protasul-10 ball hips had a displacement rate of > 0.2 mm/year. However, 95% of all ceramic ball prostheses had a displacement rate of < 0.02 mm/year, with 63% having a displacement rate of < 0.1 mm/year. **C-rated.**

Schuller and Marti, 1990[41]

Weber type prostheses with metal heads were inserted in 48 patients at a teaching hospital and compared with 46 similar prostheses with ceramic heads inserted at a private clinic. The mean follow-up was 10 years (range 9–11 years) for patients with osteoarthritis (age range 48–78 years). In each group, 33 patients were available for subsequent analysis. Wear for the metal-head hips was 0.96 mm and for the ceramic hips 0.26 mm ($p < 0.001$). Of the cemented hips, 9% were revised because of loosening, 12% were loose; 6% of the ceramic hips were revised because of loosening and 9% were loose (p, not significant). No analysis assessing the possible influences surrounding the different types of hospital was undertaken. **C-rated.**

Hoffman and colleagues, 1994[33]

This retrospective study involved a total of 1166 hips in 974 patients. Six surgeons (grade not specified) performed the operations either at a publicly funded hospital (66%) or at a private clinic (34%). The average age of the patients was 66 years, 89% had a diagnosis of osteoarthritis, there were slightly more males (54%) than females, and 55% of operations were to the right side of the body. This section of the report is only part of the overall study. Four types of cemented prostheses were mentioned in the study but survival information was only given for one: Müller (n = 92) – 26 failed (23 were loose), annual failure rate was 6.93%, and 11-year survival rate 63%. However, this result is confounded by the poor results of one surgeon. The ceramic prosthesis, Autophor (n = 35), had a failure rate of 15% in 3 years (approximate annual failure rate of 5%). **C-rated.**

Non-Charnley cemented versus HA-coated

Comment

The follow-up period for both these C-rated studies is very short; hence, the results should be treated cautiously. The second study used specially-designed prostheses.[12] The authors of both studies suggested that HA-coated models had more stable early fixation than cemented models. No difference in early pain scores can be substantiated. Freeman and Plante-Bordeneuve[12] suggest that there is an association between pain and the extent of early migration on radiological assessment, and that HA-coated components perform better in this respect, at least in the early postoperative period. This may be in contrast to the comparison of Charnley with HA-coated designs (see page 22) in which no differences were demonstrated. Longer-term assessments involving greater numbers of patients are required.

Karrholm and colleagues, 1994[42]

A computer program was used to randomly allocate the 64 patients (age range 58–66 years) in this RCT. The patients were stratified by various characteristics. The hips had the Ti-Fit femoral component inserted with a press-fit acetabular component by one of four surgeons at one of two hospitals. The femoral stems were inserted with either cement (n = 20) or an HA-coating (n = 23). After 2 years the cemented stems had subsided more than the HA-coated stems ($p = 0.002$). The HA-coated components also rotated less compared with the cemented stems ($p = 0.03$). The Harris Hip and

Pain Scores did not differ significantly between the groups, although the small sample sizes make this result tentative. There were no revisions within the 2 years of the study. **C-rated.**

Freeman and Plante-Bordeneuve, 1994[12]

The prosthesis in this study was specially designed to allow measurement of vertical migration. The THRs were either cemented (55 hips in 54 patients, 91% with osteoarthritis, 69% female, age range 57–83 years) or HA-coated (34 hips in 34 patients, 88% with osteoarthritis, 41% female, age range 33–76 years). The amount of migration was assessed at 2 years: cemented hips (n = 55) had a mean of 0.55 mm, while those hips with no pain (n = 52) had migrated 0.38 mm on average; all HA-coated hips (n = 34) had migrated on average 0.4 mm, the same as those hips with no pain (n = 34). At a minimum follow-up of 5 years, 7.9% of the cemented prostheses were loose. At 4-year follow-up no HA-coated hips required analgesia or had been revised. **C-rated.**

Non-Charnley cemented versus hybrid

Comment

As with the Charnley versus hybrid comparisons, hybrid prostheses appear to survive in the following studies as well as, if not slightly better than, cemented hips in the short term but it is not yet possible to comment on longer follow-up results. There may be the potential for hybrid prostheses to equal or improve on the results of cemented hips such as those reported here. As both of these studies were C-rated, higher quality studies are also required of this type of comparison.

Wixson and colleagues, 1991[43]

A total of 197 hips were implanted into 176 patients by two surgeons and, after a mean follow-up period of 2.8 years (maximum 4 years) 144 hips were available for analysis. The mean age of patients, 60% of whom were female, was 61 years; 65% had osteoarthritis and 15% rheumatoid arthritis. Various types of cemented stems were used along with either cemented cups (PCA, TiBac, Harris) or porous-coated (PCA, Harris-Galante, APR). The various combinations were categorised as cemented or hybrid as appropriate. Two cemented hips (3.8%) were revised because of loosening while one hybrid (3.7%) was revised (but not for loosening). Of the cemented cups, 12% had migrated or changed position compared with 3% of porous cups (p = ?). **C-rated.**

Callaghan and colleagues, 1995[36]

As part of a larger study, IOWA prostheses, using second generation cementing techniques (n = 187, minimum follow-up 10 years), were compared with 130 hybrid hips (Harris-Galante type 1 cup + IOWA cemented stem, 5-year follow-up) and 61 similar hybrids were used as revisions (minimum follow-up 5 years). No patient details were given. In the cemented group, cup loosening occurred in 24.5% of patients (metal-backed 17%, all-polyethylene 30%) and stem loosening in 1.2%. This is in contrast to the hybrid groups where no revisions (or re-revisions) were reported and only one migration in the revision group occurred; however, the follow-up period was only 5 years or more compared to 10 years or more for the cemented group. Wear rates were also assessed in this study with the Harris-Galante cups (28 mm head) having less wear than the other groups but, unfortunately, no details were given. **C-rated.**

Non-Charnley cemented versus modular

Comment

The fact that neither clear results nor patient details were given for the Osteonics/DuPuy model in the study below makes commenting on it difficult. The cemented hips gave fairly typical results for this type of prosthesis and appear to be superior to the modular forms but a better evaluation is required.

Chmell and colleagues, 1995[44]

Three surgeons performed all the operations in this study. No details are given about the patients involved. Three cemented prostheses were used: Aufranc-Turner (n = 778); T-28 (n = 823) and Osteonics non-modular stem with cemented cup (n = 329); these were compared with three modular prostheses (as specified by the authors): Osteonics modular stem and cemented cup (n = 233); DePuy Profile modular stem with ACS modular cup (n = 203) and Osteonics modular stem with either Osteonics or DePuy Duraloc modular cup (n = ?). The percentages needing revision for loosening in the cemented groups ranged from 2.1% of the Osteonics hips after an average follow-up of 7.5 years to 22% of the Aufranc-Turner hips after an average follow-up of 12 years, the majority of these being after the first 6 years. In the modular group, 3% of the Osteonics modular stem (with the same cemented cup as before) were revised within 6 years and 12% needed revision in the DePuy Profile/ACS hip because of linear wear or fracture (all but two of

PAPER 3

the 15 had a polyethylene thickness of less than 6 mm). No details were given for the Osteonics/DePuy hip, except that they "have not been associated with the catastrophic failure rate seen in the ACS cups". **C-rated.**

Non-Charnley cemented versus press-fit

Comment

The results from the three studies selected below are conflicting. Two (one of them A-rated) showed more problems with the press-fit than with the cemented hips. However, in the remaining study,[45,46] although the press-fit stem showed evidence of subsidence it was the cemented cups which were deemed to be loose, although this was at only 4-year follow-up. Overall, cemented types of prosthesis would appear to be superior to press-fit.

Godsiff and colleagues, 1992[25]

This RCT compared 30 cemented with 28 uncemented femoral components (Ring prosthesis) in patients, age range 55–74 years, with osteoarthritis of the hip. Both patients and the non-orthopaedic clinical assessor were blinded and surgery was by one of two surgeons. At 2 years both groups (n = 47) reported similar pain incidence, the press-fit group having had more pain at 4 and 12 months. By 2 years, 96% (cemented) and 62% (uncemented) of patients did not require walking aids ($p = 0.01$–0.05). Preliminary results indicated cemented to be superior to press-fit; however, because of unacceptable levels of femoral breakages at 3–5 years, the authors withdrew the Ring prosthesis. These failures may have been due in part to design and manufacturing factors, as reported in a Safety Notice issued by the Department of Health Medical Devices Agency (MDA SN 9520) in August 1995. The design has subsequently been modified. **A-rated.**

Bourne and colleagues, 1995;[45] Rorabeck and colleagues, 1996[46]

All patients in this RCT were operated on or supervised by two senior surgeons using the Mallory Head prosthesis, either cemented or press-fit. A total of 250 patients were originally recruited from a group with an age range of 18–75 years and were stratified by age and surgeon. Diagnosis was osteoarthritis of the hip. Clinical results had a 5-year follow-up period (n = ?) and radiographic analysis a 4-year follow-up (n = 147). Patients and clinical observers were blinded. All clinical assessments (e.g., Harris Hip Score, d'Aubigne Score and Sickness Impact Profile, among others) were almost identical for each group both pre- and postoperatively. There was no subsidence in the cemented stems but 14% of press-fit stems subsided by 3–5 mm. No revisions were required within 4 years and no press-fit components or cemented stems were loose; however, 26% of the cemented cups were described as definitely or probably loose. **C-rated.**

Krismer and colleagues, 1991[47]

Uncoated RM cups were paired with Müller stems to form the press-fit hip and 160 from 173 (mean age of patients 57 years, average follow-up 5.3 years) were assessed and compared with 263 from 309 Müller prostheses (mean age of patients 63 years, average follow-up 6.1 years). The diagnosis in 75% of patients was primary coxarthrosis. None of the cemented prostheses migrated during the study period but 25% of the press-fit migrated between 2.1 and 16 mm. After 7–8 years, 12% of the press-fit hips had been revised and 40% were loose compared with 4% and 15%, respectively, of the cemented hips. **C-rated.**

Non-Charnley cemented versus porous-coated

Comment

Ten comparisons of these types of prosthesis have been reviewed overall. In only one of the three papers selected below was an attempt made to compare the two types of prosthesis at the same time after surgery. In one study with medium-term follow-up and a fair sample size, Callaghan and colleagues[36] suggested that the cemented acetabular component performed better than the porous-coated but that the porous-coated stem was better than the cemented. As with the Charnley comparisons, porous-coated types appear to have good short-term survival results which need to be followed further.

Callaghan and colleagues, 1995[36]

As part of a larger study, IOWA prostheses, using second generation cementing techniques (n = 187, minimum follow-up period 10 years) were compared with 100 PCA prostheses (minimum follow-up 7 years). No patient details were given. In the cemented group, cup loosening occurred in 24.5% of patients (metal-backed 17%, all polyethylene 30%) and stem loosening in 1.2%. The PCA prostheses had a cup revision incidence of 4% and a migration incidence of 5%, which included two revisions. **C-rated.**

PAPER 3

Hoffman and colleagues, 1994[33]
This retrospective study involved a total of 1166 hips in 974 patients. Six surgeons (grade not specified) performed the operations at either a publicly funded hospital (66%) or a private clinic (34%). The average age of patients, of whom 89% had a diagnosis of osteoarthritis, was 66 years; there were slightly more males (54%) than females, and 55% of operations were to the right side of the body. This section of the report forms only a part of the overall study. Four types of cemented prostheses are mentioned in the study but survival information is only given for one: Müller (n = 92) – 26 failed (23 were loose), the annual failure rate was 6.93%, and the 11-year survival rate 63%. However, this result is confounded by the poor results of one surgeon. Neither the Harris-Galante (n = 105) nor the PCA (n = 38) prostheses had any revisions in approximately 3–4 years of follow-up. **C-rated.**

Hearn and colleagues, 1995[48]
A total of 36 consecutive patients underwent primary cemented THR (Charnley, Dual Lock or Pennsylvania Total Hip) followed by primary porous-coated THR (Trilock or Taperloc) of the contralateral hip (total number of hips, 72). Of these, 60 were assessed after 8.1 years (cemented hip) and 3.0 years (porous hip) ('same interval' data for the cemented hips were also compared with the porous data at 3.6 years). The patients' age range was 21–82 years, 92% had diagnoses of osteoarthritis and 8% of rheumatoid arthritis. Preoperative pain levels differed (cemented 3.1, porous 2.5, $p = 0.002$), as did the range of movement measurements at the same interval of follow-up (cemented 5.1, porous 5.6, $p = 0.002$). There was no migration or subsidence, and there were no revisions. One cemented stem was probably loose but no porous components were loose. **C-rated.**

Non-Charnley cemented versus resurfacing

Comment
All three papers summarised below report resurfacing prostheses to be inferior to all of the cemented hips with which they were compared. Thus resurfacing prostheses cannot be recommended as an alternative to cemented THRs.

Reigstad and colleagues, 1986[49]
A total of 155 Müller and 149 ICLH prostheses were implanted into 231 patients (age range 60–79 years) by 13 surgeons. All patients were diagnosed with osteoarthritis of the hip and had a mean follow-up of 48.5 months. No Müller hips were revised compared with 8.7% ICLH ($p < 0.001$) and, in addition, one component (0.6%) was loose compared with 12 (8%), respectively. Post-operatively the Müller group had consistently higher scores than the ICLH group on all three modified Merle d'Aubigne and Postel parameters and total hip function. The level of significance reached by 1 year in 3/4 parameters was $p < 0.001$. **C-rated.**

Ritter and Gioe, 1986[50]
Bilateral hips in 50 patients were replaced with one cemented T-28 prosthesis and one Indiana conservative resurfacing hip using the same anaesthetic; 45 (90%) of these patients were followed-up for a minimum of 5 years. The mean age of the patients was 62 years (range 21–87 years) and 79% were diagnosed with osteoarthritis. There was no difference in the level of pain in the non-revised hips, as recorded by the Hospital for Special Surgery Rating System. Two (4.4%) cemented hips were revised (none were loose) and 15 (33%) resurfacing hips were revised (these patients were younger, average age 55 years). **C-rated.**

Ahnfelt and colleagues, 1990[14]
Results from this retrospective study were taken from a Swedish multicentre registry. The patients had an approximate median age of 64 years for women and 66 years for men, and a main diagnosis of osteoarthritis. In all cases the hips had originally been implanted between 1979 and 1986 and had required revision. Survival without revision for loosening of the original THR was observed for various prostheses. Eight out of ten cemented prostheses had results quoted for them. The observed survival ranged from 63% (Christiansen) to 89% (Stanmore) at 10 years and from 95% (Exeter) at 5 years to 93% (Lubinus) at 9 years. This is in comparison to the Wagner resurfacing hip prosthesis which had only 28% survival at 10 years, the worst result in the study. **C-rated.**

Ceramic (cemented) versus ceramic (cementless)

Comment
No statistical analysis was performed on the revision figures in the paper below but, given the difference in follow-up time, the results may be roughly equal, suggesting no difference at short-term follow-up between the cemented and cementless methods of fixation of ceramic hips.

Riska, 1993[51]
The Ceraver Osteal aluminia on aluminia prosthesis was implanted with either a cemented ceramic cup (n = 143, mean follow-up period 6.7 years) or an uncemented ceramic cup (n = 112, mean follow-up period 3.6 years). One surgeon performed the operations on the patient group who had a mean age of 62 years and 73% of whom had a diagnosis of osteoarthritis. Prostheses with the cemented cup had 16 revisions (11.2%) and 14 were loose. Seven revisions (6.3%) were required in the uncemented cup hips and two were loose. **C-rated.**

Ceramic versus porous-coated

Comment
Very small numbers of ceramic prostheses were included in the only study permitting comparison of these two types. Strong conclusions cannot be drawn, although the results suggest that ceramic prostheses are unlikely to be superior to porous-coated over the short term.

Hoffman and colleagues, 1994[33]
Study details are given above (page 23). The ceramic prosthesis (Autophor, n = 35) had a failure rate of 15% in 3 years, giving an approximate annual rate of 5%. Two different porous-coated THRs were included in the study: Harris-Galante (n = 105) and PCA (n = 38). Both had a follow-up period of roughly 3–4 years (compared with approximately 6 years for the ceramic prosthesis) and neither design had had any failures in this time. **C-rated.**

HA-coated versus press-fit

Comment
From the two studies selected from the five available, HA-coated prostheses appear to be more stable than press-fit in the (very) short term, being associated with less migration/subsidence and pain and possibly with greater mid-term survival.

Huracek and Spirig, 1994[52]
Forty pairs of patients were retrospectively matched for various aspects from 127 possible cases. One surgeon inserted all the hips either with an HA coating or without (press-fit). All patients had primary osteoarthritis, their average age was 71 years and average length of follow-up was 4.1 years. The occurrence of pain was assessed: 59.3% of HA hips had no pain compared with only 22.5% of press-fit hips ($p < 0.0016$). No HA-coated cups showed signs of migration and 7.5% of HA-coated stems subsided. In the press-fit hips, 32.5% of cups migrated by 5 mm or more and 30% of stems subsided. There were no revisions or loose components in either group. **C-rated.**

Moilanen and colleagues, 1996[53]
The SLF cup (together with a Freeman cemented or uncemented stem) was inserted either with an HA-coat or without (press-fit). The mean age of the patients given an HA-coated cup was 59.7 years (74% with osteoarthritis, 2.3 years follow-up) while those with press-fit cups had a mean age of 62.6 years (95% with osteoarthritis, 3.4 years follow-up). There was no difference in mean migration rate between the two groups but the press-fit group had more radiolucent lines associated with them (27%) than did the HA-coated group (6%, $p < 0.05$). Two revisions were required in the HA-coated hips within 7 months of the operation, neither being caused by loosening; no hips were replaced from the press-fit group. **C-rated.**

HA- versus porous-coated

Comment
Both of these studies have only a short follow-up period. The porous-coated hips seemed to have more subsidence and did not initially fix as well as the HA-coated hips. How this would affect the longer term clinical and survival outcomes is unclear at the present time. One study suggests there are no differences in hip scores, including pain, between the two types in the early post-operative period. This is echoed by one observational study.[54] Further investigation of the two types of coating is required.

Karrholm and colleagues, 1994[42]
A computer program was used to randomly allocate the 64 patients (aged 58–66 years) in this RCT. The patients were stratified by various characteristics. The hips had the Ti-Fit femoral component inserted with a press-fit acetabular component by one of four surgeons in one of two hospitals. The femoral stems were inserted with either an HA-coating (n = 23) or were porous-coated (n = 21). After 2 years the porous-coated stems had subsided more than the HA-coated ($p = 0.02$). The Harris Hip and Pain Scores did not differ significantly between the groups. Pain or discomfort in the thigh was reported (HA-coated, n = 5; porous-coated, n = 8; $p = ?$). There were no revisions within the 2-year period. **C-rated.**

PAPER 3

McPherson and colleagues, 1995[55]
HA-coating was added to a prosthesis, this time an APR-I hip, to perform this study. Data were collected prospectively but the study groups were selected retrospectively. From 230 patients, 42 pairs were matched, giving an average age of approximately 56 years (the diagnosis of the patients is not given but was used during the matching process). Using a modification of the DeLee-Charnley Fixation Score, the authors suggest that fixation was better in the HA-coated hips ($p = 0.002$) after a minimum follow-up period of 3 years – porous-coated hips: 62% Grade IA, 33% Grade IB; HA-coated hips: 93% Grade IA, 7% Grade IB. The mechanical failure rate for both groups was 5% – one revised HA-coated hip and one HA-coated and two porous-coated hips loose. **C-rated.**

Hybrid versus porous-coated

Comment
All three papers found hybrid prostheses to be superior to porous-coated over the short term, especially as regards the stem component. The studies reported thigh pain to be more closely associated with the porous-coated hips as were movement and the need for revision.

Wixson and colleagues, 1991[43]
Originally 197 hips were implanted into 176 patients by two surgeons, and 144 hips were available for analysis after a mean follow-up time of 2.8 years (maximum 4 years). The mean age of the patients, 60% of whom were female, was 61 years; 65% were diagnosed with osteoarthritis, 15% with rheumatoid arthritis. Various types of porous-coated cups (PCA, Harris-Galante, APR) were used together with either cemented stems (PCA, SixTi/28, ATS, Harris Design 2 and CHD) or porous stems (PCA). The various combinations were categorised into hybrid or porous, as appropriate. Thigh pain was recorded at 3 years: cemented stem 3%, porous stem 13%; $p < 0.05$. There was no subsidence of the cemented stems while 5% of the uncemented ones changed position (as did 3% of the cups). More revisions occurred in the porous hips (7.7%, 4/5 loose) than the hybrid (3.7%, none loose). **C-rated.**

Goetz and colleagues, 1994[56]
One surgeon performed 255 operations on patients with an age range of 40–71 years, 95% of whom had a diagnosis of osteoarthritis. Retrospectively 82 hips (in 74 patients) were matched and compared, with an approximate 6-year follow-up period. All had a Harris-Galante cup with either a Harris-Galante stem or a Harris Precoat (cemented) stem. Osteolysis was assessed in both groups: 29% of the porous hips showed osteolysis (five were loose) while the hybrid hips showed none ($p < 0.0002$; there was no relationship between femoral head size and osteolysis). A total of 12% of porous stems were revised (4/5 were loose, eight had subsided or migrated). None of the hybrid hips required revision ($p < 0.02$), all were stable with no radiolucent lines. None of the cups in either group had migrated or been revised. **C-rated.**

Maloney and Harris, 1990[57]
Precoated cemented stems and Harris-Galante cups were compared to Harris-Galante prostheses in a retrospectively matched study of 25 pairs of hips (selected from a group of 136 hips). One surgeon performed the operations and follow-up was for 2.5–3 years. The patients' age range was 54–69 years (average 61–62 years) and 96% had osteoarthritis. Postoperative Harris Hip Scores differed between the groups (hybrid 96, porous-coated 84; $p < 0.02$) as did thigh pain (hybrid 0, porous-coated 20%; $p = ?$). Migration had occurred in 20% of the porous group stems, but not the hybrid stems (there were no radiolucent lines or migration associated with the cups of either group). Of the porous group, 16% required revision, 3/4 due to migration. No hybrid hips were revised. **C-rated.**

Modular versus modular

Comment
Comments on this one study are difficult to make as not all the data are given for the different prostheses, nor are any patient details given. The only possible comment is that the hip which combined a modular stem with a cemented cup (classed as a 'modular' type by the authors) faired better than the fully modular prosthesis on which data were given.

Chmell and colleagues, 1995[44]
Three surgeons performed all the operations in this study. No details are given about the patients. Three modular prostheses were used as a part of this study: Osteonics modular stem and cemented cup (n = 233); DePuy Profile modular stem with ACS modular cup (n = 203) and Osteonics modular stem with either Osteonics or DePuy Duraloc modular cup (n = ?). Of the Osteonics modular stem, 3% were revised within 6 years and the DePuy Profile/ACS hip needed revision in 12%

caused by liner wear or fracture (all but two of the 15 had a polyethylene thickness of less than 6 mm). No details were given about the Osteonics/DePuy hip, except that they "have not been associated with the catastrophic failure rate seen in the ACS cups". **C-rated.**

Press-fit versus press-fit

Comment

Both studies were C-rated and only one had the basic minimum numbers of patients (see page 14). It would appear that the material from which the prostheses are made might influence the results. Further work is needed to assess this more fully but, given the poor results of the press-fit types compared with other prostheses, this is probably not worthwhile.

Schreiber and colleagues, 1993[20]

The Balgrist prosthesis was used in this study with either an outer split ring of high-density polyethylene (61 patients had a thin (6 µm) coating of titanium on the outer surface) or of titanium alloy. The study was retrospective. The patients' age range was 23–76 years (average approximately 55 years) but diagnosis and the number of surgeons involved is unknown. From 717 hips, 606 were assessed after 4.5 years (polyethylene) or 1.3 years (titanium). During the follow-up to this study, 13% of primary and 21% of revised polyethylene types were revised, as were 0.7% primary and 5% revised titanium alloy types (p = ?). **C-rated.**

Nashed and colleagues, 1995[58]

This is included as part of a larger study in which cemented prostheses were compared with the two press-fit prostheses mentioned here. The retrospective study involved one surgeon; the patients had an approximate mean age of 50–51 years, with diagnoses of osteoarthritis and rheumatoid arthritis in 53% and 16%, respectively. The press-fit hips were both BIAS prostheses with a metal-backed cup, one with a titanium head (n = 15) and the other with a cobalt–chrome head (n = 74). The follow-up period was approximately 6 years. Osteolysis occurred in 87% of the hips with titanium-heads (87% stem, 40% cup) and in 24% with cobalt–chrome heads (22% stem, 14% cup). The incidence of osteolysis was statistically higher in the titanium group than in any other group (including the cemented hips). Hips with osteolysis were found to be more likely to require revision than those without osteolysis in the overall study ($p < 0.001$). **C-rated.**

Press-fit versus porous-coated

Comment

In the one study in which this comparison was included, the porous cup performed better than the press-fit cup. This would appear to be consistent with other studies comparing either porous or press-fit to other types where press-fit designs are uniformly inferior to other types of prosthesis.

Pupparo and Engh, 1989[59]

In this prospective study, AML stems were combined with either an S-ROM Anderson cup (smooth-threaded) or an S-ROM Super cup (porous-threaded). One surgeon, of unknown grade, performed the operations. The ages of the patients were not stated but 86% of them had a diagnosis of osteoarthritis. Approximately 67% of the hips originally recruited were available for follow-up assessment at 2–3 years. Of the hips with an Anderson cup, 29% were classed as unstable and nine had migrated by a mean of 5.5 mm, whereas the hips with a Super cup showed no migration and were all classed as stable ($p < 0.001$). No Super cup hips were revised but six of the Anderson cups were, all caused by loosening (p = ?). **C-rated.**

Porous-coated versus porous-coated

Comment

The three prostheses used in the study summarised below gave similar results over the short term. However, the number of patients was small and the ratio of disease types unusual compared with the vast majority of comparative studies used in this report. This being the case, further work is required involving larger numbers to gauge if this is a true result or if the different prostheses do differ in any way.

Hwang and Park, 1995[60]

Three types of porous-coated prosthesis are compared in this prospective study: AML (n = 90, mean follow-up period 5.2 years); PCA (n = 117, mean follow-up period 4.7 years); Harris-Galante (n = 63, mean follow-up period 3.8 years). The age range of the patients was 20–86 years (approximate mean 48 years). The diagnosis for this group of patients was very different from most other studies as the main diagnosis was of avascular necrosis (66%) with osteoarthritis in 18%. One surgeon was involved in replacing the hips. Approximately 19% of each group had thigh pain. Stem subsidence ranged from 0–8 mm, with an approximate average of 2.1 mm, and was similar for all groups, as was the number with subsidence of 3 mm or

more (AML 10%, PCA 13.7%, Harris-Galante 12.7%). Cup migration did not differ between the groups (all approximately 4.1% 0 and no revisions were reported. **C-rated.**

Porous-coated versus resurfacing

Comment

The study summarised here concentrates on heterotopic bone formation and not on the usual outcome measures. Although this study (with few patients and a higher number of patients with avascular necrosis than most studies) showed no difference between the two types of prosthesis, resurfacing replacements are not recommended for the usual indications for THR, because of the results shown in the comparisons above.

Duck and Mylod, 1992[61]

The original population from which the study group was taken was not stated but the study concentrated on 66 hips in 55 patients with a range of diagnoses, such as 34.5% osteoarthritis and 36.4% avascular necrosis. The average age of the group was 60 years (range 33–76 years). As part of a larger, retrospective study, AML porous-coated hips were compared to resurfacing prostheses (TARA and Indiana Conservative hip) 3 years after the operation (number of surgeons performing surgery unknown). The study concentrated on the occurrence of heterotopic bone formation: 59% of the uncemented total hips had heterotopic as did 56% of the resurfacing hips. The authors concluded that there was "no significant correlation between the type of procedure and the percentage bone formation". **C-rated.**

Chapter 9

Results of selected observational studies

The summaries presented in this chapter should be read in conjunction with the more detailed data tables for each type of prosthesis in the appendix to this report. The best available studies of each prosthesis type have been selected for inclusion here, as indicated.

Cemented designs

Charnley

The studies summarised in *Table 18* fulfill the following criteria:

- A or B-rated by the reviewers
- cohort size of > 200 hips followed-up
- survival or revision rate data presented.

Non-Charnley cemented designs

The studies presented in *Table 19* fulfil the same criteria as the Charnley studies presented in *Table 18*.

Comment

As a group, these selected studies of cemented prostheses show that rates of early survival (up to 10 years) are generally very good for most models;

TABLE 18 *Selected studies of the Charnley cemented prosthesis*

Study	Number of hips (follow-up period, years)	Age (years)	Results
Dall, *et al.*, 1993[62]	811 (10–12)	mean 60	87% survivorship at 10–12 years (revision rate 8%)
Eftekhar, *et al.*, 1986[63]	499 (> 10)	mean 62	Re-operation 2.2% (+1.2% pending)
Garcia-Cimbrelo & Munera, 1992[64]	680 (18)	mean 56	81% survivorship at 18 years (91.6% at 10 years)
Hamilton & Joyce, 1986[65]	230 (6)	86% over 50	Revision rate for aseptic loosening: stem 0.0%, cup 0.7%
Hamilton & Gorczyca, 1995[66]	224 (10 +)	mean 58	Stem revision rate 6.3%; cup revision rate 6.7% (12.5% cup migration rate)
Joshi, *et al.*, 1993[67]	218 (10–24)	mean 32	Stem revision rate for aseptic loosening: 3% at 10 years, 14% at 20 years; cup revision rate: 4.5% at 10 years, 16% at 20 years (osteoarthritis risk revision 20% at 10 years, 49% at 20 years)
Kobayashi, *et al.*, 1994[68,69]	326 stems, 328 cups (13)	mean 58	Stem revision rate 1.2% (4.9% failure); cup revision rate 7.4% (17% failure)
Madey, *et al.*, 1997[70]	356 (15)	mean 62	Revision rate for aseptic loosening 11% at 15 years (stem 2%, cup 10%)
Neumann, *et al.*, 1994[71]	241 (15–20)	median 62	Probability of revision 10.7% at 20 years
Older & Butorac, 1992[72]	388 (17–21)	mean 68	Revision rate 6%; 89% survivorship at 20 years
Skeie, *et al.*, 1991[73]	629 (10–15)	mean 66	92% survivorship at 13 years (7% revised)

PAPER 3

revision rates at a minimum of 10 years in age groups from mid-50s to mid-60s range from about 2% to about 13%. Revision rates in the one series of young patients (by Joshi and colleagues)[67] are moderate for such a group of young patients with a diagnosis of osteoarthritis. Given the unknown part played by potentially confounding factors, comparisons between prostheses on the basis of these observational studies can be made only tentatively and by treating the reported survival rates as estimates requiring wide confidence intervals. Taking this into account, the Howse, Exeter and Lubinus models appear to bear comparison with the Charnley at medium term (10–15 years) follow-up.

Uncemented designs

Porous-coated

Only four of the reviewed studies of cementless porous-coated technology fulfil the same criteria as the 17 studies of cemented prostheses summarised in *Tables 18* and *19*. Three of these are by the same group of authors, Engh and colleagues,[80–82] and present results for the same component, the AML straight stem. The results from the most recent of these studies[82] are summarised in *Table 20* together with those from the A-rated study by Owen and colleagues,[83] in which more than 200 hips were followed up, and from the only other study with 10-year results (Sotereanos *et al.*, 1995).[84]

The study by Owen and colleagues[83] records a steep decline in survival of cups between years 6 and 9, especially in younger patients, which is attributed to severe polyethylene wear caused by the use of the large (32 mm) stem head size. Engh and colleagues' results[82] are good for medium-term follow-up, especially when the relatively young mean age of the study group is taken into account. With the exception of these two studies and that by Holman and Tyer,[85] the numbers of hips followed up with porous-coated prostheses are very modest, with the majority being about 100 and many being fewer than this.

The results at 10 years appear to bear comparison with the cemented models for the same follow-up period, especially when account is taken of the relatively lower average age of the patient groups implanted with porous-coated models compared with those receiving cemented models.

The AML and PCA models are those for which results have been most frequently published (of the reviewed observational studies: PCA, 13 studies; AML, 8; Harris-Galante, 6). Sotereanos and colleagues' results for the AML stem are exceptionally good.[84]

Thigh pain is an issue for porous-coated implants. In the studies reviewed, reports of its prevalence range from about 2% to about 25% at 2–7 years' follow-up. Several studies report prevalences of about 25%, including in non-loose stems.

The amount of porous-coating on stem components is an issue. The majority of

TABLE 19 *Selected studies of cemented non-Charnley prostheses*

Study	Prosthesis type	Number of hips (follow-up period, years)	Age (years)	Results
August, *et al.*, 1986[74]	McKee-Farrar	230 (10–22)	mean 60	91% survivorship (revision) at 10 years, 84% at 15 years, 27.5% at 20 years
Bryant, *et al.*, 1991[75]	Ring	253 (20)	mean 63	60% survivorship (revision)
Fowler, *et al.*, 1988[76]	Exeter	241 (11–16)	mean 67	Total mechanical failure 11%
Ohlin & Onsten, 1990[77]	Lubinus	202 (3–6)	median 68	Revision rate 3% for aseptic loosening
Partio, *et al.*, 1994[78]	Lubinus	444 (8–12)	mean 64	Revision rate 11.5%; 87% survivorship at 10 years
Roberts, *et al.*, 1987[79]	Howse	265 (10–15)	mean 63	90% survivorship (revision) at 10 years, 80.8% at 15 years

TABLE 20 *Selected studies of uncemented porous-coated prostheses*

Study	Prosthesis type	Number of hips (follow-up period, years)	Age (years)	Results
Engh, *et al.*, 1997[82]	AML stem	223 (minimum 10)	mean 55	85% stem survivorship at 12 years
Owen, *et al.*, 1994[83]	PCA	241 (2–9; mean 5)	mean 47	57% survivorship at 7 years (including recommendation for revision)
Sotereanos, *et al.*, 1995[84]	BIAS and AML stems	121 and 166 (10 and 8)	mean 53–54	BIAS: revision rate 4.1% at 10 years; survivorship 95.4% at 11 years; AML: revision rate 0.6%; survivorship 99.3% at 9 years

porous-coated implants, where the information was given, had the coating on the proximal part of the stem plus the cup. In a comparative study on animal models[86] (not appraised in this review), it is suggested that total circumferential coating is associated with more bone loss than partial coating.

HA-coated

Three studies met the criteria of being A- or B-rated, including more than 200 hips and having survivorship results reported.[54,87,88] These are summarised in *Table 21* together with the two studies with the longest follow-up.[89,90]

Of the nine HA-coated studies summarised in the appendix, five report on the American Osteonics Omnifit components, thought to be the most widely used HA-coated model internationally.[91] It is clear that the numbers of patients/hips and lengths of follow-up periods are insufficient to draw even tentative conclusions about the performance of this technology on the basis of survival data. The evidence for early postoperative pain associated with this type of technology suggests mild to moderate thigh pain in between 0% and about 5% of patients at 2–5 years' follow-up. This is a relatively good result in comparison to the results for porous-coated implants.

Uncoated press-fit

Only one study in this category is A- or B-rated, follows-up more than 200 hips and presents survival results.[92] This study of Mathys 'isoelastic' cups reports a high level of revision for aseptic loosening, mostly occurring after at least 8 years of implantation. The component was abandoned. Two of the other studies reviewed reported on the Mathys RM isoelastic components.[93,94] The first[93] also showed relatively poor results for the uncoated cup (although the ages of patients were not reported)

TABLE 21 *Selected studies of uncemented HA-coated prostheses*

Study	Prosthesis type	Number of hips (follow-up period, years)	Age (years)	Results
Tonino, *et al.*, 1995[87]	ABG	222 (minimum 2, mean 2.4)	mean 63	Revision rate (mechanical) 1.4%
Koch, *et al.*, 1993[88]	Furlong	190 (2–5, mean 2.9)	?	No revision or loosening
d'Antonio, *et al.*, 1992[54]	Omnifit	320 (minimum 2)	mean 50	No revisions
Capello, 1994[89]	Omnifit stem only	151 (5)	mean 50	Revision rate (pain/aseptic loosening) 3.3%
Geesink & Hoefnagels, 1995[90]	Omnifit	100/118 (5.6–7.6)	mean 53 (31 < 50 years)	100% stem survivorship, 99% cup

TABLE 20 Selected studies of uncemented porous-coated prostheses

Study	Prosthesis type	Number of hips (follow-up period, years)	Age (years)	Results
Engh, *et al.*, 1997[82]	AML stem	223 (minimum 10)	mean 55	85% stem survivorship at 12 years
Owen, *et al.*, 1994[83]	PCA	241 (2–9; mean 5)	mean 47	57% survivorship at 7 years (including recommendation for revision)
Sotereanos, *et al.*, 1995[84]	BIAS and AML stems	121 and 166 (10 and 8)	mean 53–54	BIAS: revision rate 4.1% at 10 years; survivorship 95.4% at 11 years; AML: revision rate 0.6%; survivorship 99.3% at 9 years

porous-coated implants, where the information was given, had the coating on the proximal part of the stem plus the cup. In a comparative study on animal models[86] (not appraised in this review), it is suggested that total circumferential coating is associated with more bone loss than partial coating.

HA-coated

Three studies met the criteria of being A- or B-rated, including more than 200 hips and having survivorship results reported.[54,87,88] These are summarised in *Table 21* together with the two studies with the longest follow-up.[89,90]

Of the nine HA-coated studies summarised in the appendix, five report on the American Osteonics Omnifit components, thought to be the most widely used HA-coated model internationally.[91] It is clear that the numbers of patients/hips and lengths of follow-up periods are insufficient to draw even tentative conclusions about the performance of this technology on the basis of survival data. The evidence for early postoperative pain associated with this type of technology suggests mild to moderate thigh pain in between 0% and about 5% of patients at 2–5 years' follow-up. This is a relatively good result in comparison to the results for porous-coated implants.

Uncoated press-fit

Only one study in this category is A- or B-rated, follows-up more than 200 hips and presents survival results.[92] This study of Mathys 'isoelastic' cups reports a high level of revision for aseptic loosening, mostly occurring after at least 8 years of implantation. The component was abandoned. Two of the other studies reviewed reported on the Mathys RM isoelastic components.[93,94] The first[93] also showed relatively poor results for the uncoated cup (although the ages of patients were not reported)

TABLE 21 Selected studies of uncemented HA-coated prostheses

Study	Prosthesis type	Number of hips (follow-up period, years)	Age (years)	Results
Tonino, *et al.*, 1995[87]	ABG	222 (minimum 2, mean 2.4)	mean 63	Revision rate (mechanical) 1.4%
Koch, *et al.*, 1993[88]	Furlong	190 (2–5, mean 2.9)	?	No revision or loosening
d'Antonio, *et al.*, 1992[54]	Omnifit	320 (minimum 2)	mean 50	No revisions
Capello, 1994[89]	Omnifit stem only	151 (5)	mean 50	Revision rate (pain/aseptic loosening) 3.3%
Geesink & Hoefnagels, 1995[90]	Omnifit	100/118 (5.6–7.6)	mean 53 (31 < 50 years)	100% stem survivorship, 99% cup

PAPER 3

at 9 years, and the second[94] recorded relatively poor clinical and survival results for the RM stem in small numbers of patients followed-up for 7–9 years).

Other studies of uncoated press-fit cups, mostly with peg or screw-enhanced fixation, show generally poor results (see press-fit data table in the appendix). The one exception is that by Kennedy[95] who, in a C-rated study (ages not specified), reported good results at 3–6 years' follow-up for the Arthropor cup; however, this success is attributed to the exact reaming surgical technique used, rather than to the prosthesis design. In studies of the Ring prosthesis, adequate numbers of patients are followed-up but their ages are not reported.

In general, early clinical and survival results for the press-fit stems are not encouraging in comparison to either uncemented coated or cemented models. Results for threaded cups have generally been poor, and the design has been largely abandoned (a selection of observational study results are given in the data tables in the appendix).

Hybrid designs

In this category only one study, by Helfen and colleagues,[96] follows-up more than 200 hips and presents survival results. The maximum follow-up period in published studies is about 7–8 years. The study by Helfen and colleagues is summarised in *Table 22,* together with two other studies with the longest follow-up periods. Helfen and colleagues' study suggests good early clinical and survival results in patients who are probably somewhat younger than average for THR. Given wide confidence intervals, this type of design can be regarded as comparable with the best cemented designs for early survival results.

TABLE 22 *Selected studies of hybrid prostheses*

Study	Prosthesis type	Number of hips (follow-up period, years)	Age (years)	Results
Helfen, *et al.*, 1993[96]	Marburg	212 (3–6)	mean 60	n = 1, revision for loosening
Schmalzried & Harris, 1993[97]	Two stem-collared, screw-fix cup models	97 (mean 6.5)	mean 61	n = 1, revision for stem loose (in a custom component); n = 1, cup revision
Mohler, *et al.*, 1995[98]	Harris-Galante	120 (mean 5.2)	mean 67	No revisions

Chapter 10

Summary of results in relation to key issues

The results included in this chapter are taken from RCTS and comparative studies only. First clinically important issues are considered:

- thigh pain, fracture, dislocation and bearing surface materials.

Then studies of factors that affect the performance of prostheses are considered, together with potentially confounding factors in the interpretation of study results:

- hospitals/surgeons, and patient ages and body mass.

Finally, studies are considered that report aspects of surgical (cementation) technique, fixation and one unusual prosthetic design.

Thigh pain

While thigh pain has been identified as a problem in users of uncemented coated femoral components, it is reported on less frequently as an separate outcome from general hip scale scores in observational studies of Charnley and other cemented stems. However, it is possible to comment on this issue on the basis of the comparative studies and trials reviewed here.

In summary, each of the five studies in which the thigh pain associated with porous-coated stems is compared with either cemented,[42,43] press-fit[99] or hybrid designs,[57,100] shows a higher and clinically more significant incidence in porous-coated models. Thigh pain has also been found to be significantly higher in uncoated press-fit compared with Charnley cemented prostheses.[37] The one study in which different porous-coated models are compared shows fairly consistent levels of thigh pain (range 17–21%) between them.[60]

Fracture

In the following study, the effects of fractures on clinical outcome were investigated by comparing uncemented hips with different types of fracture (according to the position of the fracture on the femoral shaft) with similar prostheses which had not fractured.

Mallory and colleagues, 1989[101]

Within a 4-year period, 56 femoral fractures occurred in various types of cementless total hip arthroplasties. These were divided into three groups by the authors: Type I (80%), Type II (16%) and Type III (4%). A total of 96% of the fractures occurred intra-operatively and 4% from post-operative trauma. The average age of the patients was 50.4 years (range 21–81 years) and 55% were female; 61% of THRs were primary replacements, the remaining 39% being revisions (91% of these for loosening). The control group comprised randomly selected patients with cementless THRs without intraoperative fractures, whose prostheses were implanted during the same period. There were no statistically significant variations between the groups. Types I and II were compared to the controls for "improvement by operation" and their modified d'Aubigne–Harris scores. No statistical differences were found. **C-rated.**

Comments

The authors concluded that long-term problems were not associated with Type I fractures (proximal zone) and possibly not with Type II (middle zone), although the numbers in the study were too small to be sure. The number of Type III fractures (distal zone) (n = 2) was too small to draw any conclusions. This subject requires further investigation to determine more precisely the prognosis for such fractures.

Dislocation

The reasons for dislocations occurring in some hips were investigated in two studies which considered the same problem but from different angles: in one the effect of patient variables was assessed while in the other CT scans were used to assess the prosthesis components.

Hedlundh and Fredin, 1995[102]

The median age in this group of patients was 70 years (range 22–94 years), approximately 68% were female and approximately 48% had osteoarthritis and 23% rheumatoid arthritis.

PAPER 3

Out of 1838 patients who had received a Charnley prosthesis, 60 hips had dislocated; these were matched with 120 non-dislocated hips, which formed the control group. Mortality was higher in the patients in the dislocated group (53%) compared with those in the control group (24.5%, $p < 0.001$), although the median age at death was similar. In a logistic regression none of the tested factors proved to be related to dislocation; however, alcohol abuse in men was more common in the dislocation group (50%) than in the control group (18%) ($p = 0.01$). **C-rated.**

Pierchon and colleagues, 1994[103]

Within a 2-year period, 38 patients with dislocations were treated; 53% were women, the average age was 57 years and 66% had a diagnosis of osteoarthritis. Müller prostheses were used in most cases (29 self-locking femoral types, five dysplasia types; 12 cups were cemented and 26 uncemented). Of the 38 patients, 11 had been operated on by the same surgeon on the contralateral side and had not dislocated; details of three further prostheses were added to these to form a control group (further details of types of prosthesis or patients details were not given). Component alignment analysis was by CT scan. No differences in mean cup abduction, cup anteversion or femoral neck anteversion was found. In seven of the dislocated hips which underwent re-operation, the possible reasons for dislocation, as diagnosed by CT scan, were only confirmed in two cases. In the other five cases, instability of the hip was caused by lack of tension in the soft tissues. **C-rated.**

Comments

Hedlundh and colleagues[102] reported an increased mortality rate among those with a dislocation but this was thought to be caused by lack of muscular strength and decreased coordination rather than by old age. Pierchon and colleagues[103] also thought that dislocation was caused by lack of tension in the soft tissues. One interesting observation was the association between dislocation rates and alcoholic abuse in men.

Bearing surface materials

Some comparative studies have addressed this aspect of hip replacement results. Those reviewed here were given a low rating in the appraisal. The wear of different materials is implicated in the production of particulate debris which, in turn, is associated with osteolysis and loosening of the prosthesis. The results for different bearing surface combinations are commented on below.

Ceramic on ceramic

Only one study,[51] which was C-rated, considered ceramic on ceramic bearing surfaces. The study compared cemented cups to uncemented screw cups in ceramic prostheses. Over a follow-up period of 1–12 years, 9% of all the prostheses required revision. No comments or analysis of the bearing surfaces are made in this study. This is a rare form of design.

Ceramic on polyethylene

Prostheses with metal femoral heads and polyethylene cups were compared with those with ceramic femoral heads and similar cups in two, C-rated studies. In both studies the ceramic– polyethylene combination gave superior results to the metal–polyethylene combination. In Müller-type prostheses,[40] 95% of the ceramic head hips had a wear rate of less than 0.2 mm/year compared with only 64–77% of the metal version (range of follow-up period, 2.5–9 years). A mean wear value of 0.26 mm was found in Weber-type prostheses with a ceramic rotating head over a mean follow-up period of 10 years;[41] this was less than the mean wear found in the same type of prosthesis but with a metal rotating head (0.96 mm; $p < 0.001$). However, this last result should be viewed with caution as the two prostheses were implanted in different hospital settings.

Metal on polyethylene

The main metal used for femoral heads now is stainless steel, superseding the titanium alloy and cobalt–chrome alloy to which many published studies refer. It is difficult to comment on the metals used as a bearing surface because of many other factors being present in the published studies. For example, in one C-rated study,[58] a cobalt–chrome BIAS stem (with an uncemented metal-backed cup) was compared with three titanium BIAS stems with different cups. The titanium stem with cemented polyethylene cup gave the best results, with the least amount of wear and no osteolysis, but a titanium stem with an uncemented metal-backed cup gave the worst results. It would have been useful to compare the results with a cobalt–chrome stem with polyethylene cup but this was not included in the study.

However, some general comments can be made. There appears to be less wear in those prostheses with a complete polyethylene cup compared with those with a polyethylene liner. Linear wear rates for the all-polyethylene cups tend to be about 0.05–0.1 mm/year.[41,58,104,105] Details of polyethylene-liner wear related to porous-coated prostheses were given in two further studies: both the mean wear (0.73 mm)[106] and wear rates (0.6–0.8 mm/year)[60] were

higher than the results above and so may need to be considered when choosing this type of prosthesis.

Comments

Further analysis needs to be performed in order to gain a better understanding of the optimal materials for bearing surfaces, as wear and debris-mediated osteolysis are considered to be important reasons for loss of fixation and subsequent failure.

Inter-surgeon and inter-hospital comparisons

In multi-surgeon and multicentre studies confounding effects might be introduced by systematic differences between surgeons and/or between hospitals (for example, in stocking some prosthesis designs but not others). A number of the RCTs and comparative studies illustrate these points.

Surgeons

Marston and colleagues, 1996[107]

Surgeons in training performed 15/16 primary procedures, which subsequently required revision, using (in 14/16 cases) the anterolateral approach. (Difference in revision rates between experienced and trainee surgeons, $p = 0.005$; relative risk of requiring revision at 5–10 years postoperatively, 11.47 times greater for trainees, 95% confidence interval 1.53–86.06). Surgeons in training only performed the operation unsupervised after being considered fully competent by the consultants who had taught the technique. Technical errors were identifiable in 11 cases. No significant difference was found between the Stanmore and Charnley prostheses. **C-rated.**

Ahnfelt and colleagues, 1990[14]

Information was available on 37 surgeons (mainly working in two hospitals), all of whom were categorised as experienced surgeons. Two of 33 surgeons had fewer complications of aseptic loosening than the others and one had more ($p < 0.001$) re-operations than any of the others. However, if complications were analysed taking into account the number of primary operations performed per year, there were no statistical differences. **C-rated.**

Hoffman and colleagues, 1994[33]

Nine different prostheses of various types with different lengths of follow-up, implanted by six surgeons, were studied in multi-variable regression analysis. Prosthesis type was not significant. The surgeon performing the operation was a significant factor, particularly if the prosthesis used was standardised to the Charnley ($p < 0.001$). However, the results of a single surgeon were not as satisfactory as the overall results and this factor masked any difference attributable to prosthesis in the two main groups (Charnley, Müller). Unsupervised registrars performed no worse than other grades of staff. **C-rated.**

Hospitals

Ahnfelt and colleagues, 1990[14]

Rates of revision for aseptic loosening and deep infection were compared between different types of hospitals. Statistical differences were found between university (tertiary), regional (secondary) and community (primary) hospitals. University hospitals reported more infection than the others. This could be caused by the selection of patients with special problems, which demanded lengthy and extensive procedures.

In a comparison of aseptic loosening in all prostheses (except the Christiansen prosthesis and the surfacing replacements), regional hospital results were better than the others ($p < 0.001$). **C-rated.**

Hoffman and colleagues, 1994[33]

Nine different prostheses were used by six surgeons at either a publicly funded hospital or a private clinic. Prostheses implanted in the private hospital survived only 70% as long as those performed in the public hospital ($p < 0.001$). The two patient groups were of a similar age, sex and natural life-expectancy. Average period of attendance at follow-up clinics was shorter in the private group than in the public group and may be a contributing factor but the multi-variable analysis was unable to explain the difference in survival of prostheses between these types of hospital. **C-rated.**

Schuller and Marti, 1990[41]

The use of the same type of prosthesis was compared in a teaching hospital environment (with a metal head) and in a private clinic (with a ceramic head). The amount of wear measured differed significantly ($p < 0.001$) but the possible confounding effect of differences attributable to the hospital setting was not addressed – patient groups were considered clinically comparable although the "main differences between the groups were socio-economic". **C-rated.**

Comments

The evidence from these studies for the comparative effect of grade/experience of surgical staff upon prosthesis longevity is conflicting. Multi-variable analysis of variance was not possible in the study by Ahnfelt and colleagues,[14] so the respective

contribution of surgeon, hospital and patient-related factors could not be estimated. This study suggests that the more specialised centres have better results overall, measured in terms of aseptic loosening and revision.

Body mass

It is rare that studies such as the following A-rated study are performed even though patient variables are important to outcomes of THR. Patient characteristics should be researched more fully in hip prosthesis survival studies.

Lehman and colleagues, 1994[27]

In this retrospective study, primary THRs without cement, implanted over a 7-year period, were divided into two groups dependent on the body-mass index of the patient concerned. Normal weight patients had an index of between 20 and 30 (n = 142 hips), while obese patients had an index of > 30 (n = 60 hips). The obese group had a subsection within it of those who were morbidly obese – body mass index of 40 or more (n = 8 hips). Those with a body-mass index of less than 20 were excluded. The patients, 30% of whom were female, had an approximate average age of 50 years, and 62% had a diagnosis of osteoarthritis. Normal weight and non-morbidly obese groups had a significant increase in each functional measure between prostheses- and postoperative evaluations ($p < 0.001$). The morbidly obese group also had increases, although smaller, in most of the measures ($p = 0.01$–0.05). In the normal weight group, 7% of cups and 7.7% of stems were either loose or revised, compared with 8% of cups and 1.7% of stems in the non-morbidly obese group (p, not significant). The morbidly obese group had no loose components and none required revision. **A-rated.**

Comments

This authors of this study concluded that obese patients (those with a body-mass index of > 30) could benefit from primary total hip arthroplasties without cement and that obesity did not markedly increase the operative risk. However, they do point out that "substantial differences might occur with long-term follow-up". This needs to be researched more fully.

Age groups

Neumann and colleagues, 1996[108]

One surgeon performed 240 Charnley hip arthroplasties in 211 patients in just over 6 years and data on the patients were collected prospectively. A total of 52 hips were implanted in patients aged between 34 and 55 years, 37 (71%) of whom were available for follow-up after approximately 17 years. Of patients aged over 55 years, 41% were also available for follow-up after a similar period (n = 77/188). A diagnosis of osteoarthritis was made in 79% of cases. The only difference seen in Charnley Hip Scores was in the Function section, where the older group had slightly reduced scores. This was thought to be caused by a deterioration in general health. The number of revisions and loose components were higher in the younger group but this was not statistically significant. Thus the probability of survival at 20 years did not differ between the two groups (younger group = 88.3%, older group = 89.3%). **B-rated.**

Comments

There is conflicting evidence on the performance of different prostheses in different age groups. Age is used as a proxy for physical activity levels but this is not a straightforward assumption. The study above concluded that Charnley low friction arthroplasties can be used for younger patients with "excellent long-term results" comparable to those in an elderly age group. However, in the C-rated study by Hoffman and colleagues,[33] in which various types of prostheses were assessed, the hips were reported to survive longer when implanted into older patients. Hips in patients over the age of 66 survived longer than those in younger patients ($p < 0.05$). More studies such as that by Neumann and colleagues[108] are needed, in which one type of prosthesis is compared in different age but otherwise matched groups, in preference to studies involving many types of THR from which only generalised conclusions can be drawn.

Cement types

The following paper investigated a new bone cement (Boneloc®) which had been developed to reduce both the leakage of chemicals and the curing temperature, both considered to be possible reasons for the failure of cemented prostheses. The new cement was compared to a conventional polymethyl methacrylate cement (Palacos®). The mechanical and chemical properties of Boneloc were assessed during laboratory tests and presented with the clinical results. The study reported the new Boneloc to have "inferior fixation" to the conventional Palacos, giving indications of increased risk of loosening. The authors suggest that this was probably caused by its mechanical properties and possibly by other mechanisms such as an increased release of monomers.

Thanner and colleagues, 1995[109]

This was a comparison of two types of cement – Boneloc and Palacos – involving 30 hips in 30 patients, aged 63–76 years, 27 of whom had primary osteoarthritis. Full radiostereometric analysis was possible in 24 patients only at 1 year (one Boneloc patient had died). Palacos fixed cups had "a small" lateral migration while cups with Boneloc migrated medially ($p = 0.03$) and proximally ($p = 0.04$); 1/16 Palacos stems subsided 0.27 mm while 6/13 Boneloc stems subsided 0.22–1.0 mm ($p = 0.005$). Increased acetabular radiolucent lines and femoral "relative cement– cortical bone contact" occurred in the Boneloc group compared with Palacos ($p = 0.04$ and $p = 0.03$, respectively). Harris Hip and Pain Scores and a Visual Analogue Scale for pain improved postoperatively ($p = 0.0004$–0.002) but did not differ between the groups (p, not significant). **C-rated.**

Cementing techniques

Cementing techniques in Charnley prostheses have been assessed most frequently, as in the first two papers summarised below. In the third paper cementing techniques are not compared (the authors state that they did not differ greatly between the two groups) but differences between Charnley designs are assessed over the same period; thus, the results may impinge on the other studies.

Cornell and Ranawat, 1986[110]

Early cementing techniques were used to implant four different prostheses in 62 hips between 1971 and 1978 and modern cementing techniques were used in 16 hips (two types of prosthesis, 1979–80). The hips were followed-up retrospectively after 5 years. The patients had a mean age of 48 years, 79% having a diagnosis of osteoarthritis, and 55% were women. There was a lower incidence of radiolucent lines around the cups plus lower radiolucent scores in the modern technique group ($p < 0.025$ for both). There were no revisions in either group. The modern technique group had no loose components by 5 years. By 10 years the early technique group had three cups loose. **C-rated.**

Ranawat and colleagues, 1988[111]

One surgeon performed 155 operations using cemented prostheses and, from these, 100 were matched for age, sex, diagnosis and body weight. Between 1970 and 1975, 50 operations were performed using early cementing techniques; the rest were implanted after 1979 using modified techniques. After a 5-year follow-up, 8% of the early group had migrated compared with none of the modified group ($p = ?$) and the cumulative radiolucent score was found to be lower in the modified group ($p = 0.0005$). Within the 5 years no early technique hips were found to be loose or require revision. None of the cups in the modified technique group were loose or revised but one stem required revision because of loosening. **B-rated.**

Dall and colleagues, 1993[112]

Between 1970 and 1986 a variety of surgeons implanted 1309 Charnley low friction arthroplasties in 1809 patients. From this group 666 hips were assessed after approximately 8 years: 264 early generation design (1970–77) and 402 second generation design (1975–86). Approximately 77% of the patients had osteoarthritis, their approximate mean age was 60 years and 60% were women. The probability of survival with respect to loosening at 10 years was reported to be 99.35% for the early hips and 86.8% for the second generation hips ($p < 0.0001$). The revision rates for both were similar: 8% early, 9% second generation. **C-rated.**

Other studies

Other prostheses have also been assessed, as part of other studies, with respect to cementing techniques. Stanmore hips had a 10-year survival without revision probability of 91.6%, when first generation techniques were used, compared with 97.4% for second generation ($p = 0.005$).[32] In another study, 307 T-28 and 162 TR-28 hips were implanted using early techniques and 99 MOSC hips were inserted using modern techniques. An increased incidence of femoral subsidence of > 5 mm in the T-28 and TR-28 hips compared with the MOSC hips (p, 0.004–0.0075) was attributed by the authors to the different methods of fixation.

Comments

The two studies which assessed cementing techniques in Charnley both showed lower radiolucent line scores and incidence to be associated with the modern cementing techniques. Over the 5-year follow-up period in both studies, this did not translate into higher revision or loosening rates but this may occur later. However, the second generation Charnley design hips in the third paper had a lower probability of survival, compared to the first generation and, if these hips were used together with the modern cementing techniques, the longer-term results might not be so clear.

Aseptic loosening

The study below was designed to assess the possible reasons for loosening within the Stanmore cemented prosthesis.

Kristiansen and Steen Jensen, 1985[113]
A total of 33 Stanmore hips with aseptic loosening were compared with a matched control series without loosening. The diagnosis for 94% of the patients was osteoarthritis, their mean age was 64 years and the study had a mean follow-up period of 36 months. Previous operations had been undertaken in four of the revision group compared with none in the control group ($p < 0.05$). Loosening occurred more often when calcar bone stock was thin prior to surgery ($p < 0.001$); insufficient packing was found in 88% of the loose hips and 39% of the stable hips. A varus position of the stem was associated with loosening as opposed to the neutral or valgus positions. **B-rated.**

Comment
The factors associated with failure possibly contributed to the loosening of prostheses. There is scope for review of further good quality studies on the mechanics of loosening in other patients and in other types of prostheses.

Wire versus cable

One study compared these methods of fixation.

Kelley and Johnstone, 1992[26]
Either a Charnley or an IOWA stem was paired with either a Charnley (22 mm) cup, an all-polyethylene (28 mm) cup or a metal-backed (22 mm) cup. Two methods of fixation were used: stainless steel wire (n = 162) or cobalt–chrome cable (n = 160); follow-up period was approximately 6 years. The patients' approximate mean age was 66 years, 52% were women and 81% had a diagnosis of osteoarthritis. Trochanteric union rates were 75% for the wired hips and 79% for those with cable. Breakage of the entire trochanteric fixation construction (all three wires or cables) occurred in 43% wire and 12% cable (the cables in 56% of the hips unravelled, 47% of these had no broken cables). Analysis of roentgenographs (performed independently of the surgeon and blinded where possible) showed loosening of the cup in 12% of wired hips and 23% of the cabled ones. The difference in cup loosening, adjusted for cup type, was significant ($p = 0.003$). **A-rated.**

Comments
Cable was introduced to improve trochanteric union rates but this study did not show any significant results in this area. Bone destruction occurred more frequently with cable ($p < 0.001$) and was associated with debris coming from the cables themselves. Debris may be responsible for the higher incidence of cup loosening in those hips with cable. These results suggest that there is no advantage of cable over wire and, as the authors point out, caution should be used when considering the use of cable for trochanteric fixation.

Isoelastic hip versus porous-coated prosthesis

This paper was not included with other prosthesis comparisons because the Butel prosthesis is of a different type to all the others.

Jacobsson and colleagues, 1994[114]
Two senior surgeons operated on 56 patients (24 women, 32 men, mean age 52 years), of whom 75% had osteoarthritis; the rest had a variety of reasons for the unilateral hip operation. Patients were matched in pairs for sex, age, weight and radiographic appearance before being randomly selected (no details of method) to have a Butel (stem made of four rods for flexibility) or a PCA (rigid) stem (three different press-fit cups were used). Each pair was operated on by the same surgeon and was followed-up for 3 years. The PCA stem gave better results, as assessed by Harris Hip Scores (mean 94.4 compared with 78.5 Butel, *p*-value not given) and the number of prostheses definitely or probably loose (PCA 18%, Butel 86%). Both groups required three hips to be revised because of loosening (one further Butel hip was revised for other reasons). **C-rated.**

Comment
The Butel was "designed to obtain flexibility similar to the proximal femur by using four rods (titanium alloy) connected proximally and distally" and was supposed to "exhibit fewer signs of stress-shielding". However, the porous-coated prosthesis gave better Harris Hip scores and was far more stable than the Butel system. Although the numbers of revisions were similar in both groups during the 3-year follow-up period, the increased number of Butel hips which were definitely or probably loose may indicate that more of these prostheses would require revision in later years than the PCAs.

Chapter 11

Economic model

Introduction

The aim in this chapter is to estimate the costs of THR using an economic model. The focus is on the model developed to incorporate relevant characteristics and costs which enables comparisons to be made of different prostheses. Results are presented by applying this model using available data. The model can also be applied as new data become available on the survival of existing prostheses, on new prostheses, on changes in costs or local data. Thus the results give an assessment of the state-of-the-art now, and the model enables these findings to be revised as new data become available. The model itself is therefore an equally important product of the research as are the results themselves.

Methods

The obvious costs of a THR are those of primary replacement. However, the total expected costs are, in fact, greater than this and may include the costs of revision.

The concept of total expected costs is based on those costs expected to be incurred over a number of years. In the case of THR, total expected costs are the sum of the primary replacement costs and the expected costs of revision. These expected costs of revision are the actual cost of revision multiplied by the probability that a revision will be performed. Thus a combination of a low revision rate and a low revision cost will result in a low expected cost of revision; likewise, a high revision rate combined with a high revision cost will result in a high expected cost of revision. Given a population who have had primary THRs, a number of revisions will be required in each future year and associated costs will therefore be incurred in each future year. For comparability, these expected costs are converted into their value now (present value). This conversion is required because a given quantity of money has different values in different years in the future. The basic principle is that the present value of £1 in the future is less than the value of £1 now. The conventional method of converting costs into their present value is called discounting and is explained in more detail below.

By calculating the total expected costs of THRs, comparisons between different prostheses can be made. Assume a choice of two prostheses: Y needs no revisions over 20 years and X has, say, a 1% per annum revision rate. Prosthesis Y costs £1000 more than prosthesis X. A purchaser, making a decision based on expected costs only, would chose Y rather than X if the expected costs of revisions of prosthesis X over the next 20 years were more than £1000 (and vice versa). Thus for equivalent total expected costs over 20 years, primary plus expected costs of revisions of prosthesis X must equal primary costs of prosthesis Y.

Comparisons of different prostheses can thus be made if details of primary and revision costs and survival data are known for both. The method can also be used for making comparisons for which costs are known but survival data are not. For example, assume that there are no survival data for prosthesis A but the costs of a primary replacement are known, and both costs and survival data are known for prosthesis B. If the primary replacement costs of prosthesis A are greater than the total expected costs of prosthesis B, then it is inevitable that the total expected costs for prosthesis A will be greater than those for prosthesis B. Even if prosthesis B had a higher revision rate than prosthesis A, B would still be preferred if a decision was based on cost alone.

This model calculates expected costs over 20 years and assumes that the quality of life of recipients is equal however many revisions are undertaken. Obviously, in terms of benefits to individual patients, any prosthesis with a lower revision rate would be preferred, as patients would not need to undergo repeated surgery. If long-term quality-of-life data were available for various prostheses, it might be possible to undertake a cost–utility analysis to compare properly the costs and benefits of THR. These data are, however, not available. Care should therefore be taken to consider the dis-benefits of repeated revisions when making choices between prostheses.

Empirical data are not readily available on all the costs which contribute to the total cost of a THR or on survival rates. When there are empirical data, these often indicate wide variations. Hence,

PAPER 3

using mean values may give misleading results for the expected costs of revisions and of the choice between prostheses for different orthopaedic surgeons. When the data on which the calculation depends are subject to a degree of uncertainty, it is vital to undertake a sensitivity analysis in which 'high' and 'low' estimates are stated for each component that is subject to some imprecision. These high and low estimates are substituted in place of the original values and the effects on the final outcome examined. The input factors that have the greatest effect on the level of total expected costs can then be investigated to see if they would change the relative total expected costs.

The model

Previous articles based on similar methods have been published by Daellenbach and colleagues in 1990[23] and by Gillespie and colleagues in 1995.[115] Daellenbach and colleagues developed a mathematical model based on costs and patient survival for comparative economic appraisal of cemented and cementless prostheses. Their results suggested the numbers of additional years a cementless prosthesis needed to last, above that of a cemented prosthesis, to justify its extra cost of NZ$1200. The figures were given for a range of additional costs of revision.

Gillespie and colleagues[115] used Swedish and Australian data to estimate the potential cost-effectiveness of new prostheses with unknown outcomes for different age groups and mortality rates. The present value of the future costs of a prosthesis of known cost and survivorship was compared to the theoretical present value of a new prosthesis with known cost but unknown outcome. Their results indicate that possible future savings resulting from increased survival and lower revision costs do not justify the use of prostheses which cost substantially more than a conventional component.

Similarly, in 1996, Pynsent and colleagues[116] suggested a model for purchasers based on a "lifetime care package". For a given initial outlay, a purchaser would buy a primary replacement and any subsequent THR revisions. The initial cost would take into account, among other factors, the quality of prosthesis in terms of expected revision rate. This computer-based model takes account of prosthesis failure, death of the recipient and re-revision rate. Its conclusion is that if this method of pricing lifetime care according to quality of prosthesis was adopted, then monitoring, and thus the availability of survival data, would improve. Additionally, there would be a disincentive to suppliers to publish overoptimistic survival rates. This is because the supplier charges a fixed cost for the care package and would thus incur a loss if the actual cost of lifetime care was higher than that advertised.

The model developed here is based on the equation given by Gillespie and colleagues[115] and is used to estimate the present value of expected total costs of THR over 20 years.

The equation is of the form:

$$PV_{cj} = C_j + H + \sum_{i=0}^{19} \frac{\{Lmi.\ Pc_jmi.\ (C_j + H + R)\}}{(1 + r)^i}$$

where:

C_j = cost of prosthesis j

H = hospital costs including separate categories for:
- theatre costs
- ward costs
- prophylaxis costs
- physiotherapy costs

Lmi = probability of an individual at age m when receiving a hip replacement being alive in year i

Pc_jmi = probability of prosthesis C_j in an individual aged m needing to be revised in year i

R = additional costs of a revision (i.e. additional hospital costs)

$1/(1 + r)^i$ = a discount factor where r is the discount rate, i = 0–19 where 0 is the year of the primary operation.

Essentially this gives the present value of prosthesis C as dependent upon initial prosthesis and hospital costs, plus the sum of expected future costs of revision. Future costs of revision are themselves dependent on the age of the recipient and the survival of the implant. All future costs are discounted.

Primary operation costs

Data on resource use were obtained from two collaborating hospitals in different regions of England. Both hospitals gave details of the prices they charged for primary unilateral THRs. These costs were broken down into theatre, ward, physiotherapy and prophylaxis costs for each hospital. One hospital also gave costs of revision surgery. Prosthesis costs were given separately.

For both hospitals, the cost per hour in theatre and the time spent in theatre was determined. One hospital had supplied these details in their prices. The other gave total theatre costs only,

broken down into theatre medical staff and other theatre costs. From separate data on operation and anaesthetic time, average theatre time for primary replacements was estimated and thus the cost per hour of theatre time. Data were obtained for ward costs in a similar way. Both hospitals gave total ward costs for primary replacement. One hospital gave details of cost per day and the average length of stay; for the other ward, cost per day was estimated from data on the average number of days stay and the total ward costs supplied.

By sensitivity analyses it was possible to determine the effect on the present value of total expected costs of long and short theatre times and lengths of stay.

In three studies, by Francis and colleagues,[117] Wittmann and colleagues,[118] and Sharrock and colleagues,[119] it is suggested that cementless prostheses reduce the risk of postoperative thromboembolic disease, while Laupacis and colleagues[120] found no difference in the frequency of deep vein thrombosis between two patient groups (cemented and cementless). Lieberman and Geerts[121] suggest that prophylaxis against deep vein thrombosis reduces both symptomatic thromboembolism complications and saves lives, and saves subsequent healthcare expenditure. Data on prophylaxis costs provided by one of the collaborating hospitals indicated that they were a small proportion (1.8%) of total primary replacement costs. In summary, prophylaxis costs make up only a small proportion of total costs and the evidence suggests that prophylaxis against postoperative thromboembolic disease is cost-effective.

Prophylaxis and physiotherapy costs were supplied by one of the collaborating hospitals (Hospital A). The second hospital was not able to separate these costs from other costs and therefore those for Hospital A were used as a proxy and subtracted from theatre and ward costs as appropriate. To allow cost per hour in theatre and on the ward to be calculated (see above), prophylaxis and physiotherapy costs were included in the model independently from overall ward and theatre costs.

Data supplied by one of the hospitals suggested that prices charged for a revision are 1.19 times the price of a primary unilateral replacement; that is, the additional costs of revision are 19% greater than the primary replacement costs. Using this percentage, the additional costs of revision were therefore made proportional to the primary replacement costs for both hospitals. The additional costs for revision are due, in part, to the longer operation time (on average, 40 minutes extra) and longer length of stay (approximately 3 days). The model assumes that the impact of these extra times remains proportional to the costs of primary revision. This allows the sensitivity analyses to impact on both primary and revision costs.

In 1996, Pynsent and colleagues[116] suggested that the overall costs of revision surgery were twice that of a primary replacement because of the longer operation time and length of stay in hospital. No empirical evidence, however, was presented to justify this assumption. Revision surgery can be complicated and difficult to perform, with the personnel undertaking the revisions requiring higher skill levels. As such, revision costs cannot be assumed to be equal to primary replacement costs. This issue was examined in the sensitivity analysis by altering the revision costs to double the primary costs and determining the effect on the total expected costs.

The two collaborating hospitals gave their own prosthesis costs. Both supplied the price charged to purchasers for 'a hip prosthesis'. This was the average prosthesis price for all types of prostheses used in the hospitals and was not the specific price of a Charnley or other model. The survival data and prosthesis cost of the Charnley are used as the gold standard in the model for comparison with other prostheses. We decided to model the equation using the hospitals' prosthesis costs first and then using the price of a Charnley prosthesis as quoted by Murray and colleagues.[91] The costs of other prostheses were also taken from this paper to allow comparisons to be made between different prostheses using costs that were provided for the same year. This paper reviewed all manufacturers and distributors in the UK and listed prostheses supplied by most major competitors with their market price. Where survival data were available for other (non-Charnley) prostheses, these were used with associated prices to estimate total expected costs. However, prices may vary from those stated, dependent upon quantity purchased and arrangements with purchasing organisations. Local prices should be used where known. The published prices were used to indicate a range of market prices for the sensitivity analysis.

Prosthesis survival data

The probability of an implant needing revision in year i was estimated from published sources. As mentioned above, the survival data and prosthesis cost of the Charnley are used in the model as the gold standard for comparison with other types of prostheses. The survival data for the Charnley

for up to 20 years have been published in various articles.[7,9,122] The data from these sources were collated and a best estimate of the probability of revision in each year over 20 years was calculated. Gaps in the data were filled assuming a straight line relationship between two known points. The average revision rate per year over 20 years for a Charnley prosthesis was about 1% (ranging from a low of 0.5% to a high of 3%).

To estimate the present value of expected total costs for a range of competing prostheses, the same exercise was undertaken for other types. As our review of evidence has shown, there are few data available on the survival rates of many prostheses. Often the survival rate was known at only one or two points in time, for example, after 4 or 5 years. For these prostheses, a straight line was fitted through the known point(s) and 100% survival (at time zero). This rate was then extrapolated over 20 years. Obviously, survival may not be linear, and rates of revision will increase as the number of years since replacement increases. Some implants show a dramatic rise in revision rate after about 5 years. Care should be taken when interpreting the results of prostheses with limited survival data. A rise in revision rate above the linear rate assumed in the model would result in increased expected costs.

Much of the published survival data are reports from 'centres of excellence' and may not therefore reflect common practice in the UK where revision rates may be higher. However, as long as data from centres of excellence are used consistently for all prosthesis types, the comparisons ought to be a reliable guide to the relative performance of different prostheses. What matters is relative cost. However, if the model is to be used at local level to inform purchasing decisions, it is important to model local survival rates in the equation. An increase in the probability of revision will result in an increase in the total expected costs.

The sensitivity analysis gives total expected costs for straight line revision rates over 20 years of 3% and 5% per year.

Discount factor

The conventional way of accounting for costs (and/or benefits) of a treatment occurring over a number of years is to put them on the same basis by discounting. The value of £1 in i years time is less than the value of £1 now, even after allowing for inflation. This is because costs incurred in the future are less important to us than costs now. To allow for this change in value over time, costs incurred in the future are multiplied by a weighting factor (the discount factor) thus enabling the comparison of current and future costs as if they occurred at the same time. The discount factor is $1/(1 + r)^i$, where r is the discount rate and i the year in which the costs (or benefits) occur.

The total expected costs of a hip replacement include the original replacement and the costs of any subsequent revisions. Revisions may occur in any year and the costs occur at the same time. For those revisions taking place in 15 or 20 years time, the present value of these costs will be small because the denominator of the discount factor becomes larger (as i increases). The discount factor applied to a cost 20 years into the future is about 0.39; that is, the present value of a cost incurred 20 years hence is only about 40% of its nominal value. This model estimates the total expected costs over a maximum of 20 years. Costs due to revisions will be incurred after 20 years; however, for the reasons described above, these costs will be small.

The discount rate in the original calculation was assumed to be 5%. This figure was varied in the sensitivity analysis to 0% and 6%.

Mortality data

The probability (Lmi) of a patient who received a primary hip replacement at age m being alive in year i was calculated from OPCS Mortality Statistics for 1992.[123] The probabilities for males and females were modelled separately in the equation. Median age at primary operation (70 years) was calculated from data supplied by one of the hospitals. Swedish data supplied by Malchau and colleagues[122] supports this by giving mean age at primary replacement for men as 67.5 years and for women 68.2 years. An age of 70 years and the corresponding probability of being alive in year i was used in the calculations for both hospitals.

In the absence of any other data it is assumed that individuals who have had a hip replacement have a mortality rate equal to the general population for their age group. However, the true mortality rate for individuals with osteoarthritis and rheumatoid arthritis is different from the general population.[122] Current research by the Somerset and Avon Survey of Health (Personal communication, 1996) suggests that people with a THR may have a lower life-expectancy than average. Accurate data on true mortality rates are not known. If these data were available and used in this model, and the mortality rates were higher than for the general population, then total expected costs would be lower than suggested because of the greater number of

individuals dying before their prosthesis needed replacing. A youngest age of 20 years and an oldest of 80 years was used in the sensitivity analysis. For elderly patients, the discount factor combined with the high mortality rate results in very small expected costs in 20 years time. However, for young patients, the low mortality rate and the increasing THR revision rate over time mean that a high proportion of younger individuals will survive longer than their implants and, despite the discount factor, these costs are still important. Young patients will live for more than 20 years (that is, longer than the span of the model) and this must be considered when comparing the results.

Assumptions

A spreadsheet model is used to estimate the total expected costs of one prosthesis relative to another. The results from the model are intended to further current knowledge and to aid decision-making, not to prescribe policy. This model is based on a number of simplifying assumptions.

- As explained above, the model assumes prices quoted in a 1995 paper.[9] These prices may have changed and prices of prostheses may also vary between purchasing institutions. Any increases in prosthesis price will obviously increase total expected costs. More importantly, any change in the relative prices of prostheses may change their relative cost-effectiveness.
- A further assumption is that prosthesis revision rates are linear when long-term survival data are not known. These estimates are based on trends for the years immediately following primary replacement. The assumption that the rate is linear throughout the 20 years of the model will underestimate longer-term revision rates and, thus, underestimate total expected costs.
- The model assumes that mortality rates of THR recipients are equal to those of the general population. This assumption is made because of the lack of data on actual mortality rates of THR recipients. If actual rates are higher than average, then the model will overestimate total expected costs, and this may change rankings of cost-effectiveness. This is because, if the mortality rate of THR patients increased, thus reducing length of life, there would be no change in the total expected costs of a low revision prosthesis whereas there would be a reduction in number of revisions and, hence, costs of a high revision rate prosthesis.
- No account is taken in the model of re-revisions. There is, inevitably, a cumulative effect in that a number of primary replacements will fail, be revised and fail again. Revision THRs have a greater chance of needing a further revision.[116] This is not incorporated in the model which assumes a maximum of only one revision over the 20-year period. Including re-revisions would increase the total expected costs.

