Does the National Institute for Health and Clinical Excellence take account of factors such as uncertainty and equity as well as incremental cost-effectiveness in commissioning health care services? A binary choice experiment

P Tappenden and J Brazier and J Ratcliffe

The University of Sheffield

June 2006

HEDS Discussion Paper 06/10


ScHARR
School of Health and
Related Research

The University of Sheffield

ScHARR
School of Health and Related Research

Health Economics and Decision Science

Discussion Paper Series

June 2006
Ref: 06/10

Does the National Institute for Health and Clinical Excellence take account of factors such as uncertainty and equity as well as incremental cost-effectiveness in commissioning health care services? A binary choice experiment

Paul Tappenden, John Brazier and Julie Ratcliffe

Corresponding Author:
Paul Tappenden
Health Economics and Decision Science
School of Health and Related Research
University of Sheffield
Regent Court, Sheffield, UK
S1 4DA
Email: p.tappenden@sheffield.ac.uk

Abstract

Background

NICE is an independent organisation responsible for providing national guidance on the promotion of good health and the prevention and treatment of ill health in England and Wales. One of NICE's main roles is to produce national guidance on the use of health technologies within the NHS. Despite the Institute's recent efforts to clarify the way in which its Appraisal Committees reach their recommendations concerning the use of health technologies, there remains ambiguity about how cost-effectiveness evidence is interpreted alongside other considerations such as the degree of clinical need within the patient population, and the degree of uncertainty surrounding cost-effectiveness estimates.

Objective

To explore whether the NICE takes account of factors such as uncertainty and equity as well as incremental cost-effectiveness in commissioning health care services.

Methods

A binary choice experiment was undertaken using NICE's three Appraisal Committees. The experiment included five attributes:

(1) Incremental cost-effectiveness
(2) Degree of economic uncertainty
(3) Age of the target population
(4) Baseline health-related quality of life
(5) Availability of other therapies

A choice questionnaire detailing 18 scenarios was administered to NICE's Appraisal Committees. For each scenario, respondents were asked to indicate whether they would recommend the intervention under consideration or not. The stated preference data obtained from respondents were analysed using a random effects logit regression model.

Results

A response rate of 46% was obtained from the Appraisal Committees. The regression model suggests that increases in cost-effectiveness, economic uncertainty, and the availability of other therapies are associated with statistically significant reductions in the odds of adoption ($p<0.05$). The transition from a very low to a comparatively high

level of health-related quality of life is also associated with a statistically significant reduction in the odds of a positive recommendation. Smaller changes in health-related quality of life, and the age of the target population are not associated with a statistically significant reduction in the odds of a positive recommendation. Analysis of revealed preference data indicates that the model is capable of distinguishing between those technologies which the Appraisal Committees would be highly likely to recommend, and those technologies which appear to be less attractive, although further external validation is warranted.

Conclusion

The modelling suggests that cost-effectiveness, uncertainty and certain equity concerns influence the NICE Appraisal Committees' recommendations on the use of health technologies. The modelling results appear to support Rawlins and Culyer's notion of a probabilistic cost-effectiveness threshold approach; the "mythical" £30,000 per QALY gained threshold assumed within the literature is not supported by this stated preference modelling analysis.

Acknowledgements

I would like to thank Dr Carole Longson, Professor David Barnett, Ms Emily Marschke, Professor Andrew Stevens and Professor Jeremy Wyatt at NICE, for making this study possible. I would also like to thank all the members of NICE's Appraisal Committees who participated in the choice experiment. In addition, I would like to acknowledge the Health Economics and Decision Science (HEDS) Section within ScHARR who participated in the pilot study used to inform the design of the choice questionnaire. I would also like to acknowledge the input of the other members of the ScHARR study group (Professor John Brazier and Dr Julie Ratcliffe) who provided useful advice to inform the scope and design of the choice experiment. Finally, I would like to acknowledge the technical input of Dr Terry Flynn at the Medical Research Council Health Services Research Collaboration, and Dr Stephen Walters from the HEDS Section within ScHARR.

Contents

List of tables

List of figures

Abbreviations

BLM	Binary logit model
CEAcc	Cost-effectiveness Acceptability Curve
CCOHTA	Canadian Co-ordinating Office for Health Technology Assessment
EUT	Expected Utility Theory
HEDS	Health Economics and Decision Science
HRQoL	Health-related quality of life
HSM	Health Status Measure
ICER	Incremental cost-effectiveness ratio
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
PBAC	Pharmaceutical Benefits Advisory Committee
PSS	Personal and Social Services
QALY	Quality adjusted life year
RUT	Random Utility Theory
ScHARR	School of Health and Related Research
SMC	Scottish Medicines Consortium
UREC	University Research Ethics Committee