Results

This section includes:

- an estimate of total expected costs, based on survival data for the Charnley prosthesis and on actual hospital costs
- a comparison of the total expected costs of other prostheses
- a sensitivity analysis.

An estimate of total expected costs (based on survival data of the Charnley prosthesis and actual hospital costs)

The costs provided by both Hospitals A and B are presented in *Table 23*. Published yearly survival rates for the Charnley prosthesis and the estimated revision rates used in the equation are presented in *Table 24*.

TABLE 23 *Primary unilateral replacement costs in Hospital A and Hospital B*

Type of cost	Hospital A (£)	Hospital B (£)
Prosthesis	629	700
Theatre	1197	946
Ward	1651[a]	2533
Prophylaxis	66	NA[b]
Physiotherapy	71	NA
Total for primary unilateral	3614	4179
(Additional revision costs)	(693)	(NA)
(Total costs of revision)	(4307)	(NA)

[a] *Ward costs in Hospital A include direct and indirect overheads and exclude SIFT payments.*

[b] *NA, not available.*

If the revision rates from *Table 24* and the costs from *Table 23* are used in the model, with the average age of primary implant (70 years) and a discount rate of 5%, the present value of expected revision costs for a selection of years will be those shown in *Table 25* (for each hospital and for men and women separately). The primary replacement costs are the hospital and prosthesis costs for the initial replacement.

TABLE 24 *Published yearly survival rates for the Charnley prosthesis*

Year *i*	Published survival rate (%)	Estimated revision rates*
1	99.1	0.009
2	98.4	0.007
4	97.0	0.007
5	95.9	0.011
6	95.0	0.009
8	93.6	0.007
10	89.7	0.03
13	92.0	0.005
14	79.0	0.005
15	87.3	0.005
18	81.0	0.025
20	86.8	0.018

* *To avoid negative revision rates, when survival rates rose over time, the rates were assumed to be a straight line between the years on either side.*

In the Methods section above, a choice of two implants was assumed, one needing no revisions over 20 years and the other having about a 1% revision rate (as in the above model). A prosthesis with an expected 100% survival rate over 20 years has costs over 20 years equal to primary replacement costs. A purchaser making a decision based on expected costs only would not be prepared to pay substantially more over 20 years for one prosthesis rather than the other.

The costs of primary replacement and expected revision costs over ā number of years for men and women are presented in *Table 25*. The expected costs of revision for women are slightly higher than for men because of the lower mortality rate for women; that is, fewer women die before needing a revision.

In Hospital A, for men, the difference between the cost of a primary replacement (£3614) and the expected total costs over 20 years is £297, the expected costs of revisions. For Hospital B the difference in costs is £344. The expected revision costs over 20 years for women are £371 and £431 for Hospitals A and B, respectively.

These figures imply that, for male patients in Hospital A, assuming equal hospital costs for different prostheses, a purchaser would not be prepared to pay more than £297 extra, compared with the cost of the current prosthesis, for a new type of prosthesis. Paying more than £297 extra (that is, £926 in total) for the 'no revisions' prosthesis would result in the costs over 20 years being greater than costs using the current implant. Using the costs supplied by Hospital B, the maximum extra a purchaser would be prepared to pay for a no revisions prosthesis is slightly higher at £344 (£973 in total).

TABLE 25 *Present value of expected revision costs for males and females for a selection of years (average hospital prosthesis and Charnley prosthesis costs separately)*

	Present value of total expected costs (£)							
	Hospital A				Hospital B			
Prosthesis price	**629[a]**		**353[b]**		**700[c]**		**353[b]**	
Primary replacement costs (including prosthesis)	**3614**		**3338**		**4179**		**3832**	
Expected costs of revisions	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
At end of 5th year	136	145	127	135	158	168	147	156
At end of 10th year	244	278	228	260	283	323	263	301
At end of 15th year	265	310	248	290	308	360	286	335
At end of 20th year	297	371	278	348	344	431	320	401

[a] *Average prosthesis cost in Hospital A, including cement (£71)*
[b] *Cost of Charnley prosthesis*[9]
[c] *Average prosthesis cost in Hospital B*

Similar figures using the lower prosthesis prices quoted by Murray and colleagues[9] give the maximum extra a purchaser would be prepared to pay for a no revisions prosthesis in men as £278 and £320 for Hospitals A and B, respectively (£631 and £673 in total, respectively). For women, the corresponding figures are £348 and £401, respectively (£701 and £754 in total, respectively).

Obviously, given a choice, the recipient of an implant would prefer the prosthesis with the lowest revision rate or not to have to undergo revision surgery at all. Here only the expected total costs of THR have been examined, without considering any benefits.

Comparison of the total expected costs of other prostheses

An indication is presented here of the total expected costs of a range of competing prostheses, for comparison with the above estimates which used the survival rates for the Charnley prosthesis. Published costs and estimates of survival rates were used in the model together with the cost data from Hospital A.

These results are based on published survival rates from centres of excellence which may not reflect common practice but do compare prostheses in similar settings. The expected costs of both prostheses will be greater in hospitals where the prosthesis survival rate is lower.

The results are given separately for prostheses where the longer-term survival rates are available (i.e. more than 10 years) and for prostheses where only short-term survival rates are known (i.e. less than 10 years).

Prostheses evaluated over the long term (with 10 years or more data on survival)

Survival rates of 10 years or more were used for five types of cemented prosthesis and one cementless prosthesis. These data are presented in *Table 26*, together with the price of the prosthesis, if available.

The results were derived separately for men and women, using the following assumptions:

- hospital cost data from Hospital A
- for cemented prostheses, the average cost (£629 including cement) from Hospital A was used if no prosthesis specific cost was available
- for cementless prostheses, the average price (£1150) of those listed by Murray and colleagues[9] was used if no specific prosthesis cost was available
- a recipient aged 70 years at time of implant (unless otherwise stated)
- a discount rate of 5%
- the cost of reoperation remained proportional to the cost of primary operation (excluding prosthesis cost)
- linear prosthesis revision rates (except for the McKee-Farrar prosthesis for which the actual rates were known).

Actual revision rates will probably rise over time leading to greater expected costs than those indicated here. The resulting expected costs using these linear assumptions are presented in *Table 27*.

TABLE 26 *Published prices of prostheses and their survival rates*

Make/type	Price (£)	5-year survival rate (%)	10-year survival rate (%)	15-year survival rate (%)	20-year survival rate (%)
Cemented					
McKee-Farrar[74]	–	94.7	91.0	84.3	48.9
Stanmore[91]	285 (250–320)	–	94.0	91.0	–
Howse[79]	–	–	90.0	80.0	–
Exeter	340	98.0[9] (4 years)	–	89.0[76] (13.4 years)	–
Cemented alumina-alumina (age ≤ 50 years)[124]	–	95.0	94.7	–	–
Cemented alumina-alumina (age > 50 years)[124]	–	95.0	80.4	–	–
Cementless					
AML[82]	799	–	85.0 (12 years)	–	–

PAPER 3

TABLE 27 *Expected total costs of Charnley prosthesis and seven comparison prostheses*

Make/type of prostheses	Assumed price (£)	Expected total costs (£)					
		Over 10 years		Over 15 years		Over 20 years	
		Male	Female	Male	Female	Male	Female
Charnley	**353**	**3566**	**3598**	**3586**	**3628**	**3616**	**3686**
Cemented							
Stanmore	356	3489	3506	3514	3543	3522	3560
Exeter	411	3579	3604	3621	3667	3636	3697
Cemented alumina-alumina (age ≤ 50 years)[a]	629	3813	3814	3874	3875	3919	3922
Howse	629	3877	3907	3966	4040	3998	4102
McKee-Farrar	629	3855	3883	3909	3965	4023	4189
Cemented alumina-alumina (age > 50 years)	629	4061	4130	4149	4261	4179	4322
Cementless							
AML	799	4125	4165	4343	4470	4406	4610

[a] *Recipient assumed to be aged 40 years*

TABLE 28 *Published prices of prostheses and their survival rates*

Make/type	Price of prostheses (£)	4-year survival rate (%)	5-year survival rate (%)	7-year survival rate (%)	9-year survival rate (%)
Cemented					
Müller straight stem	334	–	99.0[91]	–	94.0[91]
CAD	–	–	–	95.1[122]	–
Lubinus IP	–	–	–	95.5[122]	–
Spectron	700	98.0[91]	97.9[122]	–	–
Lubinus SP	700	98.5[122]	–	–	–
Cementless					
PCA	–	–	94.4[122]	95.0[126]	–
Omnifit	1260	–	99.0 (5.6 years)	–	–
Harris-Galante	–	–	96.71[125]	–	–

The expected total costs of the Charnley are £3616 and £3686 per hip over 20 years for men and women, respectively. From *Table 27,* prostheses with survival data of 10 or more years, the Stanmore, with a prosthesis and cement price of £356, has expected total costs of £3522 (£3560 for women) over 20 years, almost £100 (£126) less than the Charnley. The Exeter prosthesis has similar costs over 20 years to the Charnley at £3636 (£3697), with an initial prosthesis and cement cost of £411. The one cementless prosthesis (DePuy's AML) at £799 is seemingly not a cost-effective option for either men or women given the assumptions used in this model.

Prostheses evaluated over the short term (with less than 10 years data on survival)

Survival rates of less than 10 years were used for six cemented and three cementless types of prosthesis. These data are presented in *Table 28,* together with the price of prosthesis if available. The results were derived from the model using the same assumptions as above. The resulting expected costs are given in *Table 29.*

TABLE 29 *Expected total costs of Charnley and eight comparison prostheses*

Make/type	Assumed price (£)	Expected total costs (£)							
		Over 5 years		Over 10 years		Over 15 years		Over 20 years	
		Male	Female	Male	Female	Male	Female	Male	Female
Charnley	**353**	**3465**	**3473**	**3566**	**3598**	**3586**	**3628**	**3616**	**3686**
Cemented									
Müller straight stem	405	3422	3424	3535	3561	3587	3639	3606	3676
Lubinus IP	629	3727	3733	3784	3802	3811	3842	3820	3861
CAD	629	3731	3738	3798	3819	3829	3866	3840	3887
Lubinus SP	771	3821	3825	3858	3870	3875	3895	3881	3907
Spectron	771	3831	3835	3877	3890	3896	3919	3903	3932
Cementless									
Omnifit	1260	4276	4278	4295	4302	4305	4315	4308	4321
Harris-Galante	1150	4259	4266	4329	4352	4362	4401	4374	4424
PCA	1150	4270	4278	4346	4370	4381	4423	4394	4447

Some of the figures on which *Tables 27* and *29* are based use an assumed prosthesis price and estimates of revision rates that are probably lower than actual revision rates. The results should therefore be treated with some caution. However, *Tables 27* and *29* do allow comparison of costs over a number of years between the gold standard of the Charnley and the competing prostheses.

Table 29 presents the results for prostheses with less than 10 years survival data; actual survival rates over 20 years may be different from the linear ones assumed. Only the Müller straight stem at £3606 (£3676 for women) over 20 years, for an initial prosthesis and cement cost of £405, is similar in cost to the Charnley. All of the cementless prostheses have expected costs over 20 years of about £700 more than a Charnley prosthesis and all are more costly than any of the cemented prostheses.

Sensitivity analysis

The purpose of a sensitivity analysis, as described above, is to indicate how sensitive the results are to certain key components in the model. The most important consideration here is the effect that changes in the key assumptions will have on the expected costs of revisions and on the relative cost-effectiveness of different prostheses in terms of total expected costs. For simplicity, the sensitivity analyses presented here model the equation for men and for Hospital A's costs only.

The sensitivity analyses explore three main issues.

1. What effect do changes in the different inputs to the model have on total expected costs?
2. Does the relative cost-effectiveness of different prostheses change as these inputs are changed?
3. In general, what are the relationships between prosthesis price and revision rate?

The effect of changes in the different inputs to the model on total expected costs

The purpose of this sensitivity analysis is to determine the effect on total expected costs over 20 years of different input data. For example, the model assumes that the length of stay as an inpatient after a THR is 13 days. This is the average length of stay from local data but other data suggest that the range may be from 9 days to 22 days. Clearly, substituting 22 days into the model for 13 days will raise costs. The question is – by how much over 20 years? Other data in the model are also subject to such uncertainties and, hence, are explored in this sensitivity analysis.

Using the Charnley prosthesis cost and survival data plus other assumptions as explained above, each of the main components in the model are altered, as appropriate, to a low and high estimate. Of the seven components in the model, three impact predominantly on costs in the current period, one affects future costs alone (revision costs), two affect the expectation of future costs (revision rate and recipient age) and one affects the weight given to future costs (the discount rate).

PAPER 3

The original data and the estimates used in the sensitivity analysis are presented in *Table 30*. The effects on the total expected costs over 20 years of varying the level of major current costs (that is, hospital and prosthesis costs) are shown in *Table 31*.

TABLE 30 *Estimates used in the sensitivity analysis*

	High	Hospital A	Low
Prosthesis	£1321.00[a]	£629.00	£321.00[a]
Theatre: length of stay	246 minutes[b]	144 minutes	60 minutes[b]
Ward: length of stay	22 days[c]	13.2 days	9 days[c]
Additional costs of revision	£3613.96[116]	£693.00	£0
Age at operation	40 years	70 years	80 years
Revision rate	5%	(1%)[d]	–[e]
Discount rate	0%	(5%)	6%

[a] *From Murray,* et al., *1995,*[91] *plus £71 for cement*

[b] *Data from Hospital A (Personal communication, 1997)*

[c] *Data from Trent Region for 1990 (Personal communication, 1997)*

[d] *Rates assumed in our basic model*

[e] *Zero revision rate: no extra costs*

TABLE 31 *Effects on total expected costs over 20 years of varying the level of major current costs*

	High cost (£)	Low cost (£)	Current cost difference (£) (high – low estimate)	Total expected cost difference (£) (high – low total costs)
Prosthesis	4650	3581	1000	1069
Theatre: length of stay	4831	3153	1500	1677
Ward: length of stay	5160	3314	1560	1846

The effect of a change in prosthesis price from £250 to £1250 falls on both current and future costs. From *Table 31* it can be seen that an increase of £1000 in prosthesis price results in increased expected future costs of £1069 over 20 years. Obviously £1000 of this increase occurs in the current period; thus, the present value of expected future costs is only about £70 over 20 years.

The range of time spent in theatre is taken from data provided by Hospital A (1–4 hours). From an average time of 144 minutes and total theatre costs, the cost per hour of theatre time was calculated to be about £500. Thus an increase in theatre time of 3 hours results in an increase in the cost of a primary operation of about £1500. From *Table 31,* the increase in total expected costs over 20 years of a 3-hour increase in theatre time is £1677. When these costs are compared with the increase in cost of the primary operation, it is clear that the additional future costs are relatively small (about £170).

The cost of an inpatient stay in Hospital A was calculated as about £120 per day. As above, the number of days stay as an inpatient may range from 9 days to 22 days (1990 data from Trent Region: Personal communication, 1997). Such an increase in stay (13 days) results in an increase in the cost of a primary operation of about £1560. The resulting increase in expected future costs over 20 years is £1846. Thus, it is clear that the expected future costs are about £285.

For each of the three analyses above, the impact on total expected costs is between £1000 and £2000. These are substantial increases for an operation that costs about £3500. Most of these costs are incurred in the current period and have very little impact on future expected costs.

The effects on total expected costs over 20 years of varying four factors affecting the level of expected future costs are shown in *Table 32:* discount rate, additional costs of revision, recipient age, and revision rate. The results are modelled using the assumptions for prostheses evaluated over the long term (page 51).

The discount rate affects all costs incurred; the further into the future the costs are incurred, the greater the effect (i.e. the smaller the present value of the discounted costs). Reducing the discount rate from 6% to 0% results in an increase in the current value of expected future costs of £143 when calculated over 20 years. The discount rate used in the model is 5%. The difference in costs between an assumed rate of 5% and 6% is very small (£14).

The effect of any **additional** costs of revision, that is, revision costs in excess of the costs of a primary replacement, occur, by definition, in the future. These costs are thus subject to the effects of rate of revision, survival of recipient and discount rate. Comparing an extra revision cost of £3614 to no extra cost of revision gives a small additional expected future cost of over 20 years of £249. This means that,

PAPER 3

despite the higher level of skill needed for revision surgery, the costs over 20 years are relatively low.

The probability of a recipient surviving to a time when a revision may be needed is dependent on age at operation. The increase in expected costs over 20 years for a primary replacement in a 40-year-old compared with an 80-year-old man is £378. This model incorporates a maximum of 20 years costs. Obviously, an individual aged 40 years could incur at least two 20-year periods; hence, the lifetime costs would be far greater. However, the further into the future that the costs are incurred, the lower their present value because of the discount factor (see page 48 above) and therefore the less 'important' they become.

It is clear from *Table 32* that the revision rate has the largest effect on the total expected costs over 20 years. A change in the revision rates of prostheses from the best estimate of Charnley prosthesis survival (approximately 1% revision per year) to 5% per year results in additional expected costs over 20 years of £1320. The costs over 20 years with a 3% revision rate would be £4584 (£673 more than a 1% revision rate (not shown)).

From *Tables 31* and *32* it is apparent that the factors affecting the total expected costs of THR by the greatest amounts are prosthesis and hospital costs and the revision rate.

TABLE 32 *Effects on total expected costs over 20 years of varying factors affecting future costs*

	High cost (£)	Low cost (£)	Total expected cost difference (£) (high – low total costs)
Discount rate	4036	3893	143
Additional costs of revision	4112	3863	249
Recipient age	4163	3785	378
Revision rate	5230	3911[a]	1320

[a] *Using 1% per year revision rate*

Does the relative cost-effectiveness of different prostheses change as input costs are changed?

Given that, for named prostheses, prices and revision rates are known, the variable input in the model that will have a large impact on total expected costs is hospital costs. In order to establish any changes in the relative cost-effectiveness of different prostheses, low and high estimates of hospital costs (combined theatre and ward costs) are modelled – that is, 60 minutes in theatre and 9 days as an inpatient as the low cost estimate and 246 minutes and 22 days as the high cost estimate. The results are presented in *Table 33*.

From *Table 33*, it can be seen that there are very few changes in the relative cost-effectiveness of prostheses as hospital costs are changed. The prostheses listed in the table are ranked according to the low cost estimate. The McKee-Farrar and CAD prostheses are the only ones to change order in the ranking when the higher hospital costs are modelled. The McKee-Farrar prosthesis moves down the ranking by two places and CAD by one place.

TABLE 33 *Relative cost-effectiveness of different prostheses using low and high estimates of hospital costs*

Make/type	Total expected costs (£)	
	Low estimate	High estimate
Charnley	**2262**	**5785**
Prostheses with 10 or more years survival data		
Cemented		
Stanmore	2205	5633
Exeter	2298	5781
McKee-Farrar	2327	6257
Cemented alumina–alumina (age ≤ 50 years)	2500	5941
Howse	2613	6217
Cemented alumina–alumina (age > 50 years)	2729	6503
Cementless		
AML	2944	6748
Prostheses with less than 10 years survival data		
Cemented		
Müller straight stem	2277	5736
Lubinus IP	2498	5938
CAD	2511	5970
Lubinus SP	2590	5950
Spectron	2604	5984
Cementless		
Omnifit	3040	6339
Harris-Galante	3050	6496
PCA	3063	6525

The reason for the change in relative cost-effectiveness of the McKee-Farrar is that the prosthesis survival rate falls quite markedly after about 15 years (15-year survival 84%, 20-year survival 49%). As more revisions are needed, the increased hospital costs impact to a greater extent on future costs than for the other prostheses with lower revision rates. The survival data for the McKee-Farrar prosthesis were available and have been used for each and every year up to 20 years. Such detailed data were not available for other prostheses and, hence, survival rates were assumed to be linear. It may be that the survival rates of other prostheses fall equally quickly as that of the McKee-Farrar, thus increasing the high cost estimate of other prostheses.

In the bottom half of *Table 33*, showing prostheses with less than 10-years' survival data, the CAD prosthesis moved down the ranking by one place to below the Lubinus SP when higher hospital costs were used. The assumed linear revision rates for the CAD and Lubinus SP prostheses are similar (0.007 and 0.004 per year, respectively). The reason for the Lubinus SP being more expensive using low hospital costs is that the prosthesis itself is more expensive than the assumed price of the CAD (£771 and £629, respectively, including cement). As hospital costs increase, prosthesis price becomes a smaller proportion of total costs and the higher revision rate of the CAD becomes the most influential factor on total expected costs.

Overall, the relative cost-effectiveness of different prostheses does not appear to be altered under assumptions of different hospital costs. The main finding is that if revision rates increase dramatically over time, increases in hospital costs will have a greater impact on total expected costs than lower revision rates.

The relationship between prosthesis price and revision rate

For general reference, the relationship in terms of expected total costs over 20 years between differently priced prostheses, revision rates and discount rates are shown in *Table 34*. Hospital costs in these scenarios are assumed to be the same for all prices of prosthesis. Prosthesis costs range from £400 to £2000, revision rates from zero to 5% per year, and the discount rate from zero to 6%.

Where the revision rate given in *Table 34* is zero, the total expected cost shown is the cost of primary operation only. This is not affected by changes in the discount rate as there are no future costs. The important question to be answered from this information is whether or not greater costs of prosthesis result in lower total expected costs because of lower expected revision rates.

Assuming an 'average' prosthesis price and revision rate of £700 and 2%, respectively, with a 5% discount rate, *Table 34* gives the total expected costs as £4342 over 20 years. As expected, the higher the cost of prosthesis, the lower the revision rate must be to make the THR cost-effective in terms of total expected costs. For a £1000 prosthesis to be comparable in total expected costs with such a £700 prosthesis, revision rates must be 1% or less per year over the 20 years. A £1500 prosthesis costs over £100 more for just the primary operation so would be required to have a zero revision rate even to be considered in comparison to a £700/2% revision rate prosthesis. Primary replacement costs for a £2000 prosthesis are equivalent to total expected costs over 20 years of a £700 prosthesis with a 4% revision rate.

The cementless prostheses listed in *Tables 27* and *29* and the more expensive prostheses in *Table 33* have greater total expected costs under the assumptions in this model than cemented/lower priced prostheses. Cementless prostheses are generally more expensive than cemented but may last longer in younger patients. If this is the case then, in younger patients, it could be expected than a £1500 prosthesis with a 0% or 1% revision rate may be more cost-effective than a £700 prosthesis with, say, a 2% revision rate. The relative cost-effectiveness of different prices of prosthesis and revision rates in a 40-year-old patient are presented in *Table 35*.

If the revision rate of a cemented prosthesis (price £700) in a 40-year-old patient is 3% per year over 20 years, the total expected costs would be £5314. Using this as the comparison figure, it is clear that more expensive prostheses can be more cost-effective with a lower revision rate. For example, assume a cementless prosthesis is priced at £1500. With a 1% revision rate, this prosthesis would be more cost-effective than our £700/3% comparison (£5127 and £5314, respectively). Likewise, both a £1000/2% revision rate and £2000/no revisions prosthesis would be more cost-effective over 20 years (total expected costs £5146 and £4985, respectively).

The conclusions that can be drawn from *Tables 34* and *35*, under the assumptions made in this model, are that:

- for patients aged 70 years at primary THR, lower priced prostheses with 1–2% revision rates are

TABLE 34 *Relative cost-effectiveness over 20 years for a range of prosthesis costs, revision rates and discount rates in a recipient aged 70 years*

Prosthesis cost (£)	Revision rate (% per year)	Total expected costs (£)		
		(0% discount rate)	(5% discount rate)	(6% discount rate)
400	0	–	3385	–
	1	3803	3691	3675
	2	4220	3997	3965
	3	4638	4303	4255
	4	5055	4609	4545
	5	5473	4915	4834
700	0	–	3685	–
	1	4133	4014	3996
	2	4582	4342	4307
	3	5030	4671	4619
	4	5478	4999	4930
	5	5927	5328	5241
1000	0	–	3985	–
	1	4464	4336	4318
	2	4943	4687	4650
	3	5422	5038	4983
	4	5901	5390	5315
	5	6380	5741	5648
1500	0	–	4485	–
	1	5015	4874	4853
	2	5545	5262	5221
	3	6076	5651	5589
	4	6606	6040	5957
	5	7136	6428	6325
2000	0	–	4985	–
	1	5566	5411	5389
	2	6148	5837	5792
	3	6729	6264	6196
	4	7311	6690	6600
	5	7892	7116	7003

usually more cost-effective than higher priced prostheses even with lower revision rates

- for younger patients, aged 40 years at primary replacement, higher priced prostheses with low revision rates can be more cost-effective than less expensive ones with higher revision rates.

Summary and conclusions of the economic analysis

The model presented above estimates the relative cost-effectiveness of different prostheses in terms of total expected costs. The results are intended to assist decision-making, not to be a prescription for policy.

The model is dependent upon a number of simplifying assumptions because of limitations in the data. These assumptions may affect both the total expected costs and the relative cost-effectiveness of prostheses in the following ways.

Total expected costs would increase and relative cost-effectiveness may change if re-revisions were included in the model. For a prosthesis with a high revision rate, the cumulative effects of the costs of re-revision would be greater than for a prosthesis with a lower revision rate. For example, in *Tables 34* and *35*, high cost/low revision prostheses may become more cost-effective than low cost/higher revision rate prostheses.

Data on the survival of many of the prostheses are limited. Some survival data are only available for up to 4 or 5 years. The model assumes linear extrapolation of survival data to 20 years. The survival rate of many prostheses may in fact fall (sometimes

PAPER 3

TABLE 35 *Relative cost-effectiveness over 20 years for a range of prosthesis costs, revision rates and discount rate in a recipient aged 40 years*

Prosthesis cost (£)	Revision rate (% per year)	Total expected costs (£)		
		(0% discount rate)	(5% discount rate)	(6% discount rate)
400	0	–	3385	–
	1	4171	3891	3853
	2	4957	4397	4322
	3	5744	4903	4790
	4	6530	5408	5259
	5	7316	5914	5727
700	0	–	3685	–
	1	4529	4228	4188
	2	5373	4771	4691
	3	6217	5314	5194
	4	7061	5857	5697
	5	7906	6400	6200
1000	0	–	3985	–
	1	4887	4565	45,228
	2	5789	5146	5060
	3	6691	5726	5597
	4	7592	6306	6135
	5	8494	6887	6672
1500	0	–	4485	–
	1	5483	5127	5080
	2	6481	5770	5675
	3	7480	6412	6269
	4	8478	7054	6864
	5	9476	7697	7459
2000	0	–	4985	–
	1	6080	5689	5637
	2	7174	6394	6289
	3	8269	7098	6942
	4	9364	7802	7594
	5	10,458	8507	8246

dramatically) over time. Increases in revision rates to levels above those assumed will increase total expected costs. Cost-effectiveness rankings will also be changed if rates of revision of prostheses change relative to each other.

The model also assumes that individuals who have had a hip replacement have a mortality rate equal to the general population for their age group. If, however, as studies suggest, these individuals have on average a lower life-expectancy, total expected costs would be lower than estimated here because of the greater number of individuals dying before their prosthesis needs replacing.

Increases (decreases) in prosthesis costs used will increase (decrease) total expected costs. Changes in prosthesis prices relative to each other may also change relative cost-effectiveness. Costs at specific hospitals can be modelled to allow for differences in purchaser prices.

The model does not calculate costs occurring after 20 years, although the present value of future costs is reduced by the discount rate.

The general conclusions under the assumptions of this model are summarised below.

- Compared with Charnley prosthesis survival data from centres of excellence and a prosthesis cost of £353 including cement, the model suggests that a no revisions prosthesis should cost not more than about £650 to have equivalent total expected costs over 20 years (*Table 25*).
- Given the hip survival data used in the model, the Stanmore prosthesis appears to be more cost-effective over 20 years than the Charnley

prosthesis; the Exeter Polished and Müller straight stem are of similar cost-effectiveness (*Tables 27* and *29*).

- Prosthesis cost, hospital costs and revision rate are the components of the model that have the biggest impact in terms of changing total expected costs (*Tables 31* and *32*).
- Very high and very low estimates of hospital costs change the total expected costs of individual prostheses but have little effect on their relative cost-effectiveness (*Table 33*).
- In 70-year-old (men), a low price prosthesis is generally more cost-effective than a high price prosthesis, even with a very low revision rate (*Table 34*).
- In 40-year-old (men), prostheses with high prices and low revision rates can be more cost-effective than low priced prostheses with higher failure rates (*Table 35*).

Despite data of variable quality, and limited data on important characteristics such as long-term survival of prostheses, the approach based on total expected costs enables robust conclusions to be drawn on choice of prostheses. This approach also enables new prostheses to be assessed against those for which good data are available. The model allows new assessments to be made relatively easily as new data become available.

PAPER 3

Chapter 12
Conclusion

Health technology assessment deploys the perspectives and techniques of different scientific disciplines in order to produce research-based information of relevance to policy-making and practice in the health service. This review has employed health economics and systematic critical review of clinical research to examine the issues of costs and outcomes of hip prosthesis implantation. Clinical decisions of surgeons as to the choice of prosthesis for given patients take place in the context of a range of other factors beyond the evidence regarding optimal outcomes. These include, for example, the presumed ease of a revision operation offered by different technologies, and a surgeons' familiarity with certain models and the associated surgical techniques (with regard to the latter, for example, the newer cementation techniques are reported to have diffused relatively slowly into THR practice in the UK[127]).

The culture of the manufacture and clinical application of hip prostheses is characterised by a high level of innovation and experimentation. New models are proliferating and many of the new models are also among the most expensive. Clinical outcomes during early postoperative follow-up are unreliable as guides to future performance. This means that comparative performance is difficult to evaluate. The cost of the prosthesis is a significant component of the total cost of the THR procedure.

A full policy analysis of the options for containing costs and maintaining or improving quality of hip prosthesis performance is beyond the scope of this report. It can be noted that a surveillance scheme for hip implants, based on the concept of a 'recommended list', was agreed in 1981 but not implemented. There are a number of possible avenues, not necessarily mutually exclusive, which might be considered in a policy appraisal. These include:

- the concept of the lifetime care package for healthcare commissioners, which includes a quality incentive for the prosthesis supplier and an insurance element for the clinical service provider and healthcare commissioner,[116] discussed in the context of the economic model presented in this report
- nationally co-ordinated audit and the use of registers such as those employed in the Scandinavian countries
- the role of the national Medical Devices Agency, providers and purchasers specifically in requiring reports of adverse incidents and generally in the use of mechanisms for monitoring outcomes in terms of hip scores and prosthesis longevity
- implant standardisation programmes,[128] which can show cost savings and quality maintenance using patient scoring systems to match prosthesis type to expected demand placed upon the prosthesis
- use of competitive bidding practices by hospitals or consortia
- restriction of substantially new technologies to high-quality multicentre controlled trials
- audit of standards of surgical training and experience, especially with newer types of prosthesis which require familiarity with new instrumentation and techniques.

Improving effectiveness and cost-effectiveness of total hip prostheses requires major commitment from the many disciplines involved in health technology assessment in this area, and the health service and policy users of these assessments.

PAPER 3

Chapter 13

Summary and recommendations

In this summary the findings of our review of evidence and the main results of our economic model in estimating the cost-effectiveness of alternative hip prostheses are presented. It includes recommendations for future research and suggestions regarding THR policy in the NHS. Policy options relating to monitoring, innovation and the diffusion of hip prosthesis technology were suggested in the previous chapter, although a full and detailed consideration of these awaits further dedicated investigation.

Clinical research on hip prosthesis technology

- Clinical outcomes of THR are measured by prosthesis survival, radiographic measurement and hip scoring. Clinical hip scoring performed by clinicians is likely to underestimate the qualitative significance of pain and functional impairment for patients receiving hip implants.[28] Incorporation of patient perspectives is inadequate. Different studies define clinical outcome measures in somewhat different ways, making comparison of studies difficult.
- The great majority of studies appear in a small number of specialist orthopaedic journals and emanate from specialist orthopaedic centres and departments, mainly in teaching hospitals. This has a bearing on the generalisability of the results of individual studies. The cemented Charnley prosthesis is the most studied single model.
- About 12% of the reviewed studies originate in the UK. Length of follow-up is inadequate for the full evaluation of the longevity of more recently introduced types of prosthesis. Evidence for association between early radiographic signs of loosening and migration and long-term prosthesis survival is equivocal, although there is some evidence that early radiographically defined failure predicts later requirement for revision of the prosthesis.
- The majority of studies of the outcomes of hip prostheses in primary THR are observational in design. Few RCTs have been published. This review maximises the use of studies with an element of comparison between prosthesis types.
- Critical appraisal of relevant studies shows that, with some exceptions, the methodological quality of studies is generally poor. Especially notable is the lack of reporting of a sample size calculation in any of the reviewed studies. Sample sizes actually reported in most studies are notably smaller than would be recommended to achieve valid generalisable results.
- Given the generally poor methodological quality of the reviewed studies, results for different types of prostheses should be treated as estimates with wide confidence intervals.

Comparative evidence for hip prosthesis technologies

Taking the above points into account, the following conclusions can be drawn about the performance of different types of prosthetic hip technology on the basis of this review. (Definitions of the different types of prosthesis can be found in chapter 3.)

- **Cemented** designs in general show good survival results at 10–15 years and beyond. Models with good, published, comparable results (at 10 years or more) include the Stanmore, Howse, Lubinus, Exeter and Charnley. The rate of acetabular revision in cemented implants remains problematical. There is some evidence that all-polyethylene acetabular components are preferable to metal-backed designs in terms of longevity of the implant. Newer (second-generation) cementation techniques in general provide better results than traditional techniques.
- Evidence of short- to medium-term comparisons of prosthesis longevity between **non-cemented porous-coated** designs and **cemented** designs is equivocal. The first 10-year survival results for porous-coated models appear to bear comparison with the cemented models after the same follow-up period, especially taking into account the relatively lower average age of the patient groups implanted with the porous-coated models. One comparative radiographic study suggests that cemented acetabular components performed better than porous-coated designs but that porous-coated stems performed better than cemented models.
- The comparative evidence suggests strongly that thigh pain is a problem associated with **cementless porous-coated** implants, to which **cemented** designs are not prone. In the observational studies of porous-coated implants reviewed here,

reports of its prevalence range from about 2% to about 25% at 2–7 years follow-up, with several studies reporting prevalences around the higher 25% level, even in non-loose stems.
- In contrast to porous-coated models, the small number of studies of **cementless HA-coated** models report mild to moderate thigh pain ranging from 0% to about 5% of patients at 2–5 years follow-up. This is a relatively good result in comparison to reports of porous-coated implants and requires further investigation.
- Radiographic studies of **cemented** versus **HA-coated** designs suggest that HA-coating has better early fixation and less migration than cemented models. The lesser migration of HA-coated models may be associated with less early postoperative pain, according to one comparative study. However, with maximum follow-up periods of only 3–4 years for this form of fixation, longer-term study of survival and clinical results is required.
- **Hybrid** prostheses appear to do well in the short term but the available studies cannot give any indications for their mid- or long-term performance. Given wide confidence intervals, this type of design can be regarded as comparable with the best cemented designs for early (6–7 years) survival. Their early survival is superior to uncemented porous-coated implants, and early thigh pain in cemented stem components in hybrid implants is minimal or absent compared with porous-coated designs. Longer follow-up, especially of the coated acetabular component of hybrid implants, is required to ascertain the medium and long-term performance of this design.
- Little evidence is available about **fully modular** prostheses. Theoretically, modularity permits greater intra-operative flexibility for the surgeon and potentially better component fit but further evidence, especially comparison with cemented implants, is required. One comparative study suggests that a fully modular stem has performed less well than cemented stems. Laboratory analysis of retrieved components suggests that mixed-alloy components are more prone to corrosion than single alloy devices.[29]
- The implications of laboratory studies of alternative bearing surface materials require further assessment. The small amount of evidence for the performance of hips with **ceramic** bearing surfaces is equivocal. Wear rates are less than for other materials at the articulating surface of the joint. Comparative studies have suggested either lower or equivalent revision rates for ceramic versus cemented implants at medium-term follow-up. Ceramic heads are common, but major manufacturers are currently developing **metal–metal** versions of common designs, for which published evidence is lacking.
- The **uncoated press-fit** and **resurfacing** types of hip prosthesis generally have survival results notably inferior to the other types of design available.

Economic modelling

- The economic model developed in this study and presented in the report can be used to model the cost-effectiveness of different hip prostheses under any different resource and clinical outcome assumptions which healthcare practitioners and decision-makers might foresee.

The general conclusions under our assumptions used in this model are summarised below.

- Prosthesis cost, hospital costs and the revision rate are the components of the model that have the biggest impact in terms of changing total expected costs for THR procedures.
- Very high and very low estimates of hospital costs change individual prostheses' total expected costs but have little effect on their cost-effectiveness relative to each other.
- Compared with survival data for the Charnley cemented prosthesis from centres of excellence, and assuming a prosthesis cost of £353 including cement, the model suggests that even a no revisions prosthesis should cost no more than about £650 currently (1997 prices) to have equivalent total expected costs over 20 years. Only cemented prostheses are currently available at this price.
- In 70-year-olds (men), a low price prosthesis is generally more cost-effective than a high price prosthesis, even with a very low revision rate.
- In 40-year-olds (men), prostheses with high prices and low revision rates can be more cost-effective than low-priced prostheses with higher failure rates.

Policy/service implications

The authors suggest that:

- mechanisms for improving support for the use of appropriate prostheses could be examined in a wide-ranging policy analysis, to include combinations of local contracting, coordinated audit or monitoring, central UK registers, and regulation of new technologies via coordinated trials
- healthcare commissioners could consider modelling costs of alternative prosthesis designs and

models using their own local input resource assumptions and outcome data, along the lines of the model demonstrated in this report.

Given the variation in effectiveness of prosthesis types, the authors suggest that commissioners and providers could consider the following monitoring issues when developing policy.

- The range and extent of use of prostheses known to have results poorer than the best cemented designs, such as the Stanmore, Howse, Lubinus, Exeter and Charnley prostheses.
- In the case of substantively new designs, appropriate monitoring and evaluation (including cost dimensions) prior to diffusion into routine practice.
- The extent of implantation of different design types (such as cemented, hybrid, porous) in relation to age-groups of patients, seeking audit of clinical and patient outcomes.
- Routine rates for different types of prosthesis – including revision (and re-revision) rates – as proportions of total THR rates for the provider/NHS Trust, taking into account status as general or specialist tertiary referral centres.

Recommendations for further research

General

- Improvements should be sought in the design and reporting of the generality of research studies in this area. Notable aspects are sample sizes, reporting of data on characteristics of the study group or groups, use of blinding or independent evaluation as appropriate, and reporting of patient selection criteria and procedures.
- Further inclusion of patient-derived quality-of-life measures in studies of hip prosthesis performance is essential. Clinical hip scoring systems do not take account of the patient's point of view in assessing outcomes.
- Further review from existing studies of short-term hip score outcomes could yield valuable information about pain and everyday activity during the early 'settling down' postoperative period, which appears to vary between different types of prosthesis.
- The existing clinical research on THR assumes that given tolerable pain and physical function, longevity of the implant is the primary goal. This may be the case from the patients' perspective also but this has not been demonstrated. If ease of revision were an important criterion from the patient's perspective as well as from the surgeon's then the choice of implant would be affected. Patients' values and choices regarding quality of life in relation to the perceived risk of undergoing a revision operation should be investigated. This applies especially to younger and/or more active patients for whom revision is more likely.
- There is scope for review of further good quality studies, not included in this review, on the mechanics of loosening in different types of prostheses.
- New primary studies of the mechanics of loosening could employ radiographic techniques and/or autopsy limb retrieval approaches.

Prosthesis types

- Reporting of longer follow-up studies especially of the hybrid and cementless HA-coated models is required in order to assess further the early promising outcomes of these technologies. Longer follow-up of the coated acetabular component of hybrid implants is required to ascertain the medium and long-term performance of this design.
- Results for thigh pain in HA-coated models appear relatively good in comparison to reports of porous-coated implants, and this requires further examination for longer follow-up periods. The extent and significance to patients of thigh pain associated with porous- and HA-coated implants should be assessed.
- The implications of laboratory studies of alternative bearing surface materials require further review and investigation.
- Porous-coated cementless designs should be monitored further where already implanted to assess longevity.
- Fully modular designs may offer advantages in surgical procedure but the lengths of follow-up are currently insufficient to establish patient benefits or problems of this type of design.
- In the past there has appeared to be reluctance, inertia or lack of resources in many orthopaedic departments in the UK to adopt new cementation techniques for cemented prostheses.[127] More recent information is required on the use of these methods so that, given the better outcomes generally associated with them, their use can be encouraged.
- Further exploration is required of associations between radiographic signs of loosening/migration and later mechanical failure of different designs. Insufficient data exist on the predictive power of radiographic measurements.

PAPER 3

Acknowledgements

This study was supported by the National Health Service R&D Health Technology Assessment Programme of England and Wales.

The authors would like to thank Professor Ian Learmonth and the staff of the Avon Orthopaedic Centre, Bristol, for providing hospital data, and Professor Paul Gregg, Department of Trauma and Orthopaedic Surgery, University of Newcastle-upon-Tyne, and David Fender, Orthopaedic Surgery, Leicester General Hospital, for providing hospital data and information from the Trent Regional Arthroplasty database. Our thanks are also due: to Delma Cowley, formerly of the Australian Institute of Health and Welfare, for providing the bibliography of her critical review; to Judy Feinberg of Indiana State University, USA, for supplying pre-publication manuscripts; and to the orthopaedic surgeons who participated in interviews for the project.

We are also indebted to the staff of the Department of Social Medicine, University of Bristol, who provided support and advice, including, in particular, Dr Nicky Pearson, Sara Brookes, Dr Tim Peters and Professor George Davey Smith. Our thanks are also due to Jeannette Smith for her expert technical assistance with the production of this document.

Finally, we also owe our thanks to the referees for their perseverance in reading the report and the quality of their comments.

PAPER 3

References

1. Huiskes R. Failed innovation in total hip replacement: diagnosis and proposals for a cure. *Acta Orthop Scand* 1993;**64**:699–716.
2. Older J. Low friction arthroplasty of the hip. A ten to twelve year follow-up study. *Clin Orthop* 1986;(211):36.
3. Huo MH, Salvati EA, Buly RL. Wear debris in cemented total hip arthroplasty. *Orthopedics* 1991;**14**:335–40.
4. Mittelmeier H, Heisel J. Sixteen-years' experience with ceramic hip prostheses. *Clin Orthop* 1992;(282):64–72.
5. Head WC, Bauk DJ, Emerson RH. Titanium as the material of choice for cementless femoral components in total hip arthroplasty. *Clin Orthop* 1995;(311):85–90.
6. Gardiner RC, Hozack WJ. Failure of the cement–bone interface: a consequence of strengthening the cement-prosthesis interface? *J Bone Joint Surg Br* 1994;**76**:49–52.
7. NHS Centre for Reviews and Dissemination. Total hip replacement. *Effective Health Care* 1996;**2**(7):1-12.
8. Dorey F, Grigoris P, Amstutz H. Making do without randomised trials. *J Bone Joint Surg Br* 1994;**76**:1–3.
9. Murray DG, Crown RS, Dickersin K, Woods Duncan P, Epps CH, Ettinger WH, *et al.* NIH Consensus Conference: total hip replacement. *JAMA* 1995;**273**:1950–6.
10. Naylor C, Williams JI. Primary hip and knee replacement surgery: Ontario criteria for case selection and surgical priority. *Qual Health Care* 1996;**5**:20–30.
11. Hadorn DC, Holmes AC. The New Zealand priority criteria project. *BMJ* 1997;**314**:131–4.
12. Freeman MAR, Plante-Bordeneuve P. Early migration and late aseptic failure of proximal femoral prostheses. *J Bone Joint Surg Br* 1994;**76**:432–8.
13. Karrholm J, Borssen B, Lowenhielm G, Snorrason F. Does early micromotion of femoral stem prostheses matter? 4–7 year stereoradiographic follow-up of 84 cemented prostheses. *J Bone Joint Surg Br* 1994;**76**:912–17.
14. Ahnfelt L, Herberts P, Malchau H, *et al.* Prognosis of total hip replacement. A Swedish multicenter study of 4664 revisions. *Acta Orthop Scand Suppl* 1990;**238**:1–26.
15. Heaton J, Long A, Dixon P, Brettle A. Measuring the health outcomes of total hip replacement through the commissioning process. UK Clearing House on Health Outcomes, Nuffield Institute for Health, University of Leeds: Outcomes Measurement Reviews no 1; May 1995.
16. Cowley D. Technology assessment reports: reports from the Australian Institute of Health and Welfare (AIHW): prostheses for total hip replacement. *Int J Technol Assess Health Care* 1995;**11**:117–23.
17. Yahiro MA, Gantenberg JB, Nelson R, Lu HTC, Mishra NK. Comparison of the results of cemented, porous-ingrowth, and threaded acetabular cup fixation: a meta-analysis of the orthopaedic literature. *J Arthroplasty* 1995;**10**:339–50.
18. Havelin LI, Espehaug B, Vollset SE, Engesaeter LB. Early failures among 14,009 cemented and 1,326 uncemented prostheses for primary coxarthrosis: the Norwegian arthroplasty register, 1987–1992. *Acta Orthop Scand* 1994;**65**:1–6.
19. Ebramzadeh E, Sarmiento A, McKellop HA, Llinas A, Gogan W. The cement mantle in total hip arthroplasty. Analysis of long term radiographic results. *J Bone Joint Surg Am* 1994;**76**:77–87.
20. Schreiber A, Jacob HAC, Hauser R, Kern S. The self-locking Balgrist hip socket for cementless fixation: biomechanical principles and clinical results. *Keio J Med* 1993;**42**(2):53–9.
21. Daellenbach HG, Gillespie WJ, Crosbie P, Daellenbach US. Economic appraisal of new technology in the absence of survival data – the case of total hip replacement. *Soc Sci Med* 1990;**31**:1287–93.
22. Dall DM, Learmonth ID, Solomon MI. Charnley low friction arthroplasty: the yardstick (a 3–7 year follow-up). *J Bone Joint Surg Br* 1988;**70**:499.
23. Mohler CG, Callaghan EJJ, Collis DK, Oregon E, Johnston RC. Early loosening of the femoral component at the cement–prosthesis interface after total hip replacement. *J Bone Joint Surg Am* 1995;**77**:1315–22.
24. Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. *BMJ* 1998;**316**:140–4.
25. Godsiff SP, Emery RJH, Heywood-Waddington MB, Thomas TL. Cemented versus uncemented femoral components in the Ring hip prosthesis. *J Bone Joint Surg Br* 1992;**74**:822–4.

26. Kelley SS, Johnstone RC. Debris from cobalt–chrome cable may cause acetabular loosening. *Clin Orthop* 1992;(285):140–6.

27. Lehman DE, Capello WN, Feinberg JR. Total hip arthroplasty without cement in obese patients: a minimum two-year clinical and radiographic follow-up study. *J Bone Joint Surg Am* 1994;**76**:854–62.

28. Liebermann JR, Dorey F, Shekelle P, Schumacher L, Thomas BJ, Kilgus DJ, *et al.* Differences between patients' and physicians' evaluations of outcome after total hip arthroplasty. *J Bone Joint Surg Am* 1996;**78**:835–8.

29. Collier JP, Mayor MB, Williams IR, Surprenant VA, Surprenant HP, Currier BH. The tradeoffs associated with modular hip prostheses. *Clin Orthop* 1995;(311):91–101.

30. Hodgkinson JP, Maskell AP, Ashok P, Wroblewski BM. Flanged acetabular components in cemented Charnley hip arthroplasty. Ten-year follow-up of 350 patients. *J Bone Joint Surg Br* 1993;**75**:464–7.

31. Ritter MA. The cemented acetabular component of a total hip replacement: all polyethylene versus metal backing. *Clin Orthop* 1995;(311):69–75.

32. Britton AR, Murray DW, Bulstrode CJ, McPherson K, Denham RA. Long-term comparison of Charnley and Stanmore design total hip replacements. *J Bone Joint Surg Br* 1996;**78**:802–8.

33. Hoffman CW, Lynskey TG, Hadlow V. Twenty years of total hip arthroplasty in Taranaki: a survival analysis. *Aust N Z J Surg* 1994;**64**:538–41.

34. Bradley JG, Lee KM. Cement versus calcium hydroxyapatite ceramic – early results of a randomised prospective clinical study. In: Wallace WA, editor. Joint replacement in the 1990s. Clinical studies, financial implications and marketing approaches. Bury St Edmonds, UK: Mechanical Engineering Publications, 1992:55–60.

35. Onsten I, Carlsson AS, Ohlin A, Nilsson JA. Migration of acetabular components, inserted with and without cement, in one-stage bilateral hip arthroplasty. *J Bone Joint Surg Am* 1994;**76**:185–94.

36. Callaghan JJ, Kim YS, Brown TD, Pedersen DR, Johnston RC. Concerns and improvements with cementless metal-backed acetabular components. *Clin Orthop* 1995;(311):76–84.

37. Wykman A, Olsson E, Axdorph G, Goldie I. Total hip arthroplasty: a comparison between cemented and press-fit noncemented fixation. *J Arthroplasty* 1991;**6**:19–29.

38. Olsson E, Goldie I, Wykman A. Total hip replacement. A comparison between cemented (Charnley) and non-cemented (HP Garches) fixation by clinical assessment and objective gait analysis. *Scand J Rehabil Med* 1986;**18**:107–16.

39. Espehaug B, Havelin LI, Engesaeter LB, Vollset SE, Langeland N. Early revision among 12,179 hip prostheses. A comparison of 10 different brands reported to the Norwegian arthroplasty register, 1987–1993. *Acta Orthop Scand* 1995;**66**:487–93.

40. Zichner LP, Willert HG. Comparison of alumina–polyethylene and metal–polyethylene in clinical trials. *Clin Orthop* 1992;(282):86–94.

41. Schuller HM, Marti RK. Ten-year socket wear in 66 hip arthroplasties: ceramic versus metal heads. *Acta Orthop Scand* 1990;**61**:240–3.

42. Karrholm J, Malchau UH, Snorrason F, Herberts P. Micromotion of femoral stems in total hip arthroplasty. *J Bone Joint Surg Am* 1994;**76**:1692–705.

43. Wixson RL, Stulberg SD, Mehlhoff M. Total hip replacement with cemented, uncemented and hybrid prostheses: a comparison of clinical and radiographic results at two to four years. *J Bone Joint Surg Am* 1991;**73**:257–70.

44. Chmell MJ, Rispler D, Poss R. The impact of modularity in total hip arthroplasty. *Clin Orthop* 1995;(319):77–84.

45. Bourne RB, Rorabeck CH, Laupacis A, Feeny D, Tugwell PSL, Wong C, *et al.* Total hip replacement: the case for noncemented femoral fixation because of age. *Can J Surg* 1995;**38** suppl 1:61–6.

46. Rorabeck CH, Bourne RB, Mulliken BD, Nayak N, Laupacis A, Tugwell PSL, *et al.* The Nicolas Andry Award. Comparative results of cemented and cementless total hip arthroplasty. *Clin Orthop* 1996;(325):330–44.

47. Krismer M, Fischer M, Klestil T, Frischhut B. Uncoated polyethylene RM acetabular component versus Müller cemented acetabular component – a 4–8 year follow-up study. *Arch Orthop Trauma Surg* 1991;**110**:195–9.

48. Hearn SL, Bicalho PS, Eng K, Booth REJ, Hozack WJ, Rothman RH. Comparison of cemented and cementless total hip arthroplasty in patients with bilateral hip arthroplasties. *J Arthroplasty* 1995;**10**:603–8.

49. Reigstad A, Brandt M, Hetland KR. Total hip replacement with Muller prosthesis and ICLH double cup: 2–6 year results of a prospective comparative study. *Arch Orthop Trauma Surg* 1986;**105**:175–82.

50. Ritter MA, Gioe TJ. Conventional versus resurfacing total hip arthroplasty. A long-term prospective study of concomitant bilateral implantation of prostheses. *J Bone Joint Surg Am* 1986;**68**:216–25.

51. Riska EB. Ceramic endoprosthesis in total hip arthroplasty. *Clin Orthop* 1993;(297):87–94.

52. Huracek J, Spirig P. The effect of hydroxyapatite coating on the fixation of hip prostheses. A comparison of clinical and radiographic results of hip replacement in a matched pair study. *Arch Orthop Trauma Surg* 1994;**113**:72–7.

PAPER 3

53. Moilanen T, Stocks GW, Freeman MAR, Scott G, Goodier WD, Evans SJW. Hydroxyapaptite coating of an acetabular prosthesis. *J Bone Joint Surg Br* 1996;**78**:200–5.

54. d'Antonio JA, Capello WN, Jaffe WL. Hydroxyapatite-coated hip implants: multicenter three-year clinical and roetgenographic results. *Clin Orthop* 1992;(285):102–15.

55. McPherson EJ, Dorr LD, Gruen TA, Saberi MT. Hydroxyapatite-coated proximal ingrowth femoral stems: a matched pair control study. *Clin Orthop* 1995;(315):223–30.

56. Goetz DD, Smith EJ, Harris WH. The prevalence of femoral osteolysis associated with components inserted with or without cement in total hip replacements: a retrospective matched-pair series. *J Bone Joint Surg Am* 1994;**76**:1121–9.

57. Maloney WJ, Harris WH. Comparison of a hybrid with an uncemented total hip replacement. A retrospective matched-pair study. *J Bone Joint Surg Am* 1990;**72**:1349–52.

58. Nashed RS, Becker DA, Gustilo RB. Are cementless acetabular components the cause of excess wear and osteolysis in total hip arthroplasty? *Clin Orthop* 1995;(317):19–28.

59. Pupparo F, Engh CA. Comparison of porous-threaded and smooth threaded acetabular components of identical design: 2–4 year results. *Clin Orthop* 1992;(271):201–6.

60. Hwang SK, Park JS. Cementless total hip arthroplasty with AML, PCA and HGP prostheses. *Int Orthop* 1995;**19**:77–83.

61. Duck HJ, Mylod AG Jr. Heterotopic bone in hip arthroplasties: cemented versus noncemented. *Clin Orthop* 1992;(282):145–53.

62. Dall DM, Learmonth ID, Solomon M, Davenport DM. Charnley hips followed for 3–17 years. *Acta Orthop Scand* 1993;**64**:252–6.

63. Eftekhar NS, Tzitzikalakis G. Failures and re-operations following low friction arthroplasty of the hip. *Clin Orthop* 1986;(211):65.

64. Garcia-Cimbrelo E, Munera L. Early and late loosening of the acetabular cup after low-friction arthroplasty. *J Bone Joint Surg Am* 1992;**74**:1119–29.

65. Hamilton HW, Joyce M. Long term results of low friction arthroplasty performed in a community hospital including a radiologic review. *Clin Orthop* 1986;(211):55.

66. Hamilton HW, Gorczyca J. Low friction arthroplasty at 10 to 20 years: consequences of plastic wear. *Clin Orthop* 1995;(318):160–6.

67. Joshi AB, Porter ML, Trail IA, Hunt LP, Murphy JCM, Hardinge K. Long-term results of Charnley low friction arthroplasty in young patients. *J Bone Joint Surg Br* 1993;**75**:616–23.

68. Kobayashi S, Eftekhar NS, Terayama K. Predisposing factors in fixation failure of femoral prostheses following primary Charnley low friction arthroplasty: a 10- to 20-year follow-up study. *Clin Orthop* 1994;(306):73–83.

69. Kobayashi S, Eftekhar NS, Terayama K, Iorio R. Risk factors affecting radiological failure of the socket in primary Charnley low friction arthroplasty: a 10- to 20-year follow-up study. *Clin Orthop* 1994;(306):84–96.

70. Madey SM, Callaghan JJ, Olejniczak JP, Goetz DD, Johnston RC. Charnley total hip arthroplasty with use of improved techniques of cementing. *J Bone Joint Surg Am* 1997;**79**:53–64.

71. Neumann L, Freund KG, Sorenson KH. Long-term results of Charnley total hip replacement: review of 92 patients at 15 to 20 years. *J Bone Joint Surg Br* 1994;**76**:245–51.

72. Older J, Butorac R. Charnley low friction arthroplasty (LFA): a 17–21 year follow-up study. *J Bone Joint Surg Br* 1992;**74** suppl 3:251.

73. Skeie S, Lende S, Sjoberg EJ, Vollset SE. Survival of the Charnley hip in coxarthrosis. A 10–15 year follow-up of 629 cases. *Acta Orthop Scand* 1991;**62**:98–101.

74. August AC, Aldam CH, Pynsent PB. The McKee-Farrar hip arthroplasty: a long-term study. *J Bone Joint Surg Br* 1986;**68**:520–7.

75. Bryant MJ, Mollan RAB, Nixon JR. Survivorship analysis of the ring hip arthroplasty. *J Arthroplasty* 1991;**6** suppl:S5–10.

76. Fowler JL, Gie GA, Lee AJC, Ling RSM. Experience with the Exeter total hip replacement since 1970. *Orthop Clin N Am* 1988;**19**:477–89.

77. Ohlin A, Onsten I. Loosening of the Lubinus hip: 202 cases followed for 3–6 years. *Acta Orthop Scand* 1990;**61**:244–7.

78. Partio E, Von Bonsdorff H, Wirta J, Avikainen V. Survival of the Lubinus hip prosthesis: an 8- to 12-year follow-up evaluation of 444 cases. *Clin Orthop* 1994;(303):140–6.

79. Roberts JA, Finlayson DF, Freeman PA. The long term results of the Howse total hip arthroplasty – with particular reference to those requiring revision. *J Bone Joint Surg Br* 1987;**69**:545.

80. Engh CA, Massin P. Cementless total hip arthroplasty using the AML stem. Results using a survivorship analysis. *Clin Orthop* 1989;(249):141–58.

81. Engh CA, Glassman AH, Suthers KE. The case for porous-coated hip implants: the femoral side. *Clin Orthop* 1990;(261):63–81.

82. Engh CAJ, Culpepper II WJ, Engh CA. Long-term results of the Anatomic Medullary Locking prosthesis in total hip arthroplasty. *J Bone Joint Surg Am* 1997;**79**:169–76.

PAPER 3

83. Owen TD, Moran CG, Smith SR, Pinder IM. Results of uncemented porous-coated anatomic total hip replacement. *J Bone Joint Surg Br* 1994;**76**:258–62.

84. Sotereanos NG, Engh CA, Glassman AH, Macalino GE, Engh CA. Cementless femoral components should be made from cobalt chrome. *Clin Orthop* 1995;(313):146–53.

85. Holman PR, Tyer HDD. Porous coated anatomic non-cemented total hip arthroplasty. *Aust N Z J Surg* 1992;**62**:56–9.

86. Turner TM, Sumner DR, Urban RM, Rivero DP, Galante JO. A comparative study of porous coatings in a weight-bearing total hip arthroplasty model. *J Bone Joint Surg Am* 1986;**68**:1396–409.

87. Tonino AJ, Romanini L, Rossi P, Borroni M, Greco F, Garcia-Araujo C, *et al.* Hydroxyapatite-coated hip prostheses: early results from an international study. *Clin Orthop* 1995;(312):211–25.

88. Koch FW, Messler HH, Wagner U, Meyer HJ. Short term results (2–5 years) with Furlong type hydroxyapatite-coated total hip replacements. *Z Orthop Ihre Grenzgeb* 1993;**131**:562–7.

89. Capello WN. Hydroxyapatite in total hip arthroplasty: five years clinical experience. *Orthopedics* 1994;**17**:781–92.

90. Geesink RGT, Hoefnagels NHM. Six-year results of hydroxyapatite-coated total hip replacement. *J Bone Joint Surg Br* 1995;**77**:534–47.

91. Murray DW, Carr AJ, Bulstrode CJ. Which primary total hip replacement? *J Bone Joint Surg Br* 1995;**77**:520–7.

92. Wilson-MacDonald J, Morscher E, Masar Z. Cementless uncoated polyethylene acetabular components in total hip replacement. Review of five to ten-year results. *J Bone Joint Surg Br* 1990;**72**:423.

93. Roffman M, Juhn A. Nine year follow-up of the Mathys uncemented acetabular cup: a radiological evaluation. *Bull Hosp Jt Dis Orthop Inst* 1993;**53**:51–3.

94. Niinimaki T, Puranen J, Jalovaara P. Total hip arthroplasty using isoelastic femoral stems: a seven to nine-year follow-up in 108 *patients. J Bone Joint Surg Br* 1994;**76**:413–18.

95. Kennedy WR. Acetabular component fixation in the 1990s: cementless components: the good, the bad, and the ugly. *Orthopedics* 1994;**17**:783–4.

96. Helfen M, Malzer U, Peters P, Griss P, Himmelmann G, Weber E. Clinical results with a hybrid total hip arthroplasty: a 3–6 year follow-up [abstract]. *Z Orthop Ihre Grenzgeb* 1993;**131**:578–84.

97. Schmalzried TP, Harris WH. Hybrid total hip replacement. A 6.5 year follow-up study. *J Bone Joint Surg Br* 1993;**75**:608–15.

98. Mohler CG, Kull LR, Martell JM, Rosenberg AG, Galante JO. Total hip replacement with insertion of an acetabular component without cement and a femoral component with cement. *J Bone Joint Surg Am* 1995;**77**:86–96.

99. Burkart BC, Bourne RB, Rorabeck CH, Kirk PG. Thigh pain in cementless total hip arthroplasty: a comparison of two systems at 2 years follow up. *Orthop Clin N Am* 1993;**24**:645–53.

100. Moskal JT, Shaffrey CI, Ripley LP. Prospective analysis of uncemented and hybrid primary porous coated anatomic total hip arthroplasties in a community setting. *Clin Orthop* 1994;(304):139–47.

101. Mallory TH, Kraus TJ, Vaughn BK. Intraoperative femoral fractures associated with cementless total hip arthroplasty. *Orthopedics* 1989;**12**:231–9.

102. Hedlundh U, Fredin H. Patient characteristics in dislocations after primary total hip arthroplasty: 60 patients compared with a control group. *Acta Orthop Scand* 1995;**66**:225–8.

103. Pierchon F, Pasquier G, Cotten A, Fontaine C, Clarisse J, Duquennoy A. Causes of dislocation of total hip arthroplasty: CT study of component alignment. *J Bone Joint Surg Br* 1994;**76**:45–8.

104. Bankston AB, Faris PM, Keating EM, Ritter MA. Polyethylene wear in total hip arthroplasty in patient matched groups. A comparison of stainless steel, cobalt chrome and titanium-bearing surfaces. *J Arthroplasty* 1993;**8**:315–22.

105. Bankston AB, Keating EM, Ranawat C, Faris PM, Ritter MA. Comparison of polyethylene wear in machined versus molded polyethylene. *Clin Orthop* 1995;(317):37–43.

106. Hernandez JR, Keating EM, Faris PM, *et al.* Polyethylene wear in uncemented acetabular components. *J Bone Joint Surg Br* 1994;**76**:263–6.

107. Marston RA, Cobb AG, Bentley G. Stanmore compared with Charnley total hip replacement. A prospective study of 413 arthroplasties. *J Bone Joint Surg Br* 1996;**78**:178–84.

108. Neumann L, Freund KG, Sorensen KH. Total hip arthroplasty with the Charnley prosthesis in patients fifty-five years old and less. Fifteen to twenty-one year results. *J Bone Joint Surg Am* 1996;**78**:73–9.

109. Thanner J, Freij-Larsson C, Karrholm J, Malchau H, Wesslen B. Evaluation of Boneloc: chemical and mechanical properties, and a randomized clinical study of 30 total hip arthroplasties. *Acta Orthop Scand* 1995;**66**:207–14.

110. Cornell CN, Ranawat CS. The impact of modern cement techniques on acetabular fixation in cemented total hip replacement. *J Arthroplasty* 1986;**1**:197–202.

111. Ranawat CS, Rawlins BA, Harju VT. Effect of modern cement technique on acetabular fixation total hip arthroplasty – a retrospective study in matched pairs. *Orthop Clin N Am* 1988;**19**:599–603.

112. Dall DM, Learmonth ID, Solomon MI, Miles AW, Davenport JM. Fracture and loosening of Charnley femoral stems: comparison between first-generation and subsequent designs. *J Bone Joint Surg Br* 1993;**75**:259–65.

113. Kristiansen B, Steen Jensen J. Biomechanical factors in loosening of the Stanmore hip. *Acta Orthop Scand* 1985;**56**:21–4.

114. Jacobsson SA, Djerf K, Gillquist J, Svedberg J. Tc-scintimetry in 56 cementless hip arthroplasties: a prospective, randomized comparison of two femoral components. *Acta Orthop Scand* 1994;**65**:418–23.

115. Gillespie WJ, Pekarsky B, O'Connell DL. Evaluation of new technologies for total hip replacement: economic modelling and clinical trials. *J Bone Joint Surg Br* 1995;**77**:528–33.

116. Pynsent PB, Carter SR, Bulstrode CJK. The total cost of hip-joint replacement; a model for purchasers. *J Public Health Med* 1996;**18**:157–68.

117. Francis CW, Marder VJ, Evarts CM. Lower risk of thromboembolic disease after total hip replacement with non-cemented than with cemented prostheses. *Lancet* 1986;**i**:769–71.

118. Wittman FW, Ring PA. Anaesthesia for hip replacement in ankylosing spondylitis. *J R Soc Med* 1986;**79**:457–9.

119. Sharrock NE, Go G, Harpel PC, Ranawat CS, Sculco TP, Salvati EA. The John Charnley award. Thrombogenesis during total hip arthroplasty. *Clin Orthop* 1995;(319):16–27.

120. Laupacis A, Rorabeck C, Bourne R, Tugwell P, Bullas R, Rankin R, *et al.* The frequency of venous thrombosis in cemented and non-cemented hip arthroplasty. *J Bone Joint Surg Br* 1996;**78**:210–12.

121. Lieberman JR, Geerts WH. Current concepts review: prevention of venous thromboembolism after total hip and knee arthroplasty. *J Bone Joint Surg Am* 1994;**76**:1239–50.

122. Malchau H, Herberts P, Ahnfelt L. Prognosis of total hip replacement in Sweden. Follow-up of 92,675 operations performed in 1978–1990. *Acta Orthop Scand* 1993;**64**:497–506.

123. Office of Population Censuses and Surveys. Mortality statistics. Series DH1, No. 27. London: HMSO, 1992.

124. Nizard RS, Sedel L, Christel P, *et al.* Ten year survivorship of cemented ceramic-ceramic total hip prosthesis. *Clin Orthop* 1992;(282):53–63.

125. Martell JM, Pierson RH, Jacobs JJ, Rosenberg AG, Maley M, Galante JO. Primary total hip reconstruction with a titanium fiber-coated prosthesis inserted without cement. *J Bone Joint Surg Am* 1993;**75**:554–71.

126. Xenos JS, Hopkinson WJ, Callaghan JJ, Heekin RD, Savory CG. Osteolysis around an uncemented cobalt chrome total hip arthroplasty. *Clin Orthop* 1995;(317):29–36.

127. Hashemi-Nejad A, Birch NC, Goddard NJ. Current attitudes to cementing techniques in British hip surgery. *Ann R Coll Surg Engl* 1994;**76**:396–400.

128. Healy WL, Kirven FM, Iorio R, Patch DA, Pfeifer BA. Implant standardization for total hip arthroplasty: an implant selection and a cost reduction program. *J Arthroplasty* 1995;**10**:177–83.

Studies appraised in the review

Studies appraised in this review are listed, in alphabetical order of author, by type of study design. While some of them have also been listed numerically in the main reference list, all of these studies are included in the Data Tables and Appraisal Tables in the appendix.

RCTs

Bourne RB, Rorabeck CH, Laupacis A, Feeney D, Wong C, Tugwell PSL, *et al.*, 1994. A randomized clinical trial comparing cemented to cementless total hip replacement in 250 osteoarthritic patients: the impact on health related quality of life and cost effectiveness. *Iowa Orthop J*;**14**:108–14.

Bourne RB, Rorabeck CH, Laupacis A, Feeney D, Tugwell PSL, Wong C, *et al.*, 1995. Total hip replacement: the case for noncemented femoral fixation because of age. *Can J Surg*;**38** suppl 1:S61–6.

Bradley JG, Lee KM, 1992. Cement versus calcium hydroxyapatite ceramic – early results of a randomised prospective clinical study. In: Wallace WA, editor. Joint replacement in the 1990s. Clinical studies, financial implications and marketing approaches. Bury St Edmonds, UK: Mechanical Engineering Publications:55–60.

Carlsson AS, Rydberg J, Onnerfalt R, 1995. A large collar increases neck resorption in total hip replacement: 204 hips evaluated during 5 years. *Acta Orthop Scand*;**66**:339–42.

Ciccotti MG, Rothman RH, Hozack WJ, Moriarty L, 1994. Clinical and roentgenographic evaluation of hydroxyapatite-augmented and nonaugmented porous total hip arthroplasty. *J Arthroplasty*;**9**:631–9.

Godsiff SP, Emery RJH, Heywood-Waddington MB, Thomas TL, 1992. Cemented versus uncemented femoral components in the ring hip prosthesis. *J Bone Joint Surg Br*;**74**:822–4.

Jacobsson SA, Djerf K, Gillquist J, Hammerby S, Ivarsson I, 1993. A prospective comparison of Butel and PCA hip arthroplasty. *J Bone Joint Surg Br*;**75**:624–9.

Jacobsson SA, Djerf K, Gillquist J, Svedberg J, 1994. Tc-scintimetry in 56 cementless hip arthroplasties: a prospective, randomized comparison of two femoral components. *Acta Orthop Scand*;**65**:418–23.

Karrholm J, Malchau UH, Snorrason F, Herberts P, 1994. Micromotion of femoral stems in total hip arthroplasty. *J Bone Joint Surg Am*;**76**:1692–705.

PAPER 3

Kelley SS, Fitzgerald RH, Rand JA, Ilstrup DM, 1993. A prospective randomized study of a collar versus a collarless femoral prosthesis. *Clin Orthop*;(294):114–22.

Krismer M, Fischer M, Mayrhofer P, Stockl F, Bittner C, Trojer C, 1994. A prospective study of the migration of two acetabular components: PCA versus RM cups. *Int Orthop*;**18**:23–8.

Marston RA, Cobb AG, Bentley G, 1996. Stanmore compared with Charnley total hip replacement. A prospective study of 413 arthroplasties. *J Bone Joint Surg Br*;**78**:178–84.

Mulliken BD, Nayak N, Bourne RB, Rorabeck CH, Bullas R, 1996. Early radiographic results comparing cemented and cementless total hip arthroplasty. *J Arthroplasty*; **11**:24–33.

Olsson E, Goldie I, Wykman A, 1986. Total hip replacement. A comparison between cemented (Charnley) and non-cemented (HP Garches) fixation by clinical assessment and objective gait analysis. *Scand J Rehabil Med*;**18**(3):107–16.

Onsten I, Carlsson AS, 1994. Cemented versus uncemented socket in hip arthroplasty. *Acta Orthop Scand*;**65**:517–21.

Onsten I, Carlsson AS, Ohlin A, Nilsson JA, 1994. Migration of acetabular components, inserted with and without cement, in one-stage bilateral hip arthroplasty. *J Bone Joint Surg Am*;**76**:185–94.

Reigstad A, Brandt M, Hetland KR, 1986. Total hip replacement with Muller prosthesis and ICLH double cup. 2–6 year results of a prospective comparative study. *Arch Orthop Trauma Surg*;**105**:175–82.

Rorabeck CH, Bourne RB, Laupacis A, *et al.*, 1994. A double blind study of 250 cases comparing cemented with cementless total hip arthroplasty: cost effectiveness and its impact on health-related quality of life. *Clin Orthop*;**298**:156–64.

Rorabeck CH, Bourne RB, Mulliken BD, Nayak N, Laupacis A, Tugwell PSL, *et al.*, 1996. The Nicolas Andry Award. Comparative results of cemented and cementless total hip arthroplasty. *Clin Orthop*;**325**:330–44.

Søballe K, Toksvig-Larsen S, Gelineck J, *et al.*, 1993. Migration of hydroxyapatite coated femoral prostheses. *J Bone Joint Surg Br*;**75**:681–7.

Thanner J, Freij-Larsson C, Karrholm J, Malchau H, Wesslen B, 1995. Evaluation of Boneloc: chemical and mechanical properties, and a randomized clinical study of 30 total hip arthroplasties. *Acta Orthop Scand*;**66**:207–14.

Wilson LF, Nolan JF, Heywood-Waddington MB, 1992. Fracture of the femoral stem of the Ring TCH hip prosthesis. *J Bone Joint Surg Br*;**74**:725.

Wykman A, Goldie I, 1984. A prospective randomized study comparing results of cemented and non-cemented endoprosthetic fixation in the surgical treatment of osteoarthritis of the hip. *Eng Med*;**13**:181–4.

Wykman A, Goldie I, 1988. A study of cemented and non-cemented prostheses. *Rev Chir Orthop Reparatrice Appar Mot*;**74** suppl 2:276–7.

Wykman A, Olsson E, Axdorph G, Goldie I, 1991. Total hip arthroplasty: a comparison between cemented and press-fit noncemented fixation. *J Arthroplasty*;**6**:19–29.

Non-controlled comparative studies

Abrahams TG, Crothers OD, 1992. Radiographic analysis of an investigational hydroxyapatite-coated total hip replacement. *Invest Radiol*;**27**:779–84.

Ahnfelt L, Herberts P, Malchau H, *et al.*, 1990. Prognosis of total hip replacement. A Swedish multicentre study of 4664 revisions. *Acta Orthop Scand Suppl*;**238**:1–26.

Bankston AB, Faris PM, Keating EM, Ritter MA, 1993. Polyethylene wear in total hip arthroplasty in patient matched groups. A comparison of stainless steel, cobalt chrome and titanium-bearing surfaces. *J Arthroplasty*;**8**:315–22.

Bankston AB, Keating EM, Ranawat C, Faris PM, Ritter MA, 1995. Comparison of polyethylene wear in machined versus molded polyethylene. *Clin Orthop*;(317):37–43.

Bertin KC, Freeman MAR, Morscher E, *et al.*, 1985. Cementless acetabular replacement using a pegged polyethylene prosthesis. *Arch Orthop Trauma Surg*;**104**:251–61.

Bourne RB, Rorabeck CH, Ghazal ME, Lee MH, 1994. Pain in the thigh following total hip replacement with a porous-coated anatomic prosthesis for osteoarthrosis. *J Bone Joint Surg Am*;**76**:1464–70.

Britton AR, Murray DW, Bulstrode CJ, McPherson K, Denham RA, 1996. Long-term comparison of Charnley and Stanmore design total hip replacements. *J Bone Joint Surg Br*;**78**:802–8.

Burkart BC, Bourne RB, Rorabeck CH, Kirk PG, 1993. Thigh pain in cementless total hip arthroplasty: a comparison of two systems at 2 years follow up. *Orthop Clin N Am*;**24**:645–53.

Callaghan JJ, Kim YS, Brown TD, Pedersen DR, Johnston RC, 1995. Concerns and improvements with cementless metal-backed acetabular components. *Clin Orthop*;(311):76–84.

Carlsson AS, Gentz CF, 1984. Radiographic versus clinical loosening of the acetabular component in noninfected total hip arthroplasty. *Clin Orthop*;(185):145–50.

Chmell MJ, Rispler D, Poss R, 1995. The impact of modularity in total hip arthroplasty. *Clin Orthop*;(319):77–84.

Cornell CN, Ranawat CS, 1986. The impact of modern cement techniques on acetabular fixation in cemented total hip replacement. *J Arthroplasty*;**1**:197–202.

Dall DM, Learmonth ID, Solomon MI, Miles AW, Davenport JM, 1993. Fracture and loosening of Charnley femoral stems: comparison between first-generation and subsequent designs. *J Bone Joint Surg Br*;**75**:259–65.

Djerf K, Wahlstrom O, 1986. Total hip replacement. Comparison between the McKee-Farrar and Charnley prostheses in a 5-year follow-up study. *Arch Orthop Trauma Surg*;**105**:158–62.

PAPER 3

Duck HJ, Mylod AG Jr, 1992. Heterotopic bone in hip arthroplasties: cemented versus noncemented. *Clin Orthop*;**282**:145–53.

Ebramzadeh E, Sarmiento A, McKellop HA, Llinas A, Gogan W, 1994. The cement mantle in total hip arthroplasty. Analysis of long term radiographic results. *J Bone Joint Surg Am*;**76**:77–87.

Espehaug B, Havelin LI, Engesaeter LB, Vollset SE, Langeland N, 1995. Early revision among 12,179 hip prostheses. A comparison of 10 different brands reported to the Norwegian arthroplasty register, 1987–1993. *Acta Orthop Scand*;**66**:487–93.

Freeman MAR, Plante-Bordeneuve P, 1994. Early migration and late aseptic failure of proximal femoral prostheses. *J Bone Joint Surg Br*;**76**:432–8.

Goetz DD, Smith EJ, Harris WH, 1994. The prevalence of femoral osteolysis associated with components inserted with or without cement in total hip replacements: a retrospective matched-pair series. *J Bone Joint Surg Am*;**76**:1121–9.

Hamada Y, Akamatsu N, Nakajima I, *et al.*, 1993. Thigh pain in cementless total hip replacement. *J Jpn Orthop Assoc*;**67**:561–71.

Havelin LI, Espehaug B, Vollset SE, Engesaeter LB, 1994. Early failures among 14,009 cemented and 1,326 uncemented prostheses for primary coxarthrosis: the Norwegian arthroplasty register, 1987–1992. *Acta Orthop Scand*;**65**:1–6.

Havelin LI, Espehaug B, Vollset SE, Engesaeter LB, 1995. Early aseptic loosening of uncemented femoral components in primary total hip replacement: a review based on the Norwegian arthroplasty register. *J Bone Joint Surg Br*;**77**:11–17.

Hearn SL, Bicalho PS, Eng K, Booth REJ, Hozack WJ, Rothman RH, 1995. Comparison of cemented and cementless total hip arthroplasty in patients with bilateral hip arthroplasties. *J Arthroplasty*;**10**:603–8.

Hedlundh U, Fredin H, 1995. Patient characteristics in dislocations after primary total hip arthroplasty: 60 patients compared with a control group. *Acta Orthop Scand*;**66**:225–8.

Hernandez JR, Keating EM, Faris PM, *et al.*, 1994. Polyethylene wear in uncemented acetabular components. *J Bone Joint Surg Br*;**76**:263–6.

Hodgkinson JP, Maskell AP, Ashok P, *et al.*, 1993. Flanged acetabular components in cemented Charnley hip arthroplasty. Ten-year follow-up of 350 patients. *J Bone Joint Surg Br*;**75**:464–7.

Hoffman CW, Lynskey TG, Hadlow V, 1994. Twenty years of total hip arthroplasty in Taranaki: a survival analysis. *Aust N Z J Surg*;**64**:538–41.

Horikoshi M, Macaulay W, Booth RE, Crossett LS, Rubash HE, 1994. Comparison of interface membranes obtained from failed cemented and cementless hip and knee prostheses. *Clin Orthop*;(309):69–87.

Hozack W, Gardiner R, Hearn S, Eng K, Rothman R, 1994. Taperloc femoral component: a 2–6 year study of the first 100 consecutive cases. *J Arthroplasty*;**9**:489–93.

Hozack WJ, Rothman RH, Booth REJ, Balderstone RA, 1993. Cemented versus cementless total hip arthroplasty: a comparative study of equivalent patient populations. *Clin Orthop*;(289):161–5.

Huracek J, Spirig P, 1994. The effect of hydroxyapatite coating on the fixation of hip prostheses. A comparison of clinical and radiographic results of hip replacement in a matched pair study. *Arch Orthop Trauma Surg*;**113**:72–7.

Hwang SK, Park JS, 1995. Cementless total hip arthroplasty with AML, PCA and HGP prostheses. *Int Orthop*;**19**:77–83.

Jacobsson S, Djerf K, Wahlstrom O, 1990. A comparative study between McKee-Farrar and Charnley arthroplasties with long-term follow up periods. *J Arthroplasty*;**5**:9–14.

Kelley SS, Johnstone RC, 1992. Debris from cobalt–chrome cable may cause acetabular loosening. *Clin Orthop*;(285):140–6.

Krismer M, Fischer M, Klestil T, Frischut B, 1991a. Uncoated polyethylene RM acetabular component versus Müller cemented acetabular component – a 4–8 year follow-up study. *Arch Orthop Trauma Surg*;**110**:195–9.

Krismer M, Klar M, Klestil T, Frischut B, 1991b. Aseptic loosening of straight and curved-stem Müller femoral prosthesis. *Arch Orthop Trauma Surg*;**110**:190–4.

Kristiansen B, Steen Jensen J, 1985. Biomechanical factors in loosening of the Stanmore hip. *Acta Orthop Scand*;**56**:21–4.

Lehman DE, Capello WN, Feinberg JR, 1994. Total hip arthroplasty without cement in obese patients: a minimum two-year clinical and radiographic follow-up study. *J Bone Joint Surg Am*;**76**:854–62.

Mallory TH, Kraus TJ, Vaughn BK, 1989. Intraoperative femoral fractures associated with cementless total hip arthroplasty. *Orthopedics*;**12**:231–9.

Maloney WJ, Harris WH, 1990. Comparison of a hybrid with an uncemented total hip replacement. A retrospective matched-pair study. *J Bone Joint Surg Am*;**72**:1349–52.

Markel DC, Huo MH, Katkin PD, Salvati EA, 1995. Use of cemented all-polyethylene and metal-backed acetabular components in total hip arthroplasty. A comparative study. *J Arthroplasty*;**10** suppl:S1–7.

McPherson EJ, Dorr LD, Gruen TA, Saberi MT, 1995. Hydroxyapatite-coated proximal ingrowth femoral stems: a matched pair control study. *Clin Orthop*;(315):223–30.

Moilanen T, Stocks GW, Freeman MAR, Scott G, Goodier WD, Evans SJW, 1996. Hydroxyapaptite coating of an acetabular prosthesis. *J Bone Joint Surg Br*;**78**:200–5.

Moskal JT, Shaffrey CI, Ripley LP, 1994. Prospective analysis of uncemented and hybrid primary porous coated anatomic total hip arthroplasties in a community setting. *Clin Orthop*;(304):139–47.

PAPER 3

Nashed RS, Becker DA, Gustilo RB, 1995. Are cementless acetabular components the cause of excess wear and osteolysis in total hip arthroplasty? *Clin Orthop*;(317):19–28.

Neumann L, Freund KG, Sorensen KH, 1996. Total hip arthroplasty with the Charnley prosthesis in patients fifty-five years old and less. Fifteen to twenty-one year results. *J Bone Joint Surg Am*;**78**:73–9.

Pierchon F, Pasquier G, Cotten A, Fontaine C, Clarisse J, Duquennoy A, 1994. Causes of CT study of component alignment. *J Bone Joint Surg Br*;**76**:45–8.

Pritchett JW, 1995. Femoral bone loss following hip replacement: a comparative study. *Clin Orthop*;(314):156–61.

Pupparo F, Engh CA, 1992. Comparison of porous-threaded and smooth threaded acetabular components of identical design: 2–4 year results. *Clin Orthop*;(271): 201–6.

Ranawat CS, Rawlins BA, Harju VT, 1988. Effect of modern cement technique on acetabular fixation total hip arthroplasty – a retrospective study in matched pairs. *Orthop Clin N Am*;**19**:599–603.

Rand JA, Ilstrup DM, 1993. Comparison of Charnley and T-28 total hip arthroplasty. *Clin Orthop*;(180):201–5.

Riska EB, 1993. Ceramic endoprosthesis in total hip arthroplasty. *Clin Orthop*;(297):87–94.

Ritter MA, 1995. The cemented acetabular component of a total hip replacement: all polyethylene versus metal backing. *Clin Orthop*;(311):69–75.

Ritter MA, Gioe TJ, 1986. Conventional versus resurfacing total hip arthroplasty. A long-term prospective study of concomitant bilateral implantation of prostheses. *J Bone Joint Surg Am*;**68**:216–25.

Schreiber A, Jacob HAC, Hauser R, Kern S, 1993. The self-locking Balgrist hip socket for cementless fixation: biomechanical principles and clinical results. *Keio J Med*;**42**(2):53–9.

Schuller HM, Marti RK, 1990. Ten-year socket wear in 66 hip arthroplasties: ceramic versus metal heads. *Acta Orthop Scand*;**61**:240–3.

Turner RS, 1994. Postoperative total hip prosthetic femoral head dislocations: incidence, etiologic factors, and management. *Clin Orthop*;(301):196–204.

Visuri T, Pulkkinen P, Turula KB, Paavolainen P, Koskenvuo M, 1994. Life expectancy after hip arthroplasty: case control study of 1018 cases of primary arthrosis. *Acta Orthop Scand*;**65**:9–11.

Walker PS, Mai SF, Cobb AG, Bentley G, Hua J, 1995. Prediction of clinical outcome of THR from migration measurements on standard radiographs: a study of cemented Charnley and Stanmore femoral stems. *J Bone Joint Surg Br*;**77**:705–14.

Wilson-MacDonald J, Morscher E, 1989. Comparison between straight- and curved-stem Müller femoral prostheses: 5- to 10-year results of 545 total hip replacements. *Arch Orthop Trauma Surg*;**109**:14–20.

Wixson RL, Stulberg SD, Mehlhoff M, 1991. Total hip replacement with cemented, uncemented and hybrid prostheses: a comparison of clinical and radiographic results at two to four years. *J Bone Joint Surg Am*;**73**:257–70.

Yahiro MA, Gantenberg JB, Nelson R, *et al.*, 1995. Comparison of the results of cemented, porous-ingrowth, and threaded acetabular cup fixation: a meta-analysis of the orthopaedic literature. *J Arthroplasty*;**10**:339–50.

Zicat B, Engh CA, Gokcen E, 1995. Patterns of osteolysis around total hip components inserted with and without cement. *J Bone Joint Surg Am*;**77**:432–9.

Zichner LP, Willert HG, 1992. Comparison of alumina–polyethylene and metal–polyethylene in clinical trials. *Clin Orthop*;(282):86–94.

Observational/descriptive studies

Ahnfelt L, Herberts P, Malchau H, *et al.*, 1990. Prognosis of total hip replacement. A Swedish multicenter study of 4664 revisions. *Acta Orthop Scand Suppl*;**238**:1–26.

Alsema R, Deutman R, Mulder TJ, 1994. Stanmore total hip replacement: a 15 to 16 year clinical and radiographic follow up. *J Bone Joint Surg Br*;**76**:240–4.

August AC, Aldam CH, Pynsent PB, 1986. The McKee-Farrar hip arthroplasty: a long-term study. *J Bone Joint Surg Br*;**68**:520–7.

Ballard WT, Callaghan JJ, Sullivan PM, Johnston RC, 1994. The results of improved cementing techniques for total hip arthroplasty in patients less than fifty years old. *J Bone Joint Surg Am*;**76**:959–64.

Bhamra M, Hulme AL, Hutton PAN, Andrews BG, Muirhead-Allwood WFG, 1992. Failure of cementless hip replacements. *J Bone Joint Surg Br*;**74**:932.

Blaha JD, Gruen TA, Grappiolo G, Mancinello CA, Spotorno L, Romagnoli S, *et al.*, 1994. Femoral component fixation in the 1990s: porous coating: do we need it? *Orthopedics*;**17**:779–80.

Bohler M, Knahr K, Reigler M, *et al.*, 1994. Ten year results of cemented Weller type total hip arthroplasties. *Arch Orthop Trauma Surg*;**113**(2):57–60.

Bosco JA, Lachiewicz PF, DeMasi R, 1993. Survivorship analysis of cemented high modulus total hip arthroplasty. *Clin Orthop*;(294):131–9.

Bourne RB, Rorabeck CH, Ghazal ME, Lee MH, 1994. Pain in the thigh following total hip replacement with a porous-coated anatomic prosthesis for osteoarthrosis. *J Bone Joint Surg Am*;**76**:1464–70.

Brady LP, McCutchen JW, 1986. A ten year follow-up study of 170 Charnley total hip arthroplasties. *Clin Orthop*;(211):51.

Bryant MJ, Mollan RAB, Nixon JR, 1991. Survivorship analysis of the Ring hip arthroplasty. *J Arthroplasty*; **6** suppl:S5–10.

Callaghan JJ, Heekin RD, Savory CG, 1992. Evaluation of the learning curve associated with uncemented primary porous-coated anatomic total hip arthroplasty. *Clin Orthop*;(282):132–44.

Capello WN, 1994. Hydroxyapatite in total hip arthroplasty: five-year clinical experience. *Orthopedics*;**17**:781–92.

Capello WN, Sallay PI, Feinberg JR, 1994. Omniflex modular femoral component. 2–5 year results. *Clin Orthop*;(298):54–9.

Carlsson AS, Gentz GF, Sanzen L, 1986. Socket loosening after hip arthroplasty radiograph observations in 241 cases up to 15 years. *Acta Orthop Belg*;**57**:97–100.

Carter SR, Pynsent PB, McMinn DWJ, 1991. Greater than ten year survivorship of Charnley low fraction arthroplasty. *J Bone Joint Surg Br*;**73**:71.

Collis DK, 1988. Long-term results of an individual surgeon. *Orthop Clin N Am*;**19**:541–50.

Cordero-Ampuero J, Garcia-Cimbrelo E, Munuera L, 1994. Fixation of cementless acetabular cups: a radiographic 4–8 year study of 102 porous-coated components. *Acta Orthop Scand*;**65**:263–6.

Cracchiolo A, Severt R, Moreland J, 1992. Uncemented total hip arthroplasty in rheumatoid arthritis diseases. A two- to six-year follow-up study. *Clin Orthop*;(277):166–74.

d'Antonio JA, Capello WN, Crothers OD, Jaffe WL, Manley MT, 1992a. Early clinical experience with hydroxyapatite-coated femoral implants. *J Bone Joint Surg Am*;**74**:995–1008.

d'Antonio JA, Capello WN, Jaffe WL, 1992b. Hydroxyapatite-coated hip implants: multicenter three-year clinical and roetgenographic results. *Clin Orthop*;(285):102–15.

Dall DM, Learmonth ID, Dall G, 1988a. Long-term results of Charnley low-friction arthroplasty of the hip. *S Afr Med J*;**73**:89–92.

Dall DM, Learmonth ID, Solomon MI, 1988b. Charnley low friction arthroplasty: the yardstick (a 3–7 year follow-up). *J Bone Joint Surg Br*;**70**:499.

Dall DM, Learmonth ID, Solomon M, Davenport JM, 1993. Charnley hips followed for 3–17 years. *Acta Orthop Scand*;**64**:252–6.

Dorr LD, Kane TJ, Conaty JP, 1994. Long-term results of cemented total hip arthroplasty in patients 45 years old or younger: a 16-year follow-up study. *J Arthroplasty*;**9**:453–6.

Drucker DA, Capello WN, d'Antonio JA, Hile HA, 1991. Total hip arthroplasty using a hydroxyapatite-coated acetabular and femoral component. *Orthop Rev*;**20**:179–92.

Duparc J, Massin P, 1992. Results of 203 total hip replacements using a smooth, cementless femoral component. *J Bone Joint Surg Br*;**74**:251–6.

Eftekhar NS, 1987. Long term results of cemented total hip arthroplasty. *Clin Orthop*;(225):207–17.

Eftekhar NS, Tzitzikalakis G, 1986. Failures and reoperations following low friction arthroplasty of the hip. *Clin Orthop*;(211):65.

Engh CA, 1994. Femoral component fixation in the 1990s: porous-coated fixation forever. *Orthopedics*;**17**:775–6.

Engh CA, Massin P, 1989. Cementless total hip arthroplasty using the AML stem. Results using a survivorship analysis. *Clin Orthop*;(249):141–58.

Engh CA, Glassman AH, Suthers KE, 1990. The case for porous-coated hip implants: the femoral side. *Clin Orthop*;(261):63–81.

Engh CA, Hooten JP, Zettl-Schaffer KF, Ghaffarpour M, McGovern TF, Macalino GE, *et al.*, 1994. Porous-coated total hip replacement. *Clin Orthop*;(298): 89–96.

Engh CA Jr, Culpepper II WJ, Engh CA, 1997. Long-term results of the Anatomic Medullary Locking prosthesis in total hip arthroplasty. *J Bone Joint Surg Am*;**79**:169–76.

Fowler JL, Gie GA, Lee AJC, Ling RSM, 1988. Experience with the Exeter total hip replacement since 1970. *Orthop Clin N Am*;**19**:477–89.

Garcia-Cimbrelo E, Munera L, 1992. Early and late loosening of the acetabular cup after low-friction arthroplasty. *J Bone Joint Surg Am*;**74**:1119–29.

Geesink RGT, Hoefnagels NHM, 1995. Six-year results of hydroxyapatite-coated total hip replacement. *J Bone Joint Surg Br*;**77**:534–47.

Geesink RGT, 1990. Hydroxyapatite-coated total hip prostheses: two-year clinical and roentgenographic results of 100 cases. *Clin Orthop*;(261):39–58.

Glorion C, Dijan P, Goutallier D, 1994. Uncemented acetabular components. Roentgenological changes in 77 total hip prostheses after a mean follow-up of 3.5 years. *Rev Rhum Engl Ed*;**61**:106–44.

Groher W, 1983. Uncemented total hip replacement. *Can J Surg*;**26**:534–6.

Gudmundsson GH, Hedeboe J, Kjaer J, 1985. Mechanical loosening after hip replacement: incidence after 10 years in 125 patients. *Acta Orthop Scand*;**56**:314–17.

Haddad RJ, Skalley TC, Cook SD, Brinker MR, Cheramie J, Meyer R, *et al.*, 1990. Clinical and roentgenographic evaluation of noncemented porous-coated Anatomic Medullary Locking (AML) and Porous-Coated Anatomic (PCA) total hip arthroplasties. *Clin Orthop*;(258):176–82.

Hamilton HW, Gorczyca J, 1995. Low friction arthroplasty at 10 to 20 years: consequences of plastic wear. *Clin Orthop*;(318):160–6.

Hamilton HW, Joyce M, 1986. Long term results of low friction arthroplasty performed in a community hospital including a radiologic review. *Clin Orthop*;(211):55.

Harper GD, Bull T, Cobb AG, Bentley G, 1995. Failure of the Ring polyethylene uncemented acetabular cup. *J Bone Joint Surg Br*;**77**:557–61.

Harris WH, Maloney WJ, 1989. Hybrid total hip arthroplasty. *Clin Orthop*;(249):21–9.

PAPER 3

Harris WH, Penenberg BL, 1987. Further follow-up on socket fixation using a metal-backed acetabular component for total hip replacement. A minimum ten-year follow-up. *J Bone Joint Surg Am*;**67**:1140.

Hartofilakidis G, Stamos K, Ioannidis TT, 1989. Fifteen years experience with Charnley low friction arthroplasty. *Clin Orthop*;(246):48–56.

Heekin RD, Callaghan JJ, Hopkinson WJ, Savory CG, Xenos JS, 1993. The porous coated anatomic total hip prosthesis, inserted without cement. Results after 5–7 years in a prospective study. *J Bone Joint Surg Am*;**75**:77–91.

Helfen M, Malzer U, Peters P, Griss P, Himmelmann G, Weber E, 1993. Clinical results with a hybrid total hip arthroplasty: a 3–6 year follow-up [abstract]. *Z Orthop Ihre Grenzgeb*;**131**:578–84.

Hellman EJ, Capello WN, Feinberg JR, 1997. Porous coated Omnifit total hip arthroplasty: 5–11 year results. In press.

Hirose I, Capello WN, Feinberg JR, 1995. Primary cemented total hip arthroplasty: five to twelve year clinical and radiographic follow-up. *Iowa Orthop*;**15**:43–7.

Hodgkinson JP, Maskell AP, Ashok P, Wroblewski BM, 1993. Flanged acetabular components in cemented Charnley hip arthroplasty. Ten-year follow-up of 350 patients. *J Bone Joint Surg Br*;**75**:464–7.

Holman PR, Tyer HDD, 1992. Porous coated anatomic non-cemented total hip arthroplasty. *Aust N Z J Surg*;**62**:56–9.

Incavo SJ, DiFazio FA, Howe JG, 1993. Cementless hemispheric acetabular components: 2–4 year results. *J Arthroplasty*;**8**:573–80.

Ivory JP, Choudhry RR, Stoyle TF, 1992. Thigh pain after hip arthroplasty. *J Bone Joint Surg Br*;**74**:934.

Jansson V, Refior HJ, 1992. Clinical results and radiologic findings after cementless implantation of PCA stems in total hip replacement. *Arch Orthop Trauma Surg*;**111**:305–8.

Jantsch S, Schwagerl W, Zenz P, Semlitsch M, Fertschak W, 1991. Long-term results after implantation of McKee-Farrar total hip prostheses. *Arch Orthop Trauma Surg*;**110**:230–7.

Johnsson R, Thorngren KG, Persson BM, 1988. Revision of total hip replacement for primary osteoarthritis. *J Bone Joint Surg Br*;**70**:56–62.

Johnston RC, Crowninshield RD, 1983. Roentgenologic results of total hip arthroplasty. A ten year follow-up study. *Clin Orthop*;(181):92–8.

Joshi AB, Porter ML, Trail IA, Hunt LP, Murphy JCM, Hardinge K, 1993. Long-term results of Charnley low friction arthroplasty in young patients. *J Bone Joint Surg Br*;**75**:616–23.

Karachalios T, Hartofilakidis G, Zacharakis N, 1993. A 12 to 18 year radiographic follow-up study of Charnley low friction arthroplasty: the role of the center of rotation. *Clin Orthop*;(296):140–7.

Karrholm J, Borssen B, Lowenhielm G, Snorrason F, 1994. Does early micromotion of femoral stem prostheses matter? 4–7-year stereoradiographic follow-up of 84 cemented prostheses. *J Bone Joint Surg Br*;**76**:912–17.

Kavanagh BF, Dewitz MA, Ilstrup DM, Stauffer RN, Coventry MB, 1989. Charnley total hip replacement with cement: fifteen year results. *J Bone Joint Surg Am*;**71**:1496–503.

Kennedy WR, 1994. Acetabular component fixation in the 1990s: cementless components: the good, the bad, and the ugly. *Orthopedics*;**17**:783–4.

Kienapfel H, Martell J, Rosenberg A, Galante J, 1991. Cementless Gustilo-Kyle and BIAS total hip arthroplasty: 2–5 year results. *Arch Orthop Trauma Surg*;**110**:179–86.

Kienapfel H, Pitzer W, Griss P, 1992. Three- to five-year results with the cementless Harris-Galante acetabular component used in hybrid total hip replacement. *Arch Orthop Trauma Surg*;**111**:142–7.

Kim YH, Kim VEM, 1992. Results of the Harris-Galante cementless total hip prosthesis. *J Bone Joint Surg Br*;**74**:83–7.

Kim YH, Kim VEM, 1993. Uncemented porous coated anatomic total hip replacement. *J Bone Joint Surg Br*;**75**:6–13.

Kobayashi S, Eftekhar NS, Terayama K, *et al.*, 1994a. Risk factors affecting radiological failure of the socket in primary Charnley low friction arthroplasty: a 10- to 20-year follow-up study. *Clin Orthop*;(306):84–96.

Kobayashi S, Eftekhar NS, Terayama K, 1994b. Predisposing factors in fixation failure of femoral prostheses following primary Charnley low friction arthroplasty: a 10- to 20-year follow-up study. *Clin Orthop*;(306):73–83.

Koch FW, Messler HH, Wagner U, Meyer HJ, 1993. Short term results (2–5 years) with Furlong type hydroxyapatite-coated total hip replacements. *Z Orthop Ihre Grenzgeb*;**131**:562–7.

Kutschera HP, Eyb R, Schartelmuller T, Toma C, Zweymuller K, 1993. Cementless hip replacement – Zweymuller. 5 years follow-up. *Z Orthop Ihre Grenzgeb*;**131**:513–17.

Lachiewicz PF, 1994. Porous-coated total hip arthroplasty in rheumatoid arthritis. *J Arthroplasty*;**9**:9–15.

Lachiewicz PF, Rosenstein BD, 1986. Long-term results of Harris total hip replacement. *J Arthroplasty*;**1**:229–36.

Langlais F, Benkalfate T, Thomazeau H, Abboud A, Lancien G, Chauvel JJ, 1995. Femoral loosening of THR due to trochanteric non unions. *Rev Chir Orthop Reparatrice Appar Mot*;**81**:95–105.

Lautiainen IA, Joukainen J, Makela EA, 1994. Clinical and roentgenographic results of cementless total hip arthroplasty. *J Arthroplasty*;**9**:653–60.

Learmonth ID, Grobler GP, Dall DM, Jandera V, 1995. Loss of bone stock with cementless hip arthroplasty. *J Arthroplasty*;**10**:257–63.

Madey SM, Callaghan JJ, Olejniczak JP, 1997. Charnley total hip arthroplasty with use of improved techniques of cementing. *J Bone Joint Surg Am*;**79**:53–64.

Maloney WJ, Davey JR, Harris WH, 1992. Bead loosening from a porous-coated acetabular component: a follow up note. *Clin Orthop*;(281):112–14.

Martell JM, Pierson RH, Jacobs JJ, Rosenberg AG, Maley M, Galante JO, *et al.*, 1993. Primary total hip reconstruction with a titanium fiber-coated prosthesis inserted without cement. *J Bone Joint Surg Am*;**75**:554–71.

McCoy TH, Salvati EA, Ranawat CS, Wilson PD, 1988. A fifteen year follow up study of one hundred Charnley low-friction arthroplasties. *Orthop Clin N Am*;**19**:467–76.

Mohler CG, Callaghan EJJ, Collis DK, Oregon E, Johnston RC, 1995a. Early loosening of the femoral component at the cement-prosthesis interface after total hip replacement. *J Bone Joint Surg Am*;**77**:1315–22.

Mohler CG, Kull LR, Martell JM, Rosenberg AG, Galante JO, 1995b. Total hip replacement with insertion of an acetabular component without cement and a femoral component with cement. *J Bone Joint Surg Am*;**77**:86–96.

Moskal JT, Shaffrey CI, Ripley LP, 1994. Prospective analysis of uncemented and hybrid primary porous coated anatomic total hip arthroplasties in a community setting. *Clin Orthop*;(304):139–47.

Negre J, Henry F, 1995. Bone reaction on a blasted titanium surface. Results of 101 THP with six year follow-up. *Rev Chir Orthop Reparatrice Appar Mot*;**81**:106–13.

Neumann L, Freund KG, Sorenson KH, 1994. Long-term results of Charnley total hip replacement: review of 92 patients at 15 to 20 years. *J Bone Joint Surg Br*;**76**:245–51.

Neumann L, Freund KG, Sorensen KH, 1996. Total hip arthroplasty with the Charnley prosthesis in patients fifty-five years old and less. Fifteen to twenty-one year results. *J Bone Joint Surg Am*;**78**:73–9.

Nicholson OR, 1992. The Charnley hip replacement: a longish term review. *J Bone Joint Surg Br*;**74** suppl 1:37.

Niinimaki T, Puranen J, Jalovaara P, 1994. Total hip arthroplasty using isoelastic femoral stems: a seven- to nine-year follow-up in 108 patients. *J Bone Joint Surg Br*;**76**:413–18.

Nizard RS, Sedel L, Christel P, Meunier A, Soudry M, Witvoet J, 1992. Ten year survivorship of cemented ceramic-ceramic total hip prosthesis. *Clin Orthop*;(282):53–63.

Ohlin A, 1990. Failure of the Christiansen hip: survival analysis of 265 cases. *Acta Orthop Scand*;**61**:7–11.

Ohlin A, Onsten I, 1990. Loosening of the Lubinus hip: 202 cases followed for 3–6 years. *Acta Orthop Scand*;**61**:244–7.

Oishi CS, Walker RH, Colwell CW, 1994. The femoral component in total hip arthroplasty: six- to eight-year follow-up of one hundred consecutive patients after use of a third-generation cementing technique. *J Bone Joint Surg Am*;**76**:1130–5.

Older J, 1986. Low friction arthroplasty of the hip. A ten to twelve year follow-up study. *Clin Orthop*;(211):36.

Older J, Butorac R, 1992. Charnley low friction arthroplasty (LFA): a 17–21 year follow-up study. *J Bone Joint Surg Br*;**74** suppl 3:251.

Owen TD, Moran CG, Smith SR, Pinder IM, 1994. Results of uncemented porous-coated anatomic total hip replacement. *J Bone Joint Surg Br*;**76**:258–62.

Papenfus K, Clayton ML, Dennis DA, Stringer TL, Stringer EA, 1992. The Clayton total hip arthroplasty: a ten year follow-up study. *Clin Orthop*;(281):133–9.

Partio E, von Bonsdorff H, Wirta J, Avikainen V, 1994. Survival of the Lubinus hip prosthesis: an eight- to 12-year follow-up evaluation of 444 cases. *Clin Orthop*;(303):140–6.

Pearse MF, Langkamer VG, Fiallos JM, Smith EJ, Atkins RM, Solomon L, 1992. Hybrid hip arthroplasty: 5 year results. *J Orthop Rheumatol*;**5**:85–96.

Pelligrini VD, Hughes SS, McCollister Evarts C, 1992. A collarless cobalt-chrome femoral component in uncemented total hip arthroplasty: 5–8 year follow-up. *J Bone Joint Surg Br*;**74**:814–21.

Picault C, Michel F, 1995. A 15- to 23-year follow-up of LFA Charnley hip arthroplasty. *J Jpn Orthop Assoc*;**69**:S704.

Ranawat CS, Hansraj KK, Neves MC, 1989. A seventeen-year survivorship study of Charnley total hip replacement. *Clin Exp Rheumatol*;**7** suppl 3:153–4.

Rasmussen S, Tromberg HB, Vestergaard F, 1991. Survival of primary Charnley THA in patients under age 55. *Acta Orthop Scand Suppl*;**62**(246):43.

Ring PA, 1978. Five to fourteen years interim results of uncemented total hip arthroplasty. *Clin Orthop*;(137):87.

Ring PA, 1987. Long term results of uncemented hip replacement. In: Noble J, Galasko CSB, editors. Recent developments in orthopaedic surgery. Manchester: Manchester University Press:214–21.

Riska EB, 1993. Ceramic endoprosthesis in total hip arthroplasty. *Clin Orthop*;(297):87–94.

Roberts JA, Finlayson DF, Freeman PA, 1987. The long term results of the Howse total hip arthroplasty: with particular reference to those requiring revision. *J Bone Joint Surg Br*;**69**:545.

Rockborn P, Olsson SS, 1993. Loosening and bone resorption in Exeter hip arthroplasties: a review at a minimum of five years. *J Bone Joint Surg Br*;**75**:865–8.

Roffman M, Juhn A, 1993. Nine year follow-up of the Mathys uncemented acetabular cup: a radiological evaluation. *Bull Hosp Jt Dis Orthop Inst*;**53**(1):51–3.

Rossi P, Sibelli P, Fumero S, Crua E, 1995. Short-term results of hydroxyapatite-coated primary total hip arthroplasty. *Clin Orthop*;(310):98–102.

Russotti GM, Coventry MB, Stauffer RN, 1988. Cemented THA using contemporary techniques – a minimum five year follow-up study. Proceedings of symposium: Cement vs. cementless total hip arthroplasty is the wrong question: 1988 September 28–October 1: Boston, Mass.

PAPER 3

Saito M, Saito S, Ohzono K, 1992. Efficacy of alumina ceramic heads for cemented total hip arthroplasty. *Clin Orthop*;(283):171–7.

Schmalzried TP, Harris WH, 1992. The Harris-Galante porous-coated acetabular component with screw fixation. *J Bone Joint Surg Am*;**74**:1130–9.

Schmalzried TP, Harris WH, 1993. Hybrid total hip replacement. A 6.5 year follow up study. *J Bone Joint Surg Am*;**75**:608–15.

Schulte KR, Callaghan JJ, Kelley SS, Johnston RC, 1993. The outcome of Charnley hip arthroplasty with cement after a minimum 20 year follow-up. The results of one surgeon. *J Bone Joint Surg Am*;**75**:961–75.

Seral F, Villar JF, Estellar A, Vivar FG, Abad I, Grande MM, *et al.*, 1992. 5 year follow-up evaluation of the noncemented press-fit titanium hip-joint endoprosthesis. *Clin Orthop*;(283):49–56.

Shaw JA, Bruna A, Paul EM, 1992. The influence of age, sex and initial fit on bony ingrowth stabilization with the AML femoral component in primary THA. *Orthopedics*;**15**:687–92.

Skeie S, Lende S, Sjoberg EJ, Vollset SE, 1991. Survival of the Charnley hip in coxarthrosis. A 10–15 year follow-up of 629 cases. *Acta Orthop Scand*;**62**:98–101.

Smith SE, Garvin KL, Jardon M, Kaplan PA, 1991. Uncemented total hip arthroplasty. Prospective analysis of the Tri-Lock femoral component. *Clin Orthop*;(269):43–50.

Solomon MI, Dall DM, Learmonth ID, Davenport JM, 1992. Survivorship of cemented total hip arthroplasty in patients 50 years of age or younger. *J Arthroplasty*; **7** suppl:347–52.

Sotereanos NG, Engh CA, Glassman AH, Macalino GE, Engh CA, 1995. Cementless femoral components should be made from cobalt–chrome. *Clin Orthop*;(313):146–53.

Stauffer RN, 1982. A ten year follow-up study of total hip replacement: with particular reference to roentgenographic loosening of the components. *J Bone Joint Surg Am*;**64**:983–90.

Stern LLD, Cabadas MI, Fdez-Arroyo JMF, Zarzosa R, Sanchez-Barbero JC, Granado J, 1992. LD hip arthroplasty: design concepts and first clinical trials of a new modular system. *Clin Orthop*;(283):39–48.

Stockley I, McLean L, Gross AE, 1992. Clinical and radiographic results of the Muller straight stem used as a press-fit. *J Arthroplasty*;**7**:477–82.

Sullivan PM, MacKenzie JR, Callaghan JJ, Johnston RC, 1994. Total hip arthroplasty with cement in patients who are less than fifty years old: a sixteen to twenty-two year follow-up study. *J Bone Joint Surg Am*;**76**:863–9.

Tang-Kue L, 1995. A 5 to 9 year follow-up of cementless porous-coated anatomic primary hip arthroplasty using modified lateral approach. *J Jpn Orthop Assoc*;**69**:S713.

Terayana K, 1986. Experience with Charnley low friction arthroplasty in Japan. *Clin Orthop*;(211):79.

Thomas AMC, McMinn DJW, 1991. Techniques for the enhancement of the bone cement interface in joint replacement. *J Jpn Orthop Assoc*;**65**:S381.

Thomas BJ, Salvati EA, Small RD, 1986. The CAD hip arthroplasty: 5–10 year follow-up. *J Bone Joint Surg Am*;**68**:640–6.

Tompkins GS, Lachiewicz PF, deMasi R, 1994. A prospective study of a titanium femoral component for cemented total hip arthroplasty. *J Arthroplasty*;**9**:623–30.

Tonino AJ, Romanini L, Rossi P, Borroni M, Greco F, Garcia-Araujo C, 1995. Hydroxyapatite-coated hip prostheses: early results from an international study. *Clin Orthop*;(312):211–25.

Warren PJ, O'Doherty D, Scott G, Greenough CG, 1993. The Furlong total hip: satisfactory interim results of a new cemented arthroplasty. *J Orthop Rheumatol*;**6**:15–20.

Wejkner B, Stenport J, 1988. Charnley total hip arthroplasty: a 10- to 14-year follow-up study. *Clin Orthop*;(231):113–19.

Welch RB, McGann WA, Picetti GD, 1988. Charnley low friction arthroplasty: a fifteen- to seventeen-year follow-up study. *Orthop Clin N Am*;**19**:551.

Wilson-MacDonald J, Morscher E, Masar Z, 1990. Cementless uncoated polyethylene acetabular components in total hip replacement. Review of five- to ten-year results. *J Bone Joint Surg Br*;**72**:423.

Wroblewski BM, 1986. 15–21 year results of the Charnley low-friction arthroplasty. *Clin Orthop*;(211):30–5.

Wroblewski BM, Siney PD, 1993. Charnley low friction arthroplasty of the hip: long term results. *Clin Orthop*;(292):210–4.

Xenos JS, Hopkinson WJ, Callaghan JJ, Heekin RD, Savory CG, 1995. Osteolysis around an uncemented cobalt chrome total hip arthroplasty. *Clin Orthop*;(317):29–36.

Selected observational studies of threaded acetabular cups

Bruijn JD, Seelen JL, Feenstra RM, Hansen BE, Bernoski FP, 1995. Failure of the Mecring screw-ring acetabular component in total hip arthroplasty: a three to seven-year follow-up study. *J Bone Joint Surg Am*;**77**:760–6.

Fox GM, McBeath AA, Heiner JP, 1994. Hip replacement with a threaded acetabular cup. A follow-up study. *J Bone Joint Surg Am*;**76**:195–201.

Gouin F, Fechoz F, Passuti N, 1993. Uncemented threaded acetabular cups. *Int Orthop*;**17**(2):65–72.

Gut M, Hilfiker B, Schreiber A, 1990. 5–7 year results of treatment with endler cementless acetabular sockets. *Z Orthop Ihre Grenzgeb*;**128**:598–605.

Harwin SF, Stern RE, Kulick RG, 1991. Work in progress #7. Clinical experience with the accu-path threaded acetabular cup. A preliminary report. *Orthop Rev*;**20**:897–914.

Krugluger J, Eyb R, 1993. Bone reaction to uncemented threaded polyethylene components. *Int Orthopaed*;**17**:259–65.

PAPER 3

Appendix 1

Data tables and appraisal tables

The Data Tables, which include data extracted from all the studies included in this review, and the Appraisal Tables, the detailed checklist-based appraisals of these studies, form the major part of this appendix.

However, given the high value placed upon – and the relative scarcity of – RCTs in the scientific orthopaedic literature on hip prostheses, a brief summary of each of the RCTs included in this review is presented first. Details of the critical appraisal of each are in the appraisal tables. Studies with reference numbers are mentioned in the main text.

Individual summaries of RCTs

Bourne and colleagues, 1995;[45] Rorabeck and colleagues, 1996[46]

There have been a number of reports of this patient population. These two papers present the most recent results.

All patients were operated on or supervised by two senior surgeons using the Mallory Head prosthesis, either cemented or press-fit. Originally 250 patients were recruited from a group with an age range of 18–75 years, and were stratified by age and surgeon. Diagnosis was primary or secondary osteoarthritis of the hip without additional life threatening illnesses. Results of clinical aspects were reported with a 5-year follow-up (number of patients not stated) and for radiographic analysis with a 4-year follow-up (147 patients). Patients and clinical observers were blinded. All clinical assessments (for example, Harris Score, d'Aubigne Score and Sickness Impact Profile) were almost identical for each group both pre- and postoperatively. There was no subsidence in the cemented stems but 14% of press-fit stems subsided by 3–5 mm. No revisions were required within the 4 years and no press-fit components or cemented stems were loose; however, 26% of the cemented cups were termed definitely or probably loose. **C-rated.**

Bradley and Ledd, 1992[34]

In this study the Furlong (HA-coated prosthesis, n = 97) was compared with the Charnley (n = 73). The patients (with primary osteoarthritis and of average age 68 years, range 45–75 years) were randomly allocated by year of birth, even numbered years to the Furlong, odd numbers to the Charnley (any patient technically unsuitable for the HA-coated prosthesis were excluded irrespective of their year of birth). This paper reported on 139 patients with follow-up of 1 year and 74 patients with 2 years. Patients were seen pre- and postoperatively at their homes by a nurse practitioner, radiographic assessment took place at hospital at similar time intervals. The results at 1 year and 2 years were very similar in both groups (that is, numbers with no pain or limp, walking distance and use of walking aids, ability to climb stairs and put on shoes/socks). No patients were lost to follow-up but two in each group died of unrelated causes. There were no revisions or evidence of loosening. **C-rated.**

Carlsson and colleagues, 1995

Originally 352 patients were included in this study from five Swedish orthopaedic departments (separate randomisation was performed at each centre, where surgical approach and prosthesis type differed). Those with radiographic loosening and diagnoses other than primary arthrosis were excluded (among others) leaving 190 hips. Three of the prostheses studied had a large collar (Lubinus, Harris Design 2 and Scanhip) while the Exeter and Charnley were collarless. Resorption of the resected femoral neck was more often observed in prostheses with a true collar (odds ratio at 5 years postoperatively 4.1, $p < 0.001$) and to a greater extent ($p < 0.001$). No details were given either of the patients or if the assessments were performed blindly and/or independently of the operating surgeon. **C-rated.**

Ciccotti and colleagues, 1994

Primary cementless THRs were inserted into 60 patients with osteoarthritis (mean age 63–66 years). They were matched for age and weight prior to randomisation (no details) to HA- or porous-coated Taperloc prostheses **and** to postoperative weight-bearing status (at 6 or 12 weeks), thus giving four groups. Eight 0.8 mm tantalum markers were also positioned in the bone during the operation. All patients were assessed after 2 years. No differences were found between

PAPER 3

the groups either relating to the coating on the prostheses or to the timing of weight-bearing status as measured by Charnley scores (pre- or postoperatively) or by Visual Analogue Scales for thigh pain. Migration was less than 1.40 mm in all groups and no revisions were reported to be necessary within the 2 years. **C-rated.**

Godsiff and colleagues, 1992[25]

In this study 30 cemented were compared with 28 cementless femoral components (Ring prosthesis) in patients with an age range of 55–74 years and a diagnosis of primary osteoarthritis of the hip only. Patients were assigned to a group by choosing one out of two envelopes themselves and were blinded to the result. Surgery was by one of two surgeons and the clinical assessment was by an independent, non-orthopaedic medical practitioner. At 2 years the two groups (n = 47) reported a similar pain incidence, the cementless group having had more pain at 4 and 12 months. By 2 years 96% cemented and 62% cementless did not require walking aids (p, 0.01–0.05). Preliminary results seemed to indicate cemented to be superior to cementless; however, because of unacceptable levels of femoral breakages at 3–5 years, the authors withdrew the prosthesis. **A-rated.**

Jacobsson and colleagues, 1994[114]

Two senior surgeons operated on 56 patients (24 women, 32 men, mean age 52 years) of whom 75% had a diagnosis of osteoarthritis, the remainder having various reasons for the unilateral hip operation. Patients were matched in pairs for sex, age, weight and radiographic appearance before being randomly selected (no details of method) to have a Butel (stem made of four rods for flexibility) or a PCA (rigid) stem (three different press-fit cups were used). Each pair was operated on by the same surgeon and were followed-up for 3 years. The PCA stem gave better results as assessed by Harris Hip Scores (mean 94.4 compared with 78.5 for Butel, p-value not given) and the number of prostheses definitely or probably loose (PCA 18%, Butel 86%). Both groups required three hip revisions because of loosening (one further Butel hip was revised for other reasons). **C-rated.**

Karrholm and colleagues, 1994b[42]

A computer program was used to randomly allocate the 64 patients (age range 58–66 years) to three study groups. The patients were stratified by age, sex, weight, bone quality and diagnosis (primary or non-inflammatory secondary osteoarthritis). The hips had the Ti-Fit femoral component inserted with a press-fit acetabular component by one of four surgeons in one of two hospitals. The femoral stems were inserted with either cement (n = 20), an HA-coating (n = 23) or were porous-coated (n = 21). Six 0.8 mm tantalum markers were inserted into the femoral component. After 2 years the cemented and porous-coated stems had subsided more than the HA-coated (p = 0.002 and 0.02, respectively). The HA-coated components also rotated less than the cemented stems (p = 0.03). The Harris Hip and Pain Scores did not differ significantly between the groups. Pain or discomfort in the thigh was reported in two cemented, five HA-coated and eight porous-coated prostheses (p = ?). There were no revisions within the 2 years. **C-rated.**

Kelley and colleagues, 1993

A total of 84 hips in 84 patients were randomly assigned a Harris Design 2 prosthesis either with (n = 44) or without (n = 40) a collar (method of randomisation not described). The operations were performed by two surgeons. After an average follow-up period of 4.6 years (range 2–7 years) 32 patients with collarless (mean age 70 years) and 38 with collared prostheses (mean age 68 years) were available for study (six patients died from unrelated causes). Patient diagnosis was mainly degenerative joint disease (79%). Postoperative pain and mobility levels were similar in the two groups as were the Harris Hip Scores. The amount of migration or radiolucency at the bone-cement interface did not differ significantly between them but the collar seemed to alter subsidence in the hips (mean 0.5 mm, as opposed to a mean of 2 mm for the collarless prostheses, $p < 0.05$). In all, 5% of collared and 9% of collarless prostheses required revision because of aseptic loosening of the femoral component (p = ?). **C-rated.**

Krismer and colleagues, 1994

In this study, migration was compared in two acetabular components – RM cup (HA-coated with a CLS stem) and PCA cup (porous-coated with a PCA stem). The patients (age range 50–65 years) were stratified by age and the two surgeons prior to randomisation and were followed for 5.2 years (mean). Diagnosis was primary or secondary osteoarthritis only. Almost one-third of the patients were lost or excluded from radiological assessment and almost half were not clinically evaluated. The *Standard System of Terminology for Reporting Results* (SSTRR) was used to compare clinical findings such as Pain (no/mild pain: RM 94%, PCA 97%; p = 0.05) and Limp Without Support (no/slight limp: RM 90%, PCA 55%; p = 0.02). Migration assessment found greater longitudinal movement in the PCA cup. As loosening was defined as

PAPER 3

> 1 mm longitudinal migration, 12% RM and 27% PCA cups were termed loose ($p = 0.08$). **C-rated.**

Marston and colleagues, 1996[107]

Random number tables were used to assign 213 patients to a Stanmore prosthesis and 200 to a Charnley hip. For the 53 bilateral operations the second hip was also randomised giving 14 patients with two different prostheses. The mean age of the 360 patients (126 men, 234 women) was 67 years and > 85% had primary osteoarthritis. Various surgeons at various grades performed the operations using three different approaches. The mean follow-up period was 6.5 years (range 5–10 years; 76 patients died and two were lost to follow-up) and the hips were reviewed by an independent observer. There were no differences in Harris scores between the groups either pre- or post-operatively. Three Stanmore and four Charnley stems were asymptomatically loose (the Stanmore had radiolucent lines > 2 mm around the component). Revision rates did not vary greatly between the prostheses (4.2% Stanmore, 3.5% Charnley). However, the relative risk of requiring revision was found to be 11.47 times greater for trainee compared with qualified surgeons. **C-rated.**

Olsson and colleagues, 1985[38]

A total of 119 patients had either a cemented (Charnley, n = 61) or uncemented (Honnart Patel-Garches, n = 59) prosthesis implanted. The mean ages of the patients were 67 years and 64 years, respectively; 82% had a diagnosis of osteoarthritis and 10% of rheumatoid arthritis (with the remainder miscellaneous). The number and grade of surgeons was not specified. Clinical evaluation showed similar preoperation results but the Charnley prosthesis performed better at the 1 year assessment – Harris Hip Score and Limp (Charnley vs. Honnart Patel-Garches) $p < 0.001$; maximal walking speed (Charnley vs. Honnart Patel-Garches) $p < 0.05$ (twice as many patients fitted with the latter device required an ambulatory device). A quantitative analysis of gait showed the Honnart Patel-Garches group to have slightly better preoperative results but 1 year after surgery the Charnley group showed greater improvement. No revisions were reported. **C-rated.**

Onsten and Carlsson, 1994

As a result of primary arthrosis, 60 patients (age range 40–70 years) had a unilateral hip replaced. A computer program randomised them to either a cemented, all polyethylene Charnley socket (n = 30) or a cementless, porous-coated Harris-Galante type 1 socket (n = 30). Any unstable screws were removed from the Harris-Galante prosthesis during the original operation (this did not affect subsequent results) and between five and ten 0.18 mm tantalum markers were inserted into the pelvis and socket of both types to aid movement analysis during the 2-year follow-up period. There were no revisions reported. Harris Hip and Pain Scores did not differ between the groups. There were no overall differences in either migration or rotation in any axis by 2 years but 15/27 (55.5%) Charnley and 28/30 (93.3%) Harris Galante sockets displayed "significant" movement at 2 years ($p = 0.001$). **C-rated.**

Onsten and colleagues, 1994[35]

Charnley femoral components were inserted bilaterally, under the same anaesthesia, into 29 patients with primary osteoarthritis by one of three surgeons. A Harris-Galante type 1 acetabular component was randomly implanted in one hip and a Charnley acetabular cemented component into the other. Tantalum balls (0.8 mm diameter) were inserted into the pelvis and acetabular cup during the operation. In all, 21 patients (42 hips) were studied (age range 41–76 years) for an average of 27 months. Five Charnley and three Harris-Galante cups did not move at all. The maximum migration (in any direction) was 1.7 and 2.1 mm, and maximum rotation was 2.2 and 2.0 degrees for the Harris-Galante and Charnley prostheses, respectively. There were no differences in the mean values of absolute migration between the groups in any direction ($p = 0.06$–0.98) but there were in the mean values of absolute rotation ($p = 0.08$–0.008) – Harris-Galante hips rotating the most. **C-rated.**

Reigstad and colleagues, 1986[49]

In all, 155 Müller and 149 ICLH prostheses were implanted into 231 patients (age range 60–79 years) by 13 surgeons. All patients in this study were diagnosed with osteoarthritis of the hip (excluding those with heavy bone loss in the femoral head or earlier fracture of the femoral neck) and had a mean follow-up of 48.5 months. No Müller hips were revised as opposed to 8.7% ICLH ($p < 0.001$), and, in addition, one component (0.6%) was loose compared to 12 (8%), respectively. Postoperatively, the Müller group had consistently higher scores than the ICLH group on all three modified Merle d'Aubigne and Postel parameters and total hip function. This reached a level of significance ($p < 0.001$) by 1 year in 3/4 parameters. **C-rated.**

Søballe and colleagues, 1993

Migration of titanium-coated femoral components and HA-coated stems in a Biometric prosthesis

were compared in this RCT. The same surgeon performed the surgery on all the patients (aged 48–68 years). The diagnosis was primary osteoarthritis (one patient had secondary osteoarthritis). Radiographic analysis was blinded. All components had migrated by 3 months but by 12 months the titanium-coated stems had migrated further than the HA-coated ($p = 0.02$), possibly indicating an increased risk of subsequent loosening and revision of the prosthesis. The HA-coated THR was also associated with higher Harris Hip Scores and less pain (measured by Visual Analogue Scale) at 12 months. A problem with the study was that small numbers of patients were involved – 12 titanium-coated and 14 HA-coated prostheses were available for clinical and conventional radiographic assessment, and eight and seven, respectively, for roentgen stereophotogrammetric analysis. **C-rated.**

Thanner and colleagues, 1995[108]

This comparison of two cement types – Boneloc and Palacos – involved 30 hips in 30 patients, age range 63–76 years, 27 of whom had primary osteoarthritis. Tantalum markers (0.8 mm) were inserted into the cup of the Spectrum EF prosthesis and the pelvis. Full radiostereometric analysis was possible in only 24 patients at 1 year (one Boneloc patient had died). Palacos fixed cups had 'a small' lateral migration while cups with Boneloc migrated medially ($p = 0.03$) and proximally ($p = 0.04$); 1/16 Palacos stems subsided 0.27 mm whilst 6/13 Boneloc stems subsided 0.22–1.0 mm ($p = 0.005$). Increased acetabular radiolucent lines and femoral 'relative cement–cortical bone contact' occurred in the Boneloc group compared with the Palacos group ($p = 0.04$ and $p = 0.03$, respectively). Harris Hip and Pain Scores and a Visual Analogue Scale for pain improved postoperatively ($p = 0.0004$–0.002) but did not differ between the groups (p, not significant). **C-rated.**

Wykman and colleagues, 1991[37]

A comparison of cemented (Charnley) and cementless press-fit (Honnart Patel-Garches) fixation in 150 patients; 15 in each group had bilateral arthroplasties (age range 29–82 years; diagnosis 77% osteoarthritis, 10% rheumatoid arthritis, 13% miscellaneous). The two prostheses had a similar probability of survival by 5–6 years approximately (Charnley 88%, Honnart Patel-Garches 82%; p, not significant). More revisions were required in the Honnart Patel-Garches group over 5 years (18.7%, all for loosening, all but one causing mid-thigh pain) compared to the Charnley group (11%, five for loosening, no mid-thigh pain). A further five Honnart Patel-Garches prostheses had possible need for revision due to midthigh pain (increasing the revision rate to 25%). Subsidence of more than 4 mm occurred in 5% of Charnley and 33% of Honnart Patel-Garches prostheses. **C-rated.**

PAPER 3

DATA TABLE 1 *RCTs included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
Bourne, *et al.*, 1995; Rorabeck, *et al.*, 1996 (Canada) (Previous reports: Rorabeck, *et al.*, 1994; Bourne, *et al.*, 1994; Mulliken, *et al.*, 1996)	RCT; cemented (i) vs. press-fit (ii) Mallory Head (i) cemented (ii) cementless (press-fit)	By, or supervised by, two senior surgeons Teaching hospital Direct lateral	(i) 124 in 124 patients (ii) 126 in 124 patients Radiographic data: 4.8 years (range 4–6); clinical data: 5 years 4-year follow-up: (i) 76 patients (ii) 71 patients	Mean age 65 years (at last follow-up; original range 18–75 years) Sex Diagnosis (primary or secondary osteoarthritis with exclusions)	Harris Hip Score (mean): (i) Preoperative, 43; 5 years, 96 (ii) Preoperative, 42; 5 years, 97. d'Aubigne Score: (i) Preoperative, 9; 5 years, 17.4 (ii) Preoperative, 9; 5 years, 17.5. WOMAC: (i) and (ii) Preoperative pain score, 6; 5 years, 1 (similar findings for other WOMAC dimensions). MACTAR Index: (i) Preoperative, 7.8; 5 years, 1 (ii) Preoperative, 7.7; 5 years, 0.67. Sickness Impact Profile – Global Physical Score: (i) Preoperative, 25.2; 5 years, 5.2 (ii) Preoperative, 23.3; 5 years, 5.0 (individual components gave similar results). Time trade-off: (i) Preoperative, 0.26; 5 years, 0.76 (ii) Preoperative, 0.30; 5 years, 0.61. 6-minute walk: (i) Preoperative, 227 m; 5 years, 392 m (ii) Preoperative, 229 m; 5 years, 409 m; *p*, not significant. Migration/subsidence: (i) 1 cup migrated, no subsidence (ii) 1 cup migrated; 10 (14%) stems subsided by 3–5 mm. Revisions/loosening: (i) 20 cups definitely/probably loose (26.3%); no stems loose; no revisions (ii) no revisions; no components loose.	C
Bradley & Lee, 1992 (UK)	RCT; HA (i) vs. cemented (ii) (i) Furlong (ii) Charnley	Not stated General hospital Antero-lateral	163 patients followed-up for maximum 2 years 1-year follow-up 139 patients 2-year follow-up 74 patients	Mean age 68 years (range, 45–75) Primary osteoarthritis (only those who were suitable for cementless prostheses)	Based on Harris Hip Scores: No pain: (i) 98% (ii) 96%; p = ? Absence of limp, walking aids used, walking distance, stairs, movement: uniformly good average results (p = ?). No revisions or radiographic evidence of loosening.	C
Carlsson, *et al.*, 1995 (Sweden)	RCT; all cemented (i) Charnley (no collar) (ii) Exeter (non-polished; no collar) (iii) Lubinus SP (collar) (iv) Harris Design 2 (collar) (v) Scanhip (collar)	Not stated	352 190 followed-up for 5 years, as follows: (i) 57 (ii) 58 (iii) 33 (iv) 16 (v) 26	Not specified	Resorption of the resected femoral neck: patients: (i) 19% (ii) 19% (iii) 42% (iv) 56% (v) 54%). (Odds ratio, 4.1; $p < 0.001$; more often seen in those with true collar. Amount of resorption is also larger ($p < 0.001$).)	C

continued

PAPER 3

DATA TABLE I contd *RCTs included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
Ciccotti, *et al.*, 1994 (USA)	RCT; HA vs. porous-coated Porous taperloc design: (i) HA-coated, 12-weeks post-operative weight bearing (PWB) (ii) porous-coated, 12-weeks PWB (iii) HA-coated, 6-weeks PWB (iv) porous-coated, 6-weeks PWB HA versus porous-coated	Grade not specified, but by or supervised by three MDs Teaching and general hospital Lateral with trochanteric osteotomy	60 in 60 patients 60 followed-up for 2 years	Mean age: (i) 66.3 years (ii) 66.2 years (iii) 63 years (iv) 63 years Weight Osteoarthritis	Charnley scores: Preoperative – no significant differences between groups for pain, function or motion; Postoperative – no significant difference in pain, function and motion at 2 years. Visual Analogue Scales for thigh pain: No significant difference between groups (2 years). Migration: No significant difference between groups – less than 1.40 mm for all groups. No revisions reported.	C
Godsiff, *et al.*, 1992 (UK)	RCT; cemented (i) vs. press-fit (ii) Ring Stem – (i) cemented (ii) cementless + cementless cup	Two consultants General hospital (?) Posterior	58 in 54 patients: (i) 30 (ii) 28 Followed-up for 2 years maximum (i) 23 (ii) 24	Age range 55–74 years; mean (i) 64.4 (ii) 64.5 Diagnosis (primary osteoarthritis of hip only) Sex	Authors' own 5-point scale: Pain and mobility (at 2 years) – (i) 15 pain free with no restriction (ii) 15 pain free with no restriction. Use of walking aids: (i) 22 (96%) did not use any aids (ii) 15 (62%) did not use any aids (p, 0.01–0.05). NB: At 3–5 years there were an unacceptable number of stem breakages; discontinued prosthesis use. (see Wilson, *et al.*, 1992 in RCT reference list).	A
Jacobsson, *et al.*, 1994 (Sweden) Unusual (Previous report: Jacobsson, *et al.*, 1993)	RCT; Isoelastic (i) vs. porous-coated (ii) (i) Butel stem (ii) PCA stem plus PCA, Ti-Fit or Romanus cup	Two senior surgeons Teaching hospital Dorsolateral	(i) 28 (28 patients) (ii) 28 (28 patients) Follow-up 3 years (i) 27 (ii) 26	Mean age 52 ± 8 years Sex Weight Radiographic appearance	Harris Hip Score: (i) +3 years, mean 78.5 (42–100) (ii) +3 years, mean 94.4 (59–100) p = ? Loosening: (i) 24/28 (86%) loose or possibly loose (ii) 5/28 (18%) loose or possibly loose p = ? Revisions: (i) 4 (14.3%), 3 due to loosening (ii) 3 (10.7%), all due to loosening.	C
Karrholm, *et al.*, 1994 (Sweden)	RCT; cemented (i) vs. HA-coated (ii) vs. porous-coated (iii) Ti-fit Press-fit cup plus: (i) cemented stem or (ii) HA-coated stem or (iii) porous-coated stem	Three MDs + one other (not specified) Teaching hospital Posterior or lateral	(i) 20 (ii) 23 (3) 21 in 60 patients Followed-up for 2 years 64 hips, 60 patients	Mean age (range): (i) 50 years (38–66) (ii) 56 years (38–63) (iii) 55 years (45–63) Sex Weight Diagnosis Bone quality index	Harris Hip Score (medians): (i) Preoperative, 43 (21-58); +2 years, 96 (64–100) (ii) Preoperative, 48 (14-67); +2 years, 95 (63–100) (iii) Preoperative, 39 (18-62); +2years, 94 (66–100) p, not significant. Harris Pain Score: (i) Preoperative, 10 (0–20); +2 years, 44 (30–44) (ii) Preoperative, 10 (0–30); +2 years, 40 (20–44) (iii) Preoperative, 10 (0–30); +2 years, 40 (20–44) p, not significant. 33 hips (31 patients), no or slight pain Thigh pain (pain or discomfort in the thigh): (i) 2; (ii) 5; (iii) 8; p, ? Migration of centre of stem (minima/maxima): (i) –0.8; –0.3 mm (ii) –1.2; –0.8 mm (iii) –2.7; –0.3 mm (i) & (iii) subsided more than (ii); (i) vs. (ii) p, 0.002; (ii) vs. (iii) p, 0.02. Rotation (median absolute value of anterior-posterior tilt about the transverse axis): (i) 0.4; (ii) 0.2; (iii) 0.4 degrees (i) vs. (ii) p, 0.03; (ii) vs. (iii) p, 0.07. No revisions.	C

continued

PAPER 3

DATA TABLE 1 contd *RCTs included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
Kelley, *et al.*, 1993 (USA)	RCT; cemented Harris Design 2 (i) with collar (ii) collarless	Two surgeons, grades not specified Not specified	84 in 84 patients Mean follow-up 4.6 years (range 2–7) 70 (i) 38, (ii) 32	Mean age (i) 67.7 years (± 7.7) (ii) 69.6 years (± 9.7) Sex Weight Previous surgery Diagnosis	Pain: (i) No/slight pain, 33/38 (87%) (ii) No/slight pain, 27/32 (84%) *p*, not significant. Activity patterns: (i) Able to ambulate unlimited distances, 17/38 (45%) (ii) Able to ambulate unlimited distances, 14/32 (44%) *p*, not significant. Harris Hip Score (mean): (i) 89.5; (ii) 86.7; ($p > 0.05$). Migration: no significant differences in movement or bone-cement interface radiolucency between groups. Subsidence: (i) Mean 0.5 mm (± 1.5); (ii) Mean 2 mm (± 4); ($p < 0.05$). Revisions/loosening: (i) 2 (5.3%) revisions due to loose stems (ii) 3 (9.4%) revisions due to loose stems *p*, ?	C
Krismer, *et al.*, 1994 (Austria)	RCT; HA (i) vs. porous-coated (ii) (i) RM cup (+ CLS stem) (ii) PCA cup (+ PCA stem)	Two 'experienced surgeons' Teaching hospital Anterolateral	(i) 61 (ii) 59 Follow-up: maximum 6 years; mean 5.2 years Clinical assessment: (i) 31 (51%) (ii) 33 (56%) Radiological assessment: (i) 42 (69%) (ii) 45 (76%)	Mean age (SD): (i) 58 years (4) (ii) 57 years (5) Sex Primary or secondary osteoarthritis	SSTRR system – Pain: (i) No/mild pain, 29 patients (94%) (ii) No/mild pain, 32 patients (97%) *p*, 0.05. SSTRR system – Limp without support: (i) No/slight limp, 28 patients (90%) (ii) No/slight limp, 18 patients (55%) *p*, 0.02. Migration: (i) Longitudinal: mean 0.05 mm; maximum mediolateral 0.10 mm; medial > 2 mm – 2 cups (ii) Longitudinal: mean 0.34 mm; maximum mediolateral 0.04 mm; medial > 2 mm – 2 cups. Revision/loosening: 3 revisions: 1 septic loosening, 2 CLS stems revised (i) 5 (12%) cups loose (ii) 12 (27%) cups loose *p*, 0.08.	C
Marston, *et al.*, 1996 (UK)	RCT; cemented (i) Stanmore (ii) Charnley	Various grades of surgeon General teaching unit 54% anterolateral; 40–42% McFarland-Osborne; 4–6% posterior	413 hips in 360 patients (i) 213 hips (ii) 200 hips (53 patients bilateral – 14 had two different prostheses) Follow-up, 5–10 years (mean 6.5) 413 hips (59 by questionnaire/interview, 78 patients by last clinic visit)	Mean age = 67 years (male range 18–91) (female range 30–87) Sex Diagnosis	Harris score: (i) Preoperative, 46.0; postoperative, 91.4 (ii) Preoperative, 46.0; postoperative, 91.2 *p*, not significant. Asymptomatic loosening (radiolucent lines > 2 mm): (i) 3 stems and corresponding cups (average subsidence for 57 stems at 7 years, 2.8 mm) (ii) 4 stems and corresponding cups (average subsidence for 51 stems at 7 years, 2.6 mm) *p*, not significant. Revisions (all for suspected loosening): (i) 9 (4.2%; 1 cup, 2 both, 5 stems, 1 exploration) (ii) 7 (3.5%; 5 stems, 1 both, 1 no details); relative risk of requiring revision 11.47 times greater for trainees (95% CI, 1.53–86.06); odds ratio for revision, 0.82 for (ii) vs. (i) (95% CI, 0.27–2.47).	C

continued

PAPER 3

DATA TABLE I contd *RCTs included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
Olsson, *et al.*, 1985 (Sweden)	RCT; cemented (i) vs. press-fit (ii) (i) Charnley (ii) Honnart Patel-Garches	Not specified General hospital (i) anterolateral with trochanteric osteotomy (ii) posterior without trochanteric osteotomy	(i) 61 in 61 patients (ii) 58 in 58 patients Follow-up, 1 year (i) 60–61 (ii) 55–58 (depends on assessment)	Mean age: (i) 67 years (ii) 64 years Sex Weight Height Diagnosis	Harris Hip Score: (i) Preoperative, 40; at 1 year, 89 (ii) Preoperative, 40; at 1 year, 78 Preoperative, *p*, not significant; at 1 year $p < 0.001$. Limp (moderate/severe): Preoperative: (i) 92%; (ii) 93%; *p*, not significant Postoperative: (i) 20%; (ii) 45%; $p < 0.001$. Maximal walking speed: (i) Preoperative, 94 cm/s; at 1 year, 124 cm/s (ii) Preoperative, 94 cm/s; at 1 year, 109 cm/s Preoperative, *p*, not significant; at 1 year, $p < 0.05$; twice as many in (ii) required ambulatory devices compared with (i). Quantitative Gait Analysis: Preoperative, no significant differences between groups, but (ii) consistently better results; at + 1 year, (i) had better results in all variables. No revisions reported.	C
Onsten & Carlsson, 1994 (Sweden)	RCT; cemented (i) vs. porous-coated (ii) (i) Charnley (ii) Harris-Galante type I	Not specified General hospital Supine with transtrochanteric incision	(i) 30 in 30 patients (ii) 30 in 30 patients Follow-up 2 years (i) 27 patients (ii) 30 patients	Age range 40–70 years *(i) mean 63 years (ii) mean 62 years	Harris Pain Score: (i) mean 42; (ii) mean 42. Harris Hip Score: (i) mean 91; (ii) mean 93. Migration/rotation: No overall differences between groups at 2 years. Number of sockets displaying significant movement at 2 years: (i) 15/27 (55.5%); (ii) 28/30 (93%); *p*, 0.001. No revisions necessary by 2 years.	C
Onsten, *et al.*, 1994 (Sweden)	RCT; hybrid (i) vs. cemented (ii) Charnley stem plus (i) Harris-Galante Type I cup or (ii) Charnley cup	Three MDs General hospital Lateral transtrochanteric	58 in 29 patients Follow-up 27 months (range, 23–49 months) 42 in 21 patients	Mean age 69 years (range, 41–76) Sex Primary osteoarthritis Weight	Charnley Score for both cup types: pain: mean 6 (4–6); walking: mean 6 (3–6); range of motion: mean 5 (2–6). Maximum migration: (i) 1.7 mm; (ii) 2.1 mm. Maximum rotation: (i) 2.2 degrees; (ii) 2.0 degrees; $p = ?$ (5 Charnley and 3 Harris-Galante cups did not move). Mean values of absolute migration did not differ between groups (*p*, 0.06–0.98). Mean values of absolute rotation for the cups around: transverse axis: (i) 0.7; (ii) 0.4; *p*, 0.008 sagittal axis: (i) 0.6; (ii) 0.4; *p*, 0.03 longitudinal axis: (i) 0.7; (ii) 0.4; *p*, 0.08.	C
Reigstad, *et al.*, 1986 (Norway)	RCT; cemented vs. resurfacing (i) Müller (ii) ICLH double-cap	13 surgeons, grades not specified Specialist hospital 14 ICLH – posterior-anterior; 135 ICLH and all Müller – anterolateral	313 in 231 patients (i) 155 (ii) 158 Mean follow-up 48.5 months (range 27–75) Immediately postoperation, 304 (231 patients) – (i) 155; (ii) 149 + 2 years, 296 + 5 years, 89		(i) Charnley modified d'Aubigne & Postel: Pain: (i) preoperative, 1.65; +2 years, 5.93; +5 years, 5.81 (ii) preoperative, 1.61; +2 years, 5.73*; +5 years, 5.53. Walking: (i) preoperative, 1.70; +2 years, 5.82; +5 years, 5.80 (ii) preoperative, 1.74; +2 years, 5.60**; +5 years, 5.50. Function: (i) preoperative, 6.82; +2 years, 16.41; +5 years, 16.41 (ii) preoperative, 6.74; +2 years, 15.85**; +5 years, 14.97#; *(i) vs. (ii): *p*, 0.001; **(i) vs. (ii): $p < 0.001$; #(i) vs. (ii): *p*, 0.014. Migration/subsidence: (i) cup: no migration; stem: 1 subsided 8 mm (ii) cup: no migration (except in 3 revised); stem: no subsidence in those not revised. Revision/loosening: (i) no revisions, 1 stem asymptomatically loose (ii) 13 (8.7%) revised [(i) vs. (ii): $p < 0.001$], 12 loose (12 stem, 8 cup).	C

continued

DATA TABLE 1 contd *RCTs included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
Søballe, *et al.*, 1993 (Denmark)	RCT; press-fit (i) vs. HA-coated (ii) Biometric with (i) Ti-alloy coating (ii) HA coating	One surgeon, grade not specified Teaching hospital Posterolateral	28 in 27 patients Follow-up at 1 year: clinical, 26 in 26 patients; roentgen stereographic analysis, 15 in 15 patients	Mean age (range) (i) 58.6 years (50–68) (ii) 56.8 years (48–63) Weight Charnley class	Harris Hip Score: Preoperative, (i) mean 48 (SEM 2.1); (ii) mean 56 (SEM 3.7); At 1 year, (i) 87 (SEM 3.9); (ii) 98 (SEM 0.8) *p*, ? Pain (Visual Analogue Scores): At rest: (mean (SEM)) (i) Preoperative, 5.9 (0.8); at 1 year, 2.0 (0.9) (ii) Preoperative, 5 (0.7); at 1 year, 0.42 (0.35) *p*, not significant. In function: (mean (SEM)) (i) Preoperative, 7.5 (0.67); at 1 year, 2.11 (0.8) (ii) Preoperative, 7.0 (0.5); at 1 year, 0.64 (0.35) *p*, not significant. Maximal total point motion (mm) or migration: (i) Mean 3.9 (SEM 0.8) (n = 8) (ii) Mean 1.7 (SEM 0.4) (n = 7) $p < 0.05$. Maximum subsidence in both groups – 0.2 mm. Calcar resorption – present equally in both.	C
Thanner, *et al.*, 1995 (Sweden) Unusual	RCT: cemented Spectron EF with (i)'Palacos' cement (ii)'Boneloc' cement	Grade of surgeon not specified General hospital Transgluteal lateral	30 in 30 patients Follow-up at 1 year 29 in 29 patients	Mean age 71 years (range, 63–76) Sex Diagnosis	Harris Hip Score – median (range): (i) Preoperative, 51 (24–70); at 1 year, 90 (56–97) (ii) Preoperative, 45 (22–61); at 1 year, 93 (65–99); *p*, not significant. Harris Pain Score – median (range): (i) Preoperative, 20 (10–30); at 1 year, 40 (20–44) (ii) Preoperative, 20 (10–20); at 1 year, 40 (30–44) *p*, not significant. Pain Visual Analogue Scale (mm) – median (range): (i) Preoperative, 67 (50–99); at 1 year, 6 (0–50) (ii) Preoperative, 66 (25–100); at 1 year, 3 (0–37) *p*, not significant. Harris scores and Visual Analogue Scale pre- vs. postoperative: *p*, 0.0004–0.002. Migration/subsidence: (i) Cup: small lateral migration (no value); stem: 1/16 (6.25%) subsided by 0.18 mm (ii) Cup: migrated medially and proximally (no values); stem: 6/13 (46%) subsided by 0.22–1.0 mm; Stem: (i) vs. (ii) for all observations 6 weeks–12 months: *p*, 0.005. Radiolucent lines: Cup: (i) average 11% (0–47); (ii) average 30% (0–50) of acetabular cup circumference involved (*p*, 0.04); Stem (increased relative cement-cortical bone contact): (i) median 41% (9–59); (ii) median 50% (33–65); *p*, 0.03. No revisions or loosening.	C
Wykman, *et al.*, 1991 (Sweden) (Previous reports: Wykman & Goldie, 1984; 1988)	RCT; cemented (i) vs. press-fit (ii) (i) Charnley (ii) Honnart Patel-Garches	"Experienced surgeons" General hospital? (i) Lateral with trochanteric osteotomy (ii) Posterolateral without trochanteric osteotomy	(i) 90 hips in 75 patients(?) (ii) 90 hips in 75 patients(?) Follow-up 3–5 years (i) 68 patients (ii) 70 patients	Mean age (range) (i) 67.4 years (48–82) (ii) 64.8 years (29–82) Sex Weight Diagnosis	Harris Hip Score (medians): (i) Preoperative, 37.3; 3–5 years, 95.3; 54 patients (79%) good/excellent; (ii) Preoperative, 38.1; 3–5 years, 88.7; 48 patients (70%) good/excellent; *p*, not significant. Harris Pain Score (mean): (i) Preoperative, 13.5; 3–5 years, 96% patients slight or no pain (ii) Preoperative, 14.7; 3–5 years, 86% patients slight or no pain; *p*, not significant. Mid-thigh pain: (i) none; (ii) 6 months, 46/72 (64%) had pain; > 3 years, 18/46 still in pain, all have been/will be revised. Calcar resorption > 2 mm: (i) 38%; (ii) 58%; *p*, ? Migration/subsidence: (i) cup: 59% had radiolucent zone > 2 mm between bone and cement; stem: 16% subsided > 1 mm, 5% > 4 mm (radiolucent zone > 2 mm = 16% between bone and cement); (ii) cup: 67% had non-continuous radiolucent zone between implant and bone, 9% > 2 mm; stem: 66% subsided > 1 mm, 33% > 4 mm (all had non-continuous radiolucent zone between stem and bone). "Failure events", 22 patients (15%): (i) 8 (11%), 5 due to loosening; (ii) 14 (19%), all due to loosening. Survival analysis (at 5–6 years approximately): (i) 88%; (ii) 82%.	C

PAPER 3

DATA TABLE 2 *Non-controlled comparative studies included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
Abrahams & Crothers, 1992 (USA)	Prospective Omnifit-HA (i) Press-fit (ii) Hydroxapatite-coated	Not specified Teaching and general hospital Not specified	98 in 89 patients Follow-up 1 year (i) 35 in 31 patients (ii) 63 in 58 patients	(i) 56 years (22–75) (ii) 53 years (21–73) Sex Diagnosis	Radiolucent line formation: (i) More frequent proximal line formation: (i) 20–25.7% vs. (ii) 3.2%; $p < 0.02$ (ii) More frequent distal line formation: (ii) 74.3–82.9% vs. (i) 40–42.9%; $p < 0.001$. Heterotopic bone formation: (i) No formation 39.4%; (ii) No formation, 58.%; *p*, not significant. Calcar resorption (% of cases): (i) 58.8%; (ii) 5.7%; *p*, not significant. Stem subsidence (% of cases): (i) 14.3%; (ii) 0%; *p*, not significant. Revisions: none reported.	C
Ahnfelt, *et al.*, 1990 (Sweden)	Retrospective (i) Christiansen (ii) Charnley (iii) Brunswick (iv) Lubinus IP (v) Charnley-Müller (vi) McKee-Farrar (vii) Müller, curved (viii) Wagner (ix) Stanmore (x) Müller, straight (xi) CAD (xii) Exeter (xiii) Richards II (xiv) McKee-Arden (xv) Lubinus SP plus more (not stated) Cemented: i, ii, iii, iv, v, vi, vii, ix, x, xi, xii, xv Resurfacing: viii Unknown, probably cemented: xiii, xiv	Not specified (various) Not specified (various) Not specified	Total number of reoperations, 7772; number of first reoperations, 6386; number of first revisions, 4664 Follow-up – up to 10 years, 4664 (i) 1365 (ii) 971 (iii) 483 (iv) 428 (v) 288 (vi) 250 (vii) 214 (viii) 149 (xi) 101 (x) 57 (xi) 48 (xii) 38 (xiii) 34 (xiv) 33 (xv) 18 + others	Approx. median: women 64 years; men 66 years Sex Diagnosis	Aseptic loosening main cause of both reoperation (54%) and revision (74%). Observed survival without revision for loosening: (i) 63% at 10 years (ii) 92% at 10 years (iii) not specified (iv) 93% at 9 years (v) 85% at 10 years (vi) not specified (vii) 84% at 10 years (viii) 28% at 10 years (ix) 89% at 10 years (x) 95% at 6 years (xi) 95% at 8 years (xii) 95% at 5 years (xiii) not specified (xiv) not specified (xv) not specified. Osteo/rheumatoid arthritis: Significant increase in cup loosening in rheumatoid arthritis patients > 65 years compared with osteo-arthritis ($p = ?$). Osteoarthritis: men had more loosening than women in all age groups, with the 55–64 years group having the highest risk for revision for men; women have decreasing risk of loosening with increasing age ($p < 0.001$). Rheumatoid arthritis: younger, more active patients have an increased risk of revision due to loosening. Hospitals: regional (secondary) hospitals had better results than university (tertiary) or community (primary) hospitals with respect to loosening rates ($p < 0.001$).	C
Bankston, *et al.*, 1993 (USA)	Retrospective – matched (i) T-28 (stainless steel) (ii) TR-28 (cobalt–chrome) (iii) MOSC (titanium) (cup: all-polyethylene, non-metal backed) Cementing: early technique, (i) & (ii); modern technique, (iii) Cemented	1 surgeon (grade not specified) Specialist hospital Lateral with trochanteric osteotomy	Total number, 568 (i) 307 (ii) 162 (iii) 99 Follow-up: (i) 8.0 years (ii) 7.6 years (iii) 7.9 years Matched = 231 (77 in each group)	(i) 66 years (ii) 67 years (iii) 65 years Sex Weight	Linear wear rates: (i) 0.06 mm/y; (ii) 0.05 mm/y; (iii) 0.08 mm/y; *p*, not significant. Volumetric wear rates: (i) 34.76 mm^3/y; (ii) 33.72 mm^3/y; (iii) 46.14 mm^3/y; *p*, not significant. Acetabular progressive radiolucencies in at least one zone: (i) 24.7%; (ii) 11.7%; (iii) 26.0%; (ii) < (i), *p*, 0.04; (ii) < (iii), *p*, 0.005. Incidence of complete: (i) 11.7%; (ii) 6.5%; (iii) 3.9%; between groups, *p*, not significant. Femoral osteolysis/subsidence (> 5 mm): (i) Osteolysis: 3.9%; subsid: 19.5% (ii) Osteolysis: 5.2%; subsidence: 20.8% (iii) Osteolysis: 5.2%; subsidence: 5.2%. Osteolysis: *p*, not significant. Subsidence: (iii) < (i), *p*, 0.008; (iii) < (ii), *p*, –0.004.	C

continued

DATA TABLE 2 contd *Non-controlled comparative studies included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
Bankston, *et al.*, 1995 (USA)	Retrospective (i) T-28 (cup molded, early cementing technique) (ii) Triad (cup machined, modern cementing techniques) Cemented	Two surgeons, grade not specified Hospital not specified Lateral with trochanteric osteotomy	(i) 162 in 151 patients (ii) 74 in 60 patients Follow-up: (i) 6.9 years (54 patients) (ii) 6.4 years (54 patients)	(i) 67 years (ii) 65 years Sex Weight	Wear rate: (i) 0.05 mm/y, (ii) 0.12 mm/y; $p < 0.001$. Migration/subsidence: – complete progressive radiolucencies: (i) 13.0%; (ii) 5.4%; *p*, 0.75; stem subsidence: (i) 13.6%; (ii) 4.1%; *p*, 0.03.	C
Bertin, *et al.*, 1985 (UK & Switzerland)	Prospective Pegged polyethylene prostheses designed by surgeon concerned: (i) Morscher (ii) Ring (iii) Freeman Press-fit	By or supervised by three senior surgeons University or general hospital (i) anterolateral (ii) posterolateral (iii) anterolateral	1878 (i) 788 (ii) 967 (iii) 123 Follow-up: 2 years (6 months–6 years) Total, 1724	Age range (years) 20s–90s Sex Diagnosis	No differences between results from each centre so combined results reported. Pain: 82.3% (1431), none; 14.3% (248), mild; 2.7% (47), moderate; 0.7% (15), severe. Activity level: 86.4% (1503), normal; 10.9% (190), good; 2.1% (37), fair; 0.5% (9), poor. Range of movement: 90.8% (1579), ≥ 90° flexion; 7.9% (139), 60–89°; 1.2% (20), 30–59°; 0.1% (1), 0–29°. Migration: no cup migrated more than 5 mm. Revisions/loosening: 18 (1.03%) revised; 10 due to stem loosening (2 cups revised at same time), 1 due to traumatic cup loosening.	C
Britton, *et al.*, 1996 (UK)	Prospective (i) Charnley (ii) Stanmore Cemented	By, or supervised by, one consultant surgeon Specialist or teaching hospital Posterior (Southern)	1190 (i) 208; (ii) 982 Follow-up: Median 8 years (40 hips with 16 years' follow-up) 834 patients (70%)	Not specified	Revisions/loosening: 81/1190 (7%) revised, 38 due to aseptic loosening. No significant difference in cause of failure for different implants ($p > 0.5$). Survival rate at 10 years: (i) 84% ± 6.3 for a 'revision' end-point (n = 107); 44% ± 8.7 for 'onset of slight pain' end-point; (ii) 93% ± 2.6 for a revision end-point (n = 332.5); 48% ± 4.9 for onset of slight pain end-point. Survivorship curves: similar for both (i) and (ii) up to 8 years; after this (i) significantly worse for 4/5 end-points (revision or onset of different levels of pain), *p*, 0.026–0.004. Cementing techniques: (ii) 1st generation, 1973–79 (n = 560) – 10-year survival without revision 91.6%; 2nd generation, 1979–86 (n = 422) – 10-year survival without revision 97.4%; *p*, 0.005. (i) All hips (88% pre-1977) vs. (ii) n = 280 (1973–77) – 10-year survival (no revision), (i) 79.1%, (ii) 86.3%; *p*, 0.07.	C
Burkart, *et al.*, 1993 (USA) (Some information from Bourne, *et al.*, 1994)	Prospective (i) Mallory Head – uncemented (ii) PCA (i) Press-fit (ii) Porous	Two senior surgeons (or under their supervision) Teaching hospital Direct lateral (Hardinge)	(i) 105 (100 patients) (ii) 110 (103 patients) Follow-up: 2 years (i) 94 (89.5%) (ii) ?	(i) 65 years (40–85) (ii) 61 years (26–83) Sex Side of body Osteoarthritis	Thigh pain: (i) 3% – 3 (patients (2 mild, 1 moderate); none required analgesic, none had pain at 1 year; (ii) 23%: 3 with severe pain (no further details, previously reported). Average Harris scores: (i) Thigh pain group: hip 88, pain 38; no thigh pain group: hip 96, pain 43; (ii) No details, previously reported. Radiographic analysis: (i) Positioning – neutral 63%, mildly valgus 10%, mildly varus 28%; fit – no patient had good metaphyseal fit, 27% had good isthmal fit; subsidence: 10 patients (11%); 8 patients, 0–6 months (6 patients 3–5 mm, 2 patients 6–8 mm); 1 patient, 6–12 months by 6–8 mm; 1 with thigh pain, 3–5 mm by 6 months; calcar changes: common in those with thigh pain (100%) and without (84%); cortical hypertrophy and cancellous hypertrophy uncommon in both subgroups. (ii) Positive correlations between thigh pain and following features made (reported in more detail previously): (a) tight diaphyseal fit through the isthmus; (b) subsidence > 2 mm; (c) periosteal cortical hypertrophy at stem tip; (d) cancellous hypertrophy at stem tip.	C

continued

PAPER 3

DATA TABLE 2 contd *Non-controlled comparative studies included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
Callaghan, *et al.*, 1995 (USA)	Retrospective (a) Effect of cup design (i) Charnley (ii) Charnley in patients < 50 years (iii) cemented IOWA (2nd generation cementing technique) (iv) Harris-Galante I cup + precoated cemented stem IOWA (v) PCA cementless (vi) hybrid Harris-Galante I with precoated IOWA revisions (b) Wear rates (i) Charnley 22 mm machined polyethylene (ii) Charnley 22 mm molded polyethylene (iii) all-polyethylene 28 mm cemented (iv) TiBac 28 mm metal backed cemented (v) Harris-Galante I 28 mm cementless, metal backed Cemented, hybrid, porous	Not specified	(a) n = 897 (i) 330 (> 20 years) (ii) 89 (16–22 years) (iii) 187 (> 10 years) (iv) 130 (5 years) (v) 100 (> 7 years) (vi) 61 (> 5 years) (b) n = 210 (i) ? (20 years) (ii) ? (15 years) (iii) ? (10 years) (iv) ? (7 years) (v) ? (5 years)	Not specified	(a) Effect of cup design: (i) cup revision incidence 10.6% – definitely/probably loose 12.8%; stem revision incidence 3.2% – definitely/probably loose 4.3%; (ii) cup revision incidence 13% – definitely/probably loose 37%; stem revision incidence 2.2% – definitely/probably loose 6.1%; (iii) cup loosening 24.5% (metal-backed 17%, all-polyethylene 30%); stem loosening 1.2%; (iv) no revisions or loosening; (v) cup revision incidence 4% (migration incidence 5% – includes 2 revised cases); (vi) no revisions (1 migration); (b) Wear rates: Less wear in Harris-Galante I component (28 mm head) than in other cohorts (no details); *p*, ?	C
Carlsson & Gentz, 1982 (Sweden)	Retrospective Revisions of: (i) Charnley (ii) Brunswik (iii) Christiansen (iv) Lubinus (v) others Cemented	Not specified	183 Follow-up: 54 months (range, 2–158) 100 (i) 45; (ii) 38; (iii) 6; (iv) 3; (vi) 8	Not specified	100 revisions performed, 87 due to suspected loosening (7 had no radiolucent lines around the socket and were stable at surgery). Charnley radiographic classification (modified by author): Grade 1 – total number sockets 7; number of loose sockets 0 (0%); Grade 2 – total 31; loose 4 (13%); Grade 3 – total 28; loose 4 (14%); Grade 4 – total 34; loose 22 (65%); (or Grade 4 – total 7, loose 4 (57%); and Grade 5 – total 27, loose 18 (67%)). 12/34 (35%) sockets had obvious migration or change in position on the radiographs but were stable at surgery.	C

continued

PAPER 3

DATA TABLE 2 contd *Non-controlled comparative studies included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
Chmell, *et al.*, 1995 (USA)	Retrospective (i) Aufranc-Turner (ii) T-28 (iii) Osteonics nonmodular stem/ cemented cup (iv) Osteonics modular stem/ cemented cup (v) DePuy Profile modular stem/ ACS modular cup (vi) Osteonics modular stem/ cup, Dupuy Dura-loc modular cup (i) Cemented (ii) Cemented (iii) Cemented (iv) Modular (v) Modular (vi) Modular	Three MDs General hospital Not specified	(i) 778 (ii) 823 (iii) 329 (iv) 233 (v) 203 (vi) ? Follow-up: (i) Average 12 years (ii) ? (58 patients, 10–14 years) (iii) average 7.5 years? (range 5–8) (iv) 6 years (v) minimum 5 years (vi) ? (i) 336 (ii) ? (subgroup 58) (iii) 329 (iv) 233 (v) 125 (vi) ?	Not specified	(i) 22% rate of revision for aseptic loosening, most after 6 years; loosening due to progressive bone-cement radiolucencies. In absence of loosening, bone loss or osteolysis not seen. (ii) Stem loosening greater with 1st generation cementing techniques than 2nd but osteolysis not seen in either group unless loosening occurred. In 58 patients (follow-up 10–14 years), 3.4% cups revised for loosening, 21% had continous radiolucencies but no osteolysis apparent. (iii) 28 mm head group: revision for loosening 2.8%, radiographic loosening 22%, osteolysis 0% after 5–8 years. For all: revision for loosening 2.1%; average time to revision 91 months. (iv) Revision for loosening 3.0%; average time to revision 71 months. (v) 75/125 had polyethylene thickness < 6 mm; 13 revised for liner wear or fracture (average 41 months), 19 with eccentric wear, 15 with osteolysis. Remaining 50/125 had liner > 6 mm; 2 revised for liner fracture, 5 for eccentric wear, 4 were osteolytic. (vi) "Have not been associated with the catastrophic failure rate seen in the ACS cups" – no further details given.	C
Cornell & Ranawat, 1986 (USA) Unusual	Retrospective (i) Charnley, CAD, Müller, T-28 (using early cementing techniques) (ii) Charnley, DF-80 (using modern cementing techniques) Cemented	Not specified Specialist hospital Posterior (38%) or trans-trochanteric (62%)	101 in 85 patients Follow-up: (i) vs. (ii) = 5 years (i) subgroup analysis 5 years vs. 10.7 years 78 (62 patients) (i) 62 (48 patients) (ii) 16 (14 patients)	(i) 48 ± 7.6 years (ii) 48 ± 9.4 years Sex Weight Diagnosis	Modified d'Aubigne–Postel scores: All patients (both groups) – excellent/good results. Cups with radiolucent lines at 5 years: (i) 71%, mean cup radiolucent score 1.15 ± 1.73; (ii) 60%, mean cup radiolucent score 0.19 ± 0.25; $p < 0.025$. Radiolucency, 10 years vs. 5 years: (i) mean 1.6:1.15 mm. Revisions/loosening: (i) no revisions, 3 cups (2 patients) loose by 10 years; (ii) no revisions, no loosening by 5 years.	C
Dall, *et al.*, 1993 (South Africa) Unusual	Retrospective Charnley (i) 1st generation design (ii) 2nd generation Cemented	Consultant, 97%; residents, 3% Specialist and general hospital Not specified (> 95% trochanteric osteotomy)	1309 in 1089 patients Follow-up: (i) 8.8 years; (ii) 7.8 years 666 (555 patients) (i) 264; (ii) 402	(i) 60.7 years (ii) 60.3 years Sex Diagnosis Charnley class	Modified d'Aubigne–Postel (grades 5–6): (i) pain 82.7%; function 73.4%; motion 75.0%; (ii) pain 83.6%; function 77.1%; motion 81.8%. p, ? (unrevised hips only; function: Charnley class C excluded). Wear (cup): (i) 1–2 mm, 15.7%; 3–4 mm, 4.2%; (ii) 1–2 mm, 11.7%; 3–4 mm, 2.5%. p, ? Resorption (stem): (i) 1–2 mm, 3.8%; 3–4 mm, 13.3%; 5 mm+, 9.4%; (ii) 1–2 mm, 4.1%; 3–4 mm, 6.2%; 5mm+, 7.4%. p, ? Radiolucency (stem): (i) 13.2%; maximum width: 2 mm, 2.9%; 3 mm+, 3.4%. (ii) 20.4%; maximum width: 2 mm, 8.1%; 3 mm+, 5.0%; p, ? Radiolucency (cup): (i) 49.4%; maximum width: 2 mm, 8.2%; 3 mm+, 2.5%. (ii) 50.7%; maximum width: 2 mm, 5.8%; 3 mm+, 3.6%; p, ? Migration/subsidence: stem: (i) 2 mm, 5.0%; 3 mm+, 7.1%; (ii) 2 mm, 2.3%; 3 mm+, 6.8%; cup: (i) 1–2 mm, 2.5%, 3 mm+, 2.0%; (ii) 1–2 mm, 2.0%; 3 mm+, 2.1%. p, ? Probable loosening (no revisions): (i) 9/264 (3.4%): cup 1.9%, stem 1.6%, both 0%; (ii) 27/402 (6.7%): cup 2%, stem 3.7%, both 1%. 10-year survival probability related to loosening: (i) 99.35%; (ii) 86.8%; (95% CI, 80.9–92.8; $p < 0.0001$). Revisions: (i) 21 (8%). Loosening: cup only 1; stem only 2; both 2; (ii) 38 (9%). Loosening: cup only 2; stem only 18; both 9.	C

continued

DATA TABLE 2 contd *Non-controlled comparative studies included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
Duck & Mylod, 1992 (USA)	Retrospective Various types of prosthesis used but a 'significant percentage' were: (i) AML Porcoat (ii) Dual-Lock (iii) TARA & Indiana Conservative Hip (i) Porous-coated (ii) Cemented (iii) Resurfacing	Not specified Teaching hospital Anterolateral 41%; posterolateral 38%; anterior 14%; trans-trochanteric 7%	66 in 55 patients Follow-up: 36 months (i) 39 (ii) 18 (iii) 9	60.3 years (33–76) Sex Diagnosis Side of body	Occurrence of heterotopic bone formation (HBF): Total hip: cemented, 15/22 (68%); noncemented, 10/17 (59%). Hemi-arthroplasty: cemented, 9/13 (69%); noncemented, 2/5 (40%). Resurfacing: 5/9 (56%); (11/17 (65%) revision cases had HBF). No significant correlation between type of procedure and % bone formation. Pain and HBF, no correlation; range of movement and HBF, trend of decreasing range of motion with increasing HBF.	C
Ebramzadeh, *et al.*, 1994 (USA)	Retrospective (i) Charnley stem (ii) STH stem plus unknown cup Cemented	One MD Teaching hospital (i) Lateral with trochanteric osteotomy (ii) Posterior without trochanteric osteotomy	857 in 720 patients Follow-up: 9 years (1–21) 836 (i) 413 (ii) 423	< 50 years (i) n = 67 (ii) n = 61 > 50 years (i) n = 346 (ii) n = 362 Sex Weight Diagnosis	Best results if: (a) > 2 mm and < 5 mm proximal medial thickness of cement mantle; (b) < 2 mm proximal medial thickness of cancellous bone; (c) stem filled more than half of distal part of medullary canal; (d) stem in neutral orientation. Worst results if: (a) cement mantle > 10 mm thick; (b) > 2 mm of proximal medial thickness of cancellous bone; (c) stem filled half or less of medullary canal; (d) varus orientation.	C
Espehaug, *et al.*, 1995 (Norway)	Retrospective Cup/stem (i) Charnley/Charnley (ii) Exeter/Exeter (polished) (iii) Titan/Titan (iv) Spectron/ITH (v) Elite/Charnley (vi) Spectron/Lubinus SP (vii) Biomet/Biomet (viii) Spectron/Biofit (ix) Lubinus SP/Lubinus SP (x) Müller type/Müller type Cemented	Not specified Various hospitals Not specified	Total number 18,848; after restrictions, 12,179 in 11,169 patients) Follow-up: mean 3.2 years, maximum 6.4 years 12,179 (i) 6694 (ii) 1665 (iii) 1333 (iv) 1034 (v) 507 (6) 302 (vii) 247 (viii) 152 (ix) 129 (x) 116	< 65 years, 15% 65–74 years 49% > 74 years 36% Sex	Survival analysis: 5-year failure rate (overall 2.5%) (i) Charnley/Charnley, 2.86% (ii) Exeter/Exeter, 2.15% (iii) Titan/Titan, 1.23% (iv) Spectron/ITH, 0.85% (v) Elite/Charnley, 9.84% (vi) Spectron/Lubinus SP, 4.96% (vii) Biomet/Biomet, 1.25% (viii) Spectron/Biofit, no revisions (ix) Lubinus SP/Lubinus SP, no revisions (x) Müller type/Müller type, 7.33%. Revisions/loosening: (i) 115 revised (1.7%), 63% due to loosening (ii) 23 revised (1.4%), 48% due to loosening (iii) 12 revised (0.9%), 83% due to loosening (iv) 4 revised (0.4%), 25% due to loosening (v) 12 revised (2.4), 25% due to loosening. Other combinations, 18 revised (1.9%), 100% loose. Bilateral vs. unilateral. Results of survival analysis similar in both.	C
Freeman & Plante-Bordeneuve, 1994 (UK)	Prospective 'Specially designed' Parts 1 & 2 (i) Press-fit (ii) Cemented Part 3 (i) Press-fit (ii) Cement (iii) Press-fit with proximal longi-tudinal ridges (iv) HA-coated	Not specified Teaching Not specified	(i) 125 in 117 patients (ii) 81 in 77 patients (iii) ? (iv) ? Follow-up: (i) & (ii), > 5 years (iii) & (iv), 2 years (i) At 2 years, 100 in 93 patients; at > 5 years, 89 in 81 patients (ii) At 2 years, 55 in 54 patients; at > 5 years, 38 in 37 patients (iii) At 2 years, 41 in 38 patients; (iv) At 2 years, 34 in 34 patients	(i) 54 years (22–84) (ii) 67 years (48–83) At 2 years: (iii) 51 years (27–73) (iv) 52 years (33–76) Sex Diagnosis Side of body	Loosening (thigh pain needing analgesic or revision): (i) 22/89 (24.7%); (ii) 3/38 (7.9%); *p*, ? Migration: Rate of migration at > 5 year follow-up – (i) 0.78 mm/y; (ii) 0.27 mm/y; $p < 0.0001$; (i) loose hips 1.5 mm/y; stable hips 0.6 mm/y; $p < 0.0001$; (ii) loose hips 3.4 mm/y; stable hips 0.2 mm/y; *p*, ? Amount of migration at 2-year follow-up – (i) all (n = 100) 1.85 mm; no pain (n = 80) 1.45 mm; (ii) all (n = 55) 0.55 mm; no pain (n = 52) 0.38 mm; (iii) all (n = 41) 1.7 mm; no pain (n = 36) 1.3 mm; (iv) all (n = 34) 0.4 mm; no pain (n = 34) 0.4 mm; *p*, ? 'Migration test': Migration rate of 1.2 mm/y had 78% sensitivity and 86% specificity for distinguishing hips which would fail. (Group 4 at 4 years: no hips termed loose or revised; techniques used in groups 1 and 3 were discontinued.)	C

continued

DATA TABLE 2 contd *Non-controlled comparative studies included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
Goetz, *et al.*, 1994 (USA)	Retrospective – matched Cup: Harris-Galante (cementless) Stem: (i) Harris-Galante (cementless) (size: 28 mm (23), 26 mm (16), 22 mm (2)) (ii) Precoat (cemented) (size: 32 mm (2), 28 mm (9), 26 mm (28), 22 mm (2)) (i) Porous-coated (ii) Hybrid	One senior surgeon Teaching/ general hospital Not specified, but same for all patients)	Total 255: (i) 88 (ii) 167 Selected 82 in 74 patients Follow-up (range): (i) 74 months (43–100) (ii) 72 months (48–94) (i) 41 (ii) 41	(i) 57 years (40–69) (ii) 61 years (40–71) Weight Diagnosis	Harris Hip Score: Preoperative: (i) 49 (33–66); (ii) 53 (32–70); *p*, ?; Latest follow-up: (i) 89 (40–100); (ii)97 (84–100); $p < 0.002$. Osteolysis: (i) stem: 12/41 (29%) (5 loose); (ii) stem: 0; $p < 0.0002$. No relationship between size of femoral head and osteolysis. Revisions/loosening: (i) stem: 5 (12%) revised (4 due to loosening); 8 subsided/migrated; cup: no migration/revisions. (ii) stem: no revisions ($p < 0.02$), all radiographically stable, no radiolucent lines; cup: no migration/ revisions.	C
Hamada, *et al.*, 1993 (Japan)	Retrospective (i) Model Y (ii) Model Y2 Press-fit	Not specified Teaching hospital Not specified	71 in 71 patients Average follow-up (maximum): (i) 4 years 7 months (6 years 9 months) (ii) 1 year 8 months (2 years 9 months) (i) 26 (ii) 25	(i)65 years (43–81) (ii)61 years (40–81) Sex Diagnosis	Extent of press-fit (contact ratio) 1 year postoperatively: (i) Excellent 7; Good 8; Fair 11; (ii) Excellent 23; Good 2; Fair 0; *p*, ? Thigh pain: (i) 11/26 (42%) patients with pain for 1–6 months and 6–24 months postoperatively; (ii) 2/25 (8%) patients with pain for 1–3 months and 1–8 months postoperatively; *p*, ?	C
Havelin, *et al.*, 1994 (Norway)	Retrospective (i) Cemented (27 cup and 22 stem types) (ii) Cementless (19 cup and 18 stem types including: smooth-surfaced, porous- and HA-coated) Cemented Press-fit Porous-coated HA-coated	Various grades of surgeon Various hospitals Not specified	15,335 (i) 14,009 (ii) 1326 Follow-up: 0–5.4 years 15,335	(i) mean 71 years (ii) mean 59 years Uncemented: < 65 years 31% > 65 years 3% Sex	Aseptic loosening (cumulative survival until revision due to loosening) caused 68% of 263 failures. (i) Cup 99.4%, stem 98.3%, after 4.5 years (ii) Cup 98.4%, stem 96.1% after 4.5 years; 2.3 times more likely than (i) to need revision because of loosening. Revisions after 4.5 years: (i) All hips 2.7%; < 65 years 3.3%; women 1.9%; men 4.5%; (ii) All hips 6.5%; < 65 years 7.9%; women 6.3%; men 6.8%. Risk of revision: **Uncemented hips** at 2.0 times higher risk than cemented when adjusted for sex and age; increase in risk for patients aged < 60 years with uncemented hips is 2.9 compared with 2.4 and 1.2 in those aged 60–64 and > 65 years, respectively.	C
Havelin, *et al.*, 1995 (Norway)	Retrospective Stem: (i) Biofit (ii) Corail (iii) Femora (iv) Harris-Galante (v) LMT (vi) RM-prosthesis (vii) Profile (viii) Zweymuller Various	Not specified – various Various hospitals Not specified	2907 in 2421 patients Follow-up range: 0–5.4 years 2907	Range 15–87 years Mean range 48–63 years Sex Diagnosis	Revision rates for aseptic loosening (overall 4.5%) and cumulative survival after 4.5 years: (i) 18.6% & 81.4%; (ii) < 1% & 99.5%; (iii) 13.6% & 86.4%; (iv) 3.6% & 96.2%; (v) < 1% & 99.5%; (vi) 5.6% & 94.4%; (vii) 0% & –; (viii) < 1% & 99.1%.	C

continued

DATA TABLE 2 contd *Non-controlled comparative studies included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
Hearn, *et al.*, 1995 (USA)	Retrospective (i) Cemented Charnley, Dual Lock, Pennsylvania Total Hip (ii) Uncemented Trilock, Taperloc (i) Cemented (ii) Porous	Not specified	72 in 36 patients Most recent visit: (i) 8.1 years (2.7–18.2) (ii) 3.0 years (2.0–5.9) At same interval: (i) 3.6 (ii) 3.0 60 in 30 patients	(i) 59 years (21–76) (ii) 63 years (25–82) Sex Diagnosis	Charnley Scores: (i) Preoperative – pain 3.1; motion 3.0; function 2.5; most recent – pain 5.6; motion 5.5; function 5.3; same interval – pain 5.7; motion 5.1; function 5.3. (ii) Preoperative – pain 2.5; motion 3.6; function 2.5; most recent/same interval: pain 5.6; motion 5.6; function 5.3. Preoperative pain (i) vs. (ii), p, 0.002. Range of movement: same interval (i) vs. (ii), p, 0.002. Harris Hip Scores (most recent visit, 19 patients only): (i) 91.6 (75.5–99.7); (ii) 91.3 (55.7–99.7); p, not significant. Patient preference: cementless 39%; cemented 22%; no preference 39%. No migration or subsidence noted. Loosening: (i) one stem probably loose; (ii) none. Revisions: none reported.	C
Hedlundh & Fredin, 1995 (Sweden) Unusual	Retrospective – matched Charnley (i) Dislocated (ii) Not dislocated Cemented	Not specified Teaching hospital Trans-trochanteric	Total 1838 (i) 60 (ii) 120 Follow-up period not specified (i) 60 (ii) 118	Total group median 70 years (22–94) (i) median 71 years (43–89) (ii) median 71 years (52–85) Sex Diagnosis Side of body	Mortality: (i) 32/60 (53%); (ii) 29/118 (24.5%); $p < 0.001$; death risk same for single and recurrent dislocations. Median age at death (range): (i) 77 years (56–90); (ii) 77 years (59–91); p, not significant. Logistic regression: gender, length, weight, obesity, previous contralateral hip surgery and previous arthrotomy of ipsilateral knee had no influence on dislocation rate. Alcoholism (in men only): (i) 10/20 (50%); (ii) 7/38 (18%); p, 0.01.	C
Hernandez, *et al.*, 1994 (USA)	Retrospective Cup: Universal cup design Stem: Bimetric (i) cemented (ii) uncemented (i) Hybrid (ii) Porous-coated	Not specified Specialist hospital Posterior without trochanteric osteotomy	231 in 203 patients Follow-up: Minimum 3 years; maximum 5 years (i) 97 (ii) 134 Also matched: (i) 66 in 58 patients (ii) 65 in 58 patients	Age not specified Sex matched but not specified Weight matched but not specified Diagnosis	Mean linear wear (range): (i) all cups: 0.42 mm (0–2.75); matched 0.47 mm (0–2.75); (ii) all cups: 0.73 mm (0–4.21); matched 0.72 mm (0–4.21); $p < 0.04$ for both. Mean linear wear rate (range): (i) all: 0.14 mm/y (0–0.92); matched: 0.15 mm/y (0–0.92); (ii) all and matched: 0.22 mm/y (0–1.41); ($p < 0.05$). Radiolucent lines: (i) 7 cup side; 1 stem side (1 stem subsided); (ii) 6 cup side; 5 stem side (no subsidence).	C
Hodgkinson, *et al.*, 1993 (UK)	Retrospective Charnley (i) flanged (ii) unflanged Cemented	Not specified but same group for each cohort Specialist hospital Standard Charnley technique	350 Follow-up: Maximum 9–11 years 313 in total Clinical data: (i) 150 (ii) 152	(i) 60.6 years (ii) 58.2 years (pre-operation or follow-up?) Sex Weight Diagnosis	Mean pain score: (i) 3.11; (ii) 3.13; p, not significant. Mean function score: (i) 2.81; (ii) 2.76; p, not significant. Mean movement score: (i) 2.58; (ii) 2.65; p, not significant. Radiolucency: Grade 0 (none): (i) 43%; (ii) 30%; Grade 1: (i) 39%; (ii) 45%; Grade 2 and above: (i) 19%; (ii) 25%; $p < 0.05$; Grade 2 and above, radiographic loosening; progression of radiolucency almost identical for both groups. Revisions: 15/350 (4.3%) revised (11 within 10 years; 4 at or after 10 years), 9 with radiographic loosening.	C

continued

DATA TABLE 2 contd *Non-controlled comparative studies included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
Hoffman, *et al.*, 1994 (New Zealand)	Retrospective (i) Charnley (74.4%) (ii) Harris-Galante (9.0%) (iii) Müller (7.9%) (iv) PCA (v) Autophor (vi) Brunswick (vii) Wrightington (vii) McKee (ix) Indiana Cemented (i) (iii) (vi–viii) Porous (ii) (iv) Ceramic (v) Resurfacing (ix)	Six surgeons, grades not specified General and private hospital Not specified	1166 in 974 patients (i) 867; (ii) 105; (iii) 92; (iv) 38; (v) 35; (vi) 14; (vii) 5; (viii) 5; (ix) 2 Maximum possible follow-up 20 years 1156 (99.1%)	Average (at operation) 66.2 years (SD 10.1) Sex Side of body Diagnosis	Multivariate regression analysis: four significant factors for survival: sex, age, surgeon, hospital. Revisions/loosening: (i) 72 failures (6 loose); (annual failure rate 1.78%; 15-year survival rate 73%); (ii) no failures; (iii) 26 failures (23 loose); (annual failure rate 6.93%; 11-year survival rate 63%); (iv) no failures (v) failure rate in first 3 years 15%. (Remainder not given as numbers too small)	C
Horikoshi, *et al.*, 1994 (USA)	Retrospective (i) Uncemented stem (as below) + uncemented cup: 4 Harris-Galante porous 4 Richards porous 4 BIAS 2 Osteonic 2 Macrofit 1 Intermedic 1 PCA (ii) Cemented stem (as below) + cemented cup: 5 T-28 4 Aufranc-Turner 3 Harris Design-2 2 Harris-Galante Porous-I 1 Biomet 1 Charnley 1 Müller 1 Buck 32 Uncemented vs. cemented	Not specified Teaching hospital Not specified	36 in 36 patients (i) 18 (ii) 18 Follow-up (range): (i) 4.9 years (1–16) (ii) 10.3 years (2–20) $p < 0.02$ (i) 18 (ii) 18	(i) 65 years (28–90) (ii) 73 years (64–86) Sex Diagnosis	Radiolucency: (i) complete radiolucent line > 2 mm, 5/18 (28%); partial radiolucency, 12 (66%); 1 migrating prosthesis (6%); (ii) complete radiolucent line > 2 mm, 12/18 (67%); partial radiolucency, 4 (22%); 2 migrating prostheses (11%); *p*, ? Intraoperative examination: all components loose; all surrounded by fibrous tissue membrane 2–15 mm thick.	C
Hozack, *et al.*, 1993 (USA)	Prospective, controlled Stem: (i) Dual-Lock (80% metal-backed) (ii) Trilock (98% metal-backed) Cup: cemented (i) Cemented (ii) Porous-coated	Not specified Teaching hospital Not specified	(i) 71 in 66 patients (ii) 70 in 61 patients Follow-up (range): (i) 4.3 years (2–6.5) (ii) 4.1 years (2–6) (i) 71? (ii) 70?	(i) 64 years (32–82) (ii) 52 years (25–72) Sex Weight Charnley class Diagnosis	Charnley Scores (preoperative): pain – (i) 3.1; (ii) 3.0; *p*, not significant; function – (i) 2.6; (ii) 2.6; *p*, not significant; motion – (i) 3.1; (ii) 3.1; *p*, not significant. Charnley Scores (postoperative): pain – (i) 5.6; (ii) 5.7; *p*, not significant; function – (i) 5.1 (ii) 5.6; $p < 0.001$; motion – (i) 5.4; (ii) 5.6; *p*, not significant. Cup migration > 5 mm (all cemented): (i) 2 (7.3%); (ii) 4 (6%); *p*, not significant. Definite/probable loosening of stem: (i) 3 (4%); (ii) 3 (4%). Revisions: (i) 1 (1.4%) for loosening; (ii) none.	C

continued

PAPER 3

DATA TABLE 2 contd *Non-controlled comparative studies included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
Hozack, *et al.*, 1994 (USA)	(i) Prospective (ii) Retrospective (i) Stem: Taperloc, Cup: cemented or uncemented (ii) (a) As for (i) vs. (b) cemented components (i) Porous-coated (ii) Porous-coated vs. cemented	(i), (ii)(a) By, or supervised by, one senior surgeon (ii)(b) ? Teaching hospital (i), (ii)(a) Either direct lateral or trans-trochanteric (ii)(b) ?	(i) (ii)(a) 100 (ii)(b) ? Follow-up (range): (i) 3.8 years (2–6) (ii)(a) 3.8 years (2–6) (ii)(b) 3.5 years (2–6) (i) 94 (ii)(a) 52 (ii)(b) 52	(i) 56 years (25–79) (ii)(a) 62 years (48–79) (ii)(b) 67 years (48–79) Sex Weight Diagnosis	(i) Charnley Scores, pre- vs. postoperative (range): pain, 3.0 (2–5) vs. 5.5 (2–6); function, 2.8 (2–6) vs. 5.4 (2–6); range of motion, 3.1 vs. 5.6; *p*, ? (i) Limp: 89% no limp; 11% mild/moderate limp. (i) Revisions/loosening: no revisions, all stems stable. (ii) Charnley Scores (pre- vs. post-operative): pain, (a) 3.0 (b) 3.0, vs. (a) 5.6 (b) 5.7; function, (a) 2.7 (b) 2.9, vs. (a) 5.5, (b) 5.5; motion, (a) 3.1, (b) 3.2, vs. (a) 5.5, (b) 5.6; (for all (a) vs. (b) comparisons, *p*, not significant). (ii) Limp: (a) no limp 88%; mild/moderate limp 12%; (b) no limp 90%; mild/moderate limp 10%. (ii) Revision/loosening: (a) and (b), no revisions, all components stable.	C
Huracek & Spirig, 1994 (Switzerland)	Retrospective, matched Mecron cementless (i) with HA-coating (ii) without HA-coating (i) HA-coating (ii) Press-fit	One surgeon, grade not specified General hospital Lateral	127 in 121 patients Follow-up: 4.1 years (i) 40 (ii) 40	71.1 years Sex All primary osteoarthritis	Harris Hip Scores (modified): (i) Pre- vs. postoperative, 48 vs. 78; (ii) Pre- vs. postoperative, 45 vs. 74; *p*, not significant. Pain: no pain – (i) 59.3%; (ii) 22.5%; $p < 0.0016$. Migration/subsidence: (i) cup, no migration; stem, 3 (7.5%) subsided/migrated; (ii) cup, 13 (32.5%) migrated by 5 mm or more; stem, 12 (30%) subsided/migrated. Revisions or loosening: 0/80.	C
Hwang & Park, 1995 (Republic of Korea)	Prospective (i) AML (ii) PCA (iii) Harris-Galante Porous Porous-coated	One surgeon Teaching hospital Direct lateral (44% PCA) Posterior (56% PCA, 100% AML, Harris-Galante Porous)	289 Follow-up (range): (i) 5.2 years (2.1–8.5) (ii) 4.7 years (3.1–8.0) (iii) 3.8 years (2.0–6.9) 270 in 214 patients (i) 90 (+5 years, 71) (ii) 117 (+5 years, 90) (iii) 63 (+5 years, 42)	(i) 51.2 years (20–79) (ii) 46.2 years (24–79) (iii) 46.3 years (25–86) Sex Diagnosis	Harris Hip Scores: (i) preoperative, 45; latest follow-up, 93; excellent, 71%; (ii) preoperative, 41; latest follow-up, 91; excellent, 76%; (iii) preoperative, 44; latest follow-up, 91; excellent, 69%; *p*, ? Thigh pain (at 5 years): (i) 17%; (ii) 21%; (iii) 19%; *p*, ? Stem orientation: (i) neutral 90%; varus 4%; valgus 6%; (ii) neutral 87%; varus 9%; valgus 4%; (iii) neutral 94%; varus 5%; valgus 1%; *p*, ? Osteolysis of the neck: (i) 8%; (ii) 15%; (iii) 10%; *p*, ? Loss of proximal bone density: (i) 7%; (ii) 20%; (iii) 13%; *p*, ? Heterotopic bone formation (mild or moderate): (i) 15.6%; (ii) 15.4%; (iii) 9.5%; *p*, ? Stem subsidence/migration, average (range): (i) subsidence – 2.1 mm (0–8), > 3 mm 10%; migration – present at 3 years 10%, progressive 3.3%; (ii) subsidence – 2.2 mm (0–7), > 3 mm 13.7%; migration – present at 3 years 13.7%, progressive 5.1%; (iii) subsidence – 1.9 mm (0–6), > 3 mm 12.7%; migration – present at 3 years 12.7%, progressive 4.8%. Cup migration (present at 1 year): (i) 3.3%; (ii) 4.3%; (iii) 4.8%; no progressive migration. Average wear rate (range) of polyethylene liner: (i) 0.7 mm/y (0.5–3.1); (ii) 0.6 mm/y (0.4–2.8); (iii) 0.8 mm/y (0.6–2.8); *p*, ? No revisions reported.	C

continued

DATA TABLE 2 contd *Non-controlled comparative studies included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
Jacobsson, *et al.*, 1990 (Sweden) (Previous report: Djerf & Walstrom, 1986)	Prospective (i) McKee-Farrar (ii) Charnley Cemented	Eight surgeons, various grades Teaching hospital (i) dorso-lateral (as decribed by McKee and Farrar) (ii) lateral with trochanteric osteotomy (as described by Charnley)	177 in 169 patients (i) 107 (ii) 70 Follow-up (range): (i) 11.9 years (10.7–13.5) (ii) 11.0 years (10.1–12.8) (i) 55 (ii) 41 (55 died, 22 revised, 4 lost to follow-up)	66.9 ± 8.1 years Sex Weight Diagnosis	Walking ability (data given for total group only): Distance: preoperative average 200 m, +2 years 2000 m (stayed constant at subsequent follow-ups). Speed: preoperative average 0.5 m/s, +1 year 1.0 m/s, +11.5 years 0.7 m/s. No requirement for aids: preoperative 3%; +5 years 45%; +11.5 years 23.3% and 50% did not use them regularly. Harris Hip Score: (i) mean 74.8 ± 17.3 (44% good/excellent); (ii) mean 72.9 ± 19.6 (51% good/excellent); *p*, ? Pain score – no or occasional pain: (i) 76%; (ii) 69%; *p*, ? Revisions: 22/118 (18.6%) – (i) 16; (ii) 6; (number caused by loosening not given). Loosening: 30/93 (32.2%) radiographic loosening: (i) 14–8 stems, 2 cups, 4 both; (ii) 16–10 stems, 2 cups, 4 both. Survivorship analysis: Mean annual rate of re-operation – (i) 1.61%; (ii) 0.91%; *p*, not significant. Cumulative numbers of survivors – (i) 82.2%; (ii) 89.5%; *p*, not significant.	C
Kelley & Johnstone, 1992 (USA) Unusual	Retrospective Stem: Charnley or Iowa Cup: 22 mm Charnley, 28 mm polyethylene, 28 mm metal-backed with: (i) Stainless steel monofilament wire (ii) Cobalt–chrome cable Cemented	One surgeon, grade not specified Not specified Transtrochanteric with trochanteric osteotomy	Total 796 patients Number with surgical approach required, 643 Follow-up: (i) 6 years 1 month (ii) 5 years 11 months (i) 162 patients (ii) 160 patients	(i) 67 years (ii) 65 years Sex Diagnosis	Wire vs. cable: Trochanteric union rates: wire 75%; cable 79%. Non-union rates: wire 13%; cable 8%; *p*, 0.36. Breakage (all three wires/cables): wire 43%; cable 12%; $p < 0.001$. Migration of wire/cable debris or fragments: Wire: to cup notch area 8%, < 2 cm 26%; Cable: to cup notch area 16%, < 2 cm 26%. Bone destruction: (i) 9%; (ii) 29%; $p < 0.001$. Rates of loosening (according to prosthesis type): Charnley: total 3/70 (4.3%); wire 2/42 (4.7%); cable 1/28 (3.6%). 28 polyethylene: total 24/68 (35.3%); wire 9/30 (30%); cable 15/38 (39.5%). 28 metal-backed: total 30/184 (16.3%); wire 9/90 (10%); cable 21/94 (22.3%). Differences in cup loosening, adjusted for type: *p*, 0.003. Revisions/loosening of total group (n = 643): (i) 4 revisions (2.4%), 2 for cup loosening; (ii) 10 revision (6.25%), 5 for cup loosening; (further surgery required: (i) 5; (ii) 5.)	A
Krismer, *et al.*, 1991 (Austria)	Retrospective (i) Müller straight-stem with (a) cementless RM cup or (b) cemented Müller standard cup (ii) Müller standard-stem Cemented	Not specified Teaching hospital Not specified	Total, 1099 (i) 422 (ii) 583 After criteria applied, 503 in 452 patients Follow-up: 5.8 years (± 1.24) 425 in 383 patients (i) 263 (ii) 162	60.9 ± 7.4 years Sex Diagnosis	Hip pain in groups 5 and 6, d'Aubigne classification: (i) 87% patients; (ii) 80% patients; *p*, not significant. Subsidence: (i) 32/260 (12%) migrated > 2 mm; (ii) 17/159 (10.6%) migrated. Loosening (maximum follow-up 7–8 years): (i) 21/260 (8%) (RM cup, 12/156, 7.7%; cemented cup, 9/104, 8.7%); (ii) 19/159 (11.9%) (cemented cup). Revisions due to loosening: stems (i) 5 (1.9%); (ii) 6 (3.7%).	C
Krismer, *et al.*, 1991 (Austria)	Retrospective (i) RM cup (uncoated) (ii) Müller cup (with Müller straight or standard stems) (i) Press-fit (ii) cemented	Not specified Teaching hospital Not specified	Original numbers: (i) 207; (ii) 892 After criteria applied: (i) 173 in 160 patients (ii) 309 in 292 patients Follow-up: (i) 5.3 ± 1.1 years (ii) 6.1 ± 1.3 years (i) 160 in 147 patients (ii) 263 in 236 patients	(i) 57.3 ± 7.2 years (ii) 62.9 ± 6.7 years Sex Diagnosis	Subjective results: Satisfied (i) 94.6%; (ii) 91.7%; *p*, not significant. Range of motion: Flexion: (i) 102 ± 16.9; (ii) 93 ± 15.3; $p < 0.001$. Gain in flexion: (i) 27 ± 24; (ii) 20 ± 26; *p*, 0.011. Migration: (i) 35/140 (25%) migrated between 2.1 and 16 mm; (ii) no migration values recorded. Revisions/loosening (after 7–8 years): (i) 12% revised, 40% loose; (ii) 4% revised; 15% loose; *p*, ?	C

continued

PAPER 3

DATA TABLE 2 contd *Non-controlled comparative studies included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
Kristiansen & Steen Jensen, 1985 (Denmark) Unusual	Retrospective, matched Stanmore (standard type, 29 mm head, small collar, 142 mm stem) Cemented	Not specified Teaching hospital Not specified	320 in 308 patients Follow-up (range): 36 months (4–68) (i) 33 revisions (due to loosening) (ii) 33 controls	64 years (48–79) at primary surgery Sex Weight Diagnosis	Cortical index (range) at the calcar femoral: (i) 0.15 (0.12–0.21); (ii) 0.18 (0.13–0.27); $p < 0.01$ (i.e. calcar bone stock thin prior to surgery). Cementation technique: Insufficient packing, (i) 29/33 (88%); (ii) 13/33 (39%); p, 0.0005. Positioning of the prosthesis: valgus, (i) 8/33 (24%); (ii) 10/33 (30%); neutral, (i) 7/33 (21%); (ii) 16/33 (48%); varus, (i) 18/33 (55%); (ii) 7/33 (21%); $p < 0.0005$; varus position is related to loosening in stem.	B
Lehman, *et al.*, 1994 (USA) Unusual	Retrospective Stem: HS2P, Omnifit, Omniflex Cup: Dual-geometry, Peripheral Self-Locking, Mecron ring, Harris-Galante cup Various uncemented	By, or supervised by, one senior surgeon Teaching hospital Not specified	324 in 284 patients; divided by body-mass index: (i) 20–29 normal (ii) 30–39 obese (iii) 40+ within group (ii) morbidly obese (< 20 excluded) Follow-up (range): (i) 49 months (24–92) (ii) 20–49 months (24–89) (iii) 45 months (25–81) (i) 142 in 127 patients (ii) 60 in 55 patients (iii) 8 in 8 patients part of (ii))	(i) 48 years (19–73) (ii) 50 years (17–67) (iii) 52 years (37–72) Sex Diagnosis Weight Body Mass Index	Clinical parameters: Pain (no/mild pain pre- vs. postoperative) – (i) 3 (2%) vs. 129 (91%); $p < 0.001$; (ii) 0 (0%) vs. 48 (92%); $p < 0.001$; (iii) 0 (0%) vs. 7 (88%); p, 0.02; p, not significant – (i) vs. (ii), (ii) vs (iii). Mobility (not needing support to walk, pre- vs. postoperative): (i) 73 (51%) vs. 130 (92%); $p < 0.001$; (ii) 23 (44%) vs. 44 (85%); $p < 0.001$; (iii) 3 (38%) vs. 5 (63%); p, not significant; (postoperative (ii) vs. (iii) $p < 0.05$; all others, p, not significant). Limp (no/slight limp) pre- vs. postoperative: (i) 46 (32%) vs. 131 (92%); $p < 0.001$; (ii) 7 (13%) vs. 44 (85%); $p < 0.001$; (iii) 2 (25%) vs. 6 (75%); p, 0.043 (p, not significant – (i) vs. (ii), (ii) vs. (iii)). Trendelenburg sign present (pre- vs. postoperative): (i) 63 (44%) vs. 126 (89%); $p < 0.001$; (ii) 20 (38%) vs. 47 (90%); $p < 0.001$; (iii) 2 (25%) vs. 7 (88%); p, ? (p, not significant – (i) vs. (ii), (ii) vs. (iii)). Heterotopic ossification: Class III: (i) 15%; (ii) 13%; (iii) 2/8 (25%); p, not significant. Class IV: no patient in any group. Osteolysis of femur: (i) 13%; (ii) 13%; p, not significant ((iii) 0%). Wear of polyethylene acetabular liner: 2 mm or more: (i) 3%; (ii) 2%; p, not significant ((iii) 0%). Revisions/loosening: (i) cup, 7 (4.9%) revised (all loose) + 3 loose; stem, 9 (6.3%) revised (6 loose) + 2 loose; (ii) cup, 3 (5%) revised (all loose) + 2 loose; stem, 1 (1.6%) revised (loose) + 0 loose; (iii) no revisions or loosening; ((i) vs. (ii), p, not significant for revision or mechanical failure rates). Failure rates of components – stem: HS2P, 5/33 (15%) (average follow-up 69 months); Omnifit, 2/83 (2%) (average follow-up 54 months); Omniflex, 5/86 (6%) (average follow-up 36 months). Failure rates of components – cup: dual-geometry, 2/165 (1%) (average follow-up 46 months); Mecron ring, 13/34 (38%) (average follow-up 69 months); Harris-Galante, 0/2 (0%) (average follow-up 44 months); Peripheral self-locking, 0/1 (0%) (average follow-up 24 months).	A

continued

DATA TABLE 2 contd *Non-controlled comparative studies included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
Mallory, *et al.*, 1989 (USA) Unusual	Retrospective, controlled (i) AML, MHP, PCA, ceramic, SRN-REV (ii) Controls, uncemented, over same period with no fractures (no further details) (i) Porous-coated, cementless, porous-coated, ceramic, cementless.	Not specified ? Modified direct lateral	(i) 56 femoral fractures (ii) 56 controls Maximum follow-up: 60 months 111	50.4 years (21–81) Sex Primary or revision operation Diagnosis	Fracture types: Type I (n = 45) – AML 33%, MHP 31%, PCA, 24%, ceramic 7%, SRN-REV 5%; Type II (n = 9) – AML 33%, MHP 33%, ceramic 22%, SRN-REV 12%; Type III (n = 2) – PCA 100%. Improvement by operation (includes fracture types I and II): (i) Great/very great, 51/54 (94.4%); (ii) Great/very great, 47/53 (88.7%); *p*, not significant. Modified d'Aubigne–Harris Score, Types I and II only (*p*, not significant): Pain (a) preoperative, (b) postoperative: (i) (a) score 1/2, 27; 3/4, 27; (b) 3/4, 13; 5/6, 40; (ii) (a) 1/2, 22; 3/4, 32; (b) 3/4, 10; 5/6, 44. Function (a) preoperative, (b) postoperative: (i) (a) 1/2, 18; 3/4, 35; (b) 3/4, 23; 5/6, 30; (ii) (a) 1/2, 11; 3/4, 43; (b) 3/4, 17; 5/6, 37. Range of motion (a) preoperative, (b) postoperative: (i) (a) 1/2, 4; 3/4, 49; (b) 3/4, 16; 5/6, 37; (ii) (a) 1/2, 6; 3/4, 47; (b) 3/4, 22; 5/6, 31.	C
Maloney & Harris, 1990 (USA)	Retrospective, matched (i) Cemented Precoat stem + Harris-Galante Porous cup (ii) Harris-Galante stem and cup (i) Hybrid (ii) Porous	One senior surgeon General/ teaching hospital Not specified, but same for both groups	(i) 67 (ii) 69 Follow-up (range): (i) 32 months (24–46) (ii) 37 months (24–57) (i) 25 (ii) 25	(i) 62 years (54–67) (ii) 61 years (55–69) Diagnosis Sex Weight	Harris Hip Scores: Matched pairs – pre- vs. postoperative mean (range): (i) 52 (38–67) vs. 96 (80–100); (ii) 48 (33–67) vs. 84 (35–100); postoperative comparison, $p < 0.02$. Overall group – pre- vs. postoperative mean (range): (i) 55 (28–70) vs. 97 (74–100); (ii) 57 (20–76) vs. 87 (35–100); (matched pair scores preoperatively did not differ significantly from equivalent original group scores). Pain: (i) 24/25 (96%) no or slight pain; (ii) 19/25 (76%) no or slight pain; *p*, ? Thigh pain: (i) None; (ii) 5/25 (20%); *p*, ? Limp: (i) 19/25 (78%) no limp, 5/25 (20%) mild limp; (ii) 11/25 (44%) no limp, 9/25 (36%) mild limp; *p*, ? Migration – cups: None or complete radiolucency in both groups. Migration – stems: (i) all radiologically stable, no migration; (ii) 5/25 (20%) radiologically migrated; *p*, ? Revisions: (i) None; (ii) 4/25 (16%) revised, 3 due to migration.	C
Markel, *et al.*, 1995 (USA)	Retrospective Cups: (i) all-polyethylene (Charnley) (ii) metal-backed (Charnley design or TiBac design) plus Charnley stem Cemented cups	One senior surgeon Specialist hospital Posterior	134 in 112 patients Mean follow-up (range): 84 months (49–120) (i) 90.6 months (ii) 78.4 months 115 in 97 patients (i) 55 (ii) 60 (21 Charnley, 39 TiBac)	(i) 62.8 years (ii) 58.6 years Sex Side of body Height Weight Diagnosis (variables listed but no data given)	Hospital for Special Surgery hip rating system (n = 115): (i) Good/excellent 55 (100%); (ii) Good/excellent 60 (100%). Rate of linear polyethylene wear: (i) 0.08 mm/y; (ii) 0.078 mm/y; *p*, not significant. Volumetric polyethylene wear rate: (i) 32.9 mm^3/y; (ii) 30.3 mm^3/y; *p*, not significant. Revisions/loosening (n = 108): No revisions, (i) 32% (16) probably loose; (ii) 16% (9) probably loose; *p*, not significant; (no stems loose).	C

continued

PAPER 3

DATA TABLE 2 contd *Non-controlled comparative studies included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
McPherson, *et al.*, 1995 (USA)	Retrospective, matching (material collected prospectively) APR-I (i) HA-coating added (ii) Porous-coat	Not specified Teaching hospital Not specified	230 patients Follow-up: Minimum 3 years 84 patients	(i) 55 years (23–73) (ii) 56.5 years (22–71) Sex Weight (Also matched for Diagnosis Charnley activity and class; bone quality and type, and surgical technique but details not given)	Harris Hip Score – average (range): (i) 95.1 (65–100); 39 (93%) excellent/good; (ii) 95.8 (59–100); 41 (98%) excellent/good; *p*, not significant. Harris Pain and Limb Scores – no significant difference (no data given). Modified Engh Radiographic Fixation Score: Grade IA, B or C at 3 years: (i) 38 (90%); (ii) 35 (83%); *p*, not significant. Modified DeLee–Charnley Fixation Score (3 years): (i) Grade IA, 39 (93%); IB, 3 (7%); (ii) Grade IA, 26 (62%); IB, 14 (33%); IC, 2 (5%); (i) vs. (ii), *p*, 0.002; HA has better fixation. Revisions/loosening: Mechanical failure rate, (i) 5% – 1 revised (due to loosening) plus 1 unstable (loose?); (ii) 5% – 2 unstable (loose?).	C
Moilanen, *et al.*, 1996 (UK)	Prospective SLF press-fit cup (i) with HA-coating (ii) without HA-coating (+ Freeman stem, cemented or uncemented) (i) HA-coating (ii) Press-fit	Not specified Teaching hospital Anterolateral	111 (i) 71 (ii) 40 Follow-up: (i) 2.3 years (ii) 3.4 years (i) 71 (two revised by 7 months and excluded from further analysis) (ii) 40	(i) 59.7 years (ii) 62.6 years Sex Diagnosis	Number with pain requiring analgesics: (i) Preoperative, 66/69 (96%); +3 years, 0/30 (0%); (ii) Preoperative, 36/40 (90%); +3 years, 2/33 (6%); *p*, not significant. Number able to walk continuously for 30 minutes: (i) Preoperative 5/69 (7%); +3 years 22/30 (73%); (ii) Preoperative 5/40 (13%); + 3 years, 26/33 (79%); *p*, not significant. Vertical linear wear at 3 years: (i) 0.07 mm ± 0.19 (n = 15); (ii) 0.10 mm ± 0.17 (n = 28); *p*, 0.61. Migration (mean): (i) rate 0.06 mm/y (ii) rate 0.20 mm/y; *p*, 0.22. Length of follow-up or migration level by 6 months neither affected the results nor predicted subsequent rate. Ceramic ((i) 40%, (ii) 23%) vs. metal femoral heads did not affect rate. Radiolucent lines: (i) 3/52 (6%); (ii) 8/30 (27%); $p < 0.05$. Revisions: (i) 2 revisions, none due to loosening; (ii) no revisions.	C
Moskal, *et al.*, 1994 (USA)	Prospective PCA (i) uncemented stem (ii) cemented stem (i) Porous-coated (ii) Hybrid	One senior surgeon Community hospital Modified direct-lateral, 97%; trans-trochanteric, 3%	137 in 122 patients Follow-up: 2–4 years 134	(i) 63 years (27–95) (ii) 75 years (51–92) Sex Height Weight	Harris Hip Score, preoperative vs. +3 years: (i) 43 (1–87) vs. 89 (51–100); (ii) 41 (1–100) vs. 86 (61–91); *p*, not significant. Harrris Pain Score, preoperative vs. +3 years: (i) 15 (0–44) vs. 41 (30–44); (ii) 15 (0–44), vs. 42 (40–44); *p*, not significant. Thigh pain: (i) +3 years, 5% incidence; (ii) no thigh pain. Limp incidence: (i) 18%;(ii) 22%; *p*, not significant. Radiolucency: 6% hips had lines in each stem zone; 7/134 (5%) cups had lines in 2/3 zones; no revisions or loosening.	C
Nashed, *et al.*, 1995 (USA)	Retrospective BIAS stem (i) Titanium head, cemented polyethylene cup (ii) Titanium head, cemented metal-backed cup (iii) Titanium head, uncemented metal-backed cup (iv) Cobalt–chrome head, uncemented metal-backed cup (i, ii) Cemented vs. (iii, iv) Press-fit	One senior surgeon General hospital Not specified	193 Follow-up (range): Total average 6.9 years (2.3–12.5) (i) 9.4 years (4.3–12.5) (ii) 7.8 years (3.3–10.5) (iii) 6.6 years (4.3–8.0) (iv) 5.5 years (2.3–8.0) Total: 175 (i) 24 (ii) 62 (iii) 15 (iv) 74	(i) 50 years (ii) 52 years (iii) 51 years (iv) 50 years Sex Weight Diagnosis	Average wear rates: (i) 0.10 mm/y; (ii) 0.13 mm/y; (iii) 0.25 mm/y; (iv) 0.17 mm/y; *p*, ? Osteolysis (%) (stem lysis (%), cup lysis (%)): (i) 0 (0,00); (ii) 31 (24, 7); (iii) 87 (87, 40); (iv) 24 (22, 14). Incidence of osteolysis statistically higher in (iii) than any other group ($p < 0.001$) and statistically lower in (i) than in any other group ($p < 0.005$). Revisions: Hips with osteolysis 44%; without osteolysis 7%; $p < 0.001$.	C

continued

DATA TABLE 2 contd *Non-controlled comparative studies included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
Neumann, *et al.*, 1996 (Denmark) Unusual	Prospective Charnley (i) Patients 55 years or younger (ii) Patients older than 55 years Cemented	One senior surgeon Teaching hospital Lateral with trochanteric osteotomy	240 in 211 patients (i) 52 (ii) 188 Median follow-up (range): (i) 17.0 years (15–20.6) (ii) 17.7 years (15.1–20.4) Total: 114 (i) 37 (71%) (ii) 77 (41%)	Overall median: 62 years (34–79) (i) 51 years (34–55) (ii) 64 years (56–79) Diagnosis Previous operation on hip	Charnley Scores: Pain – identical for both groups preoperatively and at each follow-up; Function:(i) median 5; (ii) median 44 at latest follow-up; (ii) probably due to deterioration in health Motion: no substantial differences between groups. Revisions/loosening: Revisions: (i) 5/52 (10%); (ii) 15/188 (8%); Loosening: (i) 3 (6%); (ii) 5 (3%); *p*, 0.37. Probability of survival at 20 years: (i) 88.3% (95% CI ± 9.8%); (ii) 89.3% (95% CI ± 5.8%); *p*, 0.82.	B
Pierchon, et *al.*, 1994 (France) Unusual	Retrospective (i) Dislocated prosthesis Stem: 29 Müller self-locking 5 Müller dysplasia 4 not specified Cup: 12 cemented 26 uncemented (ii) Controls (not dislocated): cup not specified Cemented and hybrid	Not specified Teaching hospital Posterolateral without trochanteric osteotomy	(i) 39 (1st dislocation, 22; recurrent, 16 + 1 exclusion) (ii) 14 (11 contralateral hips from (i)) Follow-up period not specified (i) 38 (ii) 14	(i) 57 years (17–91) Sex Side of body Diagnosis (NB: for (i) only)	Mean cup abduction: (i) 44.5° (30–68°); (ii) 43.6°; *p*, not significant. Mean cup anteversion: (i) 24.2° (–5–45°); (ii) 22.3°; *p*, not significant. Mean femoral neck anteversion: (i) 16.5° (–30–37°); (ii) 14°; *p*, not significant. Revision: (i) 7 hips, all now stable.	C
Pritchett, 1995 (USA)	Retrospective (i) Müller Straight Stem (ii) Physiological Stress Loading (iii) AML (iv) Conical Collar (v) Harris Precoat (i) Cemented, collarless (ii) Porous-coated, collared (iii) Porous-coated, collarless (iv) Material not specified, collared (v) Cemented, collarless	Not specified Teaching hospital Not specified	50 in 50 patients Follow-up: (i) 3.5–4.5 years (ii) 3–5 years (iii) 3–4 years (iv) 3–4 years (v) 3–5 years (i) 10 (ii) 15 (iii) 6 (iv) 13 (v) 6	(i) 59–73 years (ii) 60–79 years (iii) 53–70 years (iv) 58–69 years (v) 64–83 years Osteoarthritis Sex	Measured bone density loss compared with contralateral ('normal') hip: (i) –57% (–42, –85); (ii) –8% (+5, –30); (iii) –34% (–30, –60); (iv) –14 (+15, –30); (v) –43% (–36, –70). (i) vs. (iii) vs. (v), *p*, not significant; (ii) vs. (iv) *p*, not significant; (i), (iii), (v) vs. (ii), (iv) $p < 0.05$. Those with collar associated with less bone density loss in proximal femur. Bone mineral density in contralateral hips similar in all groups (72% hips within 10% of average value).	C
Pupparo & Engh, 1989 (USA)	Prospective AML stem with (i) S-ROM Anderson cup (smooth-threaded) (ii) S-ROM Super cup (porous-threaded) (i) Press-fit (ii) Porous-coated	One surgeon, grade not specified Not specified Posterolateral	(i) 82 (? patients) (ii) 62 (? patients) Follow-up (range): (i) 33.3 months (24–49) (ii) 29.5 months (25–50) (i) 56 (68%) (ii) 41 (66%)	Age not specified Weight Diagnosis	d'Aubigne Score (mean): (i) pain 5.59 ; walking 5.54; (ii) pain 5.68; walking 5.50. Migration: (i) 16 (29%) unstable, nine had migrated by mean 5.5 mm (3–11 mm); (ii) all stable, no migration; $p < 0.001$. Revisions/loosening: (i) six revised due to loosening; (ii) none revised.	C

continued

DATA TABLE 2 contd *Non-controlled comparative studies included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
Ranawat, et al., 1988 (USA) Unusual	Retrospective, matched Charnley: (i) 23; (ii) 35 T-28: (i) 17; (ii) 9 CAD Müller: (i) 6; (ii) 0 Charnley Müller: (i) 3; (ii) 2 DF-80: (i) 0; (ii) 4 Aufranc-Turner: (i) 1; (ii) 0 Cemented (conventional cementing techniques vs. modified techniques)	One senior surgeon Specialist hospital Not specified	100 in 87 patients Follow-up (range): (i) 5 years, average 10 years (8–12) (ii) 5 years 5 years, 100 hips (50 pairs) (i) 10 years, 37 hips	61 years Sex Weight Diagnosis	Migration: (i) 5 years, 4 (8%); 10 years, 5/37 (2–5 mm) (14%); (ii) 5 years, 0 (0%). Radiolucency: Cumulative score lower in (ii) than (i), p, 0.0005; (i) 7 (14%) with score of 4 or more; (ii) 1 (2%) with score of 4 or more. Revisions/loosening: (i) no cup/stem revision required 5 years (1 revision due to socket migration at 8 years); (ii) no cup revision required for loosening; 1 stem loose and revised.	B
Rand & Ilstrup, 1983 (USA)	Retrospective, matched (i) Charnley (ii) T-28 Cemented	20 surgeons, grades not specified Not specified Not specified	(i) 2388 in 2388 patients (ii) 459 in 459 patients Follow-up: (i) 5.7 years (± 0.7) (ii) 5.2 years (± 1.7) (i) 40 (ii) 40	(i) 64.7 ± 7.2 years (ii) 64.0 ± 8.1 years Sex Side of body Diagnosis Contralateral THR Previous surgery Number of trochanteric osteotomies	Pain: (i) 38/40 (95%), no/slight pain; (ii) 37/40 (92.5%), no/slight pain. Use of ambulatory aids: (i) 37/40 (92.5%), no aids; (ii) 38/40 (95%), no aids. Migration/subsidence: (i) 13 (32.5%) cup migrated > 1 mm; 8 (20%) stem subsided > 1 mm; (ii) 9 (22.5%) cup migrated > 1 mm; 8 (20%) stem subsided > 1 mm. Radiolucent lines (> 1 mm): stem – (i) 3; (ii) 5; p, not significant; cup: (i) 8; (ii) 17; p, not significant. Revisions/loosening: 1 hip in each group revised due to aseptic loosening.	C
Riska, 1993 (Finland)	Retrospective Ceraver Osteal alumina on alumina prosthesis with titanium alloy stem and: (i) cemented cup (ii) uncemented screw cup Cemented ceramic vs. uncemented ceramic	One surgeon, grade not specified Teaching hospital Anterolateral (McKee)	290 in 55 patients (i) 143 patients (ii) 112 patients (three already excluded) Follow-up (range): (i) 6.7 years (1–12) (ii) 3.6 years (1–7) (i) 143 (ii) 112	62 years (25–86) Sex Diagnosis	Charnley Scores: All without revision had excellent/good scores; scores averaged 4–6 for all sections. Revision/loosening: (i) revision, 16 (11.2%); loosening, 12 cups, 1 stem, 1 both; (ii) revision, 7 (6.3%); loosening, 2 cups; p, ?	C
Ritter & Gioe, 1986 (USA)	Prospective (i) T-28 (ii) Indiana conservative hip (i) Cemented (ii) Resurfacing	Not specified Teaching hospital Transtrochanteric (n = 85); anterior (n = 15)	100 in 50 patients Follow-up: Minimum 5 years, maximum 7 years 90 in 45 patients	62 years (21–87) Sex Diagnosis	Pain: Hospital for Special Surgery hip rating system (excluding revised hips): (i) Preoperative mean, 3.1; +5 years, 5.5; (ii) Preoperative mean 3.1; +5 years, 5.6; p, not significant. Revisions/loosening: (i) 2 revised (none loose); (ii) 15 revised (6 acetebular, 1 femoral, 4, both loose); patients with resurfaced hips requiring revisions were younger – average age 55 years.	C

continued

PAPER 3

DATA TABLE 2 contd *Non-controlled comparative studies included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
Ritter, 1995 (USA)	Retrospective Charnley (n = 260) Müller (n = 163) T-28 (n = 642) MOSC (n = 319) (a) cemented all-polyethylene cup (b) cemented metal-backed cup Cemented	One MD Specialist centre/ teaching hospital Transtrochanteric	1384 Follow-up range: Overall 1–22 years Average 8.9–12.7 years Failure analysis 1172; survival analysis 1144	Range of average ages 59–76 years Sex Diagnosis	Number failed 1-year post-surgery: Charnley, 32 (14%); Müller, 29 (20%); T-28, 57 (10%); all-polyethylene MOSC; 9 (9%); metal-backed MOSC, 28 (20%). Survival analysis – % survival < 90% by: Charnley: 10 years, 93%; 20 years, 76%; Müller: 10 years, 81%; 17 years, 56%; T-28: 10 years, 93%; 17 years, 75%; all-polyethylene MOSC: 10 years, 90%; 12 years, 87%; metal-backed MOSC: 10 years, 60%.	C
Schreiber, *et al.*, 1993 (Switzerland)	Retrospective (data collected prospectively) Balgrist with outer split ring of: (i) high density polyethylene (including 61 with 6 m of titanium) (ii) titanium alloy Press-fit	Not specified Teaching hospital Not specified	717 in 644 patients (i) 346 in 309 patients (318 primary; 28 revision) (ii) 371 in 335 patients (280 primary; 91 revision) Follow-up (range): (i) 55.4 months (1–116) (ii) 15.6 months (0.5–60) (i) 282 patients (ii) 324 patients	(i) 53.8 years (23–76) (ii) 56 years (24–57) Sex	Revisions: (i) primary: 42/317 (13%); revisions: 6/29 (21%); (ii) primary: 2/280 (0.7%); revisions: 5/91 (5%); *p*, ?	C
Schuller & Marti, 1990 (The Netherlands)	Retrospective Weber type (i) metal rotating head (ii) ceramic rotating head Cemented vs. ceramic	One senior surgeon Teaching/private hospital Not specified	(i) 48 (ii) 46 Mean follow-up (range): 10 years (9–11) (i) 33 (ii) 33	(i) 69 years (61–78) (ii) 66 years (48–78) Sex Weight Osteoarthritis	Wear of polyethylene: (i) mean 0.96 mm; (ii) mean 0.26 mm; $p < 0.001$. Revisions/loosening: (i) revision, 1 cup, 2 stems (9%) due to loosening; loose: 2 cups, 2 stems (12%); (ii) revision: 1 cup, 1 stem (6%) due to loosening; loose: 2 cups, 1 stem (9%). Rate of aseptic loosening, *p*, not significant.	C
Turner, 1994 (USA)	Retrospective (i) Aufranc-Turner (ii) Harris Design 2 (15 mm) (iii) Omnifit/ Omniflex (iv) Tharies (v) Kirschner Murray Welch (11 mm) (vi) Charnley–Müller (15 mm) (vii) Kirschner Anatomic (13 mm) (viii) Harris-Galante (11 mm) (ix) Biofit (14 mm) (x) Ring (xi) Dupuy Engh–Anderson (xii) Bichat (xiii) Intermedic (xiv) Stackhouse (xv) AML (15 mm) Various	By, or supervised by, one MD General hospital Anterolateral, 9%; posterolateral Kocher (Langenbeck) 91% 561	564 Follow-up period not specified	Age not specified Sex	Dislocation rates: (i) 4/129, 3.1%; (ii) 5/74, 6.76%; (iii) 4/74, 5.5%; (iv) 1/56, 1.8%; (v) 2/56, 3.57%; (vi) 0/53, 0%; (vii) 4/35, 11.1%; (viii) 0/34, 0%; (ix) 2/34, 5.88%; (x) 1/7, 14.3%; (xi) 1/5, 20%; (xii) 0/1, 0%; (xiii) 0/1, 0%; (xiv) 1/1, 100%; (xv) 0/1, 0%. Anterolateral, 0/53, posterolateral, 25/508 (4.9%). Primary operation: 19/477 (4%); revision, 6/84 (7%). Men: 6/215 (2.8%); women: 19/346 (5.5%); *p*, not significant.	C

continued

PAPER 3

DATA TABLE 2 contd *Non-controlled comparative studies included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
Visuri, *et al.*, 1994 (Finland)	Retrospective (i) McKee-Farrar (ii) Brunswick (iii) Lubinus Cemented	Not specified Specialist hospital Not specified	Basic material, 1863 hips/patients; study group, 1018 in 1018 patients Follow-up: 12 years Total: 1018 (i) 237 (ii) 449 (iii) 332	Men 61 years Women 63 years Sex Diagnosis	10-year survivorship of patient (not hip): (i) 85% alive (95% CI: 79–89) (n = 202/237); (ii) 82% alive (95% CI: 78–85) (n = 367/449); (iii) 82% alive (95% CI: 77–86) (n = 133/332 estimated). 10-year survivorship (sex): men 77%; women 86%. 10-year survivorship after 65th birthday: 78% (men 69%, women 83%).	C
Walker, *et al.*, 1995 (UK)	Retrospective (data collected prospectively) (i) Charnley (ii) Stanmore Cemented	Several surgeons, grade not specified Specialist hospital Not specified	Originally 403 patients (a): (i) 51, (ii) 57 (b): stable (i) 23, (ii) 23; revised (i) 17, (ii) 29; Total (i) 40, (ii) 52 Follow-up (range): (a): 5.8 years (1–12) (b): stable minimum 8 years; revised 79 months (11–218); (a): (i) 49; (ii) 55 (b): (i) 40 (23 stable, 17 revised) (ii) 52 (23 stable, 29 revised)	(a) not specified (b) 63 years Sex Diagnosis	Stem subsidence (a): Identical, mean against time, *p*, not significant (mean migration: 0–6 months 1.39 mm; 0–1 year 1.93 mm; 0–5 years 2.68 mm; 0–9 years 3.42 mm). Data given as a combined group (migration rate: 0–6 months 1.82 mm/y; 6–12 months 0.96 mm/y; 1–2 years 0.54 mm/y; 2–9 years 0.21 mm/y). Stem subsidence (b): stable: +2 years 35/46 (76%), migrated < 2 mm; revised: +2 years 7/46 (15%), migrated < 2 mm; $p < 0.001$. Radiolucent zone (around entire cement–bone interface): stable: 2%; revised: 89%; *p*, ? Migration at interfaces: stable – 7% stem–cement; 77% cement–bone; 17% both; revised – 34% stem–cement; 0% cement–bone; 66% both; *p*, ?	C
Wilson-MacDonald & Morscher, 1989 (Switzerland)	Retrospective (i) Müller standard-stem (ii) Müller straight stem (iii) collared stem derived from long-stem steel prosthesis (130 mm stem, neck shaft angle 130°) (plus RM cup in all hips) Cemented	12 senior surgeons Teaching hospital Lateral without trochanteric osteotomy	545 in 518 patients (i) 76 (14%) (ii) 370 (68%) + 11 (2%) lateralised version (iii) 88 (16%) Follow-up: 5–10 years Clinical analysis not specified; radiographic examination 411 patients	65.2 years (29–89) Sex Diagnosis	Radiographic loosening: (i) + (ii) 8%; (iii) 11%; *p*, 0.02. Subsidence > 5 mm: (i) + (ii) 2.85%; (iii) 4.5%; p, not significant; subsidence > 2 mm without radiological evidence. (ii) 10.8%, 32 hips. Radiographic loosening of RM cup: (i) + (ii) 35%; (iii) 7%; *p*, 0.03; (i) vs. (ii) *p*, 0.002. Revisions/loosening (including deceased patients): 41 revised in total, 20 stems revised; (i) revised, 5/76 (6.57%), 3 loose (3.94%); (ii) revised, 10/381 (2.62%), 6 loose (1.57%); (iii) revised, 5/88 (5.68%), 4 loose (4.54%).	C

continued

DATA TABLE 2 contd *Non-controlled comparative studies included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
Wixson, *et al.*, 1991 (USA)	Prospective Stem: Uncemented: PCA, 65 Cemented: PCA, 30; Six Ti/28, 15; ATS, 14; HD-2, 18; CDH, 2 Cup: Uncemented: PCA, 84; Harris-Galante, 6; APR, 1 Cemented: PCA, 10; TiBac, 40; Harris, 3 Cemented Porous Hybrid	Two MDs Teaching hospital Posterior	197 in 176 patients Follow-up: Mean 2.8 years (maximum 4) 144 in 131 patients	61.1 years Sex Diagnosis	Harris Hip Score: cemented – preoperative 100% fair/poor; most recent 77% excellent/good; mean at 4 years, 91; uncemented – preoperative 99% fair/poor; most recent 89% excellent/good; mean at 4 years, 90; hybrid – preoperative 100% fair/poor; most recent 89% excellent/good; mean at 4 years, 95; *p*, not significant. Harris Pain Score (preoperative vs. most recent follow-up – mean): cemented: 16 vs. 42, 84% no/slight pain; uncemented: 16 vs. 43, 98% no/slight pain; hybrid: 14 vs. 43 100% no/slight pain; *p*, ? Thigh pain at 3 years: cemented stem, 3%; uncemented stem, 13%; $p < 0.05$. Need for walking aids: cemented – preoperative 80%; most recent 23%; uncemented – preoperative 57%; most recent, 11%; hybrid – preoperative 70%; most recent 29%; *p*, ? Migration/subsidence: cemented – cup, 6 (12%) migrated or changed position; stem, no subsidence. uncemented – cup, 3 (3%) changed position; stem 3 (5%) subsided. Revisions/loosening: cemented – 2 (3.8%) revisions (due to loosening); uncemented – 5 (7.7%) revisions (4 loosening); hybrid – 1 (3.7%) revision (no loosening).	C *
Yahiro, *et al.*, 1995 (Various)	Meta-analysis (i) Threaded cup (Mecring, T-TAP, S-ROM, Accu-Path, Link V) (ii) Porous-coated prosthesis (PCA, AML, Harris-Galante, Whitesides, APR, BIAS) (iii) Cemented prosthesis (Charnley, Aufranc-Turner, Müller or Dual-Lock, Harris or HD-2, STH-2, T-28, Stanmore or PCA, AlloPro, CAD)	Not specified	(i) 1269 (ii) 1979 (iii) 10,230 Follow-up (range): (i) 2.2 years (0.5–6.3) (ii) 3.6 years (0.2–9) (iii) 7.5 years (0.2–23.1) As above	(i) 51.2 years (20–91) (ii) 50.3 years (16–92) (iii) 61.1 years (14–99) Sex Diagnosis	Diagnoses of revisions of failed previous operations: (i) 30.7%; (ii) 8.4%; (iii) 13.5%; ('significantly more in (i)'; *p*, ?). Incidence of cup failures (clinical): revision rate – (i) mean 3.58%, (ii) mean 1.44%#, (iii) mean 1.61%#; migration – (i) mean 8.85%, (ii) mean 0.64%#, (iii) mean 1.48%#; pain score – (i) mean 15.10%, (ii) mean 3.01%#; (iii) mean 4.53%#; (# all significantly different from (i) $\alpha < 0.05$). Incidence of cup failures for revision subgroup: revision rate – (i) mean 14.59%; (ii) 3.68%; (iii) 3.35%; (i) vs. (ii), and (i) vs. (iii) $\alpha < 0.5$. migration: (i) mean 17.56%; (ii) 12.50%; (iii) 8.18%; (i) vs. (iii) $\alpha < 0.5$. Incidence of cup failures (radiolucencies): progressive – (i) mean 7.64%; (ii) 1.97%; (iii) 6.08%. (i) vs. (ii), (i) vs. (iii) $\alpha < 0.5$. > 1 mm: (i) mean 55.15%; (ii) 12.49%; (iii) 19.07%; (i) vs. (iii) $\alpha < 0.5$. complete: (i) mean 3.31%; (ii) 3.58%; (iii) 2.33%. incomplete: (i) mean 52.45%; (ii) 60.70%; (iii) 27.83%; (i) vs. (iii) $\alpha < 0.5$.	Not rated

continued

DATA TABLE 2 contd *Non-controlled comparative studies included in the review*

Study (country)	Study design and prosthesis type	Grade of surgeon, type of hospital, and surgical technique	Number of THRs, length of follow-up and number of patients followed-up	Patient age; other patient variables reported	Outcome measures and results	Rating
Zicat, *et al.*, 1995 (USA)	Retrospective Stem: All AML or P10 (32 mm head) Cup: (i) cemented all-polyethylene (ii) AML cup (i) Cemented (ii) Porous-coated	One senior surgeon Specialist hospital Not specified	(i) 63 in 63 patients (ii) 74 in 74 patients Follow-up (range): (i) 107 months (54–142) (ii) 102 months (62–122) (i) 63 (ii) 74	(i) 57 years (24–85) (ii) 54 years (16–79) Sex Weight Diagnosis Charnley class Side of body	Radiolucent lines: (i) 47/51 cups in unrevised hips (92%) had line in one zone or more; highest prevalence 86%, zone 1; (ii) 14/71 cups in unrevised hips (20%) had line in one zone or more; highest prevalence 16%, zone 3; $p = ?$ Osteolysis: (i) 19/51 (37%) cups had linear or expansile osteolysis; 6/51 (12%) stems with osteolysis (all in medial region). (ii) 13/71 (18%) cups had expansile osteolysis (no linear): 23/71 (32%) stems with osteolysis (21 in medial region). % hips with osteoloysis similar for both groups. Stability: (i) 19/51 (37%) unrevised cups termed unstable; (ii) 2/71 (3%) unrevised cups termed unstable. Revisions: (i) cups: 12/63 (19%) – all due to loosening; stem: 1/63 (1.6%), due to dislocation; (ii) cups: 3/74 (4%) – 1 due to loosening; stem: 1/74 (1.4%) – due to loosening.	C
Zichner & Willert, 1992 (Germany)	Retrospective Müller-type with femoral heads: (i) Protasul-2 (ii) Protasul-10 (iii) Ceramic Cemented vs. ceramic (Metal vs. ceramic heads)	Not specified but same surgeons for each group Not specified (Probably teaching hospital, but same clinic) Not specified but same technique for all	354 in 313 patients (Original cohort number unknown) (i) 149 (ii) 105 (iii) 100 Follow-up (range): (i) 66 months (30–108) (ii) 46 months (30–84) (iii) 73 months (30–102) 354	Not specified	Displacement rates (wear rates): (i) in those requiring revision, all > 0.2 mm/y, 40% > 0.3 mm/y; in non-revision group, 40 (29.9%) > 0.2 mm/y, 11 (8.2%) > 0.3 mm/y; (ii) in those requiring revision, all > 0.2 mm, 4 > 0.4 mm/y; in non-revision group, 20 (20%) > 0.2 mm; (iii) 63% < 0.1 mm/y, 95% < 0.2 mm/y, no prosthesis > 0.3 mm; *p*, ? Revisions/loosening: (i) 15 (10.0%) revised due to loosening; (ii) 5 (4.8%) revised due to loosening; (i) + (ii), 20 (7.9%); (iii) 2 (2%) revised due to loosening; *p*, ?	C

Data tables of observational studies

Studies are grouped in the following order: Charnley studies, other cemented models, cementless porous-coated, cementless HA-coated, cementless uncoated press-fit, hybrid, cementless mixed, threaded cups.

The results presented in the tables are for the latest follow-up unless otherwise stated. Scores given for clinical rating systems (e.g. Harris) are means for the patient groups unless otherwise stated. Where numbers of hips followed-up are given in parentheses (e.g. (222)), this refers to the number in the total series, the actual number reviewed for the published study then being given separately, and not in parentheses.

The 'total mechanical failure' rate is the number of revisions plus failures as defined by radiological criteria; these vary from study to study but generally include definitions of loosening, migration, stability, and fracture of components.

DATA TABLE 3 *Observational studies: Charnley*

Study, country, rating	Number followed-up (duration of follow-up)	Age	Outcome measures; results	Notes/comments
Ahnfelt, *et al.*, 1990 Multicentre registry, Sweden **C**	15,520 Charnley only 1799 at 10 years (10 years)	Not specified for subgroup	Survival 92% at 10 years	
Brady & McCutchen, 1986 USA **C**	(170) 155 followed-up (10 years)	Not specified	Revision rate 8.8%; n = 3 revisions for loose stem	3 revisions in heavy patients at 9 years; 1st generation cementation – precise technique
Carlsson, *et al.*, 1986 Sweden **B/C**	207 68.7% (207 osteoarthritis, 34 rheumatoid arthritis) (5–12 years 5 months (osteoarthritis)) (3.4–12 years (rheumatoid arthritis))	Not specified	Osteoarthritis 26% loosening; rheumatoid arthritis 34% loosening	
Carter, *et al.*, 1991 UK **C**	1616 31% (10–20 years)	Not specified	Survival 91% at 10 years; survival 82% at 20 years	
Collis, 1988 USA **C**	180 37% (10+ years)	Not specified	3.3% revision	
Dall, *et al.*, 1988 South Africa **B**	98 (mean 12 years, range 10–14)	Mean 61 years 87% > 50 years	n = 4 (4%) stems loose; 14% revised – 4 loose cups, 1 stem loose, 7 stem fracture, 2 recurrent dislocations, hence, 5% revisions for loose + 6% radiography failures	Osteoarthritis 76%
Dall, *et al.*, 1988 South Africa **Not rated**	2059 (mean 10 years 5 months, range 3–17 years)	Not specified	9.1% total revision rate (loose cup 2.1%, loose stem 4.9%); + 6% possible radiography failure	1st generation stems fractured more frequently but more loosening in stiffer 2nd generation stems
Dall, *et al.*, 1993 South Africa **B**	811 Charnley 66.2% (10–12 years)	Mean 60 years (range not specified)	Survival 87% at 10–12 years; 8% revised	
Eftekhar, 1987 USA **C** (see also Eftekhar & Tzitzikalakis, 1986 below)	1009 (20% revision/conversions) 69% (5–15 years)	Not specified	2.0% revisions; 3.8% mechanical failure	
Eftekhar & Tzitzikalakis, 1986 USA **B** (same patients as previous)	499 primaries + 197 revisions (696) (5–15 years; just over 25% followed-up at > 10 years; mean not specified)	Mean 62.4 years; range 22–67 years	4.5% total mechanical + infection failure rate 2.2% re-operation of primaries, + 1.2% (n = 6) pending failure	48% osteoarthritis; single surgeon
Garcia-Cimbrelo & Munera, 1992 Spain **A**	680 60% at 10 years (18 years)	Mean 56 years; range 18–79 years	Survival 81% at 18 years; survival 91.6% at 10 years; pain 4.6 at 17 years (d'Aubigne–Postel and Charnley); walking 4.6 at 17 years; range of motion, 4.4 at 17 years	
Gudmundsson, *et al.*, 1985 Denmark **B**	186 67.2% (n = 125) (10–14 years)	Median 71 years; range 31–85 years	29% loosening; 86% no/slight pain; 58% normal/slightly limited range of motion	District General Hospital
Hamilton & Joyce, 1986 Canada **B** (same patients as Hamilton & Gorczyca, 1995)	(450) 230 followed up for 3–11 years about 100 at 6 years	86% > 50 years	Revision rate as loose: cup 0.7%, stem 0.0%; n = 14 stem subsidence; n = 6/230 cup migration (2.6%)	Community hospital; large percentages of death and loss to follow-up not accounted for; 61% patients dislocation/childhood conditions, etc.; single surgeon; Charnley method; flanged versions introduced during study

continued

PAPER 3

DATA TABLE 3 contd *Observational studies: Charnley*

Study, country, rating	Number followed-up (duration of follow-up)	Age	Outcome measures; results	Notes/comments
Hamilton & Gorczyca, 1995 Canada **B** (same patients as Hamilton & Joyce, 1986)	224 83% at 10 years (minimum 10 years (maximum 20 years) mean not specified)	Mean 58 years	Mean d'Aubigne score between 5 and 6 every year except 19th. 12.5% cup migration; 6.7% cup revision; 2.6% (n = 5/188) stem migration/ subsidence/ fracture, 6.3% stem revision	Shows strong association between wear and migration/revision; single surgeon
Hartofilakidis, *et al.*, 1989 Greece **B**	104 89% (10–14 years)	Mean 57 years; range 24–82 years	78.5% asymptomatic 5.5 pain 5.1 function 4.9 motion	20% total revisions (7% for aseptic loosening)
Hodgkinson, *et al.*, 1993 **C**	Cup: unflanged 152 83.5% (1–10 years)	Not specified	30.3% no radiological demarcation at 10 years	
Hodgkinson, *et al.*, 1993 (same study as above) (see Comparative studies)	Cup: flanged 150 89.3% (1–10 years)	Not specified	42.7% no radiological demarcation at 10 years; flanged compared with randomly selected unflanged. Flanged socket better than unflanged by radiography criteria at 10 years, statistically significant. Previously revised hips excluded	
Johnsson, *et al.*, 1988 Sweden **C**	204 100% (4–14 years)	Males: median 65 years, range 36–87 years; females: median 67 years, range 47–84 years	Revisions 14.7%	
Johnston & Crowninshield, 1985 USA **C**	326 55.8% (n = 182) (10 years)	Not specified	9% femoral loosening; 7.9% acetabular loosening	
Joshi, *et al.*, 1993 UK **A**	(218) 166 (mean 16 years, range 10–24 years)	Mean 32 years, range 16–40 years	At 20 years: total survey 75%; stem surgery 86%; cup surgery 84%; Aseptic loosening: stem: 3% at 10 years; 14% at 20 years; cup, 4.5% at 10 years, 16% at 20 years; 4% stem subsided > 5 mm	Wrightington; survival analysis and SEMs; significantly greater failure risk in years 10-20, and in osteoarthritis compared with rheumaoid arthritis; (osteoarthritis risk of revision 20% at 10 years, nearly 49% at 20 years); small head, tapered stem
Karachalios, *et al.*, 1993 Greece **C**	95 Charnley 57.9% (12–18 years; average 13 years 5 months)	Not specified	27.4% cups migrated; 15.8% stems subsided; survival related to centre of rotation of prosthesis and body weight	
Kavanagh, *et al.*, 1989 USA **C**	333 49.8% (15 years)	Males: mean 65 years, range 38–85 years; females: mean 64 years, range 39–84 years	Probability of failure: at 1 year, 0.9%; 5 years, 4.1%; 10 years, 8.9%; 15 years, 12.7%	
Kobayashi, *et al.*, 1994a; 1994b USA/Japan (2 studies) **A/B**	(703) 326 stems followed-up for mean 13 years 3 months 328 cups followed-up for mean 13 years 2 months (10–20 years)	Mean 58 years, range 18–79 years	Charnley mean 16.1 (max. 18) 1.2% (4) stems revised; 4.9% (16) radiologically stem fix failure. 7.4% (24) sockets revised; 17% (56) radiological failure; n = 9/703 revisions for stem fail excluded from the follow-up series	Factors in mechanical loosening study, e.g. canal width; about 1/3 congenital dysplasia, about 1/3 osteoarthritis; includes table of 12 previous > 10-year follow-up studies of radiographically diagnosed Charnley cup failures for aseptic loosening (this study is the largest)
Langlais, *et al.*, 1995 France **Not rated**	(446) 48 (mean 6 years 5 months)		11% trochanteric non-union 1.3% (6) re-operations for instability	48 followed-up for mechanism of loosening, osteolysis; stems only followed-up

continued

DATA TABLE 3 contd *Observational studies: Charnley*

Study, country, rating	Number followed-up (duration of follow-up)	Age	Outcome measures; results	Notes/comments
Madey, *et al.*, 1997 USA **A/B**	356 142 followed-up (minimum 15 years)	Mean 69 years, range 24–88; for > 15-year follow-up, mean 62 years	Total revision rate 9%, 5% for aseptic loosening. At minimum 15 years: total revision rate for aseptic loosening, 11% (stem 2%, cup 10%)	Patient satisfaction measured; survival analysis; relatively high dislocation rate attributed to small head; 2nd generation cement technique; single surgeon
McCoy, *et al.*, 1988 USA **A/B**	100 40% (15–17 years)	Mean 60 years, range 25–84 years	87.5% excellent/good, 7.5% fair, 5.0% poor; 90.8% survival at 16 years; 96% survival at 15 years, cup only	
Neumann, *et al.*, 1996 Denmark **B** (see Comparative studies)	240 (15–21 years)	Young compared with older groups (see Comparative studies)	20 year revision rate for 11.7% younger patients; 10.7% for older; no significant difference between groups	Near-complete follow-up
Neumann, *et al.*, 1994 Denmark **B** (superseded by Neumann, *et al.*, 1996)	241 96% survivors (n = 103) (15–20 years)	Median 62 years, range 34–79 years	Probability of revision 10.7% at 20 years; 7% < 3 for pain movement (Charnley score) 30% loosening	
Nicholson, 1992 New Zealand **C**	185 100% (15–22 years)	Not specified	Revision > 13%; cup loosening 17.7%, survival 90.9%; stem loosening 21.9%, survival 88.1%	
Older & Butorac, 1992 UK **B**	388 34% (17–21 years)	Mean 68 years, range 42–85 years	Revision 6%; loosening 17% cups, survival 89% at 20 years (cup and stem)	District General Hospital
Older, 1986 UK **C**	217 70.5% (n = 153) (10–12 years)	Median 64 years, range 42–55 years	88% satisfactory; 6% revision; 92% patients satisfied	
Picault & Michel, 1995 France **C?**	786 for 10–15 years 290 for 15–19 years 107 for 15–23 years (15–23 years)	Not specified	d'Aubigne (15–20 years): pain 5.8; mobility 5.7; walk 5.4; 84% pain-free at 15 years; 7.7% stem subsidence; survival 85%	
Ranawat, *et al.*, 1989 USA **Not rated**	152 (17+ years)	Not specified	72% survival (revision)	
Rasmussen, *et al.*, 1991 Denmark **B?**	95 (10 years)	Not specified	Survival 85%; 14/15 revisions for aseptic loosening; 71% pain free (82% of non-revised); stem subsidence (> 5 mm) in 9%	
Schulte, *et al.*, 1993 USA **C**	322 Charnley 98.5% (20+ years)	Mean 65 years, range 29–86 years	90% survival (retained implant); 85% pain free; 53% no walking aids; 10% revised	
Skeie, *et al.*, 1991 Norway **A**	629 Charnley 89.7% (10–15 years)	Mean 66 years, range 23–88 years	92% survival at 13 years; 86% patients good result; 7% revised	District General Hospital
Solomon, *et al.*, 1992 USA **B/C**	(156) 130 Mean 10 (3–16 years)	Mean 38 years; all < 50 years, 53% 41–50 years	Mechanical failure survival 88% at 10 years; radiological loosening in 12%; d'Aubigne mean score 14.8	Contains table (9) reviewing published cemented results in young patients; follow-up range 2 years 8 months–12 years; revisions for mechanical failure in follow-up > 5 years, 2.6–21.2%
Stauffer, 1982 USA **B**	207 90% (10 years)	Mean 64 years, range 39–84 years	Revisions 10.8%; cup loosening 11.3%, stem loosening 29.9%	
Sullivan, *et al.*, 1994 USA **A/B**	(89) 84 (mean 18 years, range 16–22 years)	Mean 42 years, range 18–49 years	Cups 13% (11) revised for aseptic loosening; stems 2% (2) for mechanical failure; survival for aseptic loosening: cup 76% ± 12; stem 92% ± 12; total mechanical failure including radiographically: cup 50%, stem 8% (> 5 mm)	Survival + CIs; good follow-up rate; polished stem; old cementation; single surgeon

continued

PAPER 3

DATA TABLE 3 contd *Observational studies: Charnley*

Study, country, rating	Number followed-up (duration of follow-up)	Age	Outcome measures; results	Notes/comments
Terayana, 1986 Japan **C**	107 (> 5 years)	Not specified	At 5 years: n = 1 revision for loosening + 1 pending; loose cup in +2, stem subsidence in +2, 2 conversions (8/107 failures, 7.5%)	Most patients women; 50% osteoarthritis secondary to congenital dysplasia
Thomas & McMinn, 1991 UK **C**	1069 Charnley (10+ years)	Not specified	92% survival at 10 years; no improvement following change of cement techniques	
Wejkner & Stenport, 1988 Sweden **B**	325 50% (10–14 years)	Mean 64 years, range < 30 to > 80 years	56% excellent, 28% good, 8% fair, 8% failure (Charnley scores)	
Welch, *et al.*, 1988 USA **B/C**	(100) 97 but small numbers followed-up (15–17 years)	Mean 65 years, range 30–88 years	16% revised; mean time to revision, 10.8 years	72% osteoarthritis
Wroblewski, 1986 UK **B**	116 Charnley (15–21 years)	Mean 53 years, range 20–71 years	85.3% pain free; 78% full range of movements; subsidence 29%; socket migration 22.5%	
Wroblewski & Siney, 1993 UK (Wrightington) **C**	1324 Charnley 193 reviewed (18–26 years, average 10 years)	Mean 47 years, range 24–68 years	Dislocation 0.63%; revisions not specified, estimated as about 13–14% from survival graph (dislocation + stem fracture + loosenings) 16 years (from 1324); infection 0.3–1.5%; pain free 85% (from 193); normal function 60%	

DATA TABLE 4 *Observational studies: cemented – non-Charnley*

Study, country, rating	Prosthesis type	Number followed-up (duration of follow-up)	Age	Outcome measures; results	Notes/comments
Ballard, *et al.*, 1994 USA **A/B**	Single surgeon; 2nd generation cemented technique; mixed prosthesis designs	42 (mean 11 years, range 10–15 years)	Mean 41 years, range 18–49 years	About 25% (n = 10) revisions (10 cups aseptic loosening, 2 stems)	Severe disease, some failures in two young patients; three patients receiving renal dialysis.
Bosco, *et al.*, 1993 USA **A/B**	CAD and HD-2; 2nd generation cemented technique	86 (48 + 38) (mean 6 years 7 months or 6 years 4 months, range 2–14 years) (67 followed-up for minimum 5 years)	Mean 55 years, range 20–80 years	Hospital for Special Surgery rating: 71% satisfactory; 5.8% (n = 5) revised for aseptic loosening (3 for both components, 1 stem, 1 cup); 10-year survival rate: age > 60 years, 57% ± 20; age < 60 years, 50% ± 22; difference not significant; radiographically: 22% definite cup failure; 30% definite/possible stem loosening.	No significant difference reported between designs but no data given; weight not associated with radiographic cup outcome; borderline association stem outcome/weight; no association age/outcome; no significant difference earlier/later implants; significant correlation radiographical criteria/clinical score.
Dorr, *et al.*, 1994 USA **C**	Various: Charnley or Charnley–Müller, Aufrance–Turner, LeGrange–Letournel	49 (mean 16 years 2 months)	Mean 31 years, range 16–45 years	d'Aubigne, 27% satisfactory; 67% revised for aseptic loosening. All patients aged < 30 years revised or pending; 13/16 cups and 3/25 stems pending revision.	20 osteoarthritis Reviews other studies with high cup failure in very young patients; recommends non-arthroplasty treatment.
Fowler, *et al.*, 1988 UK **B**	Exeter; early cementation	241 at 5–10 years 121 at 11–16 years (mean 13 years 4 months, range 11–16 years)	Mean 66 years 8 months, range 30–84 years	Total mechanical failure 11% (revision rate, not specified); 1.64% stem loose; 3.9% cup loose; 5.4% fractures (attributed to early structural defect); cups revised, n = 6; 74% stems no sign of loosening.	73% osteoarthritis; includes 5-year follow-up of 2nd generation cemented series; extensive radiographic analysis.
Harris & Penenberg, 1987 USA **B/C**	Metal-backed cup, maker not specified	(48) 29 primary (mean 11 years 5 months, range 10–13 years 5 months)	Mean 44 years, range 34–76 years	Mean Harris score, unrevised 92; no revisions for loosening; 13.8% (n = 4) radiographic loosening.	12.5% (48) revised for loosening; revisions age-related.
Hirose, *et al.*, 1995 USA **B**	Variety of designs; all stems cobalt–chrome, most with collars; cups mixed – metal-/non-metal-backed; 2nd generation cement technique	(192) 131 (mean 7 years, range 5–12 years)	Mean 65 years, range 22–85 years	Johnston: pain none/mild 95%; walk satisfactory 84%; (other factors not specified). Cup mechanical failure rate 18.4%, including 9.6% revised (minimum 5-year follow-up). Stem mechanical failure rate 3.1%, including 2.3% revised.	60% primary osteoarthritis.
Karrholm, *et al.*, 1994 Sweden **A/B**	Stem only – Lubinus SP I, plus cemented polyethylene cups	58 primary (+ 26 revisions) (median 5 years 10 months, range 4 years 9 months–7 years 10 months)	Median 68 years, range 41–83 years	Stem revisions, n = 9, (10.7%) – 7 for thigh pain + radiographic loosening (but 6 were re-revisions) + 1 primary for osteolysis at 8 years. RSA of migration: logistic regression found migration at 2 years best predictor of failure (probable revisions > 50% if subsidence > 1.2 mm at 2 years).	Discusses other studies on failure prediction; percentages of revisions may mean results not applicable to primary THR.
Mohler, *et al.*, 1995 (a) USA **Not rated**	Stem only – Iowa Hip (Zimmer); mixture of cemented/non-cemented cups	1941 (2–10 years)	(Mean 59 years – failed hips only)	1.5% (29) loose at mean of 5 years, 1.1% revised.	Study of loosening/osteolysis. This type of failure not found in polished Charnley stems in study by same authors.
Ohlin, 1990 Sweden **B**	Christiansen	265 (median 6 years)	Not specified	Radiographic survival at 10 years: stem 67%, cup 0%; 1/3 revised for aseptic loosening.	Abandoned design.
Ohlin & Onsten, 1990 Sweden **B**	Lubinus	202 for survival; 151 for clinical follow-up (mean not specified, range 3–6 years)	Median 68 years, range 29–94 years	3% revised for aseptic loosening (n = 3 cups, 2 stems, 2 both); 13% cups loose at 5 years; 10% stems loose at 5 years; clinical function not specified.	Hip dysplasia only factor associated with loosening; age < 65 years associated with higher rate of revision risk.

continued

DATA TABLE 4 contd *Observational studies: cemented – non-Charnley*

Study, country, rating	Prosthesis type	Number followed-up (duration of follow-up)	Age	Outcome measures; results	Notes/comments
Oishi, *et al.*, 1994 USA **B**	Stem only – Harris Precoat; 3rd generation cemented technique	>100 88 for clinical follow-up (mean 7 years, range 6–8 years)	Mean 71 years, range 41–92 years	Harris score 91; 97% excellent/good; thigh pain 3%. 1% (1) stem revision for loosening; no other loosening (0/81).	Good results attributed to cementing technique and precoating of prosthesis with cement to decrease risk of de-bonding; osteoarthritis 74%.
Partio, *et al.*, 1994 Finland **A**	Lubinus; traditional cemented technique; stem design changed to anatomic in 1982	444 (mean 10 years 2 months, range 8–12 years)	Mean 64 years	Total revision rate 11.5% (loosening plus technical error); estimated survival rate 87% at 10 years. Aseptic loosening: cup + stem 6.5%; stem only 2.5%; cup only 2.1%. No hip score reported.	Most frequently used cemented prosthesis in Finland; 71% osteoarthritis; no significant differences for cup/stem survival, osteo- and rheumatoid arthritis, cup size, stem design, weight or gender groups; lower survival for age < 65 years.
Roberts, *et al.*, 1987 UK **A**	Howse	(506) 265 at 10 years 34 at 15 years (mean not specified, range 10–15 years)	Mean 63 years, range 19–89 years	90% survival at 10 years; 80.8% survival at 15 years. 8.3% revised at 10 years; total revisions 54, 42 at < 10 years, 29 for aseptic loosening. Total failed including clinical/radiographic 11.8%; revisions for aseptic loosening 4.35% at 10 years; revision for stem fracture 3.16% (especially in younger males).	Osteoarthritis 60%; senior + junior surgeons.
Rockborn & Olsson, 1993 Sweden **B**	Exeter; matt stem surface; 2nd generation cementation	(143) 110 radiographic/clinical follow-up (minimum 5 years, mean not specified)	Mean 71 years, range 39–83 years	Charnley score: pain none/mild 78%. 5.6% revision rate (8/143 – 6 stems + 2 cups loose); radiographically, 21% definite/probable stem loosening, 3.6% cup loosening.	Osteoarthritis 78%; no association between loosening and age; poor stem results attributed to poor cementing and too large stem component; matt surface may prevent distal movement of stem within cement mantle.
Russotti, *et al.*, 1988 USA **A**	(Harris design); HD-2 stem; four common cemented cups	(251) (mean 5 years 6 months, range 5–7 years)	Mean 63 years, range 22–90 years	Harris score 97; 98% excellent. Stem loose (definite/probable/possible) in 2.4%; cup migration, n = 1.	
Saito, 1992 Japan **Not rated**	Bioceramic; ceramic head/UMWH cup	57 (mean 6 years 2 months, range 5–8 years)	Mean 52 years 8 months, range 31–70 years	d'Aubigne total 16.6 (pain 5.7, walk 5.2, range of motion 5.7); 93% excellent/good. 1 revision at 6 years for stem loosening; no cup revisions, no ceramic head breaks; 7% (4) cups radiographic loosening, 3.5% (2) stems loose.	All osteoarthritis secondary to congenital dysplasia – high risk for cup failure. Wear not correlated with loosening but with calcar resorption. Authors suggest same bearing surface suitable for cementless implant in younger patients. Includes table comparing with three other prostheses.
Thomas, *et al.*, 1986 USA **A/B**	CAD; (minimum stress, maximum fix area; bulky rigid stem)	(114) 74 minimum 5 years follow-up (mean 7 years 1 month, range 5–10 years)	Mean 57 years, range 20–77 years	9% (7) revisions, at 6–10 years (loose); survival 77% at 9 years (revisions) or 73.7% (revision + radiographic criteria). 87% unrevised excellent/good.	
Tompkins, *et al.*, 1994 USA **A/B**	Stem only – Triad (Johnson & Johnson); titanium stem, cobalt–chrome 28 mm head, collar	(142) 116 followed-up (mean 4 years 10 months, range 2–8 years)	Mean 63 years, range 18–88 years	Hospital for Special Surgery rating: mean 32.7; 92% excellent/good. Survival (loosening) 89% ± 3% at 4 years; 4.3% revision (done/pending).	Authors abandoned this design, quoting Russotti, 1988. Advise roughening or precoating of stem for cemented implant or choice of cobalt–chrome material; poor canal fill achieved in this series. Osteoarthritis 60%.
Warren, *et al.*, 1993 UK **A/B**	Furlong; straight stem; titanium alloy tapered in two planes	(195) 148 followed-up (mean 4 years 4 months, range 3–? years)	Mean 66 years	Harris score: 86. 6 year survival (revision) 97% (89.7–100%); 6 year survival (revised or loose) 79% (62.3–95.8%); (failures: 7 cups, 2 stems, 1 both).	All grades of staff undertook operations. Significant association between Harris scores and radiographic evaluations.

DATA TABLE 5 *Observational studies: cementless – porous-coated (some hybrid with porous-coated acetabular components)*

Study, country, rating	Prosthesis type	Number followed-up (duration of follow-up)	Age	Outcome measures; results	Notes/comments
Bourne, *et al.*, 1994 Canada **A/B**	PCA; stem	101 (5 years) Osteoarthritis only	Mean 61 years (range 26–81 years)	Harris score 90; thigh pain in 27% at 5 years.	No association between subsidence/thigh pain severity.
Cordero-Amuero, *et al.*, 1994 Spain **A/B**	PCA, Howmedica; cup	128; 113 reviewed (mean 5 years, range 4–8 years)	Mean 51 years (range 24–71 years)	9 cup revisions; radiological: 40/75 neutral cups stable, 7/27 vertical cups. Fixation improved in 2 years in 12, worsened in 26. Harris score excellent/good in 85% (good in 61/64 stable cups, and 28/40 unstable cups). Bead loosening was progressive.	Uses Kaplan-Meier; teaching hospital; Hardinge's direct lateral.
Engh & Massin, 1989 USA **B**	AML; stem (+ cups)	343; 204 for 5 years (mean 4 years 9 months) Subsets: (i) 200 with adequate fixation (canal-fill), mean follow-up 4 years 2 months; (ii) 143 without, mean follow-up 4 years 9 months	Means: (i) 58 years; (ii) 57 years	X-ray: 7% stems unstable at minimum 2 years postoperatively. Stable fix survival 94% at 5 years, 88% at 8 years. d'Aubigne-Postel:at 5 years: pain 5.7, walk 5.7, some thigh pain/limp in 9.4% patients. (i) mild pain 7.8%, moderate 0%; (ii) mild pain 20%, moderate 14.7%. Revision rate 4.4% (15, including 11 cemented cups for aseptic loosening, 3 stems, none for aseptic loosening). ** For subset < 55 years + no rheumatoid arthritis (n = 107), survival rate for stem fixation, 92% at 9 years (i.e. no difference from overall series). Combined mechanical failure rate, 6.4% at 5 years.	Attempts to control for suboptimal cup fixation in subgroup analysis. Statistically significant difference in survival canal-filling vs. under-sized stems. X-ray fixation and clinical results positively correlated.
Engh, 1993 USA **C**	AML, Dupuy; stem; (metal-backed porous-coated cups)	393/227 (mean 8 years, range 1–13 years radiographically)	Not specified	Revision rate 4.4%; revision rate porous-coated stems 1.5%; stem revision rate for loosening 0.7% (3/393); 5/227 porous-coated cups revised (2.2%); revision rate for cemented cups 7.5% (11/166), survival 81.2% at 10 years; 11-year stem survival 91.8%. Overall failure rate include radiographic failures 10.8% (18/166). Revision rate later group: stems 0.5%, 9 year survival 99.3% (1/227); overall failure rate including radiographic 1.8%.	Author is originator of the AML system; reports on 2 models; hybrid subset reported. No clinical function results.
Engh, 1994 USA **C**	AML; stem	(226) 166 complete follow-up (mean 10 years)	Not specified	1 revision – survival of 99.5% at 10 years.	Includes autopsy retrieval study. Osteo-integration stated in 98% if press-fit method correct.
Engh, *et al.*, 1997 USA **A/B**	AML; stem (+ cup); (+ porous-coated AML cups)	(223) (minimum 10 years; 174 minimum 10-year evaluation, 137 10-year X-ray)	Mean 55 years (range 16–87 years)	Re-operation rate 11.5% (20/174) – 3 loose stems/symptomatic; 3 dislocating cups; 4 cups loose/symptomatic; 10 impending cup liner wear-through – + X-ray, 2 loose stems. At 12 years: total survival 85%; stem survival 97% (SE 0.02); cup survival 92% (SE 0.03), Clinical (n = 147, 10 years): pain – 87% none or slight, 10% with limiting pain; walk – 82% without aids; thigh pain – 8.5%, 4% limiting.	Authors claim stem revision rate comparable to Charnley at 10 years (mean age 57 years) and 16 years (mean age 42 years) (cf. Sullivan, *et al.*, 1994). Patients with re-operation and/or osteolysis significantly younger than others. All loose stems were 'undersized'.

continued

PAPER 3

DATA TABLE 5 contd *Observational studies: cementless – porous-coated (some hybrid with porous-coated acetabular components)*

Study, country, rating	Prosthesis type	Number followed-up (duration of follow-up)	Age	Outcome measures; results	Notes/comments
Heekin, *et al.*, 1993 USA **A/B**	PCA; stem + cup	100 (5–7 years)	Mean 58 years (range 22–81 years)	Harris score 92 (5 years): survival 93% (at 5 years) with end-point cup migration/stem subsidence. Survival 98% revision-only end-point; revision rate 2%; stem subsidence 5%; pain – none/slight 75%; thigh pain increasing 18–26% in years 1–4, decreasing to 16% in year 7.	Detailed radiographic results; multivariable analysis of clinical vs. radiographic-results.
Hellman, *et al.*, 1997 USA **B**	Omnifit (Osteonics) cobalt–chrome; stems (mostly + cementless cups)	111; 79 reviewed (analysis suggests representativeness) > 5 years (mean 8 years 5 months, range 5 years 4 months–10 years 5 months)	Mean 45 years (range 19–71 years)	5.1% (4) stem revisions – 2.5% (2) for aseptic loosening. Pain: 96% none or slight, 4% moderate; limp: none or slight, 97.3%. Radiographic (n = 72): 70 stable (97.2%) with signs of bone ingrowth; osteolysis 12% (9) treated by bone grafting (thus total mechanical failure rate, 3.8% (3)). Survival free of aseptic loosening at 10 years 91.3% (± 5.7%).	Osteolysis possibly related to polyethylene thickness; discusses other prostheses evidence for osteolysis.
Holman & Tyer, 1992 Australia **B/C**	PCA; stem + cup	318 (1–6 years; mean not specified)	Mean 53 years (range 17–71 years)	Revisions: 1% (3) – 2 undersized stems, 1 cup loosening at 3 years in rheumatoid arthritis patient. Harrington ARS 100-point (pain/function/gait/motion/deformity): good/excellent 80%; 13% some thigh pain.	
Incavo, *et al.*, 1993 USA **B**	Harris-Galante and Optifix; cup	106 Harris-Galante 66; Optifix 40 (minimum 2 years, range 2–4 years 4 months)	Harris-Galante, mean 63 years; Optifix, mean 61 years	2 Harris-Galante cups revised, 1 migration, 1 dislocation; no other loosening. (No function measures)	No statistical correlation with migration/radiolucency of: age, sex, cup coverage, component inclination, number of fixing screws.
Jansson & Refior, 1992 Germany **B/C**	PCA; stem (+ screw cups)	81 (mean 2 years 5 months, range 1 year 2 months–3 years 4 months)	Mean 56 years	1 revision; d'Aubigne score mean 13.6.	Includes patient satisfaction (7 not satisfied); X-ray results counter theory of osseo-integration.
Kienapfel, *et al.*, 1991 Germany **B**	BIAS; stem, modular (cup Harris-Galante)	40 (2 years 6 months)	Mean 50 years	Mean Harris score 90.7% at 2 years (good/excellent 91.6%). Radiographic: 95% stable; no cup migrations.	Small sample, short follow-up; no statistically significant differences in clinical results between stable/unstable groups; various surgical approaches.
Kim & Kim, 1992 South Korea **B/C**	Harris-Galante; stem (+ cup)	82 (mean 5 years 2 months, range 5–5 years 6 months)	Mean 52 years (range 24–86 years)	Harris score, 83; 62% excellent/good. 10% stems loose (revised or to be revised); 28% (20) thigh pain in non-loose stems. Radiolucency > 2 mm in 33%.	
Kim & Kim, 1993 South Korea **A/B**	PCA; stem + cup	116 (6 years 1 month–7 years 5 months)	Mean 48 years (range 19–85 years)	Harris score 91 (latest); 88% excellent/good at 6 years. 3 cups loose (+ 20 excessively worn liners); 7 stems loose; osteolysis in 33%. 17% with good stem fit had thigh pain (17/98), 9% persistent thigh pain (7/9 stems – loose-fit).	Wear related to young age but not weight, sex, diagnosis, hip score, hip movement.
Lachiewicz, 1994 USA **B/C**	Harris-Galante, titanium alloy; stem + cup (both coated, some screw-fix)	35 (mean 4 years 5 months, range 3 years–6 years 5 months)	< 60 years; mean 41 years (range 18–59 years)	Harris score, mean 91; 81% good/excellent. No revisions for aseptic loosening. X-ray: 1 definite cup loose; 3 stems non-progressively loose.	Medication recorded; rheumatoid arthritis patients.

continued

DATA TABLE 5 contd *Observational studies: cementless – porous-coated (some hybrid with porous-coated acetabular components)*

Study, country, rating	Prosthesis type	Number followed-up (duration of follow-up)	Age	Outcome measures; results	Notes/comments
Learmonth, *et al.*, 1995 South Africa **A/B**	PCA; stems	104 (mean 4 years 2 months, range 2 years–6 years 5 months)	Mean 43.4 years (range 16–67 years)	d'Aubigne score, 94% clinically excellent; thigh pain 23% (severe in 2); revision rate 1.9% (2, both loose).	Comments on osteolysis, radiographic results.
Maloney, *et al.*, 1992 USA **C**	ARC, Howmedica; cup – beaded, screw-fix (+ cemented stems), i.e. hybrid	56 (mean 4 years 7 months)	Not specified	19.6% bead loosening (increasing over time), 1 associated with migration, 1 with broken screw.	
Moskal, *et al.*, 1994 USA **B/C**	PCA; stems (+ uncemented peg-fix PCA cups)	100 (2–4 years)	Mean 63 years (range 27–95 years) (no previous arthroplasty in followed-up group)	Harris score at 2 years, 90; thigh pain 5% at 3 years; limp in 18% (believed unrelated to prosthesis or surgical approach); 99% stable.	Conducted in community hospital; compared porous-coated to hybrid (n = 34): no statistically significant differences (in spite of mean age in hybrid group being 12 years greater). Stem head larger than used conventionally – attributed with good early results by authors.
Negre & Henry, 1995 France **B**	TA6V (authors' model); stem + cup, blasted titanium with press-fit cup	101 (6 years)	Mean 68 years (range 30–88 years)	d'Aubigne clinical: 94% 'perfect', 4% mild pain; 2 revisions for ceramic head fracture; 2 stem migrations, 3 stems loose due to poor intramedullary fit.	Theory of the design is to allow bone-ingrowth without fibrous layer between bone–metal.
Owen, *et al.*, 1994 UK **A**	PCA (Howmedica)	241 (mean 5 years, range 2–9 years)	Mean 47 years (range 18–65 years)	Overall survival (for recommendation for revision) 91% at 6 years (± 6%); 73% at 7 years (± 11%); 57% at 9 years (± 20%). 6 cup failures due to loosening in 6 years; 6 stem failures in total (one intraoperative fracture) at mean of 4 years (all had poor original intramedullary fit). Osteolysis in 36% cups (n = 99) surviving > 5 years, 13% in stems. Subsidence 4% in stems > 5 years.	Analysis with CIs. Mean age at revision 39 years; cup failure at mean of 6 years; 20/26 have widespread osteolysis; all had loose beads + excessive polyethylene wear; 12 had migrated. Low overall survival caused by huge decline in survival of cups in years 6–9: 30% (n = 95) attributed to severe polyethylene wear (large head size, 32 mm was used), osteolysis and migration. Mixed surgeons; specialist centre; lateral without trochanteric osteotomy.
Pellegrini, *et al.*, 1992 USA **B**	Tri-lock; stem, beaded	57; 51 reviewed (mean 6 years 5 months, range 5–8 years)	Mean 49 years	Harris score 84%, good/excellent; Mayo 70% good/excellent; excluding hips with previous major procedures; Harris score, 88% good/excellent. 1 revision for aseptic loosening, 1 for persistent pain; subsidence in 2 stems, 1 > 5 mm.	Small sample size; cobalt–chrome; long follow-up for beaded; patients selected for high-risk early failure cemented implant. Poorest results in hips with previous procedures.
Schmalzried & Harris, 1992 USA **B**	Harris-Galante; cup (screw-fix)	111 cups; 83 reviewed (mean 5 years 8 months, range 5–7 years)	Mean 59 years (range 23–79 years)	Harris score: mean 93 (73–100). No cup loosening, 4 cup revisions – 2 liners detached, 1 metallosis, 1 lysis. No continuous radiolucent line around whole cup.	Comparison of cemented (n = 40) vs. non-cemented (43), porous-coated stems: Harris scores 95 vs. 92 – caused by pain scores 43 cemented vs. 40 non-cemented; i.e. hybrid marginally better than non-cemented/non-cemented. Senior surgeon.
Shaw, *et al.*, 1992 USA **B**	AML; stem (bipolar cups)	178; 154 for analysis of which 122 complete (mean 3 years 4 months, minimum 2 years)	Mean 57 years 7 months	92.3% stable; 9% postoperative groin pain. Harris score: 84.	Patient satisfaction; no relationship age/Harris score; no relationship sex/type of stabilisation.
Sotereanos, *et al.*, 1995 USA **B**	Two series: (supplied separately) (i) BIAS; (ii) AML; stems (i) + cemented cup (ii) + porous-coated cup	(i) 121; (ii) 166 ((i) mean 10 years 2 months, range 7–15 years (ii) mean 8 years 3 months, range 7–12 years)	(i) mean 53 years (ii) mean 53 years 8 months	(i) 5 (4.1%) revisions at mean of 10 years 2 months (2 for late loosening); survival 95.4% at 11 years (13 cup revisions). (ii) 1 (0.6%) stem revised for loosening, 99.3% survival at 9 years. 94.6% pain-free at last follow-up; 3 stems X-rayed unstable, 2 significant osteolysis.	Patient satisfaction also measured. Pro-cobalt–chrome.

continued

DATA TABLE 5 contd *Observational studies: cementless – porous-coated (some hybrid with porous-coated acetabular components)*

Study, country, rating	Prosthesis type	Number followed-up (duration of follow-up)	Age	Outcome measures; results	Notes/comments
Tang-Kue, 1995 Japan **Not rated** (abstract only available)	PCA; stems and cups (2-peg fix)	119 (mean 7 years)	Mean 46 years (range 19–78 years)	Harris score: 95; 92.4 excellent/good; 24.4% slight or > pain walking; 5.9% considered unstable; no revisions.	
Xenos, *et al.*, 1995 USA **B**	PCA; stems and cups	100 (minimum 7 years)	Mean 58 years (range 22–81 months)	Harris score: 92.4. 5% total revision – 2% stem, 3% cup; osteolysis around stem in 11%, cup in 2%, both in 2%. Most patients with osteolysis asymptomatic.	Osteolysis study: osteolysis occurred frequently around components with no evidence of migration/subsidence; mean age of osteolytic group younger by 10 years than others.

DATA TABLE 6 *Observational studies: cementless – HA-coated*

Study, country, rating	Prosthesis type	Number followed-up (duration of follow-up)	Age	Outcome measures; results	Notes/comments
Capello, 1994 USA **C**	Osteonics; stem	(436) 151 for 5 years	Mean 50 years	Harris score, 95: pain, 93% none/slight; thigh pain, 1.3% mild/moderate. Subsidence > 3 mm, n = 2; revisions, n = 10 (5 for pain, aseptic loosening); none loose. Total mechanical failure, 0.46% (excl. pain).	Multicentre
d'Antonio, *et al.*, 1992a USA **A**	Omnifit (Osteonics); stem	(238) (92 for minimum 2 years)	Mean 48 years	Harris score, 95 at 2 years; pain, 5% mild to moderate. Revisions, 2/238 (0.8% within 2 years).	Presumably earlier set from same series as previous study. Patients stated to be more active and heavier than in most comparable studies. Similar cup comparison results as for study above (23 HA-, 69 porous-coated). 5 centres
d'Antonio, *et al.*, 1992b USA **A/B**	Osteonics; stem and cup	320 (minimum 2 years); 142 (minimum 3 years)	Mean 50 years	Harris score, 95 at 3 years; pain, 4.2% mild to moderate, thigh pain, 1.4%. No revisions; stems – aseptic loosening 2, X-ray unstable 0, total 0.46%; cups – 1% migration at 2 years.	Comparison of HA- (132) vs. porous-coated ingrowth (285) cups showed no statistically significant difference in clinical Harris scores at any time up to 3 years. Multicentre
Drucker, *et al.*, 1991 USA **C**	No model name, authors' experimental design; stem and cup	58 (6 months–2 years, mean 10 months)	Mean 53 years, range 22–73 years	Not specified	
Geesink, 1990 The Netherlands **A/B**	Omnifit (Osteonics); stem and cup	100 (85 primary reviewed) (1 year 5 months–3 years 3 months, mean 2 years)	Mean 54 years, range 21–74 years	Harris score, 97; persisting pain 4%. No loosening. Harris score by cup type: HA-coated 98 vs. non-coated 94 (but at 3 months 90 vs. 71; at 6 months 95 vs. 79).	Harris cup score comparisons. HA-coated vs. non-coated contradicts d'Antonio, *et al.*, 1992b study.
Geesink & Hoefnagels, 1995 The Netherlands **A**	Omnifit (Osteonics) stem and cup	118 stems; 100 cups, threaded design only (5 years 6 months–7 years 6 months)	Mean 53 years, range 21–65 years (31 patients < 50 years)	Survival: stem 100%, cup 99%. Harris score: at 3 months 90, at 6 years 98. d'Aubigne: at 6 years – pain 5.8; motion 5.9; walk 5.9 (total 17.6). Persisting pain 4%; no osteolysis.	No association of age, gender, surgeon or weight with Harris scores at any period. Harris scores compare with < 90 for most porous and press-fit series, author claims. Notable early pain relief.
Koch, *et al.*, 1993 Germany **B?**	Furlong	233 (190 primary) (2–5 years, mean 2 years 9 months)	Not specified	No aseptic loosening; no thigh pain. d'Aubigne: 15.76.	In German.
Rossi, *et al.*, 1995 Italy **A/B**	ABG Howmedica; stem and cup	100 (minimum 2 years)	Mean 63 years	d'Aubigne: 100% excellent/good. 0% mechanical failure (1 dislocation due to cup malpositioning). No cup migration; postoperative bone–cup gaps disappeared in 3–12 months.	More details of radiographic findings given.
Tonino, *et al.*, 1995 International (Europe) **A/B**	ABG, Howmedica; stem and cup	222 (minimum 2 years, mean 2 years 4 months)	Mean 62 years 7 months	3.6% thigh pain. 4 early + 3 late dislocations (total, 3%; 2-year mechanical revision rate, 1.4%). Activity: preoperative 14.9%, at 2 years 87.3%. d'Aubigne: mean 17.4 (max. 18). Minor stem migration in 6 (2.7%); normal bone adjacent to cup in 95%.	International study; 10 surgeons; all dislocations from same single centre. No influence on clinical scores of age, weight, gender, disease, Charnley classification. No statistical correlation between radiographic results and clinical scores. Very detailed radiographic analysis.

PAPER 3

DATA TABLE 7 *Observational studies: cementless – uncoated press-fit*

Study, country, rating	Prosthesis type	Number followed-up (duration of follow-up)	Age	Outcome measures; results	Notes/comments
Blaha, *et al.*, 1994 USA **C**	CLS stem (Protek, Switzerland); collarless, tapered wedge, grooved, rough-blasted surface	300 (minimum 5 years)	Not specified	Revision rate for mechanical loosening: 1.7% (5 – 2 cup/ 2 stem, 1 fractured ceramic head) + radiographic loosening of 2 stems (total mechanical loosening rate for stem, 1.6%). Pain: 89% none/slight; Harris score > 85 (89%); thigh pain 1.3% (4).	Author claims results as good as for porous-coated stems. Poorly reported but evaluation was by independent observers.
Duparc & Massin, 1992 France **B/C**	Bichat 3 stem (Howmedica); smooth, fluted titanium	(203) 2 years: clinical 157; radiographic 145 (46 for 4 years, maximum 6 years)	Mean 57 years (range 18–85 years)	32 revised; survival 77% at 6 years (revision end-point). d'Aubigne: at 2 years 89% excellent/ good (non-revised).	Indications: use of this design now restricted to patients in whom cement is contraindicated by history of previous infection or very young age.
Glorion, *et al.*, 1994 France **A/B**	Osteal cup (Ceraver), polyethylene screwed; stem: cemented, 32 mm head	77 (mean 3 years 6 months, range 1–7 years)	Mean 63 years (range 25–76 years)	Migration-free survival 74.5% at 9 years; revision-free survival 92%. Abandoned.	All osteoarthritis.
Harper, et *al.*, 1995 UK **B**	Ring UPM cup; wedge press-fit plus Ring uncemented stems (87) or Norwich cemented stems (39)	126 (mean 4 years 5 months, range 1–7 years 6 months; 59 for mean 6 years 5 months)	Mean 63 years (range 31–93 years)	(Total revisions 22%); 17% revised for loosening. Survival 83% (76.8-89.2) at 8 years. No function data given. Polyethylene press-fit concept abandoned.	Life-table survival analysis. Mean time to granulomatous loosening 5 years 3 months; failure attributed to polyethylene wear. Results compared with other studies of Ring prostheses.
Kennedy, 1994 USA **C**	Arthrophor I cup (Joint Medical Products); press-fit (screw/peg-free, metal-backed)	488 (most 3–6 years, minimum 2, maximum 8 years)	Not specified	No revisions for loosening. Osteolysis in 3.1%; loose beads in 3%.	Press-fit interface of 1.5 mm, reamed exactly; author attributes success to this.
Kutschera, *et al.*, 1993 Austria **Not rated**	Zweymuller peg-free stem	96 (mean 5 years 3 months, range 5–5 years 9 months)	Mean 67 years 3 months (range 41–87 years)	Mean Harris score: 87.5. 1 cup revised for aseptic loosening; no stem revisions; 1 stem subsidence of 4 mm.	Abstract only.
Seral, *et al.*, 1992 Spain **B/C**	Zweymuller peg-free stem; cup: Endler polyethylene threaded	260 (mean 5 years, range 4–6 years)	69% 50–70 years (8%, 70+ years)	Singh: 67% very good/good. Cup migration, 17.6%; stem subsidence > 4 mm, 27%. Osteoarthritis group (152): 78.5% very good. No revisions reported.	
Stockley, *et al.*, 1992 Canada **B**	Müller straight stem; designed for cementing	24 (mean 7 years 3 months, range 6 years 2 months–8 years 3 months)	Mean 61 years (range 46–77 years)	Harris score: mean 79. 5 revised for aseptic loosening. 1 failed clinically. Survival 80% at 8 years.	Very small sample. Pre-dates porous coating (1982–84). Authors recommend titanium rather than cobalt–chrome.
Wilson-MacDonald, *et al.*, 1990 Switzerland **B**	RM cup: pegged polyethylene, some screw-fixed; isoelastic; plus Müller cemented stem	445 (5–10 years)	Mean 65 years (range 29–89 years)	d'Aubigne: 86% excellent/good. Revisions for aseptic loosening, n = 32, most > 8 years; about 28% radiologically loose at 9 years. Abandoned.	Good results up to 6 years. Smaller cups and use of screws associated with more loosening; increased wear in younger patients. "Bone/polyethylene contact should be avoided."

PAPER 3

DATA TABLE 8 *Observational studies: hybrid*

Study, country, rating	Prosthesis type	Number followed-up (duration of follow-up)	Age	Outcome measures; results	Notes/comments
Harris & Maloney, 1989 USA **B**	ARC cup, n = 52, HD-2 stem (Howmedica); Harris-Galante cup/Precoat, n = 74, (Zimmer) stem	126 (mean 3 years 6 months, range 2–5 years 7 months)	Mean 63 years (range 23–83 years)	Harris score: 93. No revisions for loosening.	Harris Galante/Precoat better than ARC HD-2 clinically. Single senior surgeon.
Helfen, *et al.*, 1993 Germany **Not rated**	Marburg porous-coated cup; nails plus peg; titanium stem	212 (3–6 years)	Mean 60 years (range 33–76 years)	96% very good/good. 1 revision for loosening.	Abstract only.
Kienapfel, *et al.*, 1992 Germany **B**	Harris-Galante porous-coated cup; Griss stem (Sulzer AG), titanium with ceramic head	(40) 33 followed-up (mean 3 years 3 months, range 2 years 10 months–5 years)	Mean 55 years (range 32–70 years)	Pain: 16% mild to moderate (most post hard activity); limp slight to moderate in 24%. 1 cup possibly unstable; no revisions.	
Mohler, *et al.*, 1995 USA **A/B**	Harris-Galante cup and stem; porous-coated cup, screw-fix	153 120 clinical review 109 X-ray (mean 5 years 2 months, range 4 years–7 years 1 month)	Mean 67 years (range 39–85 years)	Harris score: 86 (90 after excluding patients with unrelated problems). Pain: none, slight, mild in 97%. No revisions; 2% (2) definite stem loosening; 2% (2) definite cup loosening, others stable. Survival 95% (95–100) at 7 years 1 month (cup 98.5%, stem 96.6%).	Authors support hybrid for older patients. 4 senior surgeons plus assistants.
Pearse, *et al.*, 1992 UK **B/C**	Harris-Galante porous-coated screwed cups; Müller straight cobalt–chrome stems	58 (mean 3 years 6 months, range 2 years 6 months–5 years 6 months)	Mean 53 years 5 months (range 28–82 years)	Harris score: 91; 91% excellent/good. 1 stem revision for loosening (in patient with previous cemented THRs). No stem migration; 1 cup progressive radiolucency, no cup migration.	
Schmalzried & Harris, 1993 USA **B**	Comparison of: (i) ARC cup, HD-2 stem (Howmedica) (ii) Harris-Galante cup, Harris Precoat stem (Zimmer) Both stems collared; both cups screw-fixed	(101) 97 followed up (i) 52; (ii) 49 (mean 6 years 5 months, range 5–8 years)	Mean 61 years (range 23–83 years)	Harris score: 93; 91% good/excellent. 90% no or slight pain; Pain less for Harris-Galante group (statistically significant). 1 revision for stem loosening – in a **custom** component; no stems loose; no Harris-Galante cup loose or revised for loosening; 1 Harris-Galante cup revised for liner failure. 2 ARC cups migrated; no cup revisions for loosening. Osteolysis in 2 ARC cups, none in Harris-Galantes.	Harris score slightly better for Harris-Galante/Precoat group. Bead loosening, etc., reported.

DATA TABLE 9 *Observational studies: cementless – mixed types*

Study, country, rating	Prosthesis type	Number followed-up (duration of follow-up)	Age	Outcome measures; results	Notes/comments
Lautiainen, *et al.*, 1994 **B/C**	Two macro-interlock designs (i) Lord Madreporic (ii) Link	(i) 49 (ii) 20 (mean 5 years 4 months, range 4 years 9 months–7 years 7 months)	Mean 58 years (range 36–76 years)	Mayo: 86.7; 78% good. Total revision rate 6.3%.	No correlation between radiographic and clinical ratings.
Niinimaki, *et al.*, 1994 Finland **B**	RM; stem only; macro-interlock	(114) 85 71 SSTRR questionnaire/ radiographic (7–9 years)	Mean 64 years (range 48–79 years)	Harris score: 43% excellent/good; Johnston: 67% pain none/slight. Revisions for loosening, n = 8, plus 1 fracture; total mechanical failure rate, 10.6%. Radiographic failure 25%; subsidence > 5 mm in 12%; osteolysis in 12%.	Osteoarthritis only.
Riska, 1993 **Not rated**	Ceraver Osteal hybrid/ceramic 2 series, cemented and uncemented cups	112 (mean 3 years 6 months, range 1–7 years)	Mean 62 years	Revision rate, 7% for uncemented cup; 1.7% revisions for aseptic loosening.	
Roffman & Juhn, 1993 Israel **C**	RM (Mathys, Switzerland); cup only; (isoelastic) (i) HMHDPE (ii) ditto HA-coated (iii) ditto titanium-coated	185 (i) 60 (ii) 96 (iii) 29 (9 years)	Not specified	No hip scores. Total revisions, 2.7% (n = 5), all in (i) (8.3% of group); none in (ii) and (iii). Migration, de-alignment or pain in same, 2.7%.	Good ingrowth in groups (ii) and (iii).
Stern, *et al.*, 1992 **C**	LD; some HA-coated, some roughened	112 (6 months–2 years; 60 for up to 2 years)	Mean 62 years	d'Aubigne: 86% excellent/good; pain 5.6%; limp 12%. Number loose not specified. 'Low' stress shielding.	Complex radiographic analysis.

DATA TABLE 10 *Observational studies: cementless – threaded cups (A sample of studies, not critically appraised; this design is now largely abandoned.)*

Study, country	Number followed-up (duration of follow-up)	Outcome measures, results
Bruijn, *et al.*, 1995 The Netherlands	411 (mean 4 years 6 months, range 3–7 years)	Clinical: 82% excellent/good. Migration 25%; 6% revised for aseptic loosening. Abandoned.
Fox, *et al.*, 1994 Canada	68 (mean 6 years, range 5–9 years)	38% failure; 17 revisions at mean of 5 years. Abandoned.
Gouin, *et al.*, 1993 France	107 (2–5 years)	Survival 75% at 5 years; revision rate 11.6%. d'Aubigne: excellent/very good/good, 62%. Abandoned.
Gut, *et al.*, 1990 Switzerland	102 (5–7 years)	33% sockets loose.
Harwin, *et al.*, 1991 USA	62 (mean 2 years 4 months)	8% re-operation, 10% failures including loosening.
Krugluger & Eyb, 1993 Austria	103 (minimum 10 years)	Revision rate 24% for loosening; loosening 33%; extensive osteolysis 31%. 5-year results had been good.

DADED 2

Appraisal Table I – RCTs

Key criteria

1. Method of randomisation identified and appropriate. 2. Patient groups balanced or effect of any difference evaluated in valid statistical analysis. 3. Patients blind to prosthesis type. 4. Assessments of clinical/radiological outcome blind to prosthesis type if possible. 5. Appropriate statistical analysis. 6. Number of patients deceased or lost to follow-up reported or included in statistical analysis. 7. Follow-up period – mean and range. 8. Prosthesis model specified. 9. Clearly defined criteria for measuring outcomes. 10. Age – mean and range.

Other criteria

11. Quantification of outcomes. 12. Follow-up data compared with preoperative data (preferably mean and range). 13. Independence of investigators (declared or no vested interest). 14. Numbers of men and women given. 15. Weight – mean and range. 16. Preoperative diagnoses with percentages/numbers of patients given. 17. Clinical evaluation independent of operating surgeon.

	Key criteria						Other criteria											
Study	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**Rating**
Bourne, *et al.*, 1995[a] Rorabeck, *et al.*, 1996[a]	n	y	y	y/y	NS	y	y	y	y	y	y[a]	y	y	y	n	y	y	C
Bradley & Lee, 1992	y	NS	NS	NS	NS	y	y	y	n	y	n	n	NS	n	n	y	y	C
Carlsson, *et al.*, 1995	n	NS	NS	NS	y	n	y	y	y	n	y	n	NS	n	n	n	NS	C
Ciccotti, *et al.*, 1994	n	y	NS	NS	y	y	y	y	y	y	n	y	NS	n	y	y	NS	C
Godsiff, *et al.*, 1992	y	y	y	y/NA	y	y	y	y	y	y	y	n	NS	y	n	y	y	A
Jacobsson, *et al.*, 1994*	n	y	NS	NS	n	y	y	y	y	y	y	n	NS	y	n	n	NS	C
Karrholm, *et al.*, 1994	y	y	NS	NS	y	n	y	y	y	y	y	y	y	y	y	y	n	C
Kelley, *et al.*, 1993	n	y	NS	NS	y	y	y	y	y	y	y	n	NS	y	y	y	NS	C
Krismer, *et al.*, 1994	n	y	NS	NS	y	y	y	y	y	y	y	n	NS	y	n	y	NS	C
Marston, *et al.*, 1996	y	NS	NS	NS	NS	y	y	y	y	y	y	y	y	y	n	y	y	C
Olsson, *et al.*, 1986	n	y	NS	NS	n	y	y	y	y	y	y	y	NS	y	y	y	NS	C
Onsten & Carlsson, 1994	n	y	NS	NS	y	y	y	y	y	y	y	n	y	y	y	y	NS	C
Onsten, *et al.*, 1994	y	y	NS	NS	y	y	y	y	y	y	y	n	NS	y	y	y	NS	C
Reigstad, *et al.*, 1986	n	y	NS	NS	n	y	y	y	y	y	y	y	NS	y	y	n	NS	C
Søballe, *et al.*, 1993	n	NS	NS	NS/y	y	y	y	y	y	y	y	y	y	n	y	y	NS	C
Thanner, *et al.*, 1995*	n	NS	NS	NS	n	y	y	y	y	y	y	y	NS	y	n	y	NS	C
Wykman, *et al.*, 1991	n	y	NS	NS	NS	y	y	y	y	y	y	y	y	y	y	y	NS	C

* *Studies not comparing typical prostheses.*

[a] *Same study group. Bourne, et al., 1995: clinical data 5-year follow-up; Rorabeck, et al., 1996: radiographic data 4-year follow-up.*

PAPER 3

Appraisal Table 2 – Non-controlled comparative studies

Key criteria

1. Method of assignment of patients to different prostheses described and appropriate. 2. Patients matched or differences evaluated in valid statistical analysis. 3. Appropriate statistical analysis undertaken. 4. Number of patients deceased or lost to follow-up reported or included in analysis. 5. Follow-up period, range and mean specified. 6. Prosthesis models specified. 7. Clearly defined criteria for measuring outcomes. 8. Age – mean and range.

Other criteria

9. If retrospective, patients selected without knowledge of outcomes. 10. In prospective studies, follow-up assessments blind to prosthesis type, if possible. 11. Results given for specific models (and sizes). 12. Quantification of outcome criteria. 13. Follow-up data compared with preoperative data (mean and range). 14. Independence of investigators (declared or no vested interest). 15. Numbers of men and women given. 16. Weight – mean and range. 17. Preoperative diagnoses with percentages/numbers of patients given. 18. Clinical evaluation independent of operating surgeon. 19. Radiological evaluation independent and blinded to clinical results.

	Key criteria								Other criteria										
Study	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**Rating**
Abrahams & Crothers, 1992	n	n	y	n	y	n	y	y		NS	y[a]	y	n	NS	y	n	y	NS	NS	C
Ahnfelt, *et al.*, 1990	n	n	y	n	n	n	n	n	n		n*	n	n	NS	y	n	y	NA	NA	C
Bankston, *et al.*, 1993	y	y	y	n	y	y	y	y	y		y	y	n	NS	y	y	n	NA	NS	C
Bankston, *et al.*, 1995	y	y	NS	n	y	y	y	y	y		y	y	n	NS	y	y	n	NA	NS	C
Bertin, *et al.*, 1985	n	n	NS	y	y	y[b]	n	n		NA	n	y	n	NS	y	n	y	NS	NS	C
Britton, *et al.*, 1996	n	NS	y	y	y	y	y	n		n	y	n	n	y	n	n	n	n	n	C
Burkart, *et al.*, 1993[c]	y	n	y	y	y	y	y	y		y	y	y	n	NS	y	n	y	NS	y	C
Callaghan, *et al.*, 1995	n	n	n	n	n	y	y	n	NS		y/n	y/n	n	y	n	n	n	NS	NS	C
Carlsson & Gentz, 1982	n	n	NA	y	y	y	y	n	n		n	y	n	NS	n	n	n	NS	y	C
Chmell, *et al.*, 1995	y	n	NS	n	n	y	n	n	NS		y	n	n	NS	n	n	n	NS	NS	C
Cornell & Ranawat, 1986*	y	y	NS	y	y	y	y	y	y		n	y	n	NS	y	y	y	NS	y	C
Dall, *et al.*, 1993*	y	n	n	y	y	y	y	y	y		y	y	n	y	y	n	y	NS	NS	C
Duck & Mylod, 1992	n	n	NS	n	y	n	y	y	NS		n	n	n	NS	y	n	y	NS	NS	C
Ebramzadeh, *et al.*, 1994	y	NA	y	y	y	y	y	n	y		n	y	n	y	y	n	y	NA	y	C
Espehaug, *et al.*, 1995	n	n	y	y	y	y	y	n	y		y	y	n	NS	y	n	n	NA	NA	C
Freeman & Plante-Bordeneuve, 1994	n	n	y	n	n	n	y	y		NS	y[a]	y	n	y	y	n	y	y	NS	C
Goetz, *et al.*, 1994	y	y	NS	n	y	y	y	y	y		y	y	y	y	n	y	y	NS	NS	C

* *Studies not comparing typical prostheses.*
[a] *Results given for type of prosthesis, not specific model.*
[b] *Unclear if the prostheses used in this study were/are now in widespread use or designed for this study only.*
[c] *Appraisal covers new data given in Burkart,* et al., *1993*[99] *only. Bourne,* et al., *1994 is appraised in Appraisal Table 5.*

continued

PAPER 3

continued

	Key criteria								Other criteria											
Study	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**Rating**
Hamada, *et al.*, 1993	y	n	n	n	y	y	y	y	y		y	y	n	NS	y	n	y	NS	NS	C
Havelin, *et al.*, 1994	n	n	y	y	y	n	y	y	y		n	y	n	NS	y	n	n	NA	NA	C
Havelin, *et al.*, 1995	n	n	y	y	y	y	y	y	y		y	y	n	NS	y	n	y	NA	NA	C
Hearn, *et al.*, 1995	y	n[d]	y	y	y	y	y	y	y		n	y	y	NS	y	n	y	y	NS	C
Hedlundh & Fredin, 1995*	n	y	y	y	n	y	y	y	n		y	n	n	NS	y	n	y	NS	NA	C
Hernandez, *et al.*, 1994	n	y	y	n	n	y	y	n	y		y	y	n	y	n	n	y	NS	NS	C
Hodgkinson, *et al.*, 1993	n	y	y	y	y	y	n	y	y		y	y	n	NS	y	y	y	NS	NS	C
Hoffman, *et al.*, 1994	n	n	y	y	n	y	y	y	y		y[e]	n	n	NS	y	n	y	NS	NA	C
Hotikoshi, *et al.*, 1994	n	n	NS	n	y	y	y	y	n		n	y	n	NS	y	n	y	NS	NS	C
Hozack, *et al.*, 1993	n	n	y	n	y	y	y	y		NS	y	y	y	NS	y	y	y	NS	NS	C
Hozack, *et al.*, 1994	n	y	y	y/n[f]	y	y/n[f]	y	y	NS	NS	n	y	y	NS	y	y	y/n[f]	NS	NS	C
Huracek & Spirig, 1994	y	y	y	y	y	y	n	y	y		y	y	n[g]	NS	y	n	y	y	NS	C
Hwang & Park, 1995	n	n	n	y	y	y	y	y		NS	y	y	y	NS	y	n	y	NS	NS	C
Jacobsson, *et al.*, 1990	n	n	y	y	y	y	y	y		NS	n	y	n	NS	y	y	n	NS	NS	C
Kelley & Johnstone, 1992*	y	y	y	y	y	y	y	y	y		y[h]	y	n	NS	y	n	y	NA	y	A
Krismer, *et al.*, 1991a	n	n	y	y	y	y	y	y	y		y	y	n	NS	y	n	y	NS	NS	C
Krismer, *et al.*, 1991b	n	n	y	y	y	y	y	y	y		y	y	n	NS	y	n	y	NS	NS	C
Kristiansen & Steen Jensen, 1985*	NA	y	y	n	y	y	y	y	n		y	y	n	NS	y	y	y	NA	NS	B
Lehman, *et al.*, 1994*	y	y	y	y	y	y	y	y	y		n[i]	y	y	y	y	y	y	NS	NS	A
Mallory, *et al.*, 1989*	y	y	y	y	n	y	y	y	n		n	y	y	NS	y	n	y	NS	NS	C
Maloney & Harris, 1990	y	y	NS	n	y	y	y	y	y		y	y	y	y	y	y	y	NS	NS	C

* *Studies not comparing typical prostheses.*
[d] *Although the same patients were evaluated for two types of prostheses, operations were, on average, 4 years apart, so there were differences in many variables.*
[e] *Detailed results given for five out of nine prosthesis models.*
[f] *First part of study prospective observation, second part retrospective comparison.*
[g] *Harris Hip Score only section to have pre- and post-operative scores compared, even though many other sections could have been assessed in this way.*
[h] *Only loosening rates given for specific models.*
[i] *Only failure rate results for specific models.*

continued

	Key criteria								Other criteria											
Study	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**Rating**
Markel, *et al.*, 1995	y	n	n	y	y	y	y	y	y		y	n	n	NS	n	n	n	n	NS	C
McPherson, *et al.*, 1995	n	y	y	n	n	y	y	y	y		y	n	n	NS	y	y	n	NS	NS	C
Moilanen, *et al.*, 1996	n	n	y	n	y	y	y	y		y	y	y	y	y	y	n	y	y	n	C
Moskal, *et al.*, 1994	y	n	NS	y	n	y	y	y		n	y	y	y	NS	y	y	n	n	NS	C
Nashed, *et al.*, 1995	n	y	n	y	y	y	y	y	y		n	n	n	NS	y	y	y	NA	NS	C
Neumann, *et al.*, 1996*	NA	NA	y	y	y	y	y	y		NA	y	y	y	y	n	n	y	NS	NS	B
Pierchon, *et al.*, 1994*	n	n	y	y	n	n	n	n	n		n	y	n	y	n	n	n	NS	NS	C
Pritchett, 1995	n	n	y	n	n	y	y	n	n		y	y	n	NS	y	n	y	NA	NS	C
Pupparo & Engh, 1989	n	n	NS	y	y	y	y	n		NS	y	y	n	NS	n	y	y	NS	NS	C
Ranawat, *et al.*, 1988*	y	y	y	y/n	y	y	y	y	y		n	y	n	NS	y	y	y	NS	NS	B
Rand & Ilstrup, 1983	n	y	y	n	y	y	y	y	y		y	y	n	NS	y	n	y	NS	NS	C
Riska, 1993	y	n	n	y	y	y	y	y	NS		y	n	n	NS	y	n	y	NS	NS	C
Ritter & Gioe, 1986	y	y	NS	y	y	y	y	y		NS	y	y	y	y	y	n	y	NS	NS	C
Ritter, 1995	n	n	y	y	y	y	y	y	y		y	y	n	NS	y	n	y	n	NA	C
Schreiber, *et al.*, 1993	y	n	y	y	y	y	y	y	y		y	y	n	NS	y	n	n	NS	NS	C
Schuller & Marti, 1990	n	n	y	y	y	y	y	y	y		y	y	n	NS	y	y	n	NS	NS	C
Turner, 1994	n	n	y	n	n	y	y	n	n		y	y	n	NS	y	n	n	NS	NS	C
Visuri, *et al.*, 1994	n	y	y	y	y	y	y	y	y		y	y	n	NS	y	n	n	NA	NA	C
Walker, *et al.*, 1995	NS[†]	n	y	n	n	y	y	n	n		n	y	n	y	y	n	y	NA	NS	C
Wilson-MacDonald & Morscher, 1989	n	n	y	y	n	y	y	y	y		y	n	n	NS	n	n	y	NS	NS	C
Wixson, *et al.*, 1991	y	n	y	y	y	y	y	y		NS	n	y	y	y	y	n	y	NS	NS	C
Yahiro, *et al.*, 1995	Meta-analysis																			
Zicat, *et al.*, 1995	y	y	y	n	y	y	y	y	y		y	y	n	y	y	y	y	NA	NS	C
Zichner & Willert, 1992	n	n	n	n	y	y	y	n	y		y	y	n	NS	n	n	n	NS	NS	C

* *Studies not comparing typical prostheses.*
† *Assignment of patients to different prosthesis described in Marston,* et al., *1996*[120] *(RCT).*

PAPER 3

Observational studies – appraisal criteria

Key criteria

1. Method of selection of patients identified. 2. Prosthesis models specified. 3. Results given for specific models. 4. Follow-up period, range/mean specified. 5. Number of patients deceased or lost to follow-up reported or included in analysis. 6. Ages – mean/range. 7. Preoperative diagnoses of reviewed patients specified with percentages/numbers. 8. Clearly defined criteria for measuring outcomes/quantification of outcomes.

Other criteria

9. Valid statistical analysis. 10. Outcome data compared with preoperative data. 11. Data given for deceased patients. 12. Clinical evaluation independent of operating surgeon. 13. Radiological evaluation independent and blinded to clinical results. 14. Numbers of men/women specified. 15. Weight range/mean specified. 16. Surgical technique/approach specified. 17. Grade/experience & number of surgeons specified. 18. Type of hospital/centre (general/specialist/teaching) specified. 19. Unilateral/bilateral results separate. 20. Independence of investigators (vested interest) specified.

An abbreviated form of these criteria is given in a footnote to *Appraisal Tables 3–9.*

APPRAISAL TABLE 3 *Observational: cemented – Charnley*

	Key criteria								Other criteria												
Study	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**Rating**
Ahnfelt, *et al.*, 1990	y	y	y	n	n	n	y		y	n	n	NA	NA	y	n	n	n	y		NA	C
Brady & McCutchen, 1986	?	y	y	y/n	NA	y	y	y	n	n	NA	n	n	y	n	y	y	n	n	n	C
Carlsson, *et al.*, 1986	?	y	y	y	y	y	y	y	?	NA	n	n	y/n[a]	n	n	y	y	y	n	n	B/C
Carter, *et al.*, 1991	n	y	y	y	y	n	n														C
Collis, 1988	y	y	y/n	y	y	n	n		n	n	n	n	n	n	n	y	y	y		n	C
Dall, *et al.*, 1988	y	y	y	y	y	y	y	y	?	n	n	n	n	y	n	n	n	y	n	n	B
Dall, *et al.*, 1988	n?	y	y	y	n?	n?	n?	n?	Abstract of conference paper												?
Dall, *et al.*, 1993	y	y	y	y	y	y	y		y	n	n	NS	NS	y	n	n	n	n		n	B
Eftekhar & Tzitzikalakis, 1986	y	y	y	y/n	y	y	y	y	?	NA	y	n	n	y	n	y	y	y	n	n	B
Eftekhar, 1987[a]	n	y	y	y	n	n	n														C
Garcia-Cimbrelo & Munera, 1992	y	y	y	y	y	y	y		y	y	n	NS	NS	y	y	y	y/n	y		y	A
Gudmundsson *et al.*, 1985	y	y	y	y	y	y	y		NS	n	n	NS	NS	y	n	y	n	y		n	B
Hamilton & Joyce, 1986	y	y	y	y	y	y/n	y	y	?	y	n	n	n	y	n	y	y	y	n	n	B
Hamilton & Gorczyca, 1995	?	y	y	y	y	y	y	y	y	n	n	n	n	y	n	y	y	y	n	n	B
Hartofilakidis, *et al.*, 1989	y	y	y	y	y	y	y		NS	n	n	n	n	y	n	y	y	y		n	B
Hodgkinson, *et al.*, 1993	see Comparative studies																				C
Johnsson, *et al.*, 1988	y	y	y/n	n	y	y	y		y	n	y	NS	NS	y	n	y	n/y	y		n	C

[a] *New data only assessed. ? Doubtful that criterion met; not clear from paper.*

Key: 1. Selection of patients. 2. Prosthesis models. 3. Results for models. 4. Follow-up period. 5. Loss to follow-up/deceased. 6. Ages. 7. Preoperative diagnoses. 8. Outcomes clear/quantified. 9. Statistical analysis. 10. Comparison with preoperative data. 11. Data on deceased. 12. Clinical evaluation independent. 13. Radiological evaluation independent. 14. M/F numbers. 15. Weight. 16. Surgical technique. 17. Surgeons' grade, etc. 18. Type of hospital. 19. Bilateral results separate. 20. Independence of investigators.

continued

PAPER 3

APPRAISAL TABLE 3 *contd* *Observational: cemented – Charnley*

	Key criteria								Other criteria											
Study	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**Rating**
Johnston & Crowninshield, 1983	y	y	y	y	y	n	n		n	n	n	NS	n	n	n	y	y	n		n	C
Joshi, *et al.*, 1993	y	y	y	y	y	y	y	y	y	y	y	n	n	y	y/n	n	y/n	y	n	y	A
Karachalios, *et al.*, 1993	y	y	y	y	y	n	y		y	n	n	NS	NS	y	n	y	y	y		n	C
Kavanagh, *et al.*, 1989	y	y	y	y	y	y	n		y	n	n	NS	NS	y	n	y	n	y		y	C
Kobayashi, *et al.*, 1994a	y	y	y	y	y	y	y	y	y	y	n	n?	y	y	n	y	y	y	n	n	A/B
Kobayashi, *et al.*, 1994b	y	y	y	y	y	y	y	y	y	y	n	n?	y	y	n	y	y	y	n	n	A/B
Langlais, *et al.*, 1995	Mechanics of loosening study – not appraised																				?
Madey, *et al.*, 1997	?	y	y	y	y	y	y	y	y	y	y	n	n	y	NS	y	y	y	n	y	A/B
McCoy, *et al.*, 1988	y	y	y	y	y	y	y		y/n	n	y	NS	NS	y	y	n	n	y		n	A/B
Neumann, *et al.*, 1994	y	y	y	y	y	y	y		y	n	n	NS	NS	y	n	y	n *	y		y	B
Neumann, *et al.*, 1996	see Comparative studies																				B
Nicholson, 1992	n	y	y	y	n	n	n									n	n/y	y		n	C
Older, 1986	y	y	y	y	y	n	y		n	n	n	y	NS	n	n	n	y	y		n	C
Older & Butorac, 1992	y	y	y	y	y	y	y		y	n	n	NS	NS	y	n	n	n?y	n		n	B
Picault & Michel, 1995	?	y	y	y/n	n	n	n	n	Abstract only												C?
Ranawat, *et al.*, 1989	Conference abstract																				?
Rasmussen, *et al.*, 1991	?	y	y	y/n	?	y/n	y	?	Abstract only published												B?
Schulte, *et al.*, 1993	y	y	y	n	y	y	y		n	n	y	NS	NS	y	n	y	y	y		y	C
Skeie, *et al.*, 1991	y	y	y	y	y	y	y		y	n	y	NA	NA	y	n	y	n/y	y		n	A
Solomon, *et al.*, 1992	?	y	y	y	NA?	y	y	y	y	n	NA?	n	n	y	n	n	n	n	n	n	B/C
Stauffer, 1982	y	y	y	y	y	y	y		n	n	n	NS	NS	y	n	y	n	y		n	B
Sullivan, *et al.*, 1994	?	y	y	y	y	y	y	y	y	n	y	n	n	y	n	y	y	y	n	y	A/B
Terayama, 1986	?	y	y	y/n	y	y	y	?													C
Thomas & McMinn, 1991	n	y	y	n	n	n	n														C
Wejkner & Stenport, 1988	y	y	y	y	y	y	y		y	n	n	NS	NS	y	n	y	y	y/n		n	B
Welch, *et al.*, 1988	y	y	y	y	y	y	y	y	n	n	n	n	n	y	n	y	n	n	n	n	B/C
Wroblewski, 1986	y	y	y	y	y	y	y		y	y	n	NS	NS	y	n	n	n	y		n	B
Wroblewski & Siney, 1993	n	y	y	y	n	y	y		n	n	n	NS	NS	y	n	n	n	y		n	C

[a] *New data only assessed. ? Doubtful that criterion met; not clear from paper.*

Key: 1. Selection of patients. 2. Prosthesis models. 3. Results for models. 4. Follow-up period. 5. Loss to follow-up/deceased. 6. Ages. 7. Preoperative diagnoses. 8. Outcomes clear/quantified. 9. Statistical analysis. 10. Comparison with preoperative data. 11. Data on deceased. 12. Clinical evaluation independent. 13. Radiological evaluation independent. 14. M/F numbers. 15. Weight. 16. Surgical technique. 17. Surgeons' grade, etc. 18. Type of hospital. 19. Bilateral results separate. 20. Independence of investigators.

PAPER 3

APPRAISAL TABLE 4 *Observational: cemented – non-Charnley*

	Key criteria								Other criteria											
Study	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**Rating**
Alsema, *et al.*, 1994	?	y	y	y	y	y/n	y	y	y	n	y	n	n	y	n	y	y	?	n	y	A/B
August, *et al.*, 1986	?	y	y	y	y	y	y	y	y	n	n	n	n	y	n	y	n	y	n	n	A/B
Ballard, *et al.*, 1994	?	y	y	y	y	y	y	y	y	n	y	n	n	y	n	y	y	y	n	y	A/B
Bohler, *et al.*, 1994	y	y	y	y	y	y	y	y	y	y	n	n	n	y	n	y	n	y	n	n	B
Bosco, *et al.*, 1993	?	y	y	y	y	y	y	y	y	n	n	y	y	y	y	y	y	y	n	n	A/B
Bryant, *et al.*, 1991	?	y	y	y	y	y	y	y	y	NA	y	n	n	y	n	n	n	y	n	n	B
Dorr, *et al.*, 1994	?	y	y	y	?	y	y	?	?	?	NA	n	n	y	n	n	n	y	n	n	C
Fowler, *et al.*, 1988	?	y	y	y	y	y	y	y	?	y	n	n	n	y	n	?	y	y	n	n	B
Harris & Penenberg, 1987	?	?	y	y	y	y	y/n	y	?	n	n	n	n	y	n	n	y	y	n	y	B/C
Helfen, *et al.*, 1993	y	y	y	y	y	y	y		y												?
Hirose, *et al.*, 1995	?	y	y	y	y	y	y	y	Abstract only												B
Jantsch, *et al.*, 1991	?	y	y	y	y	y	y	y	?	n	y	n	n	n	n	n	n	?	n	n	B
Karrholm, *et al.*, 1994	?	y	y	y	y	y	y	y	y	NA	n	NA	n	y	n	y	n	y	n	y	A/B
Lachiewicz & Rosenstein, 1986	?	y	y	y	?	y	y	y	?	n	NA	n	n	y	y	y	n	y	n	n	B/C
Mohler, *et al.*, 1995a																				y	?
Nizard, *et al.*, 1992	Mechanics of loosening study																			n	A/B
Ohlin & Onsten, 1990	?	y	y	?	y	y	y	y	y	n	y	n	n	y	y	y	n	y	n	n	B
Ohlin, 1990	y	y	y	y	y	y	y	y	y	n	n	n	n	y	y	y	y	y	n	n	B
Oishi, *et al.*, 1994	?	y	y	y	y	y	y	y	?	n	y/n	n	n	y	y	y	y	n	n	y	B
Papenfus, *et al.*, 1992	?	y	y	y	y	y	y	y	?	y	n	n	n	y	n	?	?	y	n	n	B/C
Partio, *et al.*, 1994	?	y	y	y	y	y	y	y	y	n	y	y	y	y	y	y	y	y	n	n	A
Pearse, *et al.*, 1992	y	y	y	y	y	y	y	y	n	n	y	n	n	n	n	y	y	y	n	n	B/C
Roberts, *et al.*, 1987	y	y	y	y	y	y	y	y	y	n	y	NA	NA	y	n	y	y	y	n	n	A
Rockborn & Olsson, 1993	?	y	y	?	y	y	y	y	?	n	n?	n	n	y	n	y	n	y	n	y	B
Russotti, *et al.*, 1988	y	y	y	y	y	y	y	y	y	y	NA	n	n/y	y	n	y	n	y	n	n	A
Thomas, *et al.*, 1986	y	y	y	y	y	y	y	y	y	NA	n	n	n	y	y	y	y	y	n	n	A/B
Tompkins, *et al.*, 1994	?	y	y	y	y	y	y	y	y	y	n	n	n	y	y	y	y	y	n	n	A/B
Warren, *et al.*, 1993	y	y	y	y	y	y	y	y	y	n	y	n	n	y	n	y	y	y	n	n	A/B

? Doubtful that criterion met; not clear from paper.

Key: 1. Selection of patients. 2. Prosthesis models. 3. Results for models. 4. Follow-up period. 5. Loss to follow-up/deceased. 6. Ages. 7. Preoperative diagnoses. 8. Outcomes clear/quantified. 9. Statistical analysis. 10. Comparison with preoperative data. 11. Data on deceased. 12. Clinical evaluation independent. 13. Radiological evaluation independent. 14. M/F numbers. 15. Weight. 16. Surgical technique. 17. Surgeons' grade, etc. 18. Type of hospital. 19. Bilateral results separate. 20. Independence of investigators.

APPRAISAL TABLE 5 *Observational: cementless – porous-coated*

	Key criteria								Other criteria												
Study	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**Rating**
Bhamra, *et al.*, 1992	n	y	y	y	n	n	y	y													C
Bourne, *et al.*, 1994	y	y	y	y	y	y	y	y	y	n	n	n	y	y	n	y	n	y	n	y	A/B
Callaghan, *et al.*, 1992	y	y	y	y	y	y	y	y	y	y	n	n	n	y	n	y	y	y	n	n[¨]	B
Cordero-Ampuero, *et al.*, 1994	?	y	y	y	y	y	y	y	y	y	n	n	n	y	y	y	y	y	n	n	A/B
Cracchiolo, *et al.*, 1992	y	y	n?	y	y	y	y	y	?	y	y	n	n	y	y	y	n	y	n	n	B/C
Engh & Massin, 1989	n	y	y	y	n	y	y	y	?	n	y	n	n	y	n	n	n	y	n	n	B
Engh, *et al.*, 1990	?	y	?	y	y	y	y	y	y	y	n	n	n	y	y	y	n	y	n	n	B
Engh, 1993	n	y	y	?	n	n	n	?													C
Engh, 1994	y[b]	y	y	y[c]	y/n	n	n	y													C
Engh, *et al.*, 1997[d]	y	y	y	min	y	y	y	y	y	y	y	n	n	y	y	y	n	y	n	n	A/B
Haddad, *et al.*, 1990	y	y	y	y	?	y	y	y	y	y	NA	n	y	y	n	y	y	y	n	n	A/B
Heekin, *et al.*, 1993	?	y	y	y	y	y	y	y	y	y	y	n	n	y	n	y	n	n	n	y	A/B
Hellman, *et al.*, 1997	?	y	y	y	y	y	y	y	?	n	NA	n	n	y	n	y	y	n	n	n[*]	B
Holman & Tyer, 1992	y	y	y	?	NA?	y	y	y	?	y	NA?	n	n	n	n	y	n	y	n	n	B/C
Incavo, *et al.*, 1993	?	y	y	y	NA	y	y	y	y	n	NA	n	n	y	n	n	n	y	n	n	B
Jansson & Refior, 1992	?	y	y	y	y	y	y	y	n	n	NA	n	n	n	n	n	n	y	n	n	B/C
Kienapfel, *et al.*, 1991	?	y	?[a]	y	y	y	y	y	y	y	n	n	n	y	n	y	n	?	y	n	B
Kim & Kim, 1992	?	y	y	y	?	y	y	y	?	n	NA?	n	n	y	y	y	y	y	n	y	B/C
Kim & Kim, 1993	?	y	y	y	y	y	y	y	y	y[a]	NA	n	n	y	?	y	n	y	n	y	A/B
Lachiewicz, 1994	y	y	y	y	NA	y	y	y	n	y	NA	n	n	y	y	y	y	y	n	n	B/C
Learmonth, *et al.*, 1995	y	y	y	y	y	y	y	y	y	y	n	n	n	y	n	n	n	y	n	n	A/B
Maloney, *et al.*, 1992	?	y	y	y	y	n	n	y													C
Martell, *et al.*, 1993	y	y	y	y	y	y	y	y	y	y	n	y	y	y	y[e]	y	y	y	n	n	A
Moskal, *et al.*, 1994	y	y	y	y	y	y	n	y	?	y	NA	y	y	y	n	y	n	y	n	n	B/C
Negre & Henry, 1995	?	y	y	y	y	y	y	y	?	n	n	?	?	y	n	?	n	n	n	n	B
Owen, *et al.*, 1994	?	y	y	y	y	y	y	y	y	?	y	n	n	y	n	y	y	y	n	y	A
Pellegrini, *et al.*, 1992	y	y	y	y	y	y	y	y	?	n	y	y	n	y	n	y	n	y	n	y	B
Schmalzried & Harris, 1992	?	y	y	y	y	y	y	?	?	n	n	n	n	y	n	y	y	y	n	y	B
Shaw, *et al.*, 1992	y	y	y	y	NA	y	y	y	y	y	NA	n	n	y	n	n	n	n	n	y	B
Smith, *et al.*, 1991 (M)	y	y	y	y	y	y	y	y/n	y/n	y	n	y	y	y	y	y	y	y	n	n	B/C
Sotereanos, *et al.*, 1995	y	y	y	y	y	y	y	y	?	y	?	n	n	n	n	n	n	n	n	n	B
Tang-Kuc, 1995	Abstract only available																				?
Xenos, *et al.*, 1995	?	y	y	?	NA	y	y	y	y	n	NA	n	n	y	n	n	n	y	n	y	B

[a] *Results for two models not disaggregated for clinical scores;* [b] *One hip only of bilaterals included;* [c] *Mean only;* [d] *Longest follow-up of the Engh AML studies – forthcoming 1997;* [e] *Measured but not stated.*
M = Modular component(s)

? Doubtful that criterion met; not clear from paper.

Key: 1. Selection of patients. 2. Prosthesis models. 3. Results for models. 4. Follow-up period. 5. Loss to follow-up/deceased. 6. Ages. 7. Preoperative diagnoses. 8. Outcomes clear/quantified. 9. Statistical analysis. 10. Comparison with preoperative data. 11. Data on deceased. 12. Clinical evaluation independent. 13. Radiological evaluation independent. 14. M/F numbers. 15. Weight. 16. Surgical technique. 17. Surgeons' grade, etc. 18. Type of hospital. 19. Bilateral results separate. 20. Independence of investigators.

APPRAISAL TABLE 6 *Observational: cementless – HA-coated*

	Key criteria								Other criteria												
Study	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**Rating**
Capello, 1994	?	y	y	y	y	y	y	y	?	n	n	n	n	y	n	n	n	n	n	n	C
Capello, *et al.*, 1994	y	y	y	y	y	y	y	y	y	n	n	n	n	y	y	y	n	y	n	n	B/C
d'Antonio, *et al.*, 1992a	?	y	y	?[b]	NA	y	y	y	?	y	NA	y	y	y	y	y	n	y	n	y	A
d'Antonio, *et al.*, 1992b	?	y	y	y	?	y	y	y	y	y	?	n	n	y	y	y	n	n	n	n	A/B
Drucker, *et al.*, 1991	y	y	y	y	y	y	y	y	Follow-up < 12 months												C
Geesink, 1990	y	y	y	y	y	y	y	y	y	y	?	n	y	y	n	y	n	y	n	n	A/B
Geesink & Hoernagels, 1995	?	y	y	y	y	y	y	y	y	y	NA	n	n	y	y	y	y	y	n	y	A
Koch, *et al.*, 1993	?	y	y	y	?	y	y	y	?	y											B?
Rossi, *et al.*, 1995	?	y	y	y	y	y	y	y	y	y	NA	n	n	y	n	y	n	y	n	n	A/B
Tonino, *et al.*, 1995	?	y	y	y	?	y	y	y	y	y	n	n	n	y	?[a]	y	n	n	y	n	A/B

[a] *Measured but not specified;* [b] *Only minimum follow-up period specified.*

? Doubtful that criterion met; not clear from paper.

Key: 1. Selection of patients. 2. Prosthesis models. 3. Results for models. 4. Follow-up period. 5. Loss to follow-up/deceased. 6. Ages. 7. Preoperative diagnoses. 8. Outcomes clear/quantified. 9. Statistical analysis. 10. Comparison with preoperative data. 11. Data on deceased. 12. Clinical evaluation independent. 13. Radiological evaluation independent. 14. M/F numbers. 15. Weight. 16. Surgical technique. 17. Surgeons' grade, etc. 18. Type of hospital. 19. Bilateral results separate. 20. Independence of investigators.

APPRAISAL TABLE 7 *Observational: cementless – uncoated press-fit*

	Key criteria								Other criteria												
Study	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**Rating**
Stems/stems and cups																					
Blaha, *et al.*, 1994	?	y	y	?	?[a]	n	n	y	n	n	y	y	y	n	n	n	n	n	n	n	C
Duparc & Massin, 1992	?	y	y	y	n	y	y	y	y	n	n	n	n	n	n	y	n	n	n	y	B/C
Groher, 1983[a]	y	y	y	y	n	y	y	y													C
Ivory, *et al.*, 1992[a]	n	y	n	y	n	n	n	n													C
Kutschera, *et al.*, 1993	Abstract only																				
Ring, 1978	3 different designs analysed in aggregate																				C
Ring, 1987	y	y	y	y	y	n	n	y													C
Seral, *et al.*, 1992	?	y	y	y	NA	?	y	?	?	y	NA?	n	n	y	y	y	n	n	n	n	B/C
Stockley, *et al.*, 1992	?	y	y	y	NA	y	y	y	?	y	NA	n	n	y	n	y	y	n	n	n	B
Cups																					
Glorion, *et al.*, 1994	y	y	y	y	?	y	y	y	y	y	NA	n	n	y	y	y	y	y	n	n	A/B
Harper, *et al.*, 1995	?	y	y	y	y	y	y	y	y	n	n	n	y	n	n	y	y	y	n	y	B
Kennedy, 1994	n	y	y	y	n	n	n	n													C
Wilson-MacDonald, *et al.*, 1990	?	y	y	y	y	y	y	y	y	n	n	n	n	y	n	n	n	y	n	y	B

[a] *Plus ceramic heads*

? Doubtful that criterion met; not clear from paper.

Key: 1. Selection of patients. 2. Prosthesis models. 3. Results for models. 4. Follow-up period. 5. Loss to follow-up/deceased. 6. Ages. 7. Preoperative diagnoses. 8. Outcomes clear/quantified. 9. Statistical analysis. 10. Comparison with preoperative data. 11. Data on deceased. 12. Clinical evaluation independent. 13. Radiological evaluation independent. 14. M/F numbers. 15. Weight. 16. Surgical technique. 17. Surgeons' grade, etc. 18. Type of hospital. 19. Bilateral results separate. 20. Independence of investigators.

APPRAISAL TABLE 8 *Observational – hybrid*

	Key criteria								Other criteria												
Study	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**Rating**
Harris & Maloney, 1989	y	y	?	y	y	y	y	y	?	y	n	n	n	y	n	y	y	y	n	y	B
Helfen, *et al.*, 1993	Abstract only																				?
Kienapfel, *et al.*, 1992b	y	y	y	y	y	y	y	y	?	y	n	n	n	y	n	y	n	y	n	n	B
Mohler, *et al.*, 1995b	y	y	y	y	y	y	y	y	y	y	n	n	n	y	n	y	y	y	n	y	A/B
Pearse, *et al.*, 1992	?	y	y	y	y	y	y	y	n	n	n	n	n	n	n	y	y	y	n	n	B
Schmalzried & Harris, 1993	?	y	y	y	y	y	y	y	y	n	NA	n	n	y	n	y	n	n	n	y	B

? Doubtful that criterion met; not clear from paper.

Key: 1. Selection of patients. 2. Prosthesis models. 3. Results for models. 4. Follow-up period. 5. Loss to follow-up/deceased. 6. Ages. 7. Preoperative diagnoses. 8. Outcomes clear/quantified. 9. Statistical analysis. 10. Comparison with preoperative data. 11. Data on deceased. 12. Clinical evaluation independent. 13. Radiological evaluation independent. 14. M/F numbers. 15. Weight. 16. Surgical technique. 17. Surgeons' grade, etc. 18. Type of hospital. 19. Bilateral results separate. 20. Independence of investigators.

APPRAISAL TABLE 9 *Observational: cementless – mixed*

	Key criteria								Other criteria												
Study	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**Rating**
Lautiainen, *et al.*, 1994	?	y	y	y	y	y	y	y	?	n	n	n	n	y	n	y	n	y	n	n	B/C
Niinimaki, *et al.*, 1994	?	y	y	y	y	y	y	y	?	n	n	n	n	y	n	y	y	y	n	y	B
Riska, 1993	Hybrid and cemented series/uncemented																				
Roffman & Juhn, 1993	?	y	y	?	?	n	n	n													C
Stern, *et al.*, 1992	?	y	y	y	n	y	y	y	n	y	n	n	n	y	n	y	n	y	n	n	C

? Doubtful that criterion met; not clear from paper.

Key: 1. Selection of patients. 2. Prosthesis models. 3. Results for models. 4. Follow-up period. 5. Loss to follow-up/deceased. 6. Ages. 7. Preoperative diagnoses. 8. Outcomes clear/quantified. 9. Statistical analysis. 10. Comparison with preoperative data. 11. Data on deceased. 12. Clinical evaluation independent. 13. Radiological evaluation independent. 14. M/F numbers. 15. Weight. 16. Surgical technique. 17. Surgeons' grade, etc. 18. Type of hospital. 19. Bilateral results separate. 20. Independence of investigators.

PAPER 4

Originally published 2001

Author: A. Faulkner, J. Kent.

Innovation and regulation in human implant technologies: developing comparative approaches.

Published in: Social Science & Medicine, 53, 895–913.

Acknowledgment: Elsevier Ltd. for permission to reproduce this article.

PERGAMON

Social Science & Medicine 53 (2001) 895–913

SOCIAL SCIENCE & MEDICINE

www.elsevier.com/locate/socscimed

Innovation and regulation in human implant technologies: developing comparative approaches

Alex Faulkner[a,*], Julie Kent[b]

[a] *School of Social Sciences, University of Wales Cardiff, Glamorgan Building, King Edward VII Avenue, Cardiff CF10 3WT, UK*
[b] *Faculty of Economics and Social Science, University of the West of England, Coldharbour Lane, Frenchay, Bristol BS16 1QY, UK*

Abstract

Human implant technologies are subject to continual innovation and proliferation, raising important issues for technology testing, healthcare sciences, clinical performance and risk assessment, and regulation. The regulatory environment of medical devices is being shaped by harmonisation of standards in the European Union. The aim of this paper is to compare the histories and current regulatory environment of two technologies, breast implants and artificial hips, and to consider the implications of this comparison for a sociological healthcare research agenda to investigate the issues raised. The main focus is upon developments in the United Kingdom. Major points of contrast between the two technologies include the institutional contexts in which clinical evidence has been marshalled for government attention; the relative importance of strategic alliances between clinicians and manufacturers in the innovation process; the degree of public controversy evident; the varying definitions of an 'adverse incident' within medical device vigilance systems; and in the UK the presence of a national register for breast implants but not for hip implants. Inter-national contrasts in these dimensions are noted. The analysis suggests that improved understanding is required of the institutional, organisational and professional processes involved in implant technology innovation and regulation. A comparative research agenda is proposed, focusing upon: innovativeness and proliferation; safety and technological standards; clinical and social outcomes; and consumer/user information and choice. It is concluded that research in these areas will enhance the 'evidence-base' for the evaluation of human implant technologies in the context of their innovatory and regulatory environments. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Implants; Innovation; Regulation; Medical device; Comparative study

1. Introduction — health technology and policy

There is a constant proliferation of health technologies that offer benefits to public health in maintaining or enhancing the quality of life. Continual innovation in health technologies is increasingly being seen as a double-edged sword, raising a multitude of policy issues concerned with industrial competitiveness, public risk and safety, health service effectiveness, and consumer information and choice. The advent of an era of 'evidence-based' medicine increasingly challenges the relationships between scientific and technological knowledge, and health technology policy.

The intention of this paper is to introduce the policy issues surrounding human implant technologies. Our aim is to compare the histories and current regulatory environment of two technologies, breast implants and artificial hips; and to consider the implications of this comparison for a social research agenda to investigate the issues raised. The first section of the paper outlines the policy context in which these technologies may be understood. The second section describes the history of innovation and current regulation of each group of implants in order to develop a comparative research agenda in the final section. The main focus is upon developments in the United Kingdom, set within a European and wider international context.

*Corresponding author. Tel: +44-2920-874739; fax: +44-2920-874759.

E-mail addresses: faulknerac@cf.ac.uk (A. Faulkner), julie.kent@uwe.ac.uk (J. Kent).

0277-9536/01/$ - see front matter © 2001 Elsevier Science Ltd. All rights reserved.
PII: S0277-9536(00)00389-0

PAPER 4

Within the healthcare policy arena internationally there has been a growing concern about the proliferation and cost of new technologies (Stocking, 1992; Altenstetter, 1996). The explicit drive toward a 'knowledge-based' health service (DoH, 1993) increasingly problematises issues of the performance and ongoing evaluation of innovative health technologies. Increased concern with cost and efficiency is characteristic of the 'new public management' (Hoggett, 1991). These developments place a premium upon the production of knowledge about health technologies, control over diffusion and 'quality assessment' in terms of standards of performance. In the health arena this has led to new inter-disciplinary alliances between public health, clinical specialties, epidemiology and the academic disciplines participating in 'outcome measurement'. The Health Technology Assessment (HTA) research programme in the United Kingdom arose amidst growing concerns about cost containment and clinical effectiveness of healthcare interventions, reflecting developments internationally (Banta & Luce, 1993). Part of the agenda of the national HTA programme is directed towards controlling the diffusion of new health technologies, especially where a threat to health or healthcare budgets is identified (Advisory Group on Health Technology Assessment, 1992; Faulkner, 1997). Regulatory measures are being extended from pharmaceuticals to medical devices and even surgical procedures (Gelijns, 1990; Sheldon & Faulkner, 1996; Mowatt et al., 1997). At the same time, moves toward the harmonisation of European trading standards is further increasing attention to the regulatory environment of new devices being brought to the healthcare market.

The public image of technological medicine is generally very positive, but the invention and adoption of new medical devices present perennial dilemmas of healthcare policy. New technologies may offer potential benefits in terms of health gain or the relief of suffering, but on the other hand there may be risks to health, social or economic costs may be incurred, or ethical issues raised. From a public health perspective, therefore, innovation should, in principle, be promoted and not stifled, but adequate safeguards must exist in order to prevent potentially harmful technologies being adopted. Different countries have evolved collections of mechanisms which deal with these dilemmas in different ways, with diverse levels of explicit regulatory control (Altenstetter, 1996; Faro & Huiskes, 1992). Occasionally the emergence of a particular technology heightens the level of policy concern about these matters. Perhaps the most outstanding example is the computed tomography (CT) scanner, which was associated with a major debate in the 1970s and 1980s framed in terms of a concern with expensive 'new medical technologies' (Jennett, 1986; Stocking, 1988), a debate which certainly paved the way for the emergence of HTA.

Currently, this type of debate is being given renewed impetus, especially in the context of the development of new genetic technologies and capital-intensive technologies such as 'telemedicine' and positron-emission tomography (PET) scanning. This coincides with awareness that in late modern societies subjective experience is increasingly articulated through preoccupations with issues of risk and uncertainty, as traditional forms of expertise and authority, including medical authority, are being challenged (Giddens, 1991; Beck, 1992; Gabe, 1995). However, a corollary of the challenge to traditional medical authority may be a re-positioning and re-defining of the medical and allied sciences in relation to the production of new knowledge focused around care delivery. This may be seen, for example, in the 'evidence-based medicine' movement, and also illustrated by current concerns with regulation of the practices of medical and other professions allied to medicine. While the technologies of genetic and diagnostic information clearly raise major ethical and social uncertainties, so too do medical devices — though perhaps in less obvious and less often publicised ways.

Much of the preoccupation with risk and uncertainty has focused upon the human body as a 'project' of modernity (Turner, 1992; Shilling, 1993; Williams, 1997). Medical devices may offer a variety of benefits to individuals especially in terms of increased longevity and enhanced or maintained quality of life, with consequent collective benefits for society. What we have termed 'human implant technologies' are an important group of medical devices treatment with which in principle offers benefits in maintaining or repairing the functionality and quality of many different body parts and systems, with implications for body image, self-esteem and social participation. Implanted technologies raise special questions of performance in the body and of methods of assessing that performance. Evaluation of long-term performance may require regular clinical follow-up of implantees, attempts to develop techniques to predict future performance, biocompatibility testing or material analysis of devices following explantation or death of the implantee (Fielder & Black, 1995).These technologies have been increasing rapidly in their numbers and variety in the last 20 years, and a vast array of organs, vessels, tissues and bones can now be replaced by 'artificial' or salvaged parts. While the public and mass media imaginations appear especially fascinated with organ transplantation, the numbers of people directly affected by this are small compared to those receiving surgical implants such as arterial stents, heart valves, breast implants, joint replacements, defibrillation devices, or spinal implants. Science fiction-like images of the bionic person and the cyborg are increasingly seen within the medical profession itself as becoming reality (Marinker & Peckham, 1998), and

PAPER 4

developments in tissue engineering technologies offer the possibility of 'living' implants.

It has been strongly argued, in investigating relationships between science and policy in healthcare, that medical sociology should concern itself more with the production of medical knowledge by pharmaceutical industries and governments, in addition to its other concerns (Abraham, 1997). Alongside this, there is an important opportunity for medical sociology to broaden its focus, in concert with the social study of science and technology, to include exploration of the newly emergent healthcare sciences, epitomised by the health technology assessment movement. We suggest that the uses of science and technology tests in the construction and regulation of medical devices in general, and human implant technologies in particular, are an equally if not even more neglected topic. As with pharmaceutical products, there are significant implications for both the protection and advancement of public health and for healthcare budgets, as well as broader social and economic implications. We take it that human implant technologies are 'sociotechnical' constructs (Law & Bijker, 1992) embodying scientific and technological knowledge and social processes. And in parallel with Abraham's analyses of the operation of science in pharmaceutical regulation (e.g. Abraham, 1997), broad political and economic factors influence this knowledge and shape the devices upon which we might rely, either as implantees or as consumers of other types of medical device.

Implantable medical devices — regulatory environment

The Medical Devices Agency (MDA) is the 'competent authority' with responsibility for 'ensuring the safety and quality' of all medical devices used in the UK. Human implant technologies are a category of medical device which fall within the scope of the European Medical Devices Directives (EMDD) (93/42/EEC). "Products as diverse as CT scanners, cardiac pacemakers, syringes, bandages and hospital laboratory equipment all come under the scrutiny of MDA's highly knowledgeable medical devices experts" (MDA, undated). The EMDD distinguish between active and non-active implantable medical devices. Active implants are powered devices left in the body (e.g. heart pacemakers). In these terms both breast and hip implants are non-active (though bio-active coatings and materials may be used, in which case they may be classified at a higher level of risk). However, as our discussion will suggest there are both similarities and differences in the ways in which these two groups of non-active implants have been developed and regulated.

The move from voluntary control of medical devices in the UK to a Europe-wide statutory system raises questions about the effects of harmonising standards (Altenstetter, 1996; Levidow, Carr, von Schomberg, & Wield, 1996). Since 1993 a process of harmonisation of European standards in medical devices has begun and the UK MDA has played a part in this process. The award of the 'CE' (Conformité Européen) mark now denotes a European conformity standard and eventually all medical devices will be required to carry this mark (MDA, 1993a,b). Under the CE marking system medical devices are classified according to the degree of risk assigned to them. CE marking does not guarantee safety but it does mean that the product may be freely marketed anywhere in the EC without further control (MDA, 1993a,b, 1995). Operation of a post-marketing 'vigilance system' is the responsibility of Member States of the European Union.

Under the vigilance system in the UK, the MDA operates an 'Adverse Incident Centre' (AIC) to which users (for example nurses, clinicians, patients) and manufacturers are able to report cases where a medical device has failed or produced unwanted side effects. This remains a voluntary system, though in certain circumstances manufacturers are required by law to report incidents (MDA, 1998a). Following investigation by the MDA, a number of actions may follow: either a *hazard notice* may be issued alerting others to potential danger; *a safety notice* may be issued advising users in less urgent situations; or *a device bulletin* with guidance and information for users may be produced. Finally there may be "an improvement in product design, labelling or instructions for use. In some cases the manufacturer will withdraw the products from the market" (MDA, undated). The guidelines produced by the European Commission on the vigilance system "are not enforceable by law" (MDA, 1995, p. 5; MDA, 1998b). However, within the UK the MDA has a duty under the Consumer Protection Act 1987 to enforce the essential requirements of the EMDD regulations (MDA, 1995). Nevertheless, the efficacy of reporting systems may be open to question: analysis of the post-marketing surveillance system operated in the USA by the Food and Drug Administration shows a high degree of under-reporting of adverse incidents to the regulatory agency even when users have reported problems to manufacturers (Bowsher, 1991).

There are a small number of MDA-funded implant-specific national registries — the National Pacemaker Database; the UK Heart Valve Registry; the Hydrocephalus Shunt Registry and the National Breast Implant Registry. According to the MDA "they are used to provide an early warning of device related problems and a general indication of implant usage" (MDA, 1999a). In the arena of the new healthcare sciences addressed to issues of health service effectiveness and cost effectiveness, the Health Technology Assessment programme in the UK has given priority

to a small number of implant technologies in its research programme in healthcare technologies. These include cochlear implants, intrathecal opioid pump systems, oesophageal and aortic stents, and hip and knee prostheses (National Coordinating Centre for Health Technology Assessment, 1998).

Total hip joint implants

Innovation

Artificial hips have been implanted in the UK routinely since the mid-1960s. Currently, there are over 60 different named models available in the UK. Most of these do not have published scientific evidence to support their use, and about half have been introduced within the last 7–8 years. The proliferation of different design features and materials is spectacular. The most common design uses 'cement' to fix the prosthesis, but non-cemented designs are increasingly being developed. These are more expensive, and their performance is controversial within the orthopaedic profession internationally. A recent trend is the production of modular designs, which introduce more mechanical components, with the (cl)aim of achieving better anatomical fit. This has led certain leading orthopaedic researcher-surgeons to warn against a trend toward 'designer hips' (Bulstrode, Murray, Carr, Pynsent, & Carter, 1993), and the recent attempts to improve upon the standard cemented design have been strongly criticised (Huiskes, 1993).

About 40,000 first-time implantations are carried out annually in the UK, one-fifth of these outside the NHS. Increasing numbers of younger and older (over 80 years) people generally are receiving hip implants. Total hip replacement is generally recognised to be of great benefit in the relief of arthritic pain and improvements in locomotor mobility. There is some evidence that implantees' quality of life is improved. The most common prosthesis (the 'Charnley' cemented model) is used in between 40% and 50% of primary hip replacements in the UK. However, some 70% of hospitals have reported using both cemented and uncemented models to some extent (Newman, 1993). There appears to be a regional preference for different types of hip implant, with uncemented models being more common in the south of England. In Newman's survey 30% of hospitals in England and Wales stocked at least two uncemented models (Newman, 1993). Unlike breast implants, no breakdown of numbers of people registered as receiving different types of implant is available in the UK, even on the basis of voluntary data. Prices per unit vary widely, from about £250 to £2000 (1996 prices) for the most expensive modular hip systems (Murray, Carr, & Bulstrode, 1995). The pattern of types of implant used varies widely between countries. For example in the early 1990s more than 50% of implants in Finland were non-cemented, while these models accounted for only 4% of those in Sweden and 15% in Norway (Havelin, Espehaug, & Vollset, 1993).

The main materials used in hip implants are metals, plastics and ceramics. The 'same' model is often produced by manufacturers in different metals (titanium, cobalt-chrome) and coatings (porous mesh or beading, pre-roughened, hydroxyapatite (a biologically active ceramic, usually referred to as 'HA'), and others). The bearing surface materials of the femoral head (ball) and acetabular component (socket) also vary widely: metal, ceramic and high-density plastics being used in different combinations. The arrival of superalloys and composite materials has almost eliminated mechanical breakage in normal usage of the artificial hip. However, there is continual innovation in materials and design features, with more advanced production processes being applied and rare materials being used, such as zirconium. The Norwegian national register of artificial hip implants (see below) yielded a count of over 400 different designs and sizes of socket component, and nearly 400 stem components, which was felt by the orthopaedic surgeon reporting this to be, "from a medical point of view unreasonable" (Havelin et al., 1993, p. 251). The stronger, new materials appear to have brought with them their own problems, and these may appear only two or three years after implantation. Small changes may have large consequences (e.g. the matt finish on the 'Exeter' femoral component, which proved ineffective). One surgical commentator has described the effects of the new materials thus: "We have unleashed a torrent of particles into our joints, producing a devastation far exceeding simple prosthetic failure or fragmentation of parts" (Booth, 1994). This may be a somewhat exaggerated account, but the issue has also been recognised by a National Institute of Health conference in the United States, and in the United Kingdom: "The main problems of concern related to implant design are... osteolysis due to particulate materials, biologic responses to particles of implant materials..." (Murray et al., 1995, p. 1952); "new designs and material combinations... may introduce fatigue problems which as yet have not been considered" (Styles, Evans, & Gregson, 1998). Thus there is little doubt that the clinical effects of the new combinations of materials do raise fundamental concerns about the safety implications of the technologies being employed, and possible means of evaluating them. It is certain that most implantees do not have detailed information about the design and material composition of their artificial hips.

The world annual market value for the orthopaedic implant business is estimated at around $9 billions, of which about 20% is accounted for by hips (DePuy, 1997). Six companies or groups now dominate the international market in terms of sales. There are at least

19 separate distributors or manufacturers of hip implants in the UK (Murray et al., 1995). These are mostly multinational (mainly American) companies specialising in orthopaedic and surgical products (e.g. DePuy, Zimmer, Howmedica), but also include multinational multi-sector companies (e.g. 3 M) and smaller British-based firms (e.g. Corin Medical). There are strong financial incentives to produce modified versions of designs or even entirely new models (Bulstrode et al., 1993). A typical company produces a range of different implants using different materials and design concepts. Biomet, for example (about 8% market share in orthopaedic implants generally), in March 1999 listed 17 different trademarked femoral components, and 11 socket components in its information for surgeons (Biomet, 1999). These cover different base materials such as titanium or cobalt-chrome, options for bolt or plate attachments, different sizes and stem lengths, different coatings — porous or hydroxyapatite, cemented or cementless, for primary or revision implantation, modular or 'monobloc' designs, with most being marketed for use in 'skeletally mature patients undergoing (primary) hip replacement surgery as a result of noninflammatory degenerative joint disease'.

Innovating orthopaedic surgeons, bio-engineering research laboratories and manufacturers often form strategic alliances to bring new designs to the stage of clinical experimentation. Orthopaedic surgeons may act as design advisers to an implant manufacturing company. A number of analysts of technological change have pointed to the importance of networks of interaction between different actors (included in, for example, Bijker & Law, 1992; Star, 1995; Gaudilliere & Lowy, 1998). However, while sociological historical analysis has been conducted on the innovation process in cochlear implants (Blume, 1995) and heart pacemakers (Jeffrey, 1995), little information about the interaction of social, organisational and technological processes in orthopaedic implant innovation can be found in published sources. Review of clinical research about the performance of hip implants suggests that a small proportion of orthopaedic surgeons worldwide has a financial as well as clinical interest in new designs (Faulkner et al., 1998). Anecdotal evidence suggests that provision of conference support, training courses or equipment by manufacturers may play a part in influencing some surgeons' choice of implant providers. As with pharmaceuticals, orthopaedic manufacturers employ sales representatives to promote the sale of their products.

Artificial hips became a matter of increased public concern in the early 1990s in the UK. Media coverage such as the 'High Price of Hips' (BBC2, February 26, 1993) and well-publicised scares such as the 3 M Capital Hip failures announced by the MDA in February 1998 (MDA, 1998c) caused concern focused both at the healthcare cost and at performance safety issues. The 'untested' status of many models was highlighted, and advice from consumer health organisations encouraged patients to ascertain what prosthesis is being recommended to them (e.g. Consumers'Association, 1997). The apparent successes of operations on elderly people such as the Queen Mother also highlighted this type of device. An indication of governmental concern is the launch in 1999 of an investigation into total hip replacement by the UK's National Audit Office, and in 2000 the publication of what can be regarded as quasi-regulatory guidance for the NHS by the National Institute of Clinical Excellence (NICE, 2000).

Regulation

The regulatory environment of hip prostheses in the UK has been relatively unrestrictive. A surveillance scheme based on a 'recommended list', a limited number of models agreed to be acceptable, was agreed between the government and the orthopaedic profession in 1981 (Sweetnam, 1981), but it was not implemented. Unlike some other implants there is no national registry of clinical implant information, voluntary or otherwise, although one regional health authority area (Trent) does have a monitoring system of this type, established in 1990 with support from the government Department of Health. This collects data on the brand of implant, prophylaxis, cement, theatre environment and grade of operating surgeon, plus age, sex and diagnosis details for the patient. Although orthopaedic departments and centres can be taken to record the prostheses implanted in their own patients, there is no co-ordination of these data between centres and the extent to which systematic local post-implantation monitoring occurs is unknown. This situation can be contrasted with several Scandinavian countries which have national hip implant registers which collect detailed information about the implant model, surgical staff, operating theatre environment, antibiotics and other items. These were established in Sweden in 1979, Finland in 1980 and in Norway in 1987.

Until recently, in the UK in the case of hip prostheses, only pre-clinical technological product testing has been mandatory, under 'good manufacturing practice' regulations. This called for technological bench tests and biocompatibility assessments. The manufacture of orthopaedic implants is in principle governed by a large number of national and international technical standards, produced by the British Standards Institute (BSI) and the International Standards Organisation (ISO). These cover orthopaedic implant materials in general, and some particular features of hip prosthesis implants. In recent years, under the impetus of the Medical Devices Directive, European Standards have sought convergence between national and international standards (see above). However, it is interesting to note that standards for material properties generally apply to the

PAPER 4

material as used in the production process rather than the finished product (Paul, 1997). Furthermore, some features of hip prostheses have to date proved impossible to subject to specific standards because of the great proliferation of different sizes and design features. This applies, for example, to the modular head (ball) of femoral components (Paul, 1997). From July 1998, following the EU directives, clinical investigations are required under certain conditions by the MDA, in order for a 'CE' mark (see above) to be given, enabling the device to be placed on the market. These investigations will be required for various reasons including: where there is a 'completely new concept of device...where components, features and/or methods of action are previously unknown'; in the case of 'modification of an existing device in such a way that it contains a novel feature'; or where existing materials are used in a new location in the human body (personal communication, 1997). This system thus appears to draw a distinction between devices that are in some sense new and those which can be shown to have an equivalent already on the market. This parallels closely the system based around the notion of 'substantial equivalence' used by the Food and Drug Administration in the USA, where the great majority of applications for 'new' products do not require de novo clinical investigations but rely instead upon the production of evidence that the product is technically similar to an existing device on the market, and on adequate 'clinical evidence' for that type of product (Gelijns, 1990).[1] The MDA has produced guidance for manufacturers carrying out clinical investigations in the UK, including stipulations regarding study design, sample size calculation and statistical analysis (MDA, 1998d).

However, the MDA has not produced any general reports on hip prosthesis or orthopaedic implant technology as a family of devices. Under its post-marketing surveillance responsibility, it has produced a small number of 'hazard notices' or 'safety notices' in response to notifications of failures of specific prosthesis components from hospitals. It has also, for example, issued a notice to clarify the legal responsibility for an implant in cases where an orthopaedic surgeon requested the addition of a new coating by a specialist implant coating service (MDA, 1997).[2] Total hip implants are a Class IIb (or Class III if bioactively coated) device under the MDD system which attempts to match the degree of regulation to the degree of risk (Class III being the highest risk).

As part of the EU MDD 'vigilance system', CE-marked joint replacement implants will be subject to an enhanced requirement for reporting of adverse incidents by manufacturers to the MDA. In essence, any failure of an implant attributable to premature deterioration or malfunction of the device itself will be reportable. This will include the most common cause of implant failure and need for a revision operation, known as 'aseptic loosening'. In the absence of a predicted lifespan for the device, revision within 10 years for aseptic loosening will be reportable (MDA, 1998e). Revision within this timespan will be regarded as an adverse incident and will be reportable to the MDA within a suggested elapsed time of 10 days. The MDD requires the report to be made to the MDA Adverse Incident Centre by the manufacturer. Users should thus report incidents to the manufacturer, although the facility exists also for direct reporting. However, review of the clinical research reporting the performance of hip implants shows that the attribution of responsibility for failure is often contentious. The cause of failure of hip implants in younger, more active people is especially likely to be contested (Faulkner et al., 1998). This implies that individual clinical assessment and local circumstances will be paramount in determining whether a report is actually made when revision due to *possible* device failure has been performed.

In the arena of healthcare science, the national Health Technology Assessment movement in the UK, inaugurated in 1993, made the investigation of hip prostheses one of its top 20 priorities, arrived at following a process of selection applying clinical, economic, epidemiological and other criteria to over 1000 candidate technologies. It was the subject of an 'Effective Health Care Bulletin' in 1996, which provided a review of the evidence of published clinical research studies about hip implant performance, including the comparative effectiveness of different models (NHS Centre for Reviews and Dissemination, 1996). This document advised that 'purchasers and providers should promote the use of those prostheses that have been shown to perform best in long-term follow-up', and stated that 'new prostheses should only be used after they have been thoroughly evaluated or as part of a nationally co-ordinated study'. The HTA research programme has subsequently published two 'systematic reviews' of evidence of the effectiveness and cost effectiveness of different hip prostheses (Faulkner et al., 1998; Fitzpatrick et al., 1998). These show that there is a large amount of clinical research on the clinical outcomes of hip implants published in specialist orthopaedic journals, that there are few randomised trials and that most of the studies

[1] The amount of documentation and production of evidence required for an 'equivalent' device is considerably less than for a new device, so there is clear incentive for competing manufacturers to demonstrate equivalence. As Gelijns noted in the USA, even where clinical trials have been conducted by the device developer, there may be large variations in, for example, the sample size of the population of implantees studied (Gelijns, 1990, p. 164).

[2] In such cases responsibility passes from the device manufacturer to the NHS Trust employing the surgeon.

are of relatively small numbers of prostheses. Most studies are authored by orthopaedic surgeons and studies considering the issue of the benefit of hip replacement versus non-intervention are rare. The HTA reports conclude that policy consideration should be given to different approaches to monitoring and controlling hip implants, including the use of registries along the lines of the Scandinavian models. These reports have been drawn upon in the recent guidance to the NHS issued by the National Institute of Clinical Excellence (NICE, 2000), mentioned above. The HTA reports focus almost exclusively upon clinical research results. The extent of correlation between laboratory performance or predictive studies of hip implant failure, and clinical performance once implanted, is widely debated and disputed in the orthopaedic research literature.

The professional clinical body with concern for hip implants in the UK is the British Orthopaedic Association (BOA). Affiliated to this is a specialist British Hip Society with the purpose of providing a professional forum for debating research and clinical practice. The BOA and Society have been involved in a number of activities which can be regarded as part of the self-regulatory work of the profession and as part of the 'interorganisational field' (Blume, 1992) in which the development of orthopaedic implants is embedded. Representatives of the BOA have been involved in discussion with the MDA about the implications of the EU MDD, and these have resulted in the MDA's guidance on joint replacement implants referred to above (MDA, 1998e). Currently, the BOA in principle approves of the introduction of a national register, although there is dispute about its organisation (BOA, 1998).

This brief survey of the recent science and regulation of hip implants indicates that even where a medical device is widely regarded as highly successful, there are a number of issues of uncertainty. Healthcare risk and the implications of these devices for public health are matters of negotiation amongst competing groups. Yet inter-relationships between professions, manufacturers, regulators and scientists are poorly understood. Continual technological innovation and proliferation of models, allied with an occasional controversial failure have put regulatory issues high on the policy agenda. In the European context the regulatory environment is becoming more complex and, in principle, more stringent especially with regard to vigilance systems. However, the effectiveness of vigilance systems for the timely identification of under-performing devices is in question. Areas of uncertainty, such as the definition of 'novelty' have been identified, where it is important to examine the operation of inter-organisational and disciplinary forces. The roles of professional and mass media in problematising the technology in the policy arena have been suggested. Public policy-related academic science has focused upon the issues of clinical and cost effectiveness on the basis primarily of comparative clinical research on the longevity of implants and economic analysis. A viewpoint of implantees is conspicuously absent in this science. The clinical and other evidence used in regulatory technology *adoption* decisions is not currently in the public arena. Tensions between innovation and regulation, and between the uncertainties of risk assessment and indeterminacies of evidence, are apparent.

Breast implants

Innovation

The first breast implants were a type of 'sponge' used up until the late 1950 s. Other early experiments in breast enlargement included the use of injected silicone and parafin wax (Guthrie, 1994). The late 1960s marked the beginning of commercial manufacture of silicone breast implants by Dow Corning following the work of plastic surgeons Cronin and Gerow. These 'first generation' silicone implants had a thick smooth silicone shell and were filled with silicone gel. In the mid-1970s, in an attempt to reduce the amount of capsular contracture (shrinkage and hardening of tissue around the implant) modifications in design and materials produced 'second generation' implants with thin silicone smooth shells. "It is now accepted that these are more susceptible to rupture" (IRG, 1998, p. 12, Malata, Varma, Scoot, Liston & Sharpe, 1994). Saline-filled implants have also been in use since that time and have been preferred by some plastic surgeons because they are thought to be less harmful if rupture occurs.

From the 1970s polyurethane foam coverings on implants were used but were withdrawn by manufacturers in the early 1990s after a ban on their use in the US in 1991 (Guthrie, 1994). However, the use of one polyurethane breast implant was reported in the UK during 1995/96 (NBIR, Third Annual Report, 1997; MDA, Safety Notice SN9620, 1996). A third-generation of silicone implants with a thick textured shell were developed in the mid-1980s to reduce the incidence of capsular contracture and to minimise the rate of 'gel bleed' (silicone gel leakage). Since 1995 a new type of triglyceride-filled (soya bean oil) implant with silicone shell has been available in the UK (Collagen Ltd). In the US and Canada these have not been licensed though a small clinical trial of these implants is being conducted (Studin, 1998; FDA, 1998). In addition to the variations in materials used to construct the shell (or envelope) and filling of breast implants there are other variations including size, shape, and some are expandable with valves, or have a single, double or triple lumen. Two hundred and forty two different types of named implant brands are cited in the Foreign Settlement Program

PAPER 4

discussed below (Federal Judiciary Center, 1998). While these are not all currently available as new, this indicates the range and diversity of breast implants still in use, that is implanted in women worldwide.

As with hip prostheses, manufacturers of breast implants are US-based multinational companies. Silicone gel and saline implants are produced by McGhan of the Inamed group and Mentor Corp. Hutchinson International and Poly Implants also produce saline-filled implants. Collagen Ltd are the manufacturers of Trilucent ™ soya bean oil filled implants. An estimated 11 manufacturers import breast implants to the UK.

In the UK the exact number of implants used each year is not known, despite the setting up of a National Breast Implant Registry in 1993. However, in 1996–97 12,829 implants were registered (see Table 1) used by 7136 patients. Although not identified by brand name these implants have been grouped into types according to their filling.

According to the available statistics, recipients of breast implants are usually women who range in age from 10 years old to 77 years. They may be identified as belonging to two main groups — those seeking reconstruction following mastectomy (for benign or malignant disease) and those seeking augmentation for cosmetic reasons. Two other groups may be identified — those women (or girls) seeking surgery to correct congenital deformity or developmental problems and those women who have had previous surgery but need a replacement implant (presumably due to product failure or post-operative complications) (Table 2). In addition, breast implants may be used in cases of gender reassignment (from man to woman). The types of benefit to be derived from breast implants are seen primarily to concern body image, 'looking normal', self-esteem and consequent impacts on social roles and participation. The average age for those seeking reconstruction is older than for those seeking augmentation (NBIR, 1998). By far the majority of breast implants are used for augmentation or cosmetic surgery and of those patients and implants registered between 1993 and 1997 the proportion of those seeking cosmetic surgery has increased.

The use of breast implants differs between the private and public health sectors. In 1996/97 87% of total breast implant procedures registered and carried out in the private sector were for cosmetic reasons compared to 34% of the total breast implant procedures carried out in the public sector. A greater proportion of public sector operations are for reconstruction following malignant breast disease (30%) compared with 3% in the private sector. A greater proportion of public sector operations are to replace breast implants (23%) whereas these make up a smaller proportion of the total procedures in the private sector (8%). An estimated 33% of all breast implant procedures are carried out in the public sector (NBIR, 1998, p. 10). The cost of breast augmentation in the private sector is approximately £3500.

Regulation

In order to understand the regulatory context of breast implants it is necessary to outline events in the United States that have formed a background to UK policies and practice. During the 1990s the use of silicone breast implants has been controversial. In 1991, in the US, the Food and Drug Administration (FDA) launched an inquiry into the use of silicone gel breast implants. Safety issues were raised since claims had been made that there may be a link between these implants, connective tissue disease and other adverse side effects. Since 1992, in the US, these implants can only be used as

Table 1
Implant types of those breast implants registered, 1996–97[a]

Silicone gel	Saline	Double lumen	Expandable prosthesis	Lipid or polymer	Other and not known	Total
10,220	303	8	689	1513	96	12,829

[a]Source: Adapted from data in Fourth Annual Report of National Breast Implant Registry (1998). (Note: Figures may be broken down to indicate the type of envelope which may be smooth or textured. Statistics show that the majority of implants used have a textured envelope.)

Table 2
Numbers of patients registered during 1996/97 and reasons for implantation[a]

Cosmetic	Mastectomy (malignant)	Mastectomy (benign)	Developmental	Replacement	Other	Not known	Total
4915	877	66	313	916	34	15	7136

[a]Source: Fourth Annual Report of National Breast Implant Registry (1998). (Note: The number of patients and implants registered differs significantly in any year since in some cases one implant may be used, in others two.)

PAPER 4

part of a clinical trial (Kessler, 1992; FDA, 1998). The FDA decision was criticised and a dispute about the policy ruling ensued (Council of Scientific Affairs, 1993; Brody et al., 1992; Angell, 1997; Kessler et al., 1994). At the heart of this dispute were a number of issues about the scientific evidence and the responsibilities of the regulatory authority, manufacturers of implants and clinicians. The FDA accepted "that thousands of women had reported problems with breast implants" (Gott & Tinkler, 1994, p. 12) and concluded that it was not possible to give adequate information about the risks to women of having these implants.[3] In the US today while saline-filled implants are available to everyone only those 'women with special medical needs', that is 'those who have had breast cancer, a severe injury to the breast, a birth defect that affects the breast, or a medical condition causing severe breast abnormality' and who are seeking reconstruction may have a silicone-gel-filled implant as part of a trial or adjunct study (FDA, 1998). However, it is worth noting that prior to the controversy over evidence and the introduction of the restrictive 1992 policy, breast implants had been assigned a Class 1 level of risk, in other words very low risk not requiring further research and not requiring pre-market clearance, within the FDA medical device framework introduced in 1976 (Palley, 1995). Although there are many possible reasons for this, it can perhaps be taken to reflect the general perception at the time that the predominant use of the implant was for cosmetic purposes and thus did not warrant the level of scrutiny that a device provided for the treatment of illness might require.

A growing number of litigations against breast implant manufacturers in the US led in 1993 to a setting up of a compensation fund for women with breast implants (Angell, 1997). There has ensued a lengthy, complex legal case against the manufacturers of breast implants including Dow Corning (which was the biggest exporter of breast implants to the UK), Baxter Healthcare Co, McGhan and Medical Engineering (a 3 M company). Together with other breast implant manufacturers, a Global Settlement Program was set up which included women outside the US. Some five years later the case continues, Dow Corning has declared bankruptcy but a number of other companies remain in the settlement programme. More than 2000 women in the UK have registered claims for compensation since 1994 but they have yet to receive any payments from these funds. The outcome of these proceedings remains uncertain. Other estimates of the number of women adversely affected by breast implants in the UK are as many as 60,000 (Boseley, 1998a). Rupture of these implants is now recognised as a common occurrence, though precise data on this are unavailable (Clwyd, 1996; IRG, 1998). In the UK estimates based on the number of implants replaced which are registered, even for one year, suggest that there are significant complications which lead to repeat surgery (1422 implants were replaced in 1996/97, NBIR, 1998).

In the context of this controversy, in the UK in 1992 the MDA set up an Independent Expert Advisory Group to review information on allegations that there was a link between connective tissue disease and silicone gel bleed from breast implants (Tinkler, Campbell, Senior, & Ludgate, 1992). There was growing concern amongst some commentators about the safety of silicone implants in the light of the US reports. This review was updated two years later (Gott & Tinkler, 1994). The conclusions and recommendations of both these reports were that there was no evidence to support the allegations but a National Breast Implant Registry (NBIR) was set up in 1993 to record the use of all types of breast implants on a prospective register. In contrast to the situation in the US, the use of silicone-gel- and saline-filled breast implants in the UK has continued unrestricted.

A primary task for the NBIR was to encourage participation in the registration process from hospitals and practitioners. However, registration is voluntary. By 1997 280 hospital units were participating, though the extent to which the number of registrations reflected actual procedures carried out is still a matter for conjecture. It was suggested that one way of measuring compliance would be "by matching manufacturers sales figures and registrations. ..." but "it would be difficult because of the commercially sensitive nature of the information held and the variable stock control methods (using sale or return) of many active centres" (NBIR, 1998, p. 8). There have been criticisms of the registration process for failing to promote retrospective registration of those women who had received implants prior to 1993 and the voluntary nature of the process (Clwyd, 1996; telephone interview Survivors of Silicone, 1997).

Despite the MDA's insistence in their earlier reports that there were no safety issues relating to the continued

[3] From 1985 to September 1998, 127,500 adverse reaction reports for silicone-gel-filled implants were received by the FDA and 49,661 adverse reactions to saline-filled implants (FDA, 1998). In the US ruling, a distinction was made between those women seeking reconstruction following mastectomy and those seeking cosmetic augmentation. Women in the former group are allowed access to silicone gel implants because it was assumed that the benefits of implants to them outweighed the risks. However, this distinction was criticised for the different treatment of each group and for 'federal paternalism'. It was argued that both groups of women had rights and needs that should be met and that if implants were safe then they were safe for all (Parker, 1993). On the one hand, commentators argued for greater regulation while others argued for individual freedom to choose whether breast implants were beneficial to them (Haiken, 1997).

use of silicone gel breast implants, a further review was called for in 1997. The focus of the Independent Review Group, published in 1998 was the use of silicone gel implants (IRG, 1998). No mention was made of other types of breast implants (such as triglyceride ones), not even as alternative options for women. Following an extended period of consultation the group recommended changes to the information made available to women considering implant surgery and suggested that a designated body provide such information. Changes in the consultation and decision-making process by women and doctors were also proposed and a recommendation made for a new consent form for implant surgery. Importantly, the group proposed that measures be introduced to ensure standards of care in the private sector and that prospective registration should be made *compulsory*. It was emphasised that clinicians should report adverse incidents and the MDA should advise on which incidents were to be reported. A new steering group to plan and monitor a programme of research was proposed and recommendations made that the possibility of screening to detect implant rupture should be kept under review. So far no change in the law to introduce compulsory registration is planned and a programme of further research has not yet been developed. A new Health Select Committee of Inquiry into the regulation of private and other independent health care was set up in December 1998 and is currently taking evidence (Health Committee Press Notice, Hinsliff, 1998). It seems then that the target for tighter regulation currently is those cosmetic surgeons working in the private sector (see also Boseley, 1998b; Collis & Sharpe, 1998).

Elsewhere in Europe opinion has been divided, though efforts were made to reach a consensus. The French Ministry of Health restricted use of breast implants and also commissioned an inquiry in 1996 and a further report to the European Commission in 1997 (see regulatory requirements for breast implants IRG 1998 website). In Finland there is a continuing dispute about the safety issues and arguments have been put forward for a national registry (Hovi, personal communication, 1998) while in the Netherlands the controversy is seen as an 'American thing' (Davis, personal communication, 1998). However, in 1998 there were reports of a campaign to ban silicone implants in Europe (Watson, 1998). The European Medical Devices Directives do relate to all types of breast implants. Breast implants are class IIb. Standards for breast implants have now been agreed though not yet published (EN 12180). For some time in the EU there was no agreement about "what measures are adequate to ensure the safety and performance of breast implants" (DoH, 1998). Standards relate to material and implant strength, design attributes and mechanical tests. The safety questions remain contentious. Indeed the IRG themselves point out that "safety is not a simple concept. It is widely accepted by lawyers, ethicists, doctors and regulators to mean freedom from an undue risk of harm. The word 'undue' is critical ... safety is not an absolute concept" and must be considered "in the context of individual circumstances" (IRG, 1998, p. 14). According to them risks and benefits are both to be taken into account.

For many women who have had breast implants the risks have been unknown and it is precisely this lack of information that was highlighted by the IRG. In 1998, in Britain, a woman successfully obtained legal aid to pursue a claim against the manufacturers of her breast implant which she believes has caused adverse effects to her child whom she breast-fed after having the implantation (Boseley, 1998b). Legal aid has also been granted to some women who have had implants rupture to enable them to pursue compensation claims against manufacturers in this country under the Consumer Protection Act. British lawyers have advised clients of how the product liability legislation may be applied to breast implants (Balen, personal communication, 1998; APIL, 1998). Both lawyers and members of pressure groups such as Survivors of Silicone and the Silicone Support Group have argued that the regulatory framework has failed to prevent harm and that safety issues around breast implants remain (Comber, personal communication, 1998; Watson, 1998; Boseley, 1998b, 1999). In contrast clinicians in the UK have consistently stated that "at present there is no evidence to suggest that silicone breast implants are associated with an increased incidence of breast cancer. There is also no evidence to suggest that these implants cause autoimmune diseases such as rheumatoid arthritis" (British Association of Aesthetic Plastic Surgeons, 1998; Coope & Dennison, 1998). In recent years there has been recognition that leakage of silicone can occur either as a gradual process or due to rupture but BAAPS say this has not been shown to be harmful. Other complications such as capsular contracture have also been recognised for some time (Iwuagwu & Frame, 1997). Breast implants are now seen by some clinicians in the UK as having a limited life expectancy of approximately 10 years (BAAPS Information leaflet on Breast Augmentation). However, many women have received implants unaware of these limitations.

Less than a year after the IRG report, the MDA announced that the new soya bean oil filled breast implants (Trilucent) are to be withdrawn and that no more should be implanted (MDA, 1999c). An estimated 5000 women have received these implants since they were introduced in 1995, many of whom chose these in the context of anxieties about the safety of the silicone gel implants. While the MDA announcement referred to some reports of local complications which had led to this decision 'as a precautionary measure'. They said

"there is no evidence of permanent injury or harm to general health. However, not enough is known about the long-term safety and rate of breakdown of soyabean oil in the filling and its possible effects on the body" (MDA, 1999b, 1999c; see also Iwuagwu & Frame, 1997). Media coverage of this development says that there have been 74 adverse incident reports to the MDA since the implants were introduced (Hall, 1999; Boseley, 1999; Laurance, 1999; Wilson, 1999; Ellis, 1999). Women with these implants have been advised to contact their surgeon or GP for advice, a helpline was set up for three days and further investigations into the safety and risks associated with these newest implants are now being carried out.

This discussion of breast implants outlines the changes in their design and manufacture from the earliest types of 'sponges' to the most recent soya bean oil filled, silicone gel and saline implants. The use of silicone gel implants has been especially controversial and subject to UK government funded review and inquiry. Policy regarding these has remained consistent despite continuing litigation in the US and the introduction of the EU MDD. However, after only four years on the market the newest soya bean oil filled implants have been withdrawn. There is currently no clinical data publicly available on their use. Regulation of breast implants has had a relatively high political profile and this may help to explain why a national registry was set up in 1993. However, despite the recommendations of the Independent Review Group (IRG, 1998) to introduce compulsory registration of breast implants, this remains voluntary and as yet no further research has been developed to look at implant use. The effects of harmonisation in Europe are unclear and getting agreement amongst member states about the safety and standards relating to breast implants has been a problem. No other European country has a breast implant registry. There has been organised consumer protest in the UK against the continuing availability of silicone breast implants and increasing attention focused on the activities of private cosmetic surgery clinics in a current Health Select Committee Inquiry.

Conclusion: comparative research in human implants

We conclude that our analysis raises important issues about the relationships between science, technology testing, and healthcare evaluation in relation to innovation and diffusion in these human implant technologies. The tensions between potentially beneficial innovations and state concerns with supporting this, whilst protecting the public health, are evident. Also the interaction of commercial, scientific, professional and consumer constituencies may be identified as important. Our purpose in this section is to reflect briefly upon the major contrasts in innovation and regulation evident in the two implant technologies which we have described, and to point toward the key topics for a research agenda which our analysis suggests it is now important to pursue.

The comparison of two technologies shows a number of interesting points of similarity and contrast. Implanted technologies present special problems of evaluation because long-term effects can generally not be predicted with certainty and the performance of the device is not directly observable. Perhaps most obviously, breast implants have evoked more controversy and more legal action than hip implants (e.g. Nyren et al., 1998; Shanklin & Smalley, 1998; Angell, 1997). Technological standards cover the production of both, and failures of some models of each have been reported in the UK to the governmental surveillance agency, the MDA. The definition of a 'failure' (under MDD) differs in striking ways. Safety issues arise in both cases, and are contentious. The organisation of recent policy-related clinical science for each technology has been constituted quite differently, in the one case by quasi-autonomous dedicated independent inquiries, in the other within the national health service research programme (HTA). A government-sponsored voluntary surveillance register exists for the one but not the other. International comparisons add a further dimension here: Finland, for example, currently has a national register for hip implants but not for breast implants, the opposite of the situation in the UK. Implantees in general lack information about the implanted device in both cases, but this has been much more of a public issue in the case of breast implants. Organisations of breast implantees have been formed but not hip implantees. In both cases members of the main specialty professions participate in the production of clinical evidence of long-term performance (e.g. Benediktsson, Perbeck, Geigant, & Solders, 1997; Iwuagwu & Frame, 1997; Heitmann, Schreckenberger, & Olbrisch, 1998; Coope & Dennison, 1998; Collis et al., 1998a,b, Malata, Feldberg, Coleman, Foo, & Sharpe, 1997; Coleman, Foo, & Sharpe, 1993, Nyren et al., 1998; Malchau, Herberts, & Ahnfelt, 1993). It appears that the changing meanings and dominant understandings of the different types of benefit to be derived from these implants, cosmetic or functional, may be salient in corresponding risk assessment processes and regulatory decision-making. Our limited information suggests that strategic alliances between surgeons and manufacturers may be common in the case of hips but less so for breast implant development. Design proliferation and cost issues are much more prominent as healthcare policy concerns in the case of hip implants. Both technologies have examples of anomalies in the availability of different implant designs, intra- and inter-nationally.

This brief descriptive summary based on the histories of innovation in hip and breast prostheses and their regulatory contexts raises many questions for a research agenda. In general our account shows, firstly, that there has been little study of the actors involved in innovation and the development of these technologies. Secondly, the organisational and inter-organisational processes which shape human implants and their deployment in the marketplace require investigation. There are important issues that may be raised at different points in the development of a technology. Drawing a parallel with medicines testing, the possible key stages might be identified as: pre-clinical testing, clinical testing/trial, approval/adoption and diffusion/post-marketing surveillance (though development should not be regarded as a unilinear process). We identify four related main areas for further research: (i) innovativeness and proliferation; (ii) safety and technological standards; (iii) clinical and social outcomes; (iv) consumer/user information and choice.

Innovativeness and proliferation

Professional knowledge and skills influence innovation through definition of the 'technical' issues and defining what the problem is — 'problematisation'. By examining the activity of the 'experts' — material scientists, designers and others working in both academic and commercial manufacturing settings — explanations of the diversity within product ranges will be developed. This might show, for example, the different processes affecting the development of variations in the coatings and finishes of hip prostheses. Clinicians are co-producers of knowledge about how implants work in use, so their relationships to the scientific and commercial communities are of particular importance (Blume, 1992). Their involvement in clinical trials or in state-sponsored inquiries can be expected to contribute to innovation through evaluation of clinical outcomes and clinical audit. But also they may have a key role in creating a demand for a range of prostheses and contributing to product design. Thus alliances between surgeons and manufacturers are crucial to the feasibility and forms of possible regulation.

Institutional structures and relationships also are important for understanding how certain brands or types of products become more widely available than others, or in explaining anomalies in implant availability (licensing) between countries (Altenstetter, 1996). The uneven pattern of use of hip prosthesis designs for example within the UK also has direct implications for implantees. This is of direct importance for implantees because there may be structural influences inhibiting access to 'the best' designs internationally. It has been shown in the case of pharmaceuticals, for example, that while the mass media may play a part in influencing regulatory decisions (Gabe & Bury, 1996), the structure of the regulatory process itself may be crucial in accounting for such anomalies (Abraham & Sheppard, 1998). This may apply, for example, to the non-licensing of some models of bio-actively coated implant in the USA even though they are available in Europe. Complex relationships between national politics, business, legal decisions and regulators may underlie the development of distinct national policy for breast implants in the USA (Palley, 1995).

It is important to examine the relationships between the demand for certain designs or brand of implant and clinical policy, practice or supplier/purchaser decisions at the level — in the UK — of the NHS trust or private clinic (Altenstetter, 1996). For example, what organisational (and economic) factors determined the adoption and use of the new triglyceride — oil-filled implants rather than silicone-gel-filled models? Why, if they were being promoted by manufacturers as the safer option (especially given the controversy surrounding silicone gel) were they not more widely used, at least up until the recent withdrawal of them from the UK market? The role of professional organisations in advising members on policy and practice with regard to 'innovative' products must also be taken into account in explaining different patterns of adoption and diffusion of different implants.

Safety and technological standards

Technological standards, laboratory testing and biocompatibility assessment are designed to ensure the safety of medical devices placed on the market. However, our comparison of two technologies has highlighted important contrasts. Until very recently a ruptured breast implant did not constitute an adverse incident in terms of the EC guidance on the types of incidents to be reported, neither did other commonly perceived 'complications' of breast implant surgery such as capsular contracture (EC, 1998). A serious deterioration in health or permanent impairment are criteria defining adverse incidents and in the case of breast implants the MDA had not accepted any evidence of this, even though some argue it had been provided. So while a ruptured breast implant did not previously appear to constitute a 'failure', aseptic loosening of a hip implant did. Now (since 1999) the predicted lifespan for breast implants is approximately 10 years (BAAPS), and revision within 10 years for both hip and breast implants is now seen as a reportable adverse incident in the UK. Soya bean oil filled implants were recalled after 74 adverse incidents were reported which raises questions about the number and type of reports which are received before certain policy decisions to be made. These contrasts again raise the question of whether the

different types of benefit and social meanings of different implant technologies are implicated in the different approaches to defining and assessing risk which are evident. Thus, it is important to investigate the social and organisational structures and processes that shape the definition of the key aspects of implant device regulation. This includes the processes involved in negotiating at a local level *what counts as* a reportable adverse incident, pointing towards the need for an examination of the ways in which competing accounts of what is acceptable practice are worked out. This applies also at the level of the manufacturer in deciding to report incidents to the MDA. Such an analysis would need to draw on existing sociological research which examines the relationships between science and policy making (for example, in relation to medicines (Abraham 1993, 1995) and in relation to non-medical technologies (Jasanoff, 1994; Cozzens & Woodhouse, 1995; Irwin, Rothstein, Yearley, & McCarthy, 1997)).

Involved in the negotiable definition of an adverse incident, is the question of how clinical and scientific evidence is collected and interpreted, and the institutional and professional processes in this. These processes are crucial because they may affect both the decision to report and its timing — of great importance if there is a safety issue and other implantees may be at risk if there is delay. It follows that also of importance are the issues of the construction of technological standards for implanted devices, and their operation in particular instances. Could, for example, the apparent — though contested — failure of the 'Capital Hip' (a me-too design based upon the standard 'Charnley' prosthesis) have been avoided had different biocompatibility or other test data been available to the manufacturers or the regulatory authority? The study of the practices of technological testing and the contestability of judgments made in projecting from test to 'real world' performance have been identified as important areas of concern in the study of the development of new technologies (Pinch, Ashmore, & Mulkay, 1992; Pinch, 1993). In this instance this applies to standards and test procedures for biocompatibility and other material and design properties. Although highly technical, technological standards are themselves the outcome of processes of negotiation, and are open to question and further development (e.g. Styles et al., 1998). Of interest then are both institutional politics and the cultural production of knowledge in issues of device safety. Controversies such as those surrounding the use of silicone gel, and now soya bean oil breast implants, may be approached in a number of ways and the case for an integration of a sociology of scientific knowledge approach (SSK) and a structural analysis would be fruitful here (Abraham, 1994; Martin & Richards, 1995).

Clinical and social outcomes

Surgical implants may convey a variety of individual and collective benefits affecting longevity, quality of life and social participation. The clinical and social sciences are increasingly involved in assessing these benefits and associated risks. Quality of life might be measured in relation to specific clinical interventions such as medicines or implants, and may also be assessed in terms of 'social' outcomes such as participation in social networks or employment. It has been noted especially in the case of hip implants that there has been a large amount of research knowledge examining the comparative efficacy of different models of artificial hip, whereas comparisons with alternative interventions for arthritic pain and locomotor problems, such as medicines or rehabilitative therapics, are relatively rare. This suggests that there may be systematic skewing in the setting of research agendas, possibly associated with the institutional and professional disciplinary interests which are able to establish their claims to knowledge and expertise in relation to implant technologies, as opposed to disciplines or perspectives with an interest in comparing technological to other forms of treatment.

A striking feature of our comparison of breast and hip implant technologies is the different ways in which 'clinical evidence' has been marshalled for the attention of the government and statutory authority. While there have been inquiries into silicone breast implants and a national registry set up, the process of gathering evidence about hip implants has been quite different and no registry has been set up yet (though the principle has been much debated in government and the orthopaedic profession since the 'Capital' incident). Is there a common cross-technology process at work here, reflecting distinctions between structures where safety (inherently more difficult to assure) is the prime issue rather than efficacy, as has been suggested? (Bodewitz, Buurma, & de Vries, 1987). Alternatively, it may be that the types of benefits potentially derived and the social meaning associated with different types of implant may be related to different institutional approaches to evidence-related regulation of risk. Thus the lack of attention paid to breast implants by the National Health Service HTA in the UK may be related to the fact that a large proportion of 'cosmetic' surgery is carried out in the private sector. It is also notable that while hip implants have been the object of a high level of political concern about cost effectiveness in the context of the NHS budget, this is not the case for breast implants.

Our analysis raises questions about the institutional form taken by perceived needs for surveillance of clinical outcomes. What are the institutional and professional prerequisites which predispose toward the establishment

PAPER 4

of a national registry rather than (or in addition to) other forms of regulatory structure? Are the implants regulated with registries in the UK special cases or distinctive from hip and other implants in certain respects? A comparative approach looking across the underlying processes behind several implant technologies may be especially valuable here. Of course, the existence of a registry does not guarantee or presuppose a curb on the proliferation of models (as the Norwegian hip registry noted above demonstrates). It is thus important to examine not only the causes of the creation of this type of surveillance system but also its consequences for innovation, health gain, health service policies and implantees' safety.

Under the new EU directives 'novelty' is a prerequisite for initiating clinical investigation but how this is defined, and the circumstances under which trials are required, is of crucial importance. This will influence the processes of innovation and diversification of devices. A focus here will permit examination of the relationships between clinical trials organised under different auspices (e.g. commercial, government-funded) and diffusion of the technology. For example, it is useful to know what trials have been carried out on the 'new' triglyceride breast implants in the UK or Europe, and the relation of these to other forms of evidence, and to regulatory action. The requirements and practices in the provision of clinical evidence for 'me-too' technologies, in different national bodies, where full new clinical investigations are not required, should also be investigated, using case study approaches.

User/consumer information and choice

Our comparative approach raises a number of different issues for consumers in relation to implant technologies. These relate to the role of implants in issues of personal identity, access to 'expert' information, and the question of active participation in innovatory and regulatory scientific processes.

Who are the users of implant technologies? Consumers, or those who are implanted with the prosthesis, can be taken to have only an indirect relationship to manufacturers or suppliers. The direct user is usually the surgeon. Patients are unlikely to specify or choose the type or brand of implant in other than the broadest terms. However, there may be indirect ways in which consumers shape innovation or may influence it in the future. Implant technologies can be seen as 'constructing' or 'configuring' consumers and/or users (Woolgar, 1991). The example of cochlear implants provides a striking example of a technology where the early clinical and commercial 'construction' of the end-user proved to be strongly challenged by the user community itself (Blume, 1997). Although recipients of hip prostheses are generally older and the average life expectancy of a woman with a breast implant, especially in cosmetic augmentation, is likely to be longer, both may be faced with difficult decisions when an implant fails or simply reaches the end of its lifespan. Hip implants are increasingly offered to younger people, and, given the limited lifespan of prostheses, it is important to examine the clinical and commercial construction of younger implantees in interaction with consumers' own beliefs, knowledge and values. The IRG (1998) recommends that women should have more detailed information about breast implants but what would be the consequences of providing it for this or other implanted devices? What is an acceptable lifespan for an implanted device from the implantee's point of view? Provision of more detailed information for consumers may thus have indirect implications for the processes both of innovation and of regulation.

However, the issue of provision of information raises the question of the current degree of transparency in the activities of experts involved in risk and benefit assessments. It remains the case that a culture of secrecy surrounds medical regulation in the UK, although government initiatives are beginning to address themselves to the issue. Empirical documentation of this culture has been provided in the case of medicines by Abraham and Sheppard (1997), and in the case of implant technologies it is not confined to the UK, having been discussed in the case of Finland in relation to contraceptive implants (Ollila & Hemminki, 1996). The accessibility of information may be influenced not only by trends in public governance, but also by the institutional structures through which scientific risk/benefit expertise is expressed. With the exception cited above, this appears to be an unexamined issue in the case of implant technologies. The issue of transparency is associated in turn with the related issue of consumer activism and the possible role of consumers in regulatory science.

Consumer activism and the development of social movements around health-related issues are growing phenomena, documented most fully perhaps in relation to issues of the environment and ecology (e.g. Irwin, 1995). In the case of medical science and regulation, the foremost example is certainly the case of 'AIDS activism' associated with the development of expertise and direct participation in regulatory processes by gay citizens in the USA (for a sociological view: Epstein (1996); for a Food and Drug Administration view: Edgar and Rothman, (1990)). The extent to which this may be a special case involving a controversial medical condition, high group consciousness and sense of community, and lack of effective treatment is unclear, but it does draw attention to questions of the social and institutional conditions under which active consumer participation in regulatory science might emerge. It is notable from Epstein's study that the area where active

contribution has most been accomplished is that of the science related to treatment rather than to diagnosis. Implant technologies may be seen as a technological solution offering one form of benefit, where alternative medical approaches might be envisaged in individual cases. It may be the case, also, that the different institutional relations in which the social meanings of implants are embedded have differing implications for consumer activity. Thus in our comparison it may be that the apparent relative lack of institutional affiliation between surgeons and breast implant manufacturers is associated with the relatively higher profile of breast implantee groups as a form of social movement, while in the case of hip implants the mediating position of orthopaedic surgeons contributes to a lack of a development of active consumerisation amongst these implantees.

In summary, comparative approaches to human implant technologies will assist in understanding the ways in which regulatory processes are created and applied to different groups of technologies. This in turn can contribute to policy appraisal in this complex area. While there have been sociological studies of the regulation of the pharmaceutical industry, our study of the regulation and forms of evidence applied to human implant technologies raises comparable but unexplored issues. The advent of HTA is associated with a broadening of evidence-related regulatory guidance. The contrasting histories of hip and breast implants and their geographical variations point to significant differences in the basis of regulatory decisions. A part played by healthcare science in such decisions is suggested by the HTA research programme's influence on advice disseminated in the National Health Service to restrict the range of hip prostheses used and for the government to consider the case for a national registry. However, its role in shaping national policy and regulation is less clear. In contrast, government-funded inquiry has been central to policy decisions relating to breast implants. While hip implant research has focused upon longevity and cost, the scientific discourse on breast implants has focused upon possible causal links with disease. There does not appear to be a consistent or widespread commitment to the use of registries. The operation of the 'vigilance system' also varies in each case. The regulatory definition of device failure is open to negotiation and is currently defined differently for our two exemplar technologies, resulting in different thresholds of reportability. Our analysis therefore raises important questions about post-implantation evaluation of performance for different types of devices and the processes by which evaluative standards are specified and applied to them. The necessity of revision surgery in the event of failure suggests some practical similarities, despite the different regulatory standards in place. Explanation of such variation in the UK system, we argue, needs to take into account analysis of similarities and differences between the regulatory policies, mechanisms and routines across EU member states. As in the case of pharmaceutical regulation in the European Union (Abraham & Lewis, 1999), the national implementation of EU-wide directives on safety and standards raises questions of possible competitive relationships between national regulatory bodies, and the implications of this for public health. This in turn leads to the question of the voluntary mutual recognition of standards between EU countries and between EU and other countries (Global Harmonisation Taskforce, 2000), and the extent to which such harmonisation may be in the interests of regulators, manufacturers and the public.

Finally, further study of innovation and regulation of medical devices will contribute to a critical analysis of 'evidence-based medicine'. While healthcare sciences are directed towards improving the efficacy of services, we have shown that the ways in which this expertise is enrolled, and the kinds of evidence produced, are of concern in shaping regulatory decisions. Both collaboration and conflicts between different disciplines shape the HTA agenda and healthcare knowledge (Faulkner, 1997). By comparing breast and hip implants we have begun to look at the sources and nature of the 'evidence' about such technologies and by whom it is produced. Consequently, in relation to the development of new products, we have identified a need to map out and examine the connections between clinicians, designers, manufacturers and patients stemming from this focus on the social and political processes which produce 'evidence' and healthcare knowledge. We take the view that strategic alliances between clinicians and manufacturers will impact on the diffusion of products. In addition, standard-setting and CE marking are primarily concerned with product standards and the movement of goods in the EU market. The extent to which this supports high-quality health services for patients is relatively unexplored. This is the fundamental tension between the innovative and regulatory pressures embodied in these technologies. So although it has been suggested elsewhere that commercial interests are not included in the HTA agenda (Faulkner, 1997) it is clear that they are important in shaping the proliferation, availability and use of medical devices such as hip and breast implants. These issues deserve further empirical exploration, further analysis and wider public debate. This may require a broadening of the definition of what constitutes medicine's evidence-base.

Acknowledgements

We are very grateful to Peter Glasner, Stuart Blume and an anonymous referee for comments on previous

PAPER 4

versions of this paper. Some of the material on hip implants draws on a project funded by the NHS R&D HTA Programme, 1996–97.

References

Abraham, J. (1993). Scientific standards and institutional interests: Carcinogenic risk assessment of benoxaprofen in the UK and the US. *Social Studies of Science, 23*, 387–444.

Abraham, J. (1994). Bias in science and medical knowledge: The Opren controversy. *Sociology, 28*, 717–736.

Abraham, J. (1995). *Science, politics and the pharmaceutical industry: Controversy and bias in drug regulation.* London: UCL Press.

Abraham, J. (1997). The science and politics of medicines regulation. In M.A. Elston (Ed.), *The sociology of medical science & technology* (pp. 153–182). Oxford: Blackwell Publishers

Abraham, J., & Sheppard, J. (1997). Democracy, technocracy, and the secret state of medicines control: Expert and nonexpert perspectives. *Science Technology and Human Values, 22*(2), 139–167.

Abraham, J., & Sheppard, J. (1998). International comparative analysis and explanation in medical sociology: Demystifying the Halcion anomaly. *Sociology, 32*(1), 141–162.

Abraham, J., & Lewis, G. (1999). Harmonising and competing for medicines regulation: How healthy are the European Union's systems of drug approval? *Social Science and Medicine, 48*, 1655–1667.

Advisory Group on Health Technology Assessment. (ca. 1992). *Assessing the effects of health technologies: Principles–practice–proposals.* Department of Health, London (undated).

Altenstetter, C. (1996). Regulating healthcare technologies and medical supplies in the European Economic Area. *Health Policy, 35*, 33–52.

Angell, S. (1997). *Science on trial: The clash of medical evidence and the law in the breast implant case.* New York: W.W. Norton.

Banta, H. D., & Luce, R. B. (1993). *Health care technology and its assessment: an international perspective.* New York: Oxford University Press.

BBC2. (1993). High price of hips. *Public Eye*, February 26.

Beck, U. (1992). *Risk society: towards a new modernity..* London: Sage.

Bendiktsson, K., Perbeck, L., Geigant, E., & Solders, G. (1997). Touch sensibility in the breast after subcutaneous mastectomy and immediate reconstruction with prosthesis. *British Journal of Plastic Surgery, 50*, 443–449.

Bijker, W.E., & Law, J. (Eds.). (1992). *Shaping technology/building society: Studies in sociotechnical change.* Cambridge, London: MIT Press.

Biomet. (1999). Hip systems, < http://www.biomet.com/surgeons/product/hips.html > March.

Blume, S. S. (1992). *Insight and industry: On the dynamics of technological change in medicine.* Cambridge, MA, London: MIT Press.

Blume, S.S. (1995). Cochlear implantation: Establishing clinical feasibility, 1957–1982. In N. Rosenberg, A. Gelijns, & H. Dawkins (Eds.), *Sources of medical technology, Medical innovation at the crossroads*, vol.5. (pp. 97–124). Washington, DC: Institute of Medicine/National Academy of Sciences.

Blume, S. S. (1997). The rhetoric and counter-rhetoric of a 'bionic' technology. *Science, Technology and Human Values, 22*(1), 31–56.

BOA. (1998). <http://www.kayr.demon.co.uk/bhs.htm#who > 18/12/98.

Bodewitz, H. J. H. W., Buurma, H., & de Vries, G. H. (1987). Regulatory science and the social management of trust in medicine. In W. E. Bijker, T. P. Hughes, T. Pinch (Eds.), *The social construction of technological systems: New directions in the sociology and history of technology* (pp. 243–259). Cambridge, MA, London: MIT Press.

Booth, R. E. (1994). Current concepts in joint replacement: The closing circle: Limitations of total joint arthroplasty. *Orthopedics, 17*(9), 757–759.

Boseley, S. (1998a). Clinics attacked on breast implants. *The Guardian*, July 15, 4.

Boseley, S. (1998b). Who is to blame? *The Guardian*, 21 January, 6–7.

Boseley, S. (1999). Soyabean breast implants banned. *The Guardian*, March 9,3.

Bowsher, C. A. (1991). Medical devices: Public health at risk. *Society, 28*, 74–78.

British Association of Aesthetic Plastic Surgeons (BAAPS). (1998). *Breast augmentation, patient information leaflet.* BAAPS, London.

Bulstrode, C. J. K., Murray, D. W., Carr, A. J., Pynsent, P. B., & Carter, S. R. (1993). Designer hips. *British Medical Journal, 306*, 732–733.

Coleman, D., Foo, T., & Sharpe, D. (1991). Textured or smooth implants for breast augmentation? A prospective trial. *British Journal of Plastic Surgery, 44*(6), 444–448.

Collis, N., Khoo, C., & Sharpe, D. (1998a). Media are too eager to link silicone to disease. *British Medical Journal, 316*(7129), 477.

Collis, N., & Sharpe, D. (1998b). Breast implants: A survey of cosmetic clinics. *British Journal of Plastic Surgery, 51*, 311–312.

Consumers' Association. (1997). Hip replacements. *Health Which*? February, 32–33.

Coope, C., & Dennison, E. (1998). Do silicone breast implants cause connective tissue disease? There is no clear evidence that they do. *British Medical Journal, 316*, 403–404.

Council on Scientific Affairs (1993). American Medical Association. Council on Scientific Affairs. Silicone Breast Implants. *Journal of the American Medical Association, 270*, 2602–2606.

Cozzens, S., & Woodhouse, E. (1995). Science, government and the state. In S. Jasanoff, G. Markle, J. Petersen, & T. Pinch (Eds.), *Handbook of science and technology studies.* London: Sage.

Department of Health (DoH). (1993). *Research for health.* Department of Health, Leeds.

Department of Health (DoH). (1998). Independent review Group Report <http://www.silicone. review.gov.uk/testing/index/htm > 11/4/98.

DePuy. (1997). Annual Report. DePuy International Ltd, Leeds.

PAPER 4

Edgar, H., & Rothman, D. J. (1990). New rules for new drugs: The challenge of AIDS to the regulatory process. *The Milbank Quarterly*, 68(Suppl.1), 111–142.

Ellis, R. (1999). Now women face agony over new soya bean breast implants. *The Express*, March 9, 8.

Epstein, S. (1996). *Impure science: AIDS, activism, and the politics of knowledge*. Berkeley: University of California Press.

European Commission (EC). (1998). *The medical devices vigilance system*. European Commission Guidelines, Competent Authority, UK.

Faro, L. M. C., & Huiskes, R. (1992). Regulation of orthopaedic implants: Two systems compared. *Proceedings, Institution of Mechanical Engineers, C441/035*, 43–47.

Faulkner, A. (1997). 'Strange bedfellows' in the laboratory of the NHS? An analysis of the new science of health technology assessment in the United Kingdom. In M. A. Elston (Ed.), *The sociology of medical science & technology* pp. (183–207). Oxford: Blackwell Publishers.

Faulkner, A., Kennedy, L.G., Baxter, K., Donovan, J., Wilkinson, M., & Bevan, G. (1998). Effectiveness of hip prostheses in primary total hip replacement: A critical review of evidence and an economic model. *Health Technology Assessment*, 2(6)i–iv, 1–134.

Federal Judiciary Center. (1998). <http://www.fjc.gov/BREIMLIT/mdl926.htm > 4/12/98.

Fielder, J. H., & Black, J. (1995). But doctor, it's my hip! The fate of failed medical devices. *Kennedy Institute of Ethics Journal*, 5(2), 113–131.

Fitzpatrick, R., Shortall, E., Sculpher, M., Murray, D., Morris, R., Lodge, M., Dawson, J., Carr, A., Britton, A., & Briggs, A. (1998). Primary total hip replacement surgery: A systematic review of outcomes and modelling of cost-effectiveness associated with different prostheses. *Health Technology Assessment*, 2(20)i–iv, 1–64.

Food & Drug Administration (FDA). (1998). Breast implant information pack, <http://www. fda.gov> 19/2/99.

Gabe, J. (Ed.). (1995). *Medicine, health and risk*. Oxford: Blackwell.

Gabe, J., & Bury, M. (1996). Halcion Nights: A sociological account of a medical controversy. *Sociology*, 30(1), 447–469.

Gaudilliere, J., & Lowy, I., (Eds.). (1998). *The invisible industrialist: Manufacture and the production of scientific knowledge*. Basingstoke: Macmillan.

Gelijns, A. C. (1990). Comparing the development of drugs, devices, and clinical procedures. In A. C. Gelijns (Ed.), *Modern methods of clinical investigation, Medical innovation at the crossroads*, vol. 1. pp. (147–201). Washington, DC: National Academy Press.

Giddens, A. (1991). *Modernity and self identity: Self and society in the late modern age.*. Cambridge: Polity Press.

Global Harmonisation Taskforce. (2000). Study Group 1. Role of standards in the assessment of medical devices. GHTF.SG1.NO12R10. <http://www.medical-devices.gov.uk/global.htm> 27 July 2000.

Gott, D., & Tinkler, J. (1994). *Silicone implants and connective tissue disease, evaluation of evidence for an association between the implantation of silicones and connective tissue disease*. London: Medical Devices Agency.

Guthrie, R. H. (1994). *The truth about breast implants*. New York: Wiley.

Haiken, E. (1997), *Venus Envy: A History of Cosmetic Surgery. John Hopkins University Press, USA*.

Hall, C. (1999). Breast implants ban after scare. *The Daily Telegraph*, March 9,1.

Havelin, L. I., Espehaug, B., & Vollset, S. E. (1993). The Norwegian arthroplasty register: A survey of 17,444 hip replacements 1987–1990. *Acta Orthopaedica Scandinavica*, 64(3), 245–251.

Heitmann, C., Schreckenberger, C., & Olbrisch, R. (1998). A silicone implant filled with cohesive gel: Advantages and disadvantages. *European Journal of Plastic Surgery*, 21, 329–332.

Hinsliff, G. (1998). Cosmetic surgery to face official inquiry. The Daily Mail, Decmber 9, 1–2.

Hoggett, P. (1991). A new management in the public sector. *Policy and Politics*, 19(4), 243–256.

Huiskes, R. (1993). Failed innovation in total hip replacement: Diagnosis and proposals for a cure. *Acta Orthopaedica Scandinavica*, 64(6), 699–716.

Independent Review Group (IRG). (1998). *Silicone gel breast implants*. The Report of the Independent Review Group (obtainable from <http://www.silicone-review.gov.uk> April 1998).

Irwin, A. (1995). *Citizen science: A study of people, expertise and sustainable development*. London: Routledge.

Irwin, A., Rothstein, H., Yearley, S., & McCarthy, E. (1997). Regulatory science — towards a sociological framework. *Futures*, 29(1), 17–31.

Iwuagwu, F., & Frame, J. (1997). Silicone breast implants: Complications. *British Journal of Plastic Surgery*, 50, 632–636.

Jasanoff, S. (1994). *The fifth branch, science advisers as policymakers*. London: Harvard University Press.

Jeffrey, K. (1995). Pacing the heart: Growth and redefinition of a medical technology, 1952–1975. *Technology and Culture*, 36(3), 583–624.

Jennett, B. (1986). *High technology medicine: Benefit or burden?*. Oxford: Oxford University Press.

Kessler, D. (1992). The basis of the FDAs decision on breast implants. *New England Journal of Medicine*, 326, 1713–1715.

Kessler, D. Merkatz , R., & Shapiro, R. (1993). A call for higher standards for breast implants. *Journal of the American Medical Association*, 270, 2607–2608.

Laurance, J. (1999). Poison warning to 5,000 women as breast implant banned. *The Independent*, March 9,3.

Law, J., & Bijker, W. E. (1992). Postscript: Technology, stability and social theory. In W. E. Bijker, & J. Law (Eds.), Shaping technology/building society: Studies in sociotechnical change. (pp. 290–308). Cambridge, London: MIT Press.

Levidow, L., Carr, S., von Schomberg, R., & Wield, D. (1996). Regulating agricultural biotechnology in Europe: Harmonisation difficulties, opportunities, dilemmas. *Science and Public Policy*, 23(3), 135–157.

Malata, C., Feldberg, L., Coleman, D., Foo, I., & Sharpe, D. (1997). Textured or smooth implants for breast augmentation? Three year follow-up of a prospective randomised controlled trial. *British Journal of Plastic Surgery*, 50(2), 99–105.

PAPER 4

Malata, C., Varma, S., Scoot, M., Liston, J., & Sharpe, D. (1994). Silicone breast implant rupture: Common/serious complication? *Medical Progress through Technology, 20*(34), 251–260.

Malchau, H., Herberts, P., & Ahnfelt, L. (1993). Prognosis of total hip replacement in Sweden. Follow-up of 92,675 operations performed 1978–1990. *Acta Orthopaedica Scandinavica, 64*(5), 497–506.

Marinker, M., & Peckham, M. (Eds.). (1998). *Clinical futures.* London: BMJ Publishing Group.

Martin, B., & Richards, E. (1995). Scientific knowledge controversy, and public decision making. In S. Jasanoff, G. Markle, J. Petersen, T. Pinch, (Eds.), *Handbook of science and technology studies.* London: Sage.

Medical Device Agency (1996). *Safety Notice SN9620.* Department of Health.

Medical Devices Agency (MDA). (undated). *Introduction to the adverse incident centre.* Medical Devices Agency, London.

Medical Devices Agency (MDA). (1993a). *Bulletin 2 — The CE marking.* Medical Devices Agency, London.

Medical Devices Agency (MDA). (1993b). *Bulletin 4 — Conformity assessment procedures.* Medical Devices Agency, London.

Medical Devices Agency (MDA). (1995). *Code of practice on enforcement.* Medical Devices Agency, London.

Medical Devices Agency (MDA). (1997). *Hydroxyapatite coatings for orthopaedic implants.* SN9713, November, Medical Devices Agency, London.

Medical Devices Agency (MDA). (1998a). *Bulletin 3 — The vigilance system.* Medical Devices Agency, London.

Medical Devices Agency (MDA). (1998b). *The medical devices vigilance system European Commission guidelines.* Medical Devices Agency, London.

Medical Devices Agency (MDA). (1998c). *3M Capital hip system.* HN9801, February, Medical Devices Agency, London.

Medical Devices Agency (MDA). (1998d). *Guidance notes for manufacturers on statistical considerations for clinical investigations of medical devices.* Medical Devices Agency, London.

Medical Devices Agency (MDA). (1998e). *Guidance on the medical devices vigilance system for CE marked joint replacement implants.* Medical Devices Agency, London.

Medical Device Agency (MDA). (1999a). *Guidance on the medical devices vigilance system for CE marked breast implants.*

Medical Devices Agency (MDA). (1999b). <http://www.doh.gov.mda.../DTS6... > 12/2/99.

Medical Devices Agency (MDA). (1999c). *Trilucent™ breast implants: Voluntary recall,* http://www.gov.doh.uk/ March 8th.

Mowatt, G., Bower, D. J., Brebner, J. A., Cairns, J. A., Grant, A. M., & McKee, L. (1997). When and how to assess fast-changing technologies: A comparative study of medical applications of four generic technologies. *Health Technology Assessment, 1*(14)i–iv, 1–51.

Murray, D. G., Crown, R. S., Dickersin, K., Woods Duncan, P., Epps, C. H., Ettinger, W. H., Friedlaender, G. E., Lane, J. M., Lemons, J. E., Lewis, J. L., MacKenzie, E. J., Pinals, R. S., & Treuhaft, P. S. (1995). NIH consensus conference: Total hip replacement. *Journal of the American Medical Association, 273*(24), 1950–1956.

Murray, D. W., Carr, A. J., & Bulstrode, C. J. (1995). Which primary total hip replacement? *Journal of Bone and Joint Surgery, 77-B,* 520–527.

National Breast Implant Registry (NBIR). (1998). Fourth Annual Report 1996–1997, Plastic Surgery Department, Salisbury District Hospital, Salisbury.

National Coordinating Centre for Health Technology Assessment. (1998). The Annual Report of the NHS Health Technology Assessment Programme 1998. NHS Executive, Department of Health, London.

NBIR. National Breast Implant Registry (1977). *Third Annual Report, 1995–1996.* Plastic Surgery Department, Salisbury District Hospital, Salisbury.

Newman, K. J. H. (1993). Total hip and knee replacements: A survey of 261 hospitals in England. *Journal of the Royal Society of Medicine, 86,* 527–529.

NHS Centre for Reviews and Dissemination. (1996). Total Hip Replacement. *Effective Health Care, 2*(7), Nuffield Institute for Health/NHS Centre for Reviews and Dissemination, Leeds/York.

NICE. (2000). *Selection of artifical hip joints for primary total hip replacement — guidance from NICE.* <http://www.nice.org.uk/nice-web/ > August. National Institute of Clinical Excellence.

Nyren, O., Yin, L., Josefsson, S., McLaughlin, J., Blot, W. J., Engqvist, M., Hakelius, L., Boice Jr, J. D. & Adami, H.-O. (1998). Risk of connective tissue disease and related disorders among women with breast implants: A nationwide retrospective cohort study in Sweden. *British Medical Journal, 316,* 417–422.

Ollila, E., & Hemminki, E. (1996). Secrecy in drug regulation: Licensing documentation on the Norplant contraceptive implant. *International Journal of Risk and Safety in Medicine, 9,* 161–172.

Palley, H. (1995). The evolution of FDA policy on silicone breast implants: A case study of politics, bureaucracy and business in the process of decision-making. *International Journal of Health Services, 25,* 573–591.

Parker, L. (1993). Social justice, federal paternalism, and feminism: breast implants in the cultural context of female beauty. *Kennedy Institute of Ethics Journal, 3*(1), 57–76.

Paul, J. P. (1997). Development of standards for orthopaedic implants. *Proceedings of the Institution of Mechanical Engineers, Part H, 211*(1), 119–126.

Pinch, T. (1993). "Testing-one,two,three...testing!": Toward a sociology of testing. *Science,Technology and Human Values, 18*(1), 25–41.

Pinch, T., Ashmore, M., & Mulkay, M., (1992). Technology, testing, text: Clinical budgeting in the U.K. National Health Service. In W. E. Bijker, & J. Law (Eds.), *Shaping technology/building society* (pp. 265–289). Cambridge, London: MIT Press.

Shanklin, D., & Smalley, D. (1998). Risk of connective tissue disease among women with breast implants. *British Medical Journal, 317,* 470.

Sheldon, T., & Faulkner, A. (1996). Vetting new technologies. *British Medical Journal, 313,* 508.

Shilling, C. (1993). *Body and social theory.* London: Sage.

PAPER 4

Star, S. (Ed.). (1995). *Ecologies of knowledge: Work and politics in science and technology*. Albany, New York: University of New York.

Stocking, B. (Ed.). (1988). *Expensive health technologies: Regulatory and administrative mechanisms in Europe*. Oxford: Oxford University Press.

Stocking, B. (1992). The introduction and costs of new technologies. In E. Beck, S. Lonsdale, S. Newman, & D. Patterson (Eds.), *In the best of health? The status and future of health care in the United Kingdom*. London: Chapman & Hall.

Studin, Dr. (1998). Breast implants <http://www.webplastics.-com/ silicone.htm> 6/11/98.

Styles, C. M., Evans, S. L., & Gregson, P. J. (1998). Development of fatigue lifetime predictive test methods for hip implants: Part 1. Test methodology. *Biomaterials*, *19*(11–12), 1057–1065.

Sweetnam, D. R. (1981). A surveillance scheme with 'Recommended List' of artificial joints. *Health Trends*, *13*, 43–44.

Tinkler, J., Campbell, H., Senior, J., & Ludgate, S., (1992). *Evidence for an association between the implantation of silicones and connective tissue*. Department of Health, London.

Turner, B. S. (1992). *Regulating bodies: Essays in medical sociology*. New York: Routledge.

Watson, R. (1998). Bid to ban silicone implants in Europe. *British Medical Journal*, *316*(7137), 1037.

Williams, S. J. (1997). Modern medicine and the 'uncertain body': From corporeality to hyperreality? *Social Science and Medicine*, *45*(7), 1041–1049.

Wilson, E. (1999). 'Healthier' breast implants banned over safety fears. *The Daily Mail*, March 9,5.

Woolgar, S. (1991). Configuring the user — the case of usability trials. In J. Law (Ed.), *A sociology of monsters: Essays on power, technology and domination*. London: Routledge.

PAPER 4

PAPER 5

Originally published 2000

Author: Faulkner A, Brookes S, Donovan J, Selley S, Gillatt D, Hamdy F.

The use of prostate-specific antigen testing in the detection of localised prostate cancer: current opinion and urological practice in the United Kingdom.

Published in: European Journal of Public Health, 10(4): 289-295.

Acknowledgment: Oxford University Press for permission to reproduce this article.

EUROPEAN JOURNAL OF PUBLIC HEALTH 2000; 10: 289–295

The use of prostate-specific antigen testing in the detection of localized prostate cancer

Current opinion and urological practice in the United Kingdom

ALEX FAULKNER, SARA T. BROOKES, JENNY DONOVAN, SARA SELLEY, DAVID GILLATT, FREDDIE HAMDY *

Background: The prostate-specific antigen (PSA) test and its interpretation plays a crucial role in the detection of early localized prostate cancer. However, inaccuracy of the test, inability to predict the aggressiveness of the disease and the lack of evidence about the comparative effectiveness of treatments have led to major dilemmas in considering whether to employ the PSA test and which cut-off points to use in interpreting its results. The aim of this study was to evaluate current urological practice in the UK regarding the use of PSA testing. Methods: A postal questionnaire survey of all consultant urologist members of the British Association of Urological Surgeons was conducted. Statistical analysis included proportional odds regression models to examine factors associated with urologists' preferences for different definitions of 'normal' PSA cut-off levels. Results: The survey response rate was 60%. The majority of consultant urologists applied the PSA test routinely. There was a high level of agreement amongst UK urologists on normal PSA cut-off points (<4.0 ng/ml) for asymptomatic men under 60 years of age. There was very wide variation in the definition of normal PSA cut-offs for older (≥60 years) asymptomatic men. A preference for lower cut-off points, leading to investigation with ultrasound and biopsy, was significantly associated with larger urology department size, the presence of a prostate cancer subspecialist in the department and relatively short length of specialization in urology. Conclusions: Prostate cancer screening and early detection practices and reported incidence rates of the disease are likely to be influenced by variation in urologists' interpretations of PSA. Despite increasing evidence in favour of lower PSA cut-off levels, particularly for younger men (<60 years), urologists in the UK are divided over their interpretation. Men, particularly over age 60 years, have varying chances of further investigation following PSA testing. Any trial of prostate cancer screening or treatment should take this potential variation into account. Standard protocols for PSA interpretation should be implemented.

Keywords: practice variations, prostate cancer, prostate-specific antigen testing, UK, urology

The last 10 years have seen a considerable escalation internationally in the clinical use of the prostate-specific antigen (PSA) serum test, paralleled by increases in reported incidences of prostate cancer. A substantial proportion of the increase in diagnosis is attributable to the use of this test.[1] While the test has an indisputable role in the monitoring of the disease once diagnosed, its use in early detection or screening programmes remains controversial because of the uncertainty in predicting the potential aggressiveness of the early-stage disease and benefits of radical treatment.[2] Identification of early prostate cancer in increasing numbers of men leads to persistent management dilemmas for patients and urologists alike.

Prostate cancer is a major burden. Approximately one in ten deaths from cancer in England and Wales amongst men are due to prostate cancer. It is the third most common cause of cancer death following lung and colorectal cancer. The registered number of newly diagnosed prostate cancers increased in England and Wales from 10,180 in 1986 to 13,490 in 1991, an increase of 37% over 5 years.[3] The most recently available figures show an age-standardized incidence in England and Wales of 28 per 100,000 and rates in other European countries ranging from 15 to over 60 per 100,000, the highest rates being reported in Iceland and Zurich, Switzerland.[4]

PSA is a serine protease produced almost exclusively by prostatic epithelium. It is present in blood, normally in small concentrations. The serum test is quick and easy to perform. It is used by urologists, other specialists and general practitioners. Many different assays have been developed, although there are a small number of market leaders.

While it is clear that PSA measures are able to indicate the presence of prostate cancer in many cases, there is wide variation in reported performance. Ranges from 57 to 99% were reported for sensitivity and from 59 to 97% for specificity in the authors' recent systematic review.[5] The US Congress Office of Technology Assessment[6] in the USA found positive predictive values for the test

* A. Faulkner[1], S.T. Brookes[2], J. Donovan[2], S. Selley[2], D. Gillatt[3], F. Hamdy[4]
1 School of Social Sciences, Cardiff University, Wales, UK
2 Department of Social Medicine, University of Bristol, England, UK
3 Bristol Urological Institute, Southmead Hospital, Bristol, England, UK
4 Division of Surgical Sciences and Anaesthesia, Section of Urology, University of Sheffield, Royal Hallamshire Hospital, England, UK
Correspondence: A. Faulkner, School of Social Sciences, Cardiff University, 50 Park Place, Cardiff, CF10 3AT, Wales, UK, tel. +44 29 20 875260, fax +44 29 20 874759, e-mail: FaulknerAC@cardiff.ac.uk

PAPER 5

(normal 0–4 ng/ml) ranging between 11 and 50% in 16 studies in a variety of settings with different patient selection criteria. Parkes[7] pooled published data to show a detection rate of 79% for PSA in symptomatic men with a histological diagnosis, associated with an 8% false-positive rate in asymptomatic men. Adding the PSA test to positive digital rectal examination (DRE) results increases positive predictive value by factors of between 0.5 (20–37%)[8] and 3.9 (19–75%) in asymptomatic men.[9]

These performance characteristics, particularly the relatively poor specificity, together with treatment uncertainties, make PSA currently unacceptable as a population screening test.[5] New methods of measuring or indexing PSA are continually being sought. These include age-specific reference ranges, density (ratio of PSA to prostate volume), velocity (rate of change in the level) and the ratio of 'free' to 'complexed' PSA, which currently is receiving the most attention.[10,11]

The cut-off level in PSA interpretation is critical, since this is the main factor informing the decision to investigate further with transrectal ultrasound (TRUS) and biopsy. Although 4.0 ng/ml is the most common cut-off point used for defining a 'normal' PSA, lower levels are sometimes used. There is some reason to consider lowering the cut-off level regarded as critical.[12] Catalona et al.[13] reported cancers found in a screening population of men with PSA in a range of 2.6–4.0 ng/ml: there was a 22% prevalence of prostate cancer, the 'majority' of which appeared 'medically important'.[13] Other studies reported for volunteers with PSA in the 0–4 ng/ml range percentages of 21 [14] and 29 [15] confirming prostate cancer.

Consensus about the existence and nature of the disease is partly conditional upon changing detection techniques and their interpretation. The adoption of new diagnostic technologies into clinical practice is potentially influenced by many forces.[16] The extent of diffusion of PSA technology for population screening in the USA and some European countries and its use in clinical detection of the disease suggest that there are complex international variations in the detection of prostate cancer in the male population. The social and psychological effects of the use of such diagnostic technologies, particularly in population screening programmes involving the asymptomatic public, are a major concern.[17]

Given the acknowledged role of the PSA test in the increasing incidence of prostate cancer and the ethical dilemmas surrounding early detection of the disease, it is important to know how this detection technique and its different measurement methods are being used and interpreted in everyday clinical practice. This article analyses data from a postal questionnaire survey of consultant urologists conducted in 1995 to examine urological practice and opinion in the UK regarding the use of the PSA test. This is set in the context of a review of current knowledge regarding the performance characteristics of PSA technologies. In particular, we investigated whether the detection of early prostate cancer with the aid of the PSA test is consistent between urologists and the degree to which variations in specialist opinion and practice might reflect evidence and uncertainties about the test or be associated with different characteristics of individual urologists or hospital urology departments.

METHODS

The 1995 register of members of the British Association of Urological Associations (BAUS) was used as the basis for the survey. All consultant members were sent a letter and postal questionnaire. The sampling frame was such that the exact number of active consultant urologists cannot be stated. Of the consultant members 15% are estimated to have been general surgeons[18] and approximately 400 (±10) urologists. Non-responders to the initial mailing were sent a reminder and then contacted by telephone or facsimile and sent a further questionnaire if required. The survey questions sought information about urologists' practices and opinions regarding diagnosis, management and screening for prostate cancer, using case vignettes and a mixture of closed multiple-choice and open-ended questions (full copy available from the authors). This article reports the results relevant to the use of PSA testing for detection purposes.

Statistical analyses were performed – using the STATA statistical package[19] – and included an examination of frequency distributions and the calculation of descriptive statistics. The distribution of the five-category outcome, the urologist's assessment of normal PSA cut-off points, was examined across the different levels of possible explanatory variables using the proportional odds model.[20] This considers the outcome as an ordered categorical variable and assumes that the odds ratios are the same across all of the possible binary representations of the outcome, that is <4.0 versus 4.1+, ≤6.0 versus 6.1+, ≤8.0 versus 8.1+ and ≤10.0 versus 10.1+ ng/ml. The explanatory variables considered in these models were department size, the presence of a prostate cancer subspecialist in the department, the length of time the urologist had specialized in urology, the percentage size of the urologist's case load of men having prostate cancer and the size of the local district population. These variables were first considered separately in univariable models and second, after allowing for relationships between them, in multivariable models. The analysis was carried out for each of four age groups separately (40–49, 50–59, 60–69 and 70–79 years).

Verbatim responses to open-ended questions were classified into broad categories according to the theme expressed.

RESULTS

Two hundred and forty-four practising consultant urologists responded to the survey, a response rate of approximately 60% (see Methods). The urologists varied in their level of specialist experience from 2 to 30 years (mean 14.1 years). Fifteen percent (36) of the urologists held special clinics for patients with prostate cancer, 53% (130) managed 100 or more patients and 28% (68) managed 200 or more. Two-thirds (67%,164) estimated that they had seen 20 or more new patients suspected of

having prostate cancer in the previous 3 month period, with a mean of 15.7 (median 12) newly confirmed cases per consultant over that period.

Sevety-nine percent (193) of the urologists estimated that at least 50% of patients referred to them had had a PSA test performed already by their GP and 37% reported that at least 80% of men referred had already been tested. Sixty-eight percent (207) of the urologists reported that all of their patients had received a PSA test in the process of confirming a diagnosis of prostate cancer; 84% reported that at least 50% had done so. A majority of men will thus have been PSA tested again following referral by a GP, with two-thirds of urologists testing all their patients regardless of any tests carried out in general practice.

The assay used most commonly was the Hybritech Tandem-R (50%), followed by the Abott IMx (11%) and Hybritech Tandem-E (8%). Seventeen percent used other assays and 15% reported not knowing which assay was used for their PSA tests.

Approximately one-third of the sample had the more recently introduced indexing methods available to them in addition to standard PSA measurement *(table 1)*, the ratio of free to complex PSA being least widely available (9%). The urologists' opinions of which method of PSA test indexing is the most useful clinically was divided amongst the available methods, broadly reflecting their availability, though age range was believed to be the most useful by the smallest proportion. Twenty-eight percent (68) were not able to give a clear opinion and a slightly higher percentage than had it available believed the ratio of free to complex PSA to be the most useful *(table 1)*.

The urologists were explicitly asked whether they believed age-specific cut-off points were useful in principle in the diagnosis of prostate cancer and gave open-ended comments on this issue. Opinion was greatly divided, 38% feeling this approach was useful, 33% not useful and 29% being undecided. The greatest number of those who commented (45/111) referred to clinical issues as being at least or more important than age/PSA, defining these as the size and density of the prostate (32 responses) and the totality of the clinical decision (13 responses). Further positive comments noted that age-specific ranges should not be interpreted too rigidly, but that they were valuable in averting unnecessary investigations and referrals in the older age groups, though low cut-off points might be used defensively in testing younger men.

The urologists varied widely in their interpretations of appropriate cut-off levels for PSA in asymptomatic men *(figure 1)*. Results for men over age 60 years elicited a high degree of ambivalence amongst urologists. The greater the age of the patient, the greater was the disagreement as to appropriate PSA cut-off level. While 88% believed the 4.0 ng/ml cut-off point to be appropriate for men aged 40–49 years, only 10% of urologists believed this level to be appropriate for men aged 70–79 years.

Univariable analysis *(table 2)* found that the presence of a prostate cancer subspecialist in the department was statistically significantly associated with a preference for lower cut-off points for PSA in asymptomatic men over 60 years. For men over 60 years of age or less than 50 years, a similar association was found for larger department size (defined as three or more consultants). The length of time the urologist had specialized also had a significant association with cut-off points, urologists with more than 20 years of experience being more likely to prefer a higher PSA level for asymptomatic men in the age groups less than 70 years. Neither the percentage size of the urologist's case-load of men having prostate cancer nor the size of the local district population showed any association with urologists' preferences for PSA cut-off points in any age group.

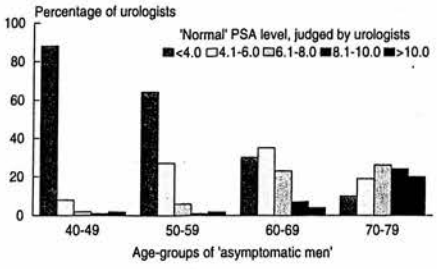

Figure 1 Variation in UK urologists' assessments of normal PSA limits for asymptomatic men in different age groups

For men aged 40–49 years, multivariable modelling *(table 3)* found that the length of specialization and the presence of a subspecialist remained statistically significant after adjusting for each other. Urologists with a subspecialist in the department were five times less likely to choose a higher cut-off point than if no subspecialist were present. Independently of this, if the urologist had specialized for more than 20 years, they were three and a half times more likely to use a higher cut-off point than if specialization were less than 10 years.

For men aged 50–59 years, size of department and length of specialization remained independently associated with cut-off point preference. A larger department size appeared to halve the likelihood of a higher cut-off point level and specialization of over 20 years more than

Table 1 Availability and perceived value of different PSA measurement techniques

Method	Percentage (number) having method available N=236		Percentage (number) stating method 'most useful' N=244	
Simple PSA	95	(225)	29	(71)
Density	34	(81)	11	(27)
Velocity	43	(102)	17	(42)
Age range	32	(75)	5	(11)
Ratio free/ complex	9	(21)	10	(25)
Do not know		–	7	(18)
Missing		–	21	(50)

PAPER 5

Table 2 Factors associated with urologists' preferences for PSA cut-off levels in asymptomatic men of different ages (odds ratios from univariable analysis)

Explanatory variable and categories	Man's age (years) 40–49 OR	p	50–59 OR	p	60–69 OR	p	70–79 OR	p
Size of department		0.98		0.11		0.018		0.01
<3 consultants	1.00		1.00		1.00		1.00	
≥3 consultants	0.99		0.64		0.56		0.56	
Subspecialist		0.0037		0.51		0.0020		0.01
Absent	1.00		1.00		1.00		1.00	
Present	0.21		0.83		0.45		0.54	
Length of specialization		0.0095		0.0012		0.037		0.59
<10 years	1.00		1.00		1.00		1.00	
10–19 years	0.86		1.60		1.49		1.21	
≥20 years	3.46		3.87		2.35		1.39	
Population		0.56		0.59		0.47		0.49
<250,000	1.00		1.00		1.00		1.00	
250,000–350,000	1.12		1.12		0.86		1.01	
≥350,000	0.65		0.80		0.68		0.74	
Case-load		0.34		0.52		0.63		0.58
<10%	1.00		1.00		1.00		1.00	
>10%	0.66		1.20		0.89		1.15	

P-values (p) given to the first two significant figures and odds ratios (OR) expressed in terms of likelihood of preference for higher PSA cut-off points

Table 3 Factors associated with urologists' preferences for PSA cut-off levels in asymptomatic men of different ages [odds ratios and confidence intervals (CI) from final multivariable regression models]

	Explanatory variable	OR	95% CI	p
Man's age (years)				
40–49	Subspecialist			0.0019
	Absent	1.00		
	Present	0.18	0.05–0.65	
	Length of specialization			0.0110
	<10 years	1.00		
	10–19 years	0.80	0.25–2.54	
	≥20 years	3.46	1.18–10.20	
50–59	Size of department			0.0240
	<3 consultants	1.00		
	≥3 consultants	0.51	0.28–0.92	
	Length of specialization			0.0003
	<10 years	1.00		
	10–19 years	1.56	0.78–3.12	
	≥20 years	4.57	2.11–9.92	
60–69	Size of department			0.0200
	<3 consultants	1.00		
	≥3 consultants	0.54	0.32–0.91	
	Subspecialist			0.0100
	Absent	1.00		
	Present	0.50	0.30–0.85	
	Length of specialization			0.0098
	<10 years	1.00		
	10–19 years	1.40	0.79–2.49	
	≥20 years	2.83	1.43–5.57	
70–79	Subspecialist			0.0160
	Absent	1.00		
	Present	0.54	0.33–0.89	

P-values (p) given to the first two significant figures and odds ratios (OR) expressed in terms of likelihood of preference for higher PSA cut-off points

PAPER 5

quadrupled the likelihood of a higher cut-off point level for these men.

For men aged 60–69 years, size of department, the presence of a subspecialist and the length of specialization in urology all remained significant after adjusting for each other. A larger department size appears to halve the likelihood of a higher cut-off point and urologists with a subspecialist in the department were half as likely to choose a higher cut-off point than if no subspecialist were present. Specialization for more than 20 years nearly tripled the likelihood of preferring a higher cut-off point, compared to specialization of less than 10 years.

For patients aged 70–79 years, only the presence of a subspecialist remained significant, halving the likelihood of preference for a higher PSA cut-off point level.

CONCLUSIONS

The results presented here are likely to be more widely applicable to consultant urologists in the UK. However, it has not been possible to compare the characteristics of the responding urologists to the total BAUS consultant membership because BAUS do not have such data. It can be stated that the range of lengths of specialization in urology, the range of sizes of department and the numbers of patients with prostate cancer seen amongst the respondents suggest that a full range is adequately represented (see Results). Other surveys of urologists in the UK with adequate response rates were sought for comparison. Only one study, a survey of urologists and general surgeons with a special interest in urology[18] conducted about the same time as the present study, was found. This survey investigated urological malignancies in general rather than prostate cancer and the two surveys collected different demographic data on the urologists, so direct comparison between the two sets of respondents was not possible. However, both surveys included a number of similar questions about treatment preferences for prostate cancer. For example, where Savage et al.[18] reported active treatment being favoured by 91% for poorly differentiated T1 disease in patients less than age 70 years, 96% of our respondents favoured this and radiotherapy was favoured by 50% in both surveys; in well-differentiated T1 disease in patients over age 70 years, Savage et al. reported 69% in favour of observation alone, comparable to 66% favouring this approach in our survey (in the case of a man aged 75 years, PSA of 20 and Gleason score 3). It seems unlikely, therefore, that major response bias was occurring amongst our 'sample' of urologists, although it is possible that some of the same urologists failed to respond to either survey.

With the reservations outlined above, we believe that our findings are more widely applicable to consultant urologists in the UK.

The PSA test is used as an aid in the detection of prostate cancer in the great majority of cases in the UK. Urologists tend to perform a PSA test irrespective of previous measurement by the GP. This may reflect 'blind' routine testing, awareness that different assays yield different results, knowledge that PSA values can fluctuate or a combination of these.

Persuasive clinical arguments are frequently advanced in support of formal screening systems using PSA for early detection of prostate cancer. However, there is acute conflict between this viewpoint and recent 'evidence-based' recommendations regarding screening. This is illustrated in the USA, where the American Cancer Society and the American Urological Association both recommend annual screening of asymptomatic men starting at age 50 years, while the US Preventive Task Force and the National Cancer Institute recommend neither PSA nor DRE as screening tools for prostate cancer detection. Despite these recommendations, use of the PSA test in screening asymptomatic men in the USA appears to be increasing.[21] Recent systematic reviews of evidence in the UK[5,22] have contributed to a policy decision by the NHS Executive not to introduce formal screening.[23] However, PSA testing is widely available and rising public awareness is likely to increase demand for the test.

The number of commercially available PSA assays is escalating. Standardization has become a major issue internationally. Differences between the sixty-plus assays currently available on the European market may create difficulties in interpretation and comparability of incidence data.[24] Three of the most common assays have been shown to detect cancer at a reasonable level of equivalence by one independent evaluation,[25] although there is conflicting evidence suggesting that one of them, the Abott IMx assay, departs significantly from the other (Hybritech Tandem-R).[26] For the best possible interpretation of results, clinicians need to know which assay is being used and to use information about the reference range for the assay in question. However, the results of the survey discussed here suggest that a notable proportion are unaware of the assay used and that standardized technology and instructions for use are not sufficient to secure standardized interpretation of PSA results across the range of clinical presentations in different specialist service settings.

There is also some suggestion of broad ethnic and/or geographic differences in average PSA levels in the male population and a cut-off level lower than 3.0 has been justified on these grounds as the level for recommendation of biopsy in an Asian (Japanese) group.[10] Oesterling et al.[27] also reported lower levels of PSA amongst Japanese men than in age-matched Whites, which was attributed to smaller gland size.

There is evidence that PSA concentration is directly related to age.[28–30] The PSA normal range of 0–4 ng/ml does not take account of either age differences or variations in prostate volume associated with age. Age-specific normal reference ranges have been suggested by Oesterling et al.[29] and Dalkin et al.,[31] with upper limits as follows: age 40–49 years 2.5/– (Dalkin et al. did not give a figure), 50–59 years 3.5/3.5, 60–69 years 4.5/5.4, and 70–79 years 6.5/6.3.

These variations in the detection of prostate cancer in relation to different cut-off levels of PSA suggest that the search for a single absolute cut-off point may be misguided

and that, as Gillatt and Reynard[32] concluded from their case-control studies of four groups of men, the 'normal range' should be adjusted depending upon the population being assessed. While this is a sensible conclusion from the available scientific evidence about PSA, we have shown in this survey how consultant urologists in the UK vary in their assessment of appropriate 'normal' cut-off points for PSA in relation to men's age group. One-third of the UK urologists assessed the normal range for men aged 60–69 years as having a higher cut-off level than Oesterling et al.[28] or Dalkin et al.[31] suggested and probably more than half did so for the 70–79 years age group. It appears that the majority of UK consultant urologists would use a higher point, with therefore lower sensitivity but higher specificity and positive predictive value, before investigating further in the case of older asymptomatic men. These judgements may be informed by beliefs regarding the value of treatment in relation to patients' age and life expectancy and knowledge of the evidence that PSA concentration is related to age. In the case of younger men, perceived detection thresholds could be subject to influence by defensive medical attitudes. These findings may reflect a more cautious attitude in the UK compared with a more aggressive surgical culture or wider diffusion of a 'screening culture' in the USA. Greater ethnic diversity in the USA may also be an underlying factor.

The implications of our results for patient care are not straightforward. Age-specific reference ranges in principle should help to prevent overinvestigation of mildly raised PSA, particularly in older men who may not benefit from aggressive treatment. Published evidence about the presence of cancer in men with PSA values less than 4.0 ng/ml, the generally accepted 'normal' level which the urologists in our sample appear to accept uncritically, suggests that for younger men this cut-off point may be too high for the purpose of detection. On the other hand, men of any age who attend larger referral centres or centres with a subspecialist are more likely to have PSA tests interpreted with reference to lower cut-off points than in other centres and, thus, proceed to further investigation by TRUS and biopsy. However, if they are managed by a consultant at an advanced stage of their urology career (≥20 years' experience), a higher PSA level is more likely to be used, thus reducing the likelihood of further investigation. The generally conservative approach to PSA interpretation in the sample as a whole appears less pronounced amongst urologists in the early stages of their consultant careers. While this may benefit detection levels in men under age 70 years, it may also imply a preference for radical treatments.[33]

These results imply that incidence rates of identification of the disease are partly dependent upon service-related, supply-induced differentials in patterns of detection based on PSA testing. While it has been known that the diffusion of PSA testing has contributed in part to the international increases in the incidence of prostate cancer, this study suggests that the overall increase masks likely interregional and intraregional variation in the stages of ascertainment of the disease. In addition, a consequence of this variation is that in any trial c screening or treatment for early prostate cancer, it i crucial to implement tightly controlled and monitore protocols for PSA interpretation in addition to stand ardized assay technology.

Urologists' interpretations of PSA results in the UK ar complex and conflicting, reflecting a mixture of servic and individual consultant variations and uncertaint about the performance of PSA technology and availabl treatments. The results of this survey support the need t control the use of PSA testing. Guidelines to reduce th scale of the routine use of PSA testing in urology referra centres have been published in the UK since the surve was conducted.[34] The effects of these guidelines shoul be evaluated by monitoring of PSA testing rates an assessment of variation as in the study reported here.

This study formed part of a project funded by the NHS R&D Healt Technology Assessment Programme.

The views expressed in this paper are those of the authors and nc necessarily those of the Standing Group on Health Technology, th HTA Commissioning Board, the HTA Panel members or the De partment of Health.

The Department of Social Medicine at the University of Bristol i part of the Medical Research Council Health Services Researcl Collaboration.

The authors are grateful to Dr Tim Peters for additional statistica advice and for the comments of three anonymous referees on previous version of this paper.

REFERENCES

1 Gann PH. Interpreting recent trends in prostate cancer incidence and mortality. Epidemiology 1997;8(2):117-20.

2 NHS Centre for Reviews and Dissemination. Screening for prostate cancer. Effectiveness Matters. York: University of York, 1997;2(2).

3 Office for National Statistics. Population and health monitor. Registrations of cancer diagnosed in 1991, England and Wales. London: Government Statistics Service, 1996.

4 Parkin DM, Whelan SL, Ferlay J, Raymond L, Young J. Cancer incidence in five continents, Vol. VII. Lyon: International Association of Cancer Registries (IARC), 1997.

5 Selley S, Donovan JL, Faulkner A, Coast J, Gillatt D. Diagnosis, management and screening of early localised prostate cancer. Health Technol Assess 1997;1(2).

6 US Congress Office of Technology Assessment. Costs and effectiveness of prostate cancer screening in elderly men. Washington: US Government Printing Office, 1995.

7 Parkes CA. An epidemiologist's viewpoint on screening. Cancer Surv 1995;23:127-40.

8 Mettlin C, Lee F, Drago J, Murphy GP. The American Cancer Society National Prostate Cancer Detection Project. Cancer 1991;67:2949-58.

9 Mettlin C. Early detection of prostate cancer following repeated examinations by multiple modalities: results of the American Cancer Society National Prostate Cancer Detection Project. Clin Invest Med 1993;16:440-7.

10 Egawa S, Soh S, Ohori M, et al. The ratio of free to total serum prostate specific antigen and its use in differential diagnosis of prostatic carcinoma in Japan. Cancer 1997;79(1):91-8.

11 Stephan C, Lein M, Jung K, Schnorr D, Loening SA. The influence of prostate volume on the ratio of free to total prostate specific antigen in serum of patients with prostate carcinoma and benign prostate hyperplasia. Cancer 1997;79(1):104-9.

12 Labrie F, Dupont A, Suburu R, et al. Serum prostate specific antigen as pre-screening test for prostate cancer. J Urol 1992;147:846-52.

13 Catalona WJ, Smith DS, Ornstein DK. Prostate cancer detection in men with serum PSA concentrations of 2.6 to 4.0

ng/ml and benign prostate examination. JAMA 1997;277(18):1452-5.
14 Catalona WJ, Richie JP, DeKernion JB, et al. Comparison of prostate specific antigen concentration versus prostate specific antigen density in the early detection of prostate cancer: receiver operating characteristic curves. J Urol 1994;152:2031-7.
15 Mettlin C, Murphy GP, Lee F, et al. Characteristics of prostate cancers detected in a multimodality early detected program. Cancer 1993;72:1701-8.
16 Blume SS. Insight and industry: on the dynamics of technological change in medicine. Cambridge, MA. and London: MIT Press, 1992.
17 Stewart-Brown S, Farmer A. Screening could seriously damage your health. BMJ 1997;314:533.
18 Savage P, Bates C, Abel P, Waxman J. British urological survey practice: 1. Prostate cancer. Br J Urol 1997;79:749-55.
19 StataCorp. Stata statistical software, release 5.0. College Station, TX: Stata Corporation, 1997.
20 Bender R, Grouven U. Ordinal logistic regression in medical research. J R Coll Phys Lond 1997;31(5):546-51.
21 McKnight JT, Tietze PH, Adcock BB, Maxwell AJ, Smith WO, Nagy MC. Screening for prostate cancer: a comparison of urologists and primary care physicians. S Med J 1997;89(9):885-8.
22 Chamberlain J, Melia J, Moss S, Brown J. The diagnosis, management, treatment and costs of prostate cancer in England and Wales. Hlth Technol Assess 1997;1(3).
23 NHS Executive. Population screening for prostate cancer. Leeds: NHS Executive, 1997.
24 Semjonow A, Brandt B, Oberpenning F, Roth S, Hertle L. Discordance of assay methods creates pitfalls for the interpretation of prostate-specific antigen values. The Prostate 1996;29(S7):3-16.
25 Oesterling JE, Moyad MA, Wright Jr GL, Beck GR. An analytical comparison of the three most commonly used prostate-specific antigen assays: Tandem-R, Tandem-E and IMx. Urology 1995;46(4):524-32.
26 Brawer MK, Bankson DD, Haver VM, Petteway JC. Comparison of three commercial PSA assays: results of restandardization of the Ciba Corning method. The Prostate 1997;30(4):269-73.
27 Oesterling JE, Kumamoto Y, Tsukamoto T, et al. Serum prostate-specific antigen in a community-based population of healthy Japanese men: lower values than for similarly aged white men. Br J Urol 1995;75:347-53.
28 Oesterling JE, Cooner WH, Jacobsen SJ, Guess HA, Lieber MM. Influence of patient age on the serum PSA concentration. Urol Clin N Am 1993;20:671-7, 680.
29 Oesterling JE, Jacobsen SJ, Chute CG, et al. Serum prostate-specific antigen in a community-based population of healthy men. JAMA 1993;270:860-5.
30 Motta J, Miller LS, Armenakas NA, Fracchia JA. The histological correlation of age and prostate volume with PSA and PSAD in 1250 men. Proc Am Urol Assoc 1995;153:418.
31 Dalkin BL, Ahmann FR, Kopp JB. Prostate-specific antigen levels in men older than 50 years without clinical evidence of prostatic carcinoma. J Urol 1993;150(6):1837-9.
32 Gillatt D, Reynard JM. What is the 'normal range' for prostate-specific antigen? Use of a receiver operating characteristic curve to evaluate a serum marker. Br J Urol 1995;75:341-6.
33 Donovan JL, Frankel SJ, Faulkner A, Selley S, Gillatt D, Hamdy FC. Dilemmas in treating early prostate cancer: the evidence and a questionnaire survey of consultant urologists in the United Kingdom. BMJ 1999;318:299-300.
34 British Association of Urological Surgeons/Royal College of Surgeons. Guidelines on management of men with lower urinary tract symptoms suggesting bladder outflow obstruction. London: Royal College of Surgeons, 1997.

Received 5 June 1998, accepted 16 September 1999

PAPER 5

PAPER 6

Originally published 1997

Author: Faulkner A, Harvey I, Peters T, Sharp D, Frankel S.

Profiling outpatient workload: practice variations between consultant firms and hospitals in South West England.

Published in: Journal of Epidemiology & Community Health, 51(3): 310-314.

Acknowledgment: BMJ Publishing Group for permission to reproduce this article.

Journal of Epidemiology and Community Health 1997;51:310–314

Profiling outpatient workload: practice variations between consultant firms and hospitals in south west England

Alex C Faulkner, Ian M Harvey, Tim J Peters, Deborah J Sharp, Stephen J Frankel

Abstract

***Objectives*—To describe the variation in outpatient new to old ratios between consultants and between providers for seven high volume specialties (four surgical, three medical).**

***Design*—This was a descriptive study at consultant and provider unit level based upon patient administration system data from the South and West Regional Health Authority for the financial year 1992–93. Additional components of variance analysis was used to distinguish individual consultant effects from host institution effects.**

***Setting*—The former South Western Regional Health Authority area from Gloucestershire to Cornwall.**

***Subjects*—Altogether 345 consultant firms in seven specialties grouped into 13 provider unit groups.**

***Main measures*—New to old ratio, omitting elective inpatients followed up as outpatients.**

***Results*—Variation between consultants is greater in surgical than in medical specialties, while absolute levels of new to old ratios tend to be higher in surgical specialties than in medical. Variation between provider unit groups is also greater in surgical specialties. Analysis of variance shows that more total variance is attributable to provider unit group in gynaecology than in other specialties.**

***Conclusions*—Within individual specialties there is evidence of substantial variation that is not attributable to underlying differences in morbidity patterns. There is evidence of marked variation in terms of both individual consultants and institutions, a finding that provides the springboard for further analytical work. Published routine outpatient activity statistics should distinguish between new referrals, inpatient follow up, and clinic rebookings.**

(*J Epidemiol Community Health* 1997;51:310–314)

The importance of outpatient services in health services policy is signalled by the UK *Patient's Charter* waiting time targets. Service rate variations are evident in many aspects of outpatient services.[1] A recent study has drawn attention to historical trends in total outpatient workload and in workload per consultant in the UK,[2] suggesting that while the total number of hospital consultants has increased over the last 40 years, numbers of both follow up and new patients seen per consultant have declined, leading overall to relatively little change in the total number of outpatients seen. The nature of the historical "outpatient problem" is changing as more flexible and complex models of care delivery are being introduced at the primary/secondary care interface.[3] Further examination of the processes of outpatient care delivery in the National Health Service is therefore timely. Routine workload activity data now available can be disaggregated to examine separately the different reasons for outpatient attendance within specialties and within provider unit groups.[4,5]

One approach to the problems of outpatient workloads and waiting lists is to examine the extent to which patients are recycled in outpatient clinics.[2,6,7] It can be assumed that a reduction in re-attendances within outpatient clinics frees time in which, in principle, additional new patients may be seen. This is indeed one reason for calls to increase the proportions of consultations conducted by consultants, who are widely assumed to be less inclined to recycle patients than are junior staff.[8] It is also part of the rationale for the *Patients' Charter* requirement that all newly referred patients be seen by a consultant. Also, the value of routine outpatient surveillance for a widening range of conditions is under scrutiny—for example, asthma, hypertension, rheumatoid arthritis, routine surgical follow up for non-complicated cases, and follow up of treated malignant conditions such as breast cancer.[9]

The ratio of new to established ("old") patients provides a measure of recycling. Variation in this ratio has not previously been documented at a hospital, specialty or consultant level. Demonstration of substantial variation beyond that expected from casemix differences, provides prima facie evidence of variation in clinical judgement and practices between individuals and institutions. This can be a springboard for further analytical work ultimately of benefit to patients.[10] For example, studies of *inpatient* practice variations in the surgical management of benign prostatic hypertrophy,[11] led to the identification of a critical level of variation, stimulating trials of the effectiveness of alternative therapies.

The aims of the present study are to describe the variation in the ratio of new referrals to re-attendances in the outpatient workload at various levels of aggregation, to estimate the

Department of Social Medicine, University of Bristol, Canynge Hall, Whiteladies Road, Bristol BS8 2PR
A C Faulkner
I M Harvey
T J Peters
D J Sharp
S J Frankel

Correspondence to:
Mr A C Faulkner.

Accepted for publication September 1996

PAPER 6

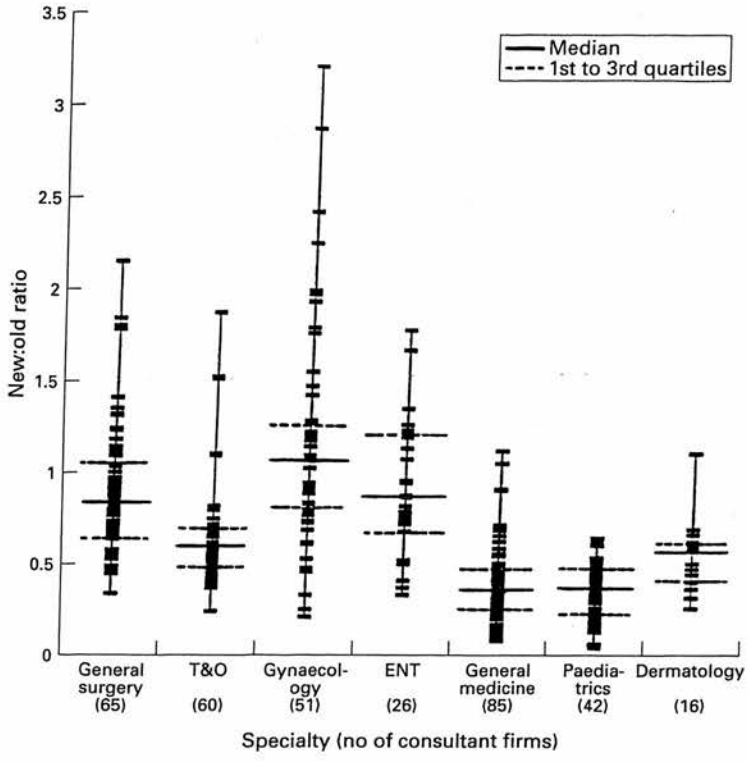

Figure 1 Variation in new to old referral ratios for outpatients in relation to consultant specialties.

respective contributions of organisational and individual differences to the total variation observed, and to suggest plausible explanations for these findings to generate hypotheses for further investigation.

Methods

DATA SOURCES

Summary data were analysed from the Patient Administration System (PAS) Outpatient module implemented by South and West Regional Health Authority. This is the routine data file used for summarising information for Körner returns (KH09) to the Department of Health. Using the unique consultant code, data were linked for the field: "Source of referral (this attendance)". The figures for total *new referrals* seen were generated from combining the following referral categories: "GP referral", "A & E referral", "Referral by other consultant", "Self referral", "Private patient (same consultant)", "Domiciliary visit (same consultant)", "Outpatient attendance, other purchaser" and "Outpatient attendance, after emergency admission". *Old* (established) patient numbers were obtained from a single code, "Outpatient attendance, this provider" (defined as attendances initiated in the consultant clinic in which the patient is again being seen). Inpatients admitted electively and subsequently followed up in outpatients were excluded from the calculation of the ratio, since their attendance represents a distinct element of outpatient workload, separate both from re-booked attendances and from new referrals. Our new to old ratio differs therefore from the new to *follow up* ratio which can be calculated from the KH09 returns. Except for inpatient follow up, these classifications to new or follow up adopt the conventions used to derive the Körner new and total follow up figures submitted by regional health authorities to the Department of Health.

Data were collected for all consultants within the former south western region from the specialties of general surgery, general medicine, trauma and orthopaedics (T&O), paediatrics, gynaecology, ear nose and throat (ENT), and dermatology for the financial year 1992–93, the most recent complete year available. These specialties were selected because they are high volume specialties spanning all ages and representing both surgical and non-surgical disciplines. Consultant codes for which fewer than 200 consultations were recorded for the whole year were excluded to avoid random variation due to small numbers.

Hospitals have been grouped together into "provider unit groups" to reflect the locations in which individual consultants predominantly or exclusively work. In cases where consultants worked in more than one provider unit group, the unit where most cases were seen was deemed to be the consultant's base for the purposes of analysis.

KEY POINTS

- Variation in the ratio of new referrals to clinic re-bookings between consultants and between provider units is greater among surgical than medical specialties.
- Patients' likelihood of being re-booked for an outpatient consultation depends partly on their place of residence within a (regional health authority) catchment area.
- Evaluation of specialty-specific outpatient service patterns should take account of institutional effects as well as practice variations between clinicians.
- Published routine outpatient activity information should allow inpatient follow up, clinic re-bookings, and new referrals to be distinguished.

STATISTICAL METHODS

Interpretation of the new to old ratio may be aided by recognising that it can be defined as: *1/mean no of re-attendances for each new patient*

Any given ratio may therefore be alternatively expressed as the number of re-attendances associated on average with each new referral attendance (in other words, the mean re-attendance). For example, a new to old ratio of 2.00 is equivalent to a mean re-attendance of 0.50.

Following descriptive analyses, the proportion of the variance in either the ratio or the mean re-attendance attributable respectively to the individual consultant/firm and to the provider unit group was assessed using a components of variance analysis[12] in *Minitab for Windows*. Prior to undertaking these analyses of variance, the validity of the assumption of a

PAPER 6

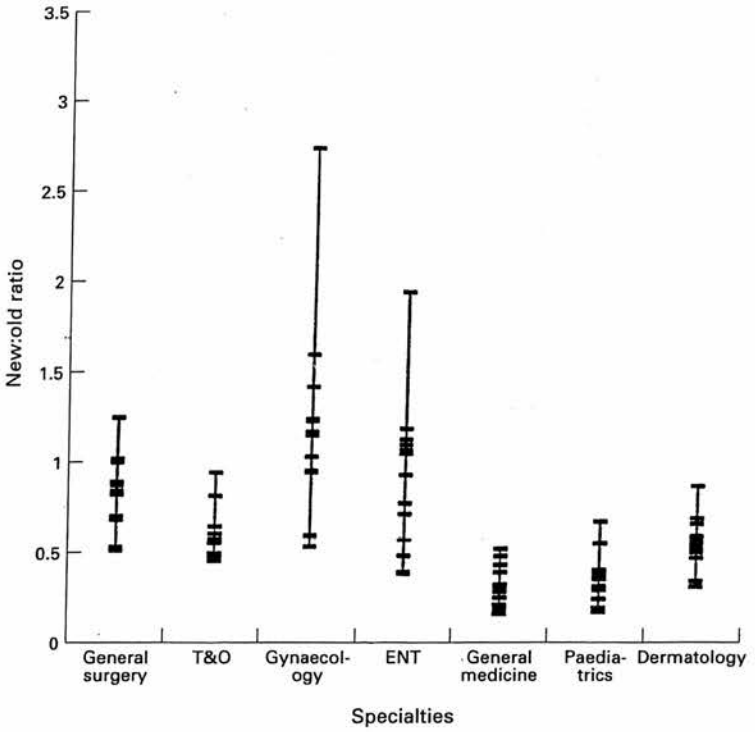

Figure 2 Variation in new:old ratios for outpatients between 13 provider groups within seven specialties.

Table 1 Ratios of new referrals seen to re-attendances initiated in the consultant clinic, south western region, 1992–93 (ranges, medians, and quartiles)

Specialty	No of consultants/firms	Range	Quartiles			
			1st	Median	3rd	Q*
General surgery	(65)	0.33–2.15	0.65	0.81	1.04	0.20
Trauma and orthopaedics	(60)	0.23–1.87	0.47	0.57	0.67	0.10
Gynaecology	(51)	0.20–3.21	0.78	1.06	1.27	0.25
ENT	(26)	0.32–1.78	0.67	0.84	1.19	0.26
General medicine	(85)	0.07–1.12	0.23	0.30	0.44	0.11
Paediatrics	(42)	0.03–0.64	0.20	0.32	0.45	0.13
Dermatology	(16)	0.25–1.11	0.40	0.58	0.60	0.10

* Q (quartile deviation) = $\frac{\text{quartile 3} - \text{quartile 1}}{2}$[14]

Table 2 Magnitudes of absolute and inter-consultant firm variance in new:old ratios

Specialty	No of consultants/ firms	Absolute inter-consultant variance	Absolute inter-provider unit group variance	Total variance
General surgery	(65)	0.110	0.024	0.134
Trauma and orthopaedics	(60)	0.066	0.014	0.080
Gynaecology	(49)*	0.143	0.113	0.256
ENT	(26)	0.078	0.071	0.149
General medicine	(85)	0.033	0.012	0.045
Paediatrics	(42)	0.022	0.003	0.025
Dermatology	(15)†	0.013	0.005	0.018

* Excluding two outliers; † excluding one outlier.

Gaussian distribution was assessed by the use of normal plots and the Shapiro–Wilk test.[13] For all specialties other than general surgery and T&O, the new to old ratio was a closer fit to a Gaussian distribution than was the mean re-attendance, although for dermatology and gynaecology it was necessary to exclude one and two (very high) outliers respectively. For general surgery and T&O, the results of the analysis of variance for the new to old ratio were nevertheless very similar to those for the mean re-attendance and hence for simplicity the results for all specialties are presented for the new to old ratio.

Given the different numbers of consultations involved for each consultant firm within a given specialty, any such components of variance models should account for the resultant differential random variation. To achieve this, as well as the unweighted analysis of new to old ratios, two further approaches were adopted. The first was an analysis of the new to old ratios using weights in proportion to the total numbers of new plus old consultations (these were heuristic weights given that no simple equation exists for the variance of such ratios). The second analysis was of the number of new consultations as a proportion *(p)* of the total new plus old consultations *(N)*, using weights proportional to the inverse of the variance of the proportion p (assuming a binomial distribution these variances were $p(1-p)/N$). These analyses are presented in table 3.

Results

VARIATION BY CONSULTANT FIRM

The main finding is the existence of marked variation between consultant firms within specialties in the ratio of new referrals to re-attendances (fig 1 and table 1). The greatest variation is seen in gynaecology, ENT, and general surgery, and the greatest overall range in gynaecology and general surgery. Relatively less variation was observed in T&O, general medicine, paediatrics, and dermatology. These findings were not altered by exclusion of consultants based in the Bristol teaching hospitals whose outpatient practice might reflect the consequences of providing a tertiary referral service.

There was also notable variation between specialties in terms of the absolute level of new to old ratios as indicated by the medians in figure 1. General surgery, gynaecology, and ENT have relatively high ratios (lower mean re-attendances), and general medicine and paediatrics have relatively low ratios (higher mean re-attendances).

VARIATION BY PROVIDER UNIT GROUP

Since each of the provider unit groups in this study is providing a general service to a local population, casemix differences that may exist between consultants can be reduced by examining data relating to units rather than to individual consultants. A degree of variation clearly persists between provider units (fig 2), although as anticipated this is less than the variation between consultants. Absolute inter-unit variation varies across specialties with the non-surgical specialties exhibiting lower variation.

It is useful in addition to examine what proportion of the variation between consultants is attributable to the effect of the unit in which they work. The results of the components of variance models are presented in table 2. Total variance is highest in gynaecology, ENT, and general surgery, consistent with the picture in figure 1. Both the absolute inter-provider unit

Table 3 Percentages of variance attributable to provider unit group, calculated by three different methods

Specialty	Percentage of variance attributable to provider unit group		
	New:old ratio*	New:old ratio†	Proportion new‡
General surgery	18.0	23.2	30.7
Trauma and orthopaedics	17.2	21.3	29.8
Gynaecology	44.1	46.2	55.4
General medicine	26.4	25.3	33.3
Paediatrics	13.6	21.9	25.5

Excluding ENT and dermatology because of relatively small numbers of consultants/consultations.
* Unweighted analysis; † weighted by total numbers of new and old outpatient consultations; ‡ weighted by the reciprocal of the variance of the proportion of new consultations (see Methods).

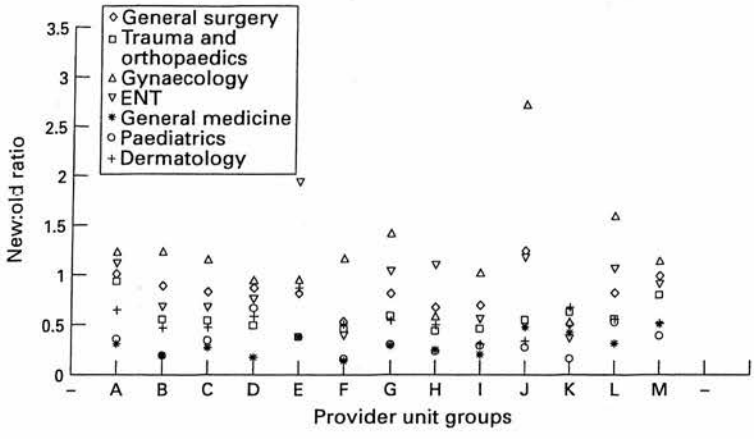

Figure 3 Variations in outpatient new:old ratios between seven specialties in 13 provider unit groups.

group variance (3rd column table 2) and the proportion of total variance attributable to provider unit group are greatest in gynaecology (table 3). This is the case regardless of whether weighted or unweighted analysis of variance is performed.

The total numbers of consultations in ENT and dermatology are too small for reliable estimates to be made of the relative size of inter-provider and intra-provider group variance. These specialties have been excluded from table 3.

Figure 3 displays the new to old ratios for all specialties in each provider unit group in the region. There is variation in the likelihood of re-attendance between the geographical areas represented by different provider unit groups, as shown for example by comparison of groups K and L.

To assess the effect of the omission of elective inpatient follow up cases from the new to old ratios, these attendances were analysed as a proportion of total consultations for the three consultant data points at each extreme of the range in each specialty. This showed that for all but one specialty there was no clear difference in the proportion of total workload constituted by elective inpatient follow up between the two extremes. The exception was ENT, where consultant firms with the highest new to old ratios also had relatively high proportions of inpatient follow up in the workload (8.1%, 10.6%, 17.6%), compared with those with the lowest new to old ratios, who conversely had consistently low proportions of inpatient follow up in their workloads (0.7%, 4.1%, 0.3%).

Discussion

The new to old ratio investigated in this paper represents the balance between the two elements of outpatient workload which account for most patients seen. The variation in the profiles of inter-consultant and inter-provider ratios across specialties suggests strongly that a specialty-specific approach is essential in examining outpatient performance data. Of the seven specialties examined here, the four with the highest inter-consultant variance are surgical, and the three with the lowest variance are broadly medical. This pattern of variation seems specific to the outpatient setting. When, for example, standardised *inpatient* discharge ratios (defined as observed to expected numbers of discharges multiplied by 100) were examined, wider range and variation were seen in general medicine than in general surgery,[15] the reverse of the pattern shown here for outpatients. This suggests that patients seen in the outpatient setting present distinctive problems of clinical management and organisation.

There are a number of possible reasons for this variation. One potential explanation is artefact due to data inaccuracy. However, the data demonstrate a reasonable and plausible level of consistency. This was seen, for example, in the relative magnitudes of new to old ratios in the surgical compared with non-surgical specialties—differences which make sense intuitively in terms of the higher proportions of chronic illness seen in the medical specialties. Hence in our view errors in data are unlikely to have had a substantial effect upon the overall profiles of variation found.

Some variation in the new to old ratio might be attributable to the classification of sub-specialisations. In gynaecology, for example, the consultants/firms with the higher ratios (seeing lower proportions of re-attendances) tended to be classified as "gynaecology only", rather than obstetrics and gynaecology combined. On the other hand, the general medicine data excluded recognised subspecialties such as cardiology, neurology, respiratory medicine, and gastroenterology. The persisting variation at provider unit group level shows that classification of subspecialties is not the full explanation of the variation, even in gynaecology.

Variation in the ratio might theoretically also be accounted for by different practice with regard to outpatient follow up of elective inpatients. The data do not in general show such an effect, except for ENT. In this specialty the effect seems to be that higher elective inpatient follow up is associated with a higher new to old ratio, suggesting that "recycling" of outpatients is inhibited by the relatively high proportions of follow up of inpatients in the ENT workload.

In summary, there is substantial variation which requires explanation. The components of variance analyses suggest that some variation is attributable to features of provider units and some to individual consultant behaviour. The absolute variation attributable to consultants is greater in surgical specialties than non-surgical specialties (table 2). A number of possible explanations can be suggested. Surgeons may in

PAPER 6

general pursue more independent, individual styles of practice than physicians due to differences in self selection into the specialties, education, or professional socialisation. Development of special interests within the surgical specialties may be greater than in the medical specialties. There was some suggestion that this was the case in gynaecology. A consultant with one of the lowest rates of new patient attendance had a particular interest in subfertility, investigation and management of which requires repeated visits. On the other hand, checking of a sample of the consultant firms in the other specialties at the extremes of the ranges of mean re-attendance did not suggest that generalist consultants (for example, general physicians) with particular specialist interests (for example, gastroenterology) tended to occupy the extremes. We conclude that this explanation does not apply generally.

A further possible explanation for inter-consultant variation is systematic variation in clinical judgement. It may be that the clinical knowledge base in the surgical specialties is less well rooted in clear evidence of effectiveness, is more complex, and is less amenable to standard-setting than is the case for medical specialties. There is certainly evidence from the south west to suggest that many more clinical guidelines have been developed in the medical than in the surgical specialties (D Baker, personal communication).

Effects at the level of the provider unit group are also substantial, the greatest being observed in gynaecology (table 2 and table 3). It is unlikely that this variation can be wholly attributed to underlying morbidity differences; the appropriate morbidity data are not available, but a study of factors predicting high mean re-attendance in another region in one specialty found that diagnosis and disease severity combined were able to account for only a small percentage (<20%) of the variance found.[20] In gynaecology, uncertain management of conditions such as menorrhagia and dysmenorrhoea may lead to hospital based conventions of practice.[16] It is likely that institutional effects are operating. The hypothesis of a consultant "signature"[17] is not sufficient to account for the total variance found in these data; there are to some degree at least "institutional signatures" which are worthy of further examination.

The currently published routine data are generally unhelpful in analysing profiles of outpatient workloads.[18] Measures specific to the complex dynamics of outpatient workloads are needed if a clear understanding of the reasons for variation in outpatient practice in different specialties is to develop. New data are not required to enable the crucial distinction to be drawn between the three key elements of outpatient workload: inpatient follow up, new referrals and clinic initiated re-attendances. These data are already available within the contract minimum data set for outpatient services. It would be straightforward to collate, present, and publish them in this form.

The existing research record has not systematically identified the key factors accounting for the high degrees of variation in the proportions of new referrals and recycled patients in outpatient workloads, though a number of contributory factors have been identified.[19 20] A complex array of factors is clearly involved. The systematic identification of these is the topic of a follow up study. The analysis presented here does not permit optimal follow up patterns to be inferred, but it does give new support to the view that substantial numbers of patients are attending unnecessarily or not frequently enough.

The authors thank Ms Jane Brooks of the South and West Regional Health Authority for assistance with extraction of the data analysed in this paper, and anonymous referees for comments on previous drafts of this paper.

Funding: the authors gratefully acknowledge the South and West Regional Research and Development Directorate for support for the research programme of which this study forms a part.
Conflicts of interest: none.

1 Faulkner A, Frankel S. *Evaluation of delivery and use of outpatient services: a review of recent published evidence and developments with possible measures for action.* Bristol: Health Care Evaluation Unit, University of Bristol, 1993.
2 Armstrong D, Nicoll M. Consultants' workload in outpatient clinics. *BMJ* 1995;**310**:581–82.
3 Frankel S, Faulkner A. The end of the outpatient problem? *BMJ* 1994;**309**:1308.
4 Volans AP. Not all specialist groups hold outpatient clinics. *BMJ* 1995;**311**:59.
5 Irvine H, Scott P. Consultants spend more time with each patient than in the past. *BMJ* 1995;**311**:59.
6 Anonymous. To come again, 3 months. *Lancet* 1976;i: 1168–9.
7 Roberts T, Heagerty AM, Foster P. Existing outpatient services could be more efficient. *BMJ* 1995;**311**:60.
8 Griffiths RW. Plastic surgery outpatient audit: principles and practice of "consultant only" clinics. *Br J Plast Surg* 1990;**43**:735–41.
9 ONCOFEM/CNR-ACRO. *Consensus conference on "Follow-up in breast cancer".* ONCOFEM and CNR-ACRO, Bari, October 8th, 1994.
10 McPherson K. Variations in hospitalisation rates: why and how to study them. In: Ham C, ed. *Health care variations: assessing the evidence.* London: King's Fund Institute, 1988: 15–20.
11 Wennberg J. Practice variations and the need for outcomes research. In: Ham C, ed. *Health care variations: assessing the evidence.* London: King's Fund Institute, 1988:32–5.
12 Armitage P, Berry G. *Statistical methods in medical research.* Oxford: Blackwell Scientific Publications, 1987.
13 Altman DG. *Practical statistics for medical research.* London: Chapman & Hall, 1991.
14 Blalock HM. *Social statistics.* New York: McGraw-Hill, 1972.
15 Price CE, Paul EA, Bevan RG, Holland WW. Equity and medical practice variation: relationships between standardised discharge ratios in total and for selected conditions in English districts. *J Epidemiol Community Health* 1992; 46:58–62.
16 Coulter A, Klassen A, Mackenzie IZ, McPherson K. Diagnostic dilatation and curettage: Is it used appropriately? *BMJ* 1993;**306**:236–9.
17 Wennberg J, Gittelsohn A. Variations in medical care among small areas. *Scientific American* 1982;**246**:100–12.
18 Government Statistical Service. *Out-patients and ward attenders.* Crown Copyright: Department of Health, 1991–95.
19 Sullivan FM, Hoare TM. New referrals to rheumatology clinics—Why do they keep coming back? *Br J Rheumatol* 1990;**29**(1):53–7.
20 Sullivan FM, Hoare T, Gilmour H. Outpatient clinic referrals and their outcome. *Br J Gen Pract* 1992;**42**:111–5.

PAPER 6

PAPER 7

Originally published 1995

Author: Faulkner A, Saltrese-Taylor A, Williams M, O'Brien J, Collins CD, & Frankel S.

Outpatients revisited: subjective views and clinical decisions in the management of general surgical outpatients in South West England.

Published in: Journal of Epidemiology & Community Health, 49:599-605.

Acknowledgment: BMJ Publishing Group for permission to reproduce this article.

527

Outpatients revisited: subjective views and clinical decisions in the management of general surgical outpatients in south west England

Alex Faulkner, Allison Saltrese-Taylor, Jim O'Brien, Mark Williams, Charles D Collins, Stephen Frankel

Abstract

Study objective **– To assess the scope for reducing unnecessary outpatient re-attendances, using as a benchmark an acute specialty at a site recognised to have an especially low ratio of repeat to new attendances.**

Design **– This was a survey of the re-attendance workload at general surgery outpatient clinics over a three month period. Patient re-booking and discharge rates for different grades of staff; clinicians' perception of the ability of the GP to have managed the patient; perception of the value of individual re-attendances; reasons given for discharging/re-booking; and outcome of attendance for patients in relation to diagnostic category were determined.**

Setting **– General surgery outpatient clinics with re-attendance rates that were 50% below average, in Taunton and Somerset Hospital, a non-teaching district general hospital.**

Patients **– Altogether 454 patients who made 470 second or subsequent visits (re-attendances) within the same episode of outpatient care.**

Main results **– Thirty eight percent (178/470) of visits were perceived as manageable by the GP, 45% (79, 17% of total re-attendances) of which were also thought to have been of marginal or little value. A substantial group of patients was being followed up largely for reasons of convention and traditional policy. Re-booking rates were higher among junior staff. Subjective views of the value of attendance at the hospital outpatient clinic and the ability of the GP to have seen the patient varied systematically between consultants and junior staff. Judgements varied to some extent according to the diagnostic group.**

Conclusion **– The numbers of patients being followed up equivocally at most general surgical outpatient departments will be 50% more on average than those in this benchmark department. A department seeing 2000 new patients per annum will have 3600 reattendances, 25·5% (918) of which may be avoidable on the basis of these results. A variety of approaches can be used to increase the proportion of patients seen appropriately by GPs. In some cases this might be achieved without the intensive commitment required to plan and develop shared care protocols or new formal discharge guidelines, but by encouraging GPs to manage some patients, increasing of hospital clinicians' access to knowledge of local general practices, and internal clinic review of 'routine' follow up policies as shown in this study. This type of review of outpatient practice can also help prioritise conditions likely to repay the effort of developing and implementing clinical management guidelines and local protocols.**

(*J Epidemiol Community Health* 1995;49:599–605)

Health Care Evaluation Unit, Department of Social Medicine, University of Bristol, Canynge Hall, Whiteladies Road, Bristol, BS8 2PR
A Faulkner
A Saltrese-Taylor
M Williams
S Frankel

Somerset Health Authority, Wellsprings Road, Taunton, Somerset TA2 7PQ
J O'Brien

Tuanton and Somerset Hospital, Musgrove Park, Taunton, Somerset, TA1 5DA
C D Collins

Correspondence to:
Mr A Faulkner.

Accepted for publication
May 1995

It is well known that repeat visits to outpatient clinics account for between 75% and 80% of all attendances recorded for this form of healthcare.[1] Perennial doubts have been expressed within the medical profession over the value of repeated reattendances both to patient and clinician. The policy issues have been highlighted by the recent National Audit Office study,[2] and subsequent review by the Parliamentary Committee of Public Accounts.[3] Innovation in National Health Service organisation, notably general practitioner (GP) fundholding, is leading to an increased scrutiny of the legitimacy of outpatient re-bookings. In spite of the many assertions, however, there has been a relative lack of research or audit based evidence about causes, reasons, and possible justifications for the high levels of patients invited to return for follow up. This type of evidence is a prerequisite of attempts to design and progress toward viable alternatives to traditional outpatient reattendance or substitute care arrangements involving primary care.

No published study has examined these issues in the high volume specialty of general surgery, with its particular case mix and typical reasons for re-attendance. Diagnostic categories within specialties have generally not been taken into account in previous analysis of reasons for reattendance, although diagnosis has been identified as one predictor of the probability of discharge after the first attendance.[4] A recent randomised trial of immediate discharge of selected surgical patients from a teaching hospital to general practice has concluded that primary care was of at least equal clinical effectiveness, and was less costly than traditional outpatient follow up.[5] This enhances the need to base discussions of al-

PAPER 7

ternatives to traditional follow up on an improved understanding of the dynamics of the outpatient clinic.

A number of possible generic explanations have been put forward for high reattendance and rebooking rates in outpatient clinics on the basis of studies in other specialties. In particular, the influential role played by junior hospital doctors in needlessly inflating outpatient clinic reattendance rates is frequently asserted. There is evidence that they do recall patients in higher proportions than do their consultants,[4 6-10] although some studies show no difference.[11-13] This suggests that there is no simple explanation of observed reattendance rates and that other factors impinging upon decisions to offer repeat visits should receive more attention. The role of clinician status in relation to the management of patients in different diagnostic categories, and clinicians' opinions on alternative management require more in depth exploration.

This study describes the casemix of the outpatient workload in general surgical clinics, explores the relationship between case specific opinions about routine follow up as expressed by members of the surgical team and their actual management decisions, and analyses the stated reasons why clinicians re-book patients whom they believe could have been managed by their GP.

The study took place in general surgery outpatient clinics at a non-teaching district general hospital which had already introduced a policy of not offering routine follow up appointments to patients after uncomplicated surgery.[14] This outpatient facility was staffed by consultants, registrars, and senior house officers, organised into four consultant "firms". The ratio of new to reattending patients at the time of the study was low (1:1·2). This compares with a ratio of approximately 1:1·8 for general surgery across the South Western Health Region as a whole at the time of the study (KH09 returns). The data analysed here thus refer to a department in a high volume acute specialty, with an exceptionally low proportion of reattendances compared with the norm, in which only marginal scope for further planned reduction in reattendances might be expected. If however, there were shown to be substantial scope for further reduction, then the implications for the nature and volume of outpatient practice in general surgery would be very large indeed.

Methods

All follow up attendances at the general surgical outpatient clinics between July and September 1990 were surveyed. Examining doctors filled in a self completion questionnaire for each patient after the consultation, and further patient details were obtained from the case notes by one of the authors (A S-T). The questionnaire was developed in collaboration with the consultants in general surgery in the unit. It was piloted on 30 patients and appropriate changes were made to ensure face validity. All relevant clinical details for the survey sample were cross checked against case notes. The records of non-attenders and patients for whom a questionnaire was not completed were checked to validate the completed questionnaire data; they were found to be representative of the broad diagnostic categories in the sample.

The questionnaire asked clinicians to indicate the main objective of each appointment, the actions taken, the type of continuation of care, and to rate, firstly, their perception of whether or not the patient's GP could have managed care and, secondly, their opinion of the overall necessity of the visit on a four point scale: "essential", "desirable but not essential", "of marginal benefit", or "not particularly useful". An open ended question then asked them to give in their own words a brief reason for the latter two judgements. During the study period, 559 follow up visits were planned. Forty seven patients (8·4%) failed to attend; questionnaires were not completed for 42 (7·5%). Data were thus obtained for 470 second or subsequent visits within the same episode of care, by 454 separate patients. These 470 reattendances form the basis for the analysis presented here.

Analysis of numeric data was by frequency counts, cross tabulation, and descriptive summary statistics (using the *Epi-Info* software package). The association of qualitative variables (grade of staff and clinician opinion; clinician opinion and perceived necessity of outpatient visit) was analysed by means of the χ^2 test with Yate's correction for two by two tables. This analysis was not extended to case mix, where aggregate of diagnostic groups would have been inappropriate. The reported textual data produced by the open ended questions were transcribed, classified by content, and hand counted by the first author (AF) and the classification decisions of the verbatim statements were checked by a second author (MW) to ensure consistency and reasonableness of interpretation.

Results

Three main clinical objectives accounted for 79·2% of outpatient reattendances in this sample:

- Postoperative follow up;
- Returning for results/continuing investigations; and
- The monitoring or treatment of chronic or malignant disease. Details of all objectives are given in table 1.

The mean number of reattendances per patient was 3·54, with a median of 2. Sixteen patients had each made more than 14 reattendances (with a maximum of 41). Table 2 gives the median (and range) reattendances for each clinical objective. It also shows for each objective the three main outcomes of appointments, namely placement on a waiting list, discharge to the GP, and a further invitation to reattend. This illustrates, notably, that patients with the stated objective of returning for results or continuing investigations are placed on a waiting list in the highest pro-

Table 1 Main clinical objectivres for reattending patients in general surgical outpatient clinic

Main clinical objective	*No*	*(%)*
Post-operative follow-up	114	(24·5)
Continuing investigations/returning for results	103	(21·9)
Monitoring or treatment of chronic/malignant condition	154	(32·8)
Treatment	41	(8·7)
Surgeon's interest	29	(6·2)
Patient's request	11	(2·3)
Miscellaneous	18	(3·8)
Total	470	100·0

portions, and that patients attending for monitoring or treatment of chronic/malignant conditions are discharged at a relatively low rate and have the highest proportion of rebooking.

One would expect the number of re-attendances per patient to be related to the diagnosis, and this is the case. Diagnoses have been grouped according to ICD-9 codes. Table 3 illustrates the variation. Patients with the highest mean re-attendances at the clinic are those suffering from malignant neoplasms, colitis or enteritis, and vascular disease. Those with hernia, other bowel disease (motility disorders, ulceration or inflammatory changes related to oesophagus, stomach or duodenum, irritable colon or appendicitis), perianal conditions (haemorrhoids, pilonidal cyst, pruritis) and venous conditions are recalled, in general, on a more short term basis. Table 3 also shows the number of cases, analysed by diagnosis, for the extremes of reattendances and management, namely patients discharged on their first or second reattendance, and patients with between three and nine reattendances who have been asked to come again. It is clear that some conditions tend to be associated with long term attendance and that others have a much faster "turnaround". The table confirms that in the former category are, most clearly, malignant neoplasms and vascular disease, and to a lesser extent colitis/enteritis; and in the latter are the specified types of bowel disease, hernia, benign neoplasms, and perianal conditions.

CLINICAL OPINIONS ON MANAGEMENT BY GP

In the survey period, reattending patients were seen in the following proportions by each grade of staff: consultants – 22·3% (105), registrars – 52·6% (247), and senior house officers – 25·1% (118). Additionally, a consultant opinion was sought during one in four appointments with registrars and one in three with senior house officers.

Overall, clinicians judged that 38% (178) of their appointments could have been seen by the patient's GP. There was no notable difference between the mean number of reattendances by patients who could have been seen by their GPs and those who could not (3·59; 3·61 visits), suggesting that the frequency of a patient's attendance is not in itself a criterion of appropriateness of reattendance for clinicians. Of the patients stated to be returning for results of tests (not for continuing investigation), it was felt that the results could have gone direct to the GP, thus avoiding the appointment, in 33% (29/88) of the cases. Of the patients returning for treatment, none was deemed substitutable for GP care. Of the post-operative patients, it was considered that 47% (54/114) could have been seen by their GP instead, as could 40% (62/154) of those with malignant or chronic conditions.

These figures differed critically according to clinician status. Figure 1 indicates the mag-

Table 2 Reattendances and appointment outcomes for each clinical objective

Main clinical objective	*Reattendances*		*Outcome of appointment*			
	Median (range)	*No*	*Waiting list (n = 63) %*	*Discharged (n = 167) %*	*re-booked (n = 221) %*	*Other (n = 19)* %*
Postoperative follow up	2 (1–30)	114	11·4	43·0	42·1	3·5
Continuing investigations/returning for results	1 (1–19)	103	27·2	42·7	24·3	5·8
Monitoring or treatment of chronic/malignant condition	3 (1–41)	154	7·8	23·4	64·9	3·9
Treatment	1 (1–11)	41	4·9	36·6	56·1	2·4
Surgeon's interest	2 (1–20)	29	6·9	37·9	51·7	3·5
Patient's request	2 (1–25)	11	18·2	18·2	63·6	0·0
Miscellaneous	1 (1–31)	18	22·2	55·6	16·7	5·5
	2 (1–41)	470	13·4	35·5	47·0	4·0

* Including 3 missing

Table 3 ICD9 outcomes of reattendance and total numbers of consultations within episodes of care in relation to diagnostic group

Diagnostic group ICD9 codes	*Mean no of reattendances*	*Median reattendances*	*Discharged 1st or 2nd reattendance*	*Discharged 3rd–9th reattendance*	*Rebooked 1st or 2nd reattendance*	*Rebooked 3rd-9th reattendance*	*10+ reattendances (Discharge Rebook)*	*Other outcomes**	*Total no of patients*
Colitis/enteritis (555–558)	7·24	4	2	0	5	6	6	6	25
Malignant neoplasm (150–151, 153–154, 157, 172–174, 194, 199, 230–234)	5·32	3	8	3	31	37	5 11	13	108
Vascular (441, 442, 443–448)	4·12	2	9	4	16	13	2 4	12	60
Skin/breast (non malignant) (707–709, 611–7)	2·71	1	4	1	6	1	0	5	17
Benign neoplasm (210–238)	2·53	2	15	4	7	4	1	3	34
Perianal (455, 685, 698)	2·23	2	17	4	15	4	1	2	43
Hernia (550–553)	2·08	1	5	1	1	1	0	4	12
Venous (451–454, 456–459)	1·63	1	9	0	8	2	0		19
Bowel disease (530–537, 560–569, 570–579, 540–543)	2·20	1	32	4	5	4	0	20	65
Miscellaneous (240–242, 285, 392, 474, 527, 590–608, 619, 741–999)	1·98	1	14	6	13	4	1	8	46
No diagnosis	3·54	2	16	2	7	5	2	9	41
							7 26		
Overall mean	3·54	2	131 (28%)	29 (6%)	114 (24%)	81 (17%)	33 (7%)	82 (17%)	470 (100%)

* Outcome values were missing for one each case of "vascular", "hernia", and "no diagnosis".

PAPER 7

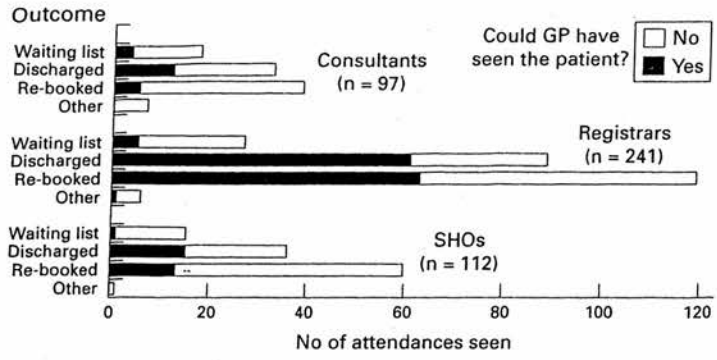

Figure 1 Outcome in relation to seniority of doctor and whether or not the patient could have been seen by a GP

	Outcome				*Total (no missing)*
	Waiting list	*Discharged*	*Re-booked*	*Other*	
Consultants:					
GP could have seen	3	12	5	0	
GP could not have seen	15	21	34	7	
					97 +8
Registrars:					
GP could have seen	5	61	62	1	
GP could not have seen	22	28	57	5	
					241 +6
Senior house officer (SHO):					
GP could have seen	1	15	13	0	
GP could not have seen	14	21	47	1	
					112 +6

χ^2: waiting list = 1·12, p = NS (0·5717); discharged = 13·82, p<0·001; re-booked = 27·56, p<0·001

nitude of the variation. Regardless of the outcome of the attendance, registrars felt the GP could have seen the patient in 54% (129/241) of reattendances overall, whereas for consultants the proportion was 21% (20/97) and that for senior house offers was 26% (29/112). The difference between the different grades of staff was statistically significant for both discharged and re-booked patients. For the two objectives of postoperative follow up and of monitoring chronic/malignant conditions, the percentages which registrars assessed as being appropriately seen by GPs were 65% (45/69) and 60% (47/78) respectively. Only in cases of colitis/enteritis and perianal conditions were the views of registrars similar to those of consultants and senior house officers in assessing that it would rarely have been appropriate for the GP to have seen them. For all other diagnostic groups, registrars viewed over half the reattendances as possibly being manageable by the GP. For patients returning for results of investigations the three grades of staff were consistent in estimating 20%–30% to have been manageable by the GP. Of those deemed to have been manageable by the general practitioner, 71% (15/21) were in fact discharged.

OUTCOMES OF ATTENDANCE: RATES OF DISCHARGE AND RE-BOOKING

Consultants saw only a marginally higher proportion of newer patients in their workload, than other grades of staff. Patients reattending for the first or second occasion constituted 68% of consultants' (71/105), 65% of registrars' (161/247), and 61% of senior house officers (72/118) workloads.

There is, however, some suggestion that they saw a higher than average proportion of certain conditions, especially in the case of enteritis/colitis (44%, 11/25, seen by consultants, although these cases amounted to less than 10% of consultants' workload in this study) and benign and malignant neoplasms (29% (10/34) and 24% (26/108), respectively, seen by consultants). Many of these patients were long term reattenders. For patients reattending for their fifth time or more (n = 96), consultants' proportion increased to 30% and registrars reduced to 44%. Within this group, proportions are the same for attendances with the explicitly stated purpose of monitoring or treatment of chronic or malignant conditions.

Of the 178 patients whom "the GP could have seen", 5% (9) were nevertheless put onto the waiting list, 49% (88) were discharged, and 45% (80) were asked to reattend again. For those who were discharged, the average number of previous attendances was 2·4; for those asked to reattend it was 4·9 visits. In other words, patients with a longer history of visits were more likely to be asked to come again. For those who clinicians felt could have been seen by the GP but who were nevertheless re-booked (80), the mean interval to the new repeat visit was 32·4 weeks (median 26); for those re-booked and not deemed manageable by the GP the interval was only 20·6 weeks (median 12).

This study does not reproduce unequivocally the common finding that junior staff discharge patients at lower rates than consultants, although the trend is in this direction. Figure 1 shows the main outcomes of appointments for each clinician status separately, with the proportions for each outcome which were assessed as manageable by a GP. The ratio of those discharged compared with those asked to re-attend varies somewhat between each clinician status (consultants 1:1·1, registrars 1:1·3, senior house officers 1:1·6). On the other hand, of the patients discharged by registrars, those they assessed as capable of having been seen by the GP outnumber those whom the GP could not have seen by more than two to one (68%, 61/89). Even among the appointments with registrars who were asked to re-attend, over half (52%, 62/119) were nevertheless felt to have been manageable by the GP. This phenomenon was much less marked in the case of consultants and senior house officers. For consultants, of those discharged 36% (12/33) and of those reattending again only 13% (5/39) were felt to have been appropriate for the GP. For senior house officers the corresponding percentages were 42% (15/36) and 22% (13/60).

OPINIONS OF VALUE OF ATTENDANCE

In order to explain why a proportion of patients whom clinicians felt could be seen by their GP were nevertheless asked to reattend, the subjective views of the clinicians in the context of their working environment have been analysed. Figure 2 indicates the outcomes of consultation for patients assessed by the clinicians either as capable or not capable of having been

PAPER 7

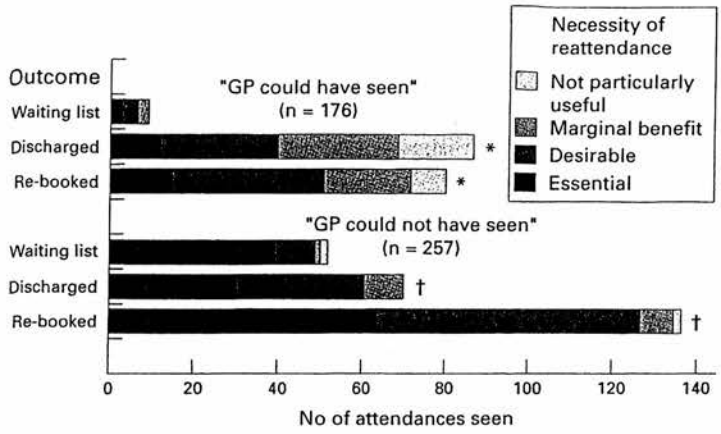

Figure 2 Outcome in relation to necessity of reattendance and whether or not the GP could have seen the patient

Outcome	*Essential*	*Desirable*	*Marginal*	*Not particularly useful*	*Total (no missing)*
GP could have seen:					
Waiting list	3	4	2	0	
Discharged*	12	28	29	18	
Re-booked*	14	37	21	8	
Other	0	0	0	0	
					176 (2)
GP could not have seen:					
Waiting list	40	9	1	1	
Discharged†	31	30	9	0	
Re-booked†	63	64	8	1	
Other	11	1	1	0	
					257 (22)
Total					446 (24)

* $\chi^2 = 4{\cdot}21$, $p < 0{\cdot}05$ (essential/desirable *v* marginal/not particularly useful)
† $\chi^2 = 1{\cdot}54$, p = NS (0·214); (essential/desirable *v* marginal/not particularly useful)

seen by their GP, and shows, for each of the three key outcomes, their opinion of the value of the attendance.

Of those reattendances which the hospital clinicians felt could have been undertaken by the GP, the current hospital outpatient appointment was considered essential or desirable for 46% of the discharged patients' appointments (40/87) and 63·8% of those invited to reattend (51/80). In other words, the correlation between the perceived ability of the GP to have seen the patient and the clinician's rating of the necessity of the appointment was low. This varied systematically according to clinician status. Consultants felt a hospital outpatient appointment to be essential/desirable regardless of whether they felt that the general practitioner could have seen the patient or not: in 93% (90/97) of cases. On the other hand, of the 128 appointments which registrars felt appropriate for GP, they rated 44% (56) as of marginal or little value, and for senior house officers the corresponding percentage was 69% (20/29).

There was some variation in these judgements according to diagnostic group, as shown in table 4. Nearly one third (10/34) of the visits of cases of malignant neoplasm deemed manageable by the GP, but nevertheless asked to reattend again, were rated as of marginal value; conversely, none of this diagnostic group (0/44), asked to reattend and *not* manageable by the GP, was regarded as of marginal value. Of those reattendances *discharged* and rated as not manageable by the GP, only nine overall were regarded as of marginal or little value; of those discharged and rated as GP-manageable, bowel disease stands out with 15/20 attendances regarded as of marginal or little value (these 15 comprised diverticulitis (3), irritable colon (3), oesophagitis (3), cholelithiasis calculus of gall bladder without cholecystitis (1), acute pancreatitis (1), gall bladder – other cholecystitis (1), peptic ulcer unspecified (1), dyspepsia/other stomach disorder (1), and anal fissure (1)). None of the attendances of the six discharged patients deemed GP manageable with malignant neoplasm was regarded as necessary. On the other hand, most of the reattendances in this group for perianal conditions were regarded as desirable (8/11); the balance of perceived necessity for outpatient attendance among the other diagnostic groups was more even.

REASONS FOR RE-BOOKING PATIENTS WHOM THE GP COULD HAVE SEEN

The most critical group in this analysis is those who were asked to re-attend even though they were assessed as suitable for management by the GP. Of this 38% (178), 55% (98) were nevertheless regarded at least as desirable clinic attenders, and 51 of these were re-booked. Overall, 80 (45%) patients whom the clinicians felt the GP could have seen were re-booked

Table 4 Perceived necessity of reattendance for different diagnosis categories in relation to outcome of reattendance †and perceived ability of GP to have seen patient*

Diagnostic group	*Outcome = discharge*				*Outcome = rebooked*				*Total*
	GP could have seen		*GP could not have seen*		*GP could have seen*		*GP could not have seen*		
	Necessary	*Marginal*	*Necessary*	*Marginal*	*Necessary*	*Marginal*	*Necessary*	*Marginal*	
Colitis/enteritis	0	1	0	1	1	0	13	3	19
Malignant neoplasm	0	6	9	0	24	10	44	0	93
Vascular	4	1	8	0	11	5	16	1	46
Skin/breast (non-malignant)	0	1	4	0	2	2	2	1	12
Benign neoplasm	4	5	8	2	3	1	6	0	30
Perianal	8	3	8	1	2	1	14	1	38
Hernia	2	2	2	0	0	1	1	0	8
Venous	4	2	3	0	2	1	7	0	19
Bowel disease	5	15	11	4	0	1	11	4	43
Miscellaneous	7	6	4	0	5	1	11	1	35
No diagnosis	6	5	4	1	1	4	7	2	30
Total	40	47	61	9	51	29	127	9	373

* Perceived necessity has been summarised as follows: "essential" or "desirable but not essential" = "necessary", "of marginal value" or not particularly useful" = "Marginal".
† Discharge and rebooking outcomes only are included in this table; of these, values were missing in 3 cases for "perceived necessity" and in 12 cases for "Could GP see?".

PAPER 7

(figs 1 and 2). Analysis of the brief reasons given in this survey by clinicians for re-booking patients who could have been seen by the GP, reveals a mixture of practical considerations, routine, tradition, and policy underlying the observed rates. Actual reasons included: "lack of a request (either from the patient or the GP) for the GP to see the patient", "tradition to follow up breast cancer forever", "education for ourselves and to keep the morale of the patient up", "I didn't ask GP to", "we like in this instance to see the result of our work", "we like to follow up abdominal aortic aneurysms ourselves", "GP did see, but we wanted to also", "we like to follow up vascular cases for a while",[11] GP lack of sigmoidoscope", "follow up of graft patency", and other more vague reasons, such as "routine", "requested for close follow up", and "agreed to attend".

It is not possible to calculate absolutely clear proportions for the different categories of reason given, because interpretation of some of the verbatim written responses was open to doubt (for example "not requested" leaves it unclear whether the GP, patient, or clinician is being referred to). However, two types of reason stood out as the most common. They can be summarised as firstly: a lack of a request either from the GP or the clinician for the GP to manage the patient: at least 22/80 responses (it is likely that some of those cases where no reason was given (n=16) also fall into this category); secondly, routine or tradition (at least 15 cases, of whom all except one had diagnoses of malignancy, peripheral vascular disease, or aneurysm). It is worth noting also that although cumbersome discharge procedures have often been suggested as the reason for many unnecessarily re-booked patients, this reason was not offered in any case by the clinicians in this group.

Discussion

This study has implications for several important aspects of the phenomenon of reattendance at outpatient clinics. Firstly, a high performing, hospital outpatient service in general surgery was re-booking a significant proportion of patients for whom the appropriate mode of management was open to doubt. The low ratio of re-bookings to new attendances in this unit suggests that the proportion of these patients being recalled in most departments of general surgery will be some 50% larger than that reported in this study. The analysis was able to show particular diagnostic categories which would be candidates for review. A proportion of cases of bowel disease in particular, amongst the diagnostic categories, might have been reviewed as possible candidates for alternative management.

The difference in reattendance ratios between this and other departments, and the proportion of patients whose appropriate management was open to doubt and were potentially dischargeable, can be used to estimate the numbers of patients potentially dischargeable in an average department. Assuming a ratio of reattendances to new referral attendances of 1·8:1 in a general surgery d partment seeing 2000 new referrals per annu there would be 3600 reattendances. Mu tiplying the proportion found in this study be both manageable by the GP and rated as marginal value or less (17%; 79/470) by 1 (proportion of reattendances 50% higher average departments) gives an estimate of 91 patient reattendances per annum (1·5 × 17% 25·5%; 25·5% × 3600) the necessity of who visit is likely to be equivocal amongst the clin cians themselves.

Secondly, the outpatient management of small but not negligible proportion of patien seemed to be perpetuated simply for lack c either part at the primary/secondary interfac making positive moves to take active re sponsibility for changing the care arrange ments.

Thirdly, some patients who had had surger for malignant or life threatening condition were being followed up essentially as a matter o routine or tradition, as the consultants involve were well aware. These patients typically ha long intervals between appointments – si months or more. Questions of alternativ methods of follow up of patients with some o these conditions are now actively being con sidered in the United Kingdom.

Fourthly, the study shows the complex relationships between clinical case based opinions, clinical management practices, diagnostic categories, and the primary/secondary care interface in influencing the outcome of outpatient reattendance. Some 45% of the 178 patients rated by clinicians as manageable by the GP were nevertheless rebooked. This phenomenon has been reported previously be Leitch *et al* in respiratory medicine[15] and Armstrong *et al* in gynaecology and general medicine.[10] The present study has been able to show in addition that many of such attendances in general surgery are, in fact, regarded as desirable by clinicians, and it has been able to describe and analyse by diagnostic category the stated reasons given for retaining patients who in principle seem to be dischargeable. Account must, therefore, be taken of clinical opinions and reasons for retaining patients in addressing the question of substituting GP care for the clinic. Subjective assessments of the ability of GPs to manage patients may not be entirely consistent, but this study does indicate some broad regularities. Data such as these provide the basis for internal audit of clinical practices and policies as well as for informed discussion between clinicians, managers, and GPs. Involvement of GPs in research, development, or audit activities in this complex area will be of benefit in strengthening the basis upon which practical changes to patterns of care might be made. The extent of GPs' willingness to participate in innovative developments can be expected to depend in part upon their involvement in this type of review process, as well as upon workloads and experience in particular conditions.

Fifthly, the study provides evidence that registrars, in particular, place high value on a relatively small proportion of their attendances

PAPER 7

in comparison with consultants and senior house officers. Consultants exhibited extremely conservative, case-specific opinions. The reasons for this are unknown. It may be that consultants have experienced instances of mismanagement by GPs, or they may be out of touch with general practice. Alternatively this may reflect the commitment generated by professional specialisation, clinical authority and responsibility, and relatively long term local ties to the hospital. In some cases, registrars may have more confidence in general practice because of greater contact with GPs through mutual education programmes, although in many instances registrars are likely to see more "routine" cases because of case selection within the clinic. Lack of contact with GPs, and greater contact with consultants in clinic may also explain why senior house officers' perceptions were closer to those of the consultants than to the registrars. Given the uncertainties in these interpretations, there is a strong case for evaluating experimentally the effect upon the patient re-booking and discharge behaviour of junior hospital doctors of experience in general practice.

A number of practical approaches to improving the "appropriateness" of reattendance at outpatient clinics have been proposed. These include making discharge policy more explicit and organising training for juniors,[10] improving the content of discharge letters to GPs especially for chronic conditions,[15] increasing consultant review of casenotes with directive plans clipped to them,[8] and the introduction of written guidelines for reattendance and discharge.[16] All these types of measures have been reported as successful, and this study provides evidence to support extension of their application in general surgery.

Further practical measures and productive avenues for research and development are suggested on the basis of the analysis presented here. The study site provides a reference point for discussion between GPs and hospital clinicians of the potential scope for discharging patients to general practitioners, which could be applied to other general surgical departments. Specific measures to consider include: the feasibility of increasing practical contact between clinic and GPs through means such as educational activities; stimulating GPs' ability and willingness to request discharge of some patients; making better information about local general practices' facilities and training of GPs readily available to clinicians; and development of more explicit, agreed procedures for the management of patients with particular conditions. Given the typically long interval until the re-booked appointment, it might be possible to book selected repeat appointments on a "pending GP intervention" basis, and notify the GP with an invitation to consider assuming responsibility for the patient.

The clear disparity between the perceptions of consultants and registrars about the value of repeat appointments and GPs' ability to see patients suggests that there is potential for general surgical departments to consider locally the issues of both formal and informal clinical policies on reattendance and discharge, and for shared interval review of clinical criteria for "routine" repeat appointments between consultants and junior staff. More generally, it suggests that there is a strong case for comparative research to examine the specific factors which may account for the discharge and re-booking behaviour of junior staff in clinics with different (high versus low) reattendance and discharge performance.

In summary, clinicians in a general surgery outpatient department perceived that there was a large proportion of re-attending patients who could appropriately have been seen by the GP, but in many cases they were equivocal in judging the appropriate site of care and acting upon these judgments. There will doubtless always be come discrepancy between the aggregate of clinicians' opinions and their clinical actions. The method of review demonstrated in this study can help prioritise conditions which may be candidates for development of alternative care arrangements across the primary/secondary care interface. Indicators of possible levels of "inappropriateness" should be taken as the opportunity to review practices in a manner sensitive to the complexities of hospital outpatient practice as shown in this study.

We are grateful for the help and co-operation of the doctors who participated in this study, Mr S Jones, Mr I Ramus, and Mr I Eyre-Brook, and that of the nursing, administrative and management staff in general surgery at Taunton & Somerset Hospital. We thank an anonymous referee for comments on a previous version of the paper.

1 Department of Health. *NHS hospital activity statistics for England 1979–1990/91.* London: Government Statistical Service, 1992. Statistical Bulletin 2(1)92.
2 National Audit Office. *NHS outpatient services.* London: HMSO, 1991.
3 Committee of Public Accounts. *NHS outpatient services.* London: HMSO, 1991.
4 Sullivan FM, Hoare T, Gilmour H. Outpatient-clinic referrals and their outcome. *Br J Gen Pract* 1992;42:111–115.
5 Du V Florey C, Yule B, Fogg A, Napier A, Orbell S, Cuschieri A. A randomised trial of immediate discharge of surgical patients to general practice. *J Public Health Med* 1994;16(4):455–64.
6 Lester JP. Why not reclaim our outpatients from hospital outpatient clinics? *J R Coll Gen Pract* 1980;30:(213):230.
7 Grace JF, Armstrong D. Reasons for referral to hospital: extent of agreement between the perceptions of patients, general practitioners and consultants. *Fam Pract* 1987; 4(3):143–147.
8 Sparks D. A firm decision to save time. *Health Services Journal* 1988;98(5090):278.
9 Jolly U. Think of the client. *Health Services Journal* 1989; 99(5178):1434–5.
10 Armstrong D, Brown K, Tatford P, Armstrong P. Inappropriate re-attendances in out-patient departments. *J Public Health Med* 1992;14(2):173–6.
11 Marsh GN. Are follow-up consultations at medical outpatient clinics futile? *BMJ* 1982;284:1176–7.
12 Bosanquet N, Fordham R. Outpatient services – a case for treatment. *Health Services Journal* 1987;97(5070):550–1.
13 Cartwright A, Windsor J. *Outpatients and their doctors: a study of patients, potential patients, general practitioners and hospital doctors.* London: HMSO, 1992.
14 McCormack TT, Collier JA, Abel PD, Collins CD, Ritchie WN. Attitudes to follow-up after uncomplicated surgery – hospital out-patients or general practitioner. *Health Trends* 1984;16:46–7.
15 Leitch AG, Parker S, Currie A, King T, McHardy GJ. Do chest physicians follow-up too many patients? *Respir Med* 1989;83:329–32.
16 Hall R, Roberts CJ, Coles GA *et al.* The impact of guidelines in clinical outpatient practice. *J R Coll Physicians Lond* 1988;22(4):244–7.

PAPER 7

PAPER 8

Originally published 1999

Author: Somerset M, Faulkner A, Shaw A, Dunn E, Sharp D.

Obstacles on the path to a primary care-led NHS: complexities of outpatient care.

Published in: Social Science & Medicine, 48: 213-225.

Acknowledgment: Elsevier Ltd. for permission to reproduce this article.

PERGAMON

Social Science & Medicine 48 (1999) 213–225

SOCIAL SCIENCE & MEDICINE

Obstacles on the path to a primary-care led National Health Service: complexities of outpatient care

Maggie Somerset[a,*], Alex Faulkner[b], Alison Shaw[a], Liz Dunn[a], Deborah J. Sharp[a]

[a]*Division of Primary Care, University of Bristol, Canynge Hall, Whiteladies Road, Bristol BS8 2PR, UK*
[b]*Department of Social Medicine, University of Bristol, Canynge Hall, Whiteladies Road, Bristol BS8 2PR, UK*

Abstract

An interpretive qualitative study was carried out as part of a large cohort study of factors affecting outpatient re-attendance. Individuals from three groups involved in the provision of care across the primary–secondary interface were interviewed: patients, general practitioners and consultants. The aim was to explore understandings concerning referral to and re-attendance at outpatients, and to elicit detailed descriptions of the complexities of the outpatient experience for both providers and recipients of care at the primary/secondary interface, given the policy commitment to a 'primary-care led National Health Service'. Semi-structured interviews were carried out with nine individuals currently attending outpatients, ten general practitioners, and ten consultants. Transcripts were analysed individually and cross-checked between analysts for validity of interpretation, to identify key themes and sub-themes. Data were compared across the three groups. Negative case analysis was employed. Seven major issues were identified, some of which could be identified with interests and experience of the three obvious groupings, and some of which were common. The three groupings are not as homogeneous as is often supposed. From the cross-group analysis common themes included: interpersonal communication, knowledge, power relations and anxiety/reassurance. Issues of trust, social status, funding and consumerism/litigation were also highlighted. The analysis has implications for altering the balance of care across the interface, for example in the finding of what could be termed a dissonance in power perceptions, in that consultants perceived general practitioners as relatively powerful and 'able to influence things', whereas general practitioners often expressed themselves as relatively powerless and unable to be proactive in 'reclaiming' their patients. The analysis highlights the complexity of the outpatient experience, drawing attention to detailed areas of contradiction, irony and conflict in the total context of outpatient care. These areas should be addressed in policy development designed to shift the balance of care further towards the primary sector. © 1998 Elsevier Science Ltd. All rights reserved.

Keywords: Outpatient care; Primary-care led NHS; Primary–secondary care interface; Qualitative study

1. Background

The notion of a primary-care led NHS is the most radical idea shaping current policy for health service provision in the United Kingdom. It promises a revolution in the organisation of healthcare which has major implications for conventional hospital-based service provision and consumption (NHS Executive, 1994). It envisages a shift in power and resources from secondary to primary care, with the aim of bringing the planning and provision of care 'closer to patients'. The White Paper, 'The New NHS' (Department of Health, 1997) provides a framework for the development of Primary Care

* Corresponding author. Tel.: +44-117-928-7282; Fax: +44-117-928-7340; E-mail: m.somerset@bristol.ac.uk.

0277-9536/98/$ - see front matter © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0277-9536(98)00338-4

PAPER 8

Groups which will commission services for local patients and improve the health of local people. Until recently outpatient services, which lie at the interface between primary and secondary care, have received relatively little direct consideration in national policy statements and debate about this radical concept for restructuring healthcare (e.g. Department of Health, 1996).

In contrast to other countries such as the USA or France where patients may self-refer to consultant specialists, in the United Kingdom, as in Australia and Canada, the General Practitioner (GP) is the gatekeeper to outpatient services. As healthcare costs have risen, policymakers in all the advanced industrialised countries have begun to examine the appropriateness of the site of healthcare delivery. The transfer of patient care from the secondary to the primary sector has been expected to contain costs and to parallel other movements toward devolution and patient empowerment.

The National Health Service (NHS) deals with approximately 40 million outpatient attendances per year in England and Wales, costing around £2.4 billion (National Audit Office, 1991), and about 20%–25% of these are new GP referrals. Attendance at specialist outpatient clinics in the UK has remained a relatively invisible issue in health services policy generally. While waiting times for inpatient services have been the subject of many research and government funding initiatives, waiting for outpatient services has until very recently been a neglected issue (Pope, 1993). There is no evidence to date from national routine data that restructuring in the NHS has had a notable effect upon outpatient discharge or re-attendance rates. Total attendances at outpatient clinics in England alone, having declined from more than 37 million in the mid-1980s to just over 36 million in the late 1980s, rose again from 1991, reaching 37.5 million in 1992–1993 and 38.2 million in 1993–1994 (Department of Health, 1993, 1994).

The White Paper 'Working For Patients' (Secretary of State for Health, 1989) attempted to highlight the idea of health professionals working in partnership in order to 'streamline' patient care, as have later statements (NHS Executive, 1996). Further NHS policy guidance has suggested mechanisms for improving the flow of patients through the outpatient system (NHS Executive, 1994). In the mid-1990s, the newly constituted NHS R&D initiative has also begun to focus some attention upon outpatient services, devoting a large programme of research to the 'balance of care' at the primary–secondary care interface. The more or less explicit agenda of this programme is to examine the feasibility and consequences of tipping the balance of care further in the direction of the primary and community sector, away from the hospital sector.

However, the benefits in terms of quality or efficiency of a primary-care led NHS are not unequivocally clear (Coulter, 1995, 1996). Both Coulter (1995) and Szecsenyi (1996) have emphasised a need to understand patients' needs, values, perceptions and attitudes in considering policy measures to improve care at the primary–secondary care interface. To this, one can add that the health professional's own values and attitudes must be considered, for example in terms of job satisfaction and perceptions of professional responsibilities.

The problem of 'unnecessary' outpatient attendance has been widely recognised (Marsh, 1982; Hartog, 1988; Samanta et al., 1991; Cartwright and Windsor, 1992; Dodd et al., 1994; Dunn et al., 1994; Ember, 1995; Neal et al., 1996; Reeve et al., 1997), and many reasons for the recycling of outpatients have been identified. While there may seem to be good clinical reasons to discharge an outpatient to general practice, in a busy outpatient clinic there may also be 'good organisational reasons' (Garfinkel, 1967) not to do so. While formal policy debate about services focuses upon clinical or economic rationale for decision making, it must also be acknowledged that outpatient services are embedded in a network of possibly competing incentives for different parties, this is likely to include non-clinical and non-economic 'real reasons' (Emerson, 1991).

It has been conventional in some studies of outpatient care to focus upon relevant issues from the perspectives of the three obvious parties involved: patients, GPs and specialists. A number of studies have examined the views and attitudes of these groups (Grace and Armstrong, 1986, 1987; Cartwright and Windsor, 1992; Dodd et al., 1994). It is unsurprising that the general conclusion of these studies is that the perspectives of the three groups diverge. However, there has been an overemphasis upon this divergence and the impasse between entrenched positions which it suggests. This has tended to lead to a portrayal of each of the groups as being relatively homogeneous and resilient to change. In this study we attempt to examine the *issues* which affect people involved in providing and receiving care at the primary–secondary care interface, as well as to identify different interests or perspectives which are important in shaping the outcome of attempts to move towards a primary-care led NHS.

The study explored the views and experience of outpatient attendance held by patients, hospital clinicians and GPs and the professional interaction between clinicians across the primary–secondary care interface, within the context of shifting resources. The study formed the preparatory hypothesis-generating phase of a cohort study of which the aim was to identify the key modifiable factors associated with long as opposed

PAPER 8

to short sequences of outpatient care following referral from general practice. This involved observation of patients newly referred from primary care to hospital outpatient specialties of general surgery, general medicine, gynaecology, ENT and paediatrics.

2. Methods

2.1. Participants

Recruitment was purposeful, the aim being to select a sample that would include participants best able to address the central research question (Marshall, 1996). A sampling framework, which identified the variables most likely to provide rich and relevant data, based on previous research experience, was drawn up. Participants were then recruited accordingly.

GPs and consultants were identified from Health Authority lists. Those who matched the selection criteria were contacted by the research team and invited to take part. There were 29 participants. This total comprised:

1. Ten GPs from a range of practice sizes (single-handed/multi-partnered), locations (urban/rural) and fund-holding status. All the GPs at times referred patients to the participating hospitals.
2. Ten consultants from a range of specialities, although working in the participating hospitals, none contributed to the main study.
3. Nine patients, all recently referred by their GP to one of the participating hospitals. Three were waiting to attend, the others had attended once or several times. Patients were recruited via their referring GP (who took part in the main study) and did not subsequently take part in the main study.

The three parties were not linked, rather the sample comprised 'critical cases' that is people who had specific, relevant experience. It was envisaged that the variation would result in greater insight and understanding of the scope of difficulties encountered throughout the referral process.

2.2. Interviews

A central principle of interpretative qualitative research is that the participant's perspective on the topic of interest is uncovered as viewed by the participant rather than the researcher (Marshall and Rossman, 1995). The interviews were semi-structured as there was a very specific focus to the study, namely the requirement to explore particular aspects of outpatient attendance. By eliciting the participant's own views and emphases while maintaining a focus on specific aspects, the semi-structured interviews comprised a bound piece of qualitative work (Miles and Huberman, 1994).

A list of broad issues to be covered was drawn up on the basis of a previous review of outpatient services research (Faulkner and Frankel, 1993). The schedule included issues relating to the initial referral, expectations and experiences of outpatient attendance, and discharge to general practice. However respondents, to a certain extent, guided the agenda of the interview by demonstration of their enthusiasm or lack of enthusiasm for topics. Later interviews were informed by themes and issues raised by other participants. By the ninth or tenth interview it became apparent that no new themes were emerging. The interviews were shared between three interviewers and lasted between 20 and 45 min. They were carried out in the workplace in the case of the hospital clinicians and GPs and at home for the patients. They were tape-recorded and transcribed, each interviewer transcribing those she had conducted.

2.3. Analysis of interviews

Initially transcriptions of the sets of interviews with patients, consultants and GPs were examined separately, each of the three researchers reading and coding her own interviews. In order to enhance validity of the process of analysis, the researchers discussed their readings and codings, and exchanged transcripts to compare codes, themes, and interpretations of the text to ensure consistency of approach. The interviews were read several times, and as segments relating either directly or indirectly to aspects of outpatient attendance were identified, they were highlighted. Gradually repeated readings enabled topics to be coded. Each researcher checked the codings and themes of the other two, and a single set of codes was finally agreed for all transcripts. The manuscripts with the marked segments and codes were then scrutinised again to identify broader themes under which the marked segments were then classified. The themes and segments were then reconsidered in context and the insight they provided on the respondent's perspective concerning each particular topic was recorded. Interpretations were checked against the transcribed texts in order to evaluate their plausibility. Negative instances were sought and by integrating further material into each theme, the emergent framework was strengthened. In the context of the larger cohort study, the themes arising from the data were used to develop hypotheses about factors influencing outpatient re-attendance.

PAPER 8

The themes and interpretations resulting from this process, as presented below, illustrate the diversity of the range of issues, perceptions and views within and across the three groups, the central aim of qualitative research being to reach improved understanding about the complexity of issues (Marshall, 1996) and their range, rather than necessarily attempting to demonstrate representativeness of a wider population to which quantitative generalisations might be drawn.

3. Findings

Participants were invited to consider their experience and views of outpatient attendance beginning with the initial referral from general practice and concluding with discharge to GP care. Several themes which were important to participants across the groups emerged from this process. These themes are presented in Table 1 below in turn, together with an interpretative commentary and illustrative extracts from the interview data.

3.1. Trust and knowledge: who is the expert?

There was a consensus amongst the patients that they had been referred appropriately to outpatients. Essentially this was because they believed that their condition required specialist advice which was best provided by a hospital consultant, rather than the GP who was trained to provide general care. To some extent, a lack of trust in the GP's knowledge was linked to the severity of the condition; that is for patients who perceived their condition to be serious, it was increasingly important to be referred. Patients often contrasted the knowledge of the GP and consultant, seeing the consultant as the 'expert able to delve deeper', in contrast to the GP who 'relies on past experience rather than knowledge of the condition'.

> If it's anything of a serious nature, I certainly wouldn't take the GP's word for it (P:S2)

> You need specialist input for some things, I don't think that GPs have enough knowledge (P:R2)

> The GP can only touch the surface ... the consultants are the ones with expertise [they] can look into it deeper (P:N1)

However, once patients had accessed the hospital outpatient services there was a range of views relating to doctors other than the consultants. The experience of some patients led them to express great confidence in the abilities of registrars and house officers, but others implied they were unhappy with levels of knowledge and expertise:

> Very competent, very much abreast with the consultant (P:R1)

> I had the impression that the registrar was not experienced because he had to leave to speak to the consultant (P:R3)

GPs acknowledged the specialist knowledge of the consultant, for example referrals were usually initiated in order to access diagnostic procedures, treatment, or management advice from a hospital clinician. Some GPs maintained they would be more likely to refer to a consultant who was known to them personally. In addition it was recognised that referral to a consultant was often a means of satisfying the patient's desire for a second opinion from an 'expert'. Trust between hospital clinicians and GPs was important when the care of a patient was transferred from one setting to another. This was particularly the case where the hospital clinician had a limited knowledge of the community setting and was uncertain of the abilities of the GP. Some GPs proposed that if consultants had more belief in their capabilities, patients could be more readily discharged back into primary care.

Table 1
Themes

Theme 1	trust and knowledge: who is the expert?
Theme 2	inter-professional communication: the key to attitudes between GPs and consultants?
Theme 3	communication with patients: a need for information?
Theme 4	non-clinical influences: do patient personality and social status count?
Theme 5	personal and professional anxiety: who needs reassurance?
Theme 6	power relations: who is in control?
Theme 7	changes in practice: is the price too high?

PAPER 8

You would just be slightly more cautious about referring to people who you've never met because I suppose you're not so confident in their knowledge and expertise (GP8)

A significant minority [of my patients] will not accept anything you say until they have had the laying on of hands by the consultant (GP10)

I think generally consultants feel the need to hang onto patients. I assume that's because they're not confident that treatment will be followed-up adequately in general practice (GP2)

[re-attendance] is to do with trust between the consultants and the GPs and knowing the GPs and their competence (GP10)

The views presented by the consultants largely confirmed the perceptions of the GPs in relation to consultants' confidence in standards of primary care. As anticipated by the GPs, there were some reservations about levels of knowledge in general practice, some consultants conceding they would be uncertain that follow-up would be carried out satisfactorily if the patient was discharged:

You lose control if you send them back to the GP for something that the GP won't have the expertise to deal with (C5)

You don't know what's out there, there are some GPs you trust more than others (C4)

At the same time, there was some acknowledgment of the advantages of the GP's knowledge, linked to the perception that patients could overestimate the outcomes of referral to the clinic and hold unrealistic expectations of hospital follow-up:

[The patients] may think they will be assuaged by coming into hospital. That is a touching faith, but it is probably better that continuity is handled by someone who knows the family background (C1)

3.2. Inter-professional communication: the key to attitudes between GPs and consultants?

While GPs identified clinical competence as a central factor in the decision to refer a patient to a particular consultant, this could only be ascertained through their knowledge of the clinician. Thus, communication was the key issue underpinning the choice of consultant for each patient. Information relating to a consultant was frequently derived from the GP's experience of referral to that individual, so the manner in which a consultant communicated with both patients and GPs was crucial.

Certain consultants are more receptive to different types of people (GP9)

The way they talk to people (GP2)

Does [that consultant take] my questions seriously as well as the patient's (GP4)

Some GPs said specifically that the information that filtered back from the hospital was frequently inadequate and they would like more detailed management advice. They pointed to the difficulties in taking over follow-up care for patients if there was a lack of such advice from the specialist. The increase in detailed management plans from clinicians at the senior registrar level, which comprised 'recipes to follow', was favoured by the GPs who had received them. The GPs also perceived the lack of communication between members of the consultant team to be a problem, with poor feedback between the consultant and junior doctors being identified as a source of inappropriate follow-up appointments.

There's absolutely no clinical information in the hospital letters whatsoever. It's completely useless (GP2)

I'm happy to have those kinds of [management] guidelines to follow (GP10)

[re-attendance] is to do with how much feedback there is from the junior members of the team to the consultant (GP5)

there's a chemistry between the consultant and his team and the patient which is working or not working and that chemistry generates a repeat appointment (GP4)

Communication was also a key issue for consultants. The main mode of communication for hospital clinicians was the GP's referral letter. On the whole, they felt that referral letters provided them with adequate information, although a minority were felt to be "bland, non-committal ... and sometimes positively unhelpful". Some consultants criticised the lack of background details which could help the consultant in the context of 'heavy-pressed clinics'. Others identified two main types of referral, for specific procedures or for diagnosis, each of which require

PAPER B

different types and amount of information from the GP. Information from the GP was felt to be less adequate in the latter type of referral which 'often will have a psychological basis which is difficult to tease out'.

> it's often peripheral but relevant information which may not be in the hospital notes ... I think sometimes they're frightened of committing to paper things which are relevant but not strictly associated with the primary complaint (C4)

3.3. Communication with patients: a need for information?

For many of the patients, the information they had been given by their GP had not been adequate to satisfy their questions about their illness. They prioritised the need to obtain more information from the hospital clinician. Several expressed the hope that during the outpatient consultation they would be given instructions to guide their behaviour at home. Usually, they expected verbal information, although some felt they would have liked written guidance from the hospital clinicians:

> a lot of the GPs have said, when I've gone up to ask about things, they've said they don't have the knowledge and they'll have to ring up and find out things (P:R2)

> Hopefully [the consultant] will tell me what I can do and what I can't do (P:N1)

> I'll probably get verbal information and the doctor will get written ... leaflets wouldn't be a bad idea though (P:N3)

Some re-attending patients were disappointed at the lack of information they had received at the outpatient clinic. This was intensified by a frustration that while investigations implied nothing was wrong, their symptoms persisted. Others expressed uncertainty about the management decisions made at the hospital, not knowing why the hospital clincians had taken certain courses of action in relation to their care. Some patients felt that the amount of information given by clinicians was partly dependent on patient 'type', specifically people who are proactive in their health and ask questions are provided with better quality information.

> they say it is not active but then why am I getting these hellish pains? ... they've never really explained it to me (P:R1)

> they still haven't told me what's wrong from the first consultation (P:N3)

> [the consultant] is referring me back to the physio again, but I've already been there and they said they couldn't do anything ... it makes you wonder what direction they're working in regarding my complaint (P:R1)

> it depends in your personality, whether you actively want to be involved in your health or whether you sit back and think "What will be will be" (P:R2)

3.4. Non-clinical influences: do patient personality and social status count?

The GPs commented on the influence of explicit and implicit patient demands on their decisions to refer. The patient's social status and ability to articulate verbally were put forward as tacit influences which affect the liklihood of referral. The reason a patient may wish to re-attend outpatients was linked to the patient's perception of the consultant as 'Dr. Big'.

> The patient factor or demand ... it's partly intelligence and that's of course linked to social status (GP4)

> the seeing of the consultant for the patients is actually the important statement ... seeing The Man (GP7)

The GPs also felt that consultants sometimes influence patient attendances at the outpatient clinic for non-clinical reasons. Specifically, whether a patient was asked to re-attend at the outpatient clinic was partly dependent on the nature of the relationship between the patient and consultant. Patients who were 'nice' and liked by the consultant would be more likely to be seen again in the outpatient setting:

> I have to say that with one or two patients you feel it's because they've got such nice personalities and the consultants enjoy seeing them (GP10)

> We've all got our rotten patients. I'm sure the consultants have got them too, I bet the rotten ones who get up your nose don't get re-attended (GP4)

The consultants similarly talked about non-clinical factors which might influence the care patients received. They recognised the implicit influence that patient demands may have on the referral process. However, some consultants acknowledged that an important and appropriate service provided by outpati-

PAPER 3

ents is to give GPs a break from more 'difficult' patients:

> it almost seems they've been slung up to hospital because the parents are sort of moaning and groaning philosophy (C2)

> Sometimes it is clear that I'm just being asked to take the patient off the GP's hands for a while and that's an appropriate function of specialist care (C9)

Interestingly, all three groups of participants, patients, GPs and consultants, indicated that patients' perceptions of their illness could in part be defined by the care they receive. For the patients, the hospital clinician's decision to discharge them to general practice was interpreted as meaning de facto that the clinical condition was at least under control. In short the prevailing view was that 'if you have been discharged you must be all right'. Similarly, both the GPs and consultants described how patients may define the extent of their illness according to their attendance at either the GP's surgery or the hospital outpatients. A common perception was that for some people outpatient attendance lends credibility to their illness:

> Hospital follow-up creates an illness mentality "I'm ill because I'm seen at the hospital all the time" (GP7)

> Many patients like to be seen at the hospital, it gives a certain credibility that they have a disease that has progressed sufficiently severely that they have to remain under surveillance at the hospital (C6)

3.5. Personal and professional anxiety: who needs reassurance?

Some patients acknowledged that one outcome of referral to outpatients was reassurance and optimism. It was assumed that if several opinions were gained this would be followed by a superior outcome, even when some of the opinions were those of medical students. Similarly being discharged back to primary care implied that the condition was improving and this again provided comfort.

> I was reassured by that because I thought, well there's 25 more opinions about me (P:N3)

> I was just relieved to be signed off because ... you think I must be alright, they're letting me out (P:N3)

GPs recognised the reassuring effect of referral and suggested that they would be more likely to refer an anxious patient, or one who was unable to accept the explanation given to them in the primary care setting. Many of the consultants also appreciated this feature of outpatient attendance:

> There are the patients who transmit anxiety to me and therefore I refer them on (GP2)

> They either don't have the confidence in me or I can't put my message across (GP4)

> I give a lot of patients the opportunity to say whether they would like to come back [to outpatients] or not, because I think re-attendance is to some extent reassurance (C7)

The consultants also agreed that it was appropriate for GPs to use outpatients to lessen their own anxiety about patients, with the proviso that the GP should explain in the referral letter that this is the main reason for requesting a consultation. This was described as a 'benign role' for them, which they were on the whole entirely happy to fulfil:

> There is nothing wrong with [the GP wanting reassurance] as long as the GP acknowledges that in the referral letter (C9)

> in general I have a very strong feeling that if a GP is worried about a patient, that is an appropriate referral, even if when I see them I am not worried about them ... that's what specialists are for (C7)

However in some cases outpatient attendance also served to alleviate the concerns of the consultant. For specialties such as gynaecology, where follow-up is less common, the paucity of re-attending patients was a potential source of anxiety as the full implications of treatment may not filter back to the hospital consultant:

> We do need feedback on what we are doing. I assume that women given hysterectomies by me go on to have a normal sex life, but would anyone tell me if this was not the case? (C8)

3.6. Power relations: who is in control?

The power relation between doctors and patients was a common theme in patients' discussions about their experiences of outpatient attendance. Many of them talked about the degree of influence they believed

PAPER 8

they had exerted in the decision to refer. Responses ranged from those who believed the decision rested entirely with the GP through to those who thought that they were primarily responsible for the referral. Appreciation was expressed by those patients who had been asked for their opinion. When considering the extent to which they were able to influence events during their outpatient attendances, some patients maintained that they would actively seek information if they were left in any doubt about their care following discharge to general practice. However, there was general acknowledgement that they did not have a sufficiently rich knowledge base from which to generate appropriate questions. So for many patients there was a general acceptance of the doctor's word, whether GP or consultant. While this was total for some, others expressed the desire to be more proactive in the decisions made about their care.

> My GP initiated everything (P:R1)
>
> Everything they have told me I have carried out, I always have done (P:S1)

But on the other hand:

> I asked my GP if I could see someone else (P:S2)
>
> My GP gave me a choice, so that was really good (P:N1)
>
> I want to be involved all along, I want to know everything, to be told everything and go over it time and time again (P:R2)

The power relation between GPs and consultants was a contentious issue for GPs. For many, there was a perceived inequality in the distribution of power between primary and secondary care clinicians. While GPs agreed with the principles of sharing patient care with hospital clinicians, there were misgivings about the extent to which this could take place. Some GPs did not feel themselves to be equal partners in the provision of shared care, but rather believed that they were viewed as the 'second rate service'. The consultant was frequently seen as the powerful figure, and there was a suggestion that the re-attendance of patients in the outpatient setting was associated with 'empire building' on the behalf of some consultants:

> The theory would be that we're on the same level [but] the consultant's up there in the clouds, we're somewhere down there on earth and the patient is sub-terranean (GP4)
>
> They like having a number of people ... I think they get a buzz out of it ... I'm sure that's a factor that's under-recognised in medicine (GP4)
>
> I don't mind shared care so long as we're treated as equals (GP5)
>
> shared care brings the tier in ... whereby I'm definitely number two ... you do tend to be seen as the second rate service within shared care (GP2)

Conversely, there was a consensus among the consultants that the GP was a relatively powerful partner. GPs were perceived to have the potential to exercise a considerable amount of authority in decisions about their patients. Indeed, some consultants expressed their own feelings of powerlessness when discharging patients back to general practice. However, while some maintained that GPs were in a good position to 'reclaim' their patients from outpatient care, GPs were also perceived as failing to exercise this power.

> GPs have that much power, and like to influence things ... You lose control if you send them back to the GP (C5)

But on the other hand:

> If nobody's stopping me seeing them again ... I carry on ... I've never met anyone trying to control the number of times I see the patient (C2)

3.7. Changes in practice: is the price too high?

For both the GPs and consultants, two recurring sub-themes in relation to perceived changes in practice were pressures associated with resource or funding issues, and the threat of litigation. Some of the GPs, particularly non-fund-holders, expressed concerns about the impact of the fund-holding system on patient care. The focus on finance over and above clinical reasons for referral or outpatient attendance was the key source of concern: that fund-holding could result in 'inappropriate' medical practice. Patients of fund-holding GPs were often felt to be at an unfair advantage in gaining access to hospital services. Some non-fund-holding GPs perceived themselves and their patients to be on the bottom tier of a 'two tier health service'. However, most of the fund-holding GPs claimed that funding arrangments had not influenced their practice with regard to referral or outpatient follow-up, although acknowledging that reductions in the use of hospital services were financially beneficial for

PAPER 8

their practice. Nevertheless, the consultants recognised an impact of funding on the practice of GPs, specifically fund-holders, which was seen potentially to affect the extent of re-attendance at outpatients:

> one is always concerned that fund-holding will lead to inappropriate medicine (GP4)

> if you're a fundholder you sharpen your focus on the kind of service you're getting a bit more, but that focus is financial rather than necessarily clinical (GP7)

> I am not a fund-holding practice, and I do have very strong feelings about what I see as the development of a two tier health service ... patients of fund-holding practices are seen sooner, my patients have to wait, I feel very resentful about that (GP2)

> fund-holding ... hasn't altered my policy as regards referral and follow-ups, but certainly our fund-holding manager is pleased that people aren't being followed up long term in the hospital because it saves money for the fund (GP5)

From the consultant's point of view:

> if patients come to see me at the hospital the payment has to be made, so particularly the fund-holding general practitioners are going to look very critically at the number of attendances (C6)

> I perceive a gradual swing away from the amount of re-attendance because GPs' minds have been concentrated on trying to save money (C7)

The resource implications of moving care from outpatients to the primary care setting was a commonly recurring issue for GPs. The majority of the GPs expressed willingness to take over the routine follow-up care of patients discharged from secondary care. However there was an important caveat: would the resumption of care of a large number of patients lead to excessive work for already overstretched GPs?

> I should be able to say I can take over again, [but] one's always slightly concerned that generates work for oneself (GP9)

> there would be this slight concern on the GP's behalf that he is going to increase his workload by saving the hospital workload (GP4)

> there is a large chunk of secondary care which is moving to primary care ... it makes a lot of sense for the patients and administratively for this to happen, but it does have major workload implications (GP7)

Concern about litigation was another important issue for both GPs and consultants. The majority of the clinicians acknowledged that the perceived or real threat of legal action from a patient had some impact on their practice. For many of the GPs, this affected their referral behaviour. They would be more likely to refer a patient if they felt there was a risk of legal action. Some of the GPs perceived consultants to be influenced by the legal implications of their practice, resulting in an increase in investigations and follow-up outpatient care. However, a minority of the GPs maintained that the threat of legal action had little impact on their practice, arguing that they did not practice 'defensive medicine'.

> I might be defensive in some positions, if you demanded a Cardiologist appointment for your nebulous chest pain, I'd be hard placed to say "Don't be daft" because I'm taking a thousand to one risk of ending up in the high court (GP4)

> I think we would tend to refer because of the fear of missing and having the medical legal implications of that ... and I would be very surprised if the consultant's investigations and follow-ups weren't affected by that as well (GP8)

But on the other hand:

> I don't feel that I practice medicine looking over my shoulder ... I don't think defensive medicine (GP10)

Many of the consultants talked at length about the effects of litigation on their work. Some of them identified a sharp increase in the amount of legal action taken against hospital clinicians, and pointed to the negative impact on the way they think about patient care. Some drew on their personal experience of official complaints to argue that the threat of litigation was leading to clinicians practising defensively. The consultants also recognised that GPs referring patients for specialist advice may be 'covering their backs':

> it's become horrendous, this has increased exponentially, I think it's having a serious effect on the way we think about our patients and our practice ... I think we're protecting ourselves more and more, and I think it's disillusioning some of us (C7)

PAPER 8

> I'm sorry to have to say and admit freely that many of my actions are taken to try and limit the possibility of litigation or complaint, because I've experienced it (C4)

> GPs feel they lack sometimes a little bit in the way of expertise and they want the cover of someone more senior to them ... there is an element of practising defensive medicine (C2)

There was a consensus in the accounts given by both groups of clinicians that the key factor in this perceived increase in the threat of litigation was the creation of a 'consumerist culture' within the NHS. Patients were held to have become customers or clients, who purchase a service, and complain when things go wrong. Public awareness about patients' rights to complain and the means for doing so were felt to have been greatly heightened. Government policy developments, as well as advertising by legal firms in hospitals, were all perceived to fuel a growth in patient expectations, which may have become unrealistically high:

> people are becoming more litigation conscious, more consumerist in their view ... they buy off the shelf and I think the tendency for people to complain ... could be a major problem ... You only need to have one complaint against you to have the whole of your clinical view coloured by that (GP7)

> the government is fermenting complaint by charter attitudes and so on, and litigation is encouraged. Our outpatient cards have a statement saying "If you have any complaints or worries please contact us" and on the back there's an address of a local solicitor (C4)

> the expectations of the patient are very high, maybe unrealistically high (GP10)

4. Conclusions

Our analysis of outpatient care identifies some of the obstacles which need to be negotiated if policy initiatives proposing a move towards a primary-care led health service are to be implemented. An important theme concerns the status of the GP. The traditional perception of hospital consultants as the premier source of knowledge about disease was largely upheld by patients and GPs. Furthermore, some consultants expressed concern about the level of expertise in the general practice setting. This stereotypical picture should however be tempered by the recognition that there is acknowledgement of the distinctive skills and understanding which GPs can provide, for example their knowledge regarding patients' social and family circumstances.

In agreement with previous research (Audit Commission, 1994; Bowling et al., 1991; Jacobs and Pringle, 1990; Lloyd and Barnett, 1993; Ong et al., 1995; Westerman et al., 1990), communication between GPs and consultants was acknowledged to be a problem, each party suggesting that the remedy rested with the other. As it would seem that the quality of communication was responsible for the opinions GPs and consultants derived of one another, poor communication resulted not only in lack of information about a particular patient but also created negative attitudes for future working relationships. This seems of crucial importance at a time when the need for health professionals to work effectively together to streamline patient care has been highlighted (Secretary of State for Health, 1989).

Patients' hopes that hospital doctors would amend the lack of information provided within general practice were frequently not fulfilled. These expectations seem to have been based on the belief that the hospital doctor had more expert knowledge than the GP. However, some patients felt they had been given a lot of information about their condition from hospital clinicians and this was often related to the extent to which they perceived they were proactive in their care. Others seemed to have little knowledge about their condition and expressed great uncertainty about their management by both GPs and hospital doctors. Patients often felt that they lacked the necessary knowledge to pose questions, were unclear who they had seen at outpatients, and (sometimes) expressed their frustration about contradictions between information from the GP and hospital.

In addition to clinical rationale, GPs and consultants agreed that reasons for referral included responding to patients' demands and expectations. This finding lends support to previous research indicating that doctors perceive pressure from patients to be a factor contributing to referral (Armstrong et al., 1991; Wilkie, 1992; Williams et al., 1995). However, each practitioner tended to identify the other party as the most likely to be influenced by a patient's personality over and above their physical condition. For many patients, outpatient attendance was regarded as an indication of more serious illness, and similarly discharge back to primary care was seen as a sign of progress. The conventional view held by doctors that outpatient attendance was a source of reassurance to patients was thus confirmed, but in addition GPs and consultants

PAPER 3

acknowledged they themselves drew comfort from their patient's outpatient attendance. If people are increasingly to be managed within primary care (NHS Management Executive, 1991), this extra buffer of reassurance may cease to exist.

Reservations have been expressed regarding the success of policies (Department of Health, 1992) directed towards transferring power from secondary to primary care and giving more power to patients (Coulter, 1995, 1996). The unequal and variable power relations which emerged from our data between and within the three groups, and the disparity in perceptions of relative power, support the case for caution regarding the pace of change.

Unequal power relations were felt particularly acutely in the context of shared care. GPs perceived consultants as the most powerful party in the GP-consultant relationship and expressed feelings of powerlessness in 'reclaiming' their patients back from secondary care. There was also perceived powerlessness to control their own workload. In contrast, the consultants did see GPs as powerful, although they suggested that GPs often failed to exercise that power.

In terms of patient empowerment, although there was evidence of proactive patients who felt they had exercised choice in the initial decision to refer, there were also passive patients who seemed content to relinquish control to their doctors. Patients might often be unaware of the extent to which they were able to influence events concerning them. In several cases, they were unquestioning of decisions because they regarded the consultant, in particular, as highly knowledgeable. This apparent discrepancy in status resulted in a lack of power which further determined the care they received. It appears that one type of measure required to foster a primary-care led NHS is to support patients and GPs in exercising the power over care which they have, but of which they may not be aware (Faulkner et al., 1995).

Recent policy developments have focused on the role of funding arrangements in the provision of efficient and effective health care (NHS Executive, 1994). In terms of perceived changes in practice related to funding issues, GPs regarded fundholding status differently according to their level of involvement. Fundholders acknowledged that they examined services more carefully for their financial costs, but maintained that the clinical needs of patients remained paramount. Meanwhile non-fund-holding GPs were concerned about the development of a two tier health service, contending that fund-holders were able to get a 'better deal' from the hospitals. Hospital consultants were aware of these issues, recognising that fund-holding GPs in particular may be less willing to pay for some hospital services, including follow-up in the outpatient setting. Consultants and GPs also acknowledged that there were resource implications for the 'sharing out' of patient care. Once more, the implications for the transfer of care to the primary sector were highlighted by both parties. While GPs expressed willingness to take over some patient care traditionally managed within the hospital setting, they were again concerned about the resource and workload implications of such changes.

Litigation by dissatisfied patients was an issue raised by both GPs and consultants, and they acknowledged that the perceived increase in legal actions had influenced their practice. However, it did not seem an important issue for the patients in our sample. This raises the question of the extent to which litigation is a threat for health professionals in today's NHS. Changes in practice brought about by fears of legal action could be a result of the broader policy changes within the NHS which emphasise patients' rights to complain about the quality of the service they receive, or they could be a reflection of wider movements towards a consumer society. In any case, the importance attached to these issues by consultants and GPs indicates a considerable resistance to systems of care which would reduce the capacity for second opinions, availability of tests or doctor-initiated referrals and repeat consultations.

From the perspective of outpatient service policy in the NHS, the high volume of outpatient attendances, and especially re-attendances, appears irrational. Certainly many attendances are unnecessary from a strictly clinical point of view, however, as this study has shown, every outpatient attendance is embedded in a network of social and professional relationships and meanings. Individual experiences of outpatient care vary widely, be they patients, GPs or consultants. Some exercise more power than others; some gain reassurance; some give or receive more effective information; patients' subjective concepts of their illness and the experience of care are intertwined; some doctors fear legal repercussions of not offering care; some deplore inequalities in access to services. It would seem that experiences of outpatient care are affected by many more or less conscious incentives – reassurance, funding, relative power and authority, consumerist culture, interpersonal perceptions, information and communication needs, professionalism and perceived authority and trust.

A move toward an NHS in which primary care plays a more pivotal role would need to bring with it major changes in attitudes, power relations and organisational responsibilities. Funding arrangements and the promotion of patients' rights to access information and receive a good quality service have effected some transfer of power, and encouraged some clinicians to be creative in their use of both primary and secondary care resources. However, while traditional views about

the relative status and expertise of secondary and primary care clinicians endure and the breakdown of communication at the interface remains a barrier to effective partnerships, the transfer of more substantial power across the interface will not be fully realised. In some aspects such as litigation, obstacles to increasing the role of primary care are actually growing. As Coulter (1996) has implied, change in culture among health professionals and the public is required if further transfer is to take place. But there may be a fundamental contradiction between the traditionally perceived trust, expertise and authority of the hospital-based consultant, and the concept of an NHS built around a consumer-oriented locally-driven system of care planning and provision.

Acknowledgements

This research was funded by the NHS Research and Development Primary–Secondary Care Interface Programme. The authors would like to thank Dr. Ian Harvey and Professor Stephen Frankel for their comments and advice on early drafts of this paper. Thanks also to the Consultants, GPs and patients who agreed to talk to us about their experiences of outpatient.

References

Armstrong, D., Fry, J., Armstrong, P., 1991. Doctor's perceptions of pressure from patients for referral. BMJ 302, 1186–1188.

Audit Commission, 1994. What Seems to be the Matter?: Communication Between Hospitals and Patients. HMSO, London.

Bowling, A., Jacobson, B., Southgate, L., Formby, J., 1991. General practitioners' views on quality specifications for outpatient referrals and care contracts. British Medical Journal 303, 292–294.

Cartwright, A., Windsor, J., 1992. Outpatients and their Doctors: A Study of Patients, Potential Patients, General Practitioners and Hospital Doctors. Department of Health, HMSO, London.

Coulter, A., 1995. Shifting the balance from secondary to primary care. British Medical Journal 311, 1447–1448.

Coulter, A., 1996. Why should health services be primary care-led? Journal of Health Services Research and Policy 1 (2), 122–124.

Department of Health, 1992. The Patient's Charter. HMSO, London.

Department of Health, 1993. Outpatient and Ward Attenders for England, Financial Year 1992–93. Department of Health Statistics Division, London.

Department of Health, 1994. Outpatient and Ward Attenders for England, Financial Year 1993–94. Department of Health Statistics Division, London.

Department of Health, 1996. Primary Care: The Future. HMSO, London.

Department of Health, 1997. The New NHS. HMSO, London.

Dodd, K.L., Rhead, S., Towey, K., 1994. Paediatric medical outpatients: are all those reviews necessary? Archives of Disease in Childhood 70, 493–496.

Dunn, J.M., Elliot, T.B., Lavy, J.A., Bell, A., Kernick, V.F.M., Campbell, W.B., 1994. Outpatient clinic review after arterial reconstruction: is it necessary? Ann. R. Coll. Surg. Engl. 76, 304–306.

Ember, M., 1995. Outpatient follow-up. Who benefits: doctors or patients? BMJ 311, 1315–1316.

Emerson, R.M., 1991. Case processing and interorganizational knowledge – detecting the real reasons for referrals. Social Problems 38, 198–212.

Faulkner, A., Frankel, S., 1993, Evaluation of Delivery and Use of Outpatient Services, 1st ed. Health Care Evaluation Unit, University of Bristol, Bristol.

Faulkner, A., Saltrese-Taylor, A., Williams, M., O'Brien, J., Collins, C.D., Frankel, S., 1995. Outpatients revisited: subjective views and clinical decisions: management of general surgical outpatients in South West England. Journal of Epidemiology and Public Health 49, 599–605.

Frankel, S., Faulkner, A., 1994. The end of the outpatient problem? BMJ 309, 1308.

Garfinkel, H., 1967. Studies in Ethnomethodology. Englewood Cliffs, NJ, Prentice Hall.

Grace, J.F., Armstrong, D., 1986. Reasons for referral to hospital: extent of agreement between the perceptions of patients, general practitioners and consultants. Family Practice 3, 143–147.

Grace, J.F., Armstrong, D., 1987. Referral to hospital: perceptions of patients, general practitioners and consultants about necessity and suitability of referral. Family Practice 4, 170–175.

Hartog, M., 1988. Medical outpatients. J. R. Coll. Physicians Lond. 22, 51.

Jacobs, L.G.H., Pringle, M.A., 1990. Referral letters and replies from orthopaedic departments: opportunities missed. British Medical Journal 301, 470–473.

Lloyd, B.W., Barnett, P., 1993. Use of problem lists in letters between hospital doctors and general practitioners. British Medical Journal 306, 247.

Marsh, G.N., 1982. Are follow-up consultations at medical outpatients futile? BMJ 284, 1176–1177.

Marshall, C., Rossman, G.B., 1995. Designing Qualitative Research, 2nd ed. Sage, London.

Marshall, M.N., 1996. Sampling for qualitative research. Family Practice 13 (6), 522–525.

Miles, M.B., Huberman, 1994. Qualitative Data Analysis. Sage, London.

National Audit Office, 1991. NHS Outpatient Services. HMSO, London.

Neal, R., Dowell, A., Heywood, P., Morley, S., 1996. Frequent attenders: who needs treatment? British Journal of General Practice, March, 131–132.

NHS Executive, 1994 Developing NHS Purchasing and GP Fund-holding: Towards a Primary Care-Led NHS. Department of Health, Leeds.

PAPER 8

NHS Executive VFM Update Plus, 1994. Outpatient Attendances: Improving the Flow, No. 1, September.

NHS Management Executive, 1991. Integrating Primary and Secondary Health Care. Department of Health, London.

Ong, L.M.L., De Haes, J.C.M.J., Hoos, A.M., Lammes, F.B., 1995. Doctor–patient communication: a review of the literature. Social Science and Medicine 40, 903–918.

Pope, C., 1993. Waiting times for outpatient appointments. British Medical Journal 306, 408–409.

Reeve, H., Baxter, K., Newton, P., Burkey, Y., Black, M., Roland, M., 1997. Long-term follow-up in outpatient clinics, 1: The view from the general practice. Family Practice 14 (1), 24–28.

Samanta, A., Haider, Y., Roffe, C., 1991. An audit of patients attending a general medical follow-up clinic. J. R. Coll. Physicians Lond. 25, 33–35.

Secretary of State for Health et al., 1989. Working For Patients, Cmnd 555. HMSO, London.

Szecsenyi, J., 1996. Improving care at the primary–secondary interface: a difficult but essential task. Quality in Health Care 5, 191–192.

Westerman, R.F., Hull, F.M., Bezemer, P.D., Gort, G., 1990. A study of communication between general practitioners and specialists. British Journal of General Practice 40, 445–449.

Wilkie, P., 1992. What does the patient want? In Referrals to Medical Outpatients: Different Agendas of Patients, General Practitioners and Hospital Physicians, ed. A. Hopkins and P. Wallace, 1st ed. RCP Publications, London.

Williams, S., Weinman, J., Dale, J., Newman, S., 1995. Patient expectations: what do primary care patients want from the GP and how far does meeting expectations affect patient satisfaction. Family Practice 12, 193–201.

PAPER 8