Chapter 1 Introduction

1.1 Background

The National Institute for Clinical Excellence was established as a Special Health Authority in 1999 following the publication of the Government White Paper *"The new NHS modern, dependable."*[1] The Institute's successor, the National Institute for Health and Clinical Excellence (NICE), is an independent organisation responsible for providing national guidance on the promotion of good health and the prevention and treatment of ill health in England and Wales.[2] One of the Institute's main functions is the appraisal of new and existing health technologies. Recommendations to the National Health Service (NHS) concerning the use of specific health technologies are prepared by one of three independent Appraisal Committees. Each Appraisal Committee consists of health professionals working in the NHS and people who are familiar with the issues affecting patients and carers.[2]

During the appraisal of each health technology, the Appraisal Committee is presented with evidence relating to the epidemiology and pathophysiology of the condition for which the technology is indicated, estimates of the clinical effectiveness and cost-effectiveness of the intervention compared to current standard treatments, the degree of uncertainty surrounding mean costs and effects, the characteristics of the relevant patient group who will benefit from the intervention, and the position of the technology within current clinical pathways.

Despite the Institute's recent efforts to clarify the way in which NICE's Appraisal Committees reach decisions concerning the recommendation for or against the use of health technologies, in particular, the role of scientific and social value judgements,[3-5] there remains ambiguity concerning how cost-effectiveness evidence is interpreted alongside other attributes such as the degree of clinical need within the patient population, and the degree of uncertainty surrounding cost-effectiveness estimates.

1.2 Study aims and objectives

The main aim of this study is to address the question: *"Does the National Institute for Health and Clinical Excellence take account of factors such as uncertainty and equity as well as incremental cost-effectiveness in commissioning health care services?"*

More specifically, the objectives of the study are:
(1) To explore current literature concerning the way in which NICE uses evidence on uncertainty and equity considerations alongside cost-effectiveness when making recommendations concerning the use of health technologies;
(2) To undertake a binary choice experiment to examine whether NICE takes into account uncertainty, age, baseline health-related quality of life (HRQoL) and the availability of other therapies alongside cost-effectiveness in commissioning health technologies;
(3) To estimate the relative importance placed on various programme attributes by NICE's Appraisal Committees.

1.3 Research design

The primary research element of the study takes the form of a binary choice experiment. The preferences of members of NICE's Appraisal Committees were elicited within a controlled experiment using five programme attributes:

(1) Incremental cost-effectiveness
(2) Degree of economic uncertainty
(3) Age of the target population
(4) Baseline HRQoL of the target population
(5) Availability of other therapies

A choice questionnaire detailing 18 scenarios was administered to each of the three NICE Appraisal Committees. For each choice scenario, respondents were asked to indicate whether they would be likely to recommend the intervention under consideration or not. The choice experiment explores *whether* and *how* other programme attributes such as uncertainty and equity considerations are traded off against cost-effectiveness when making health care commissioning decisions.

1.4 Structure of report

Chapter 2 presents a review of some existing literature relating to the decision-making criteria employed by NICE's Appraisal Committees. Evidence supporting and opposing the existence of a cost-effectiveness threshold for the adoption/rejection of health technologies is discussed. Evidence relating to the interpretation and potential influence of other attributes such as uncertainty, equity and the role of social value judgements is also presented. Gaps in current knowledge concerning the NICE Appraisal Committees' decision-making criteria are highlighted.

Chapter 3 reports the methods for the design and analysis of the binary choice experiment. The chapter outlines the identification and selection of attributes and levels used within the choice experiment. Issues surrounding sample size and sampling designs are highlighted within the chapter. Methods for random effects logit and probit regression analysis using the elicited stated preference data are detailed. An assessment of internal and external validity is also reported. Trade-offs between ideal and practical experiment designs are discussed.

Chapter 4 details the results of the binary choice experiment analysed using regression modelling. Sensitivity analysis is reported using a subset of choice scenarios included in the questionnaire. The results of the assessment of logical consistency are detailed. An assessment of external validity using a limited sample of revealed preference data from previous NICE appraisals is also presented.

Chapter 5 presents a discussion of the results of the study in the light of previous research concerning NICE's decision-making criteria. Limitations of the study are discussed. Areas in which further research is merited are highlighted.

Chapter 6 presents the conclusions of the study.

Appendix 1 presents the choice questionnaire employed within the experiment.
Appendix 2 contains the study outline and process presented to the NICE Appraisals Committees.
Appendix 3 presents the letter of approval obtained from the University of Sheffield Research Ethics Committee.

Chapter 2 Review

2.1 Introduction

The NHS was founded on three broad principles: that health care advice, treatment and care should be available to all who need it; that patients shall receive the best medical and other facilities available; and that individuals should receive these irrespective of whether they can pay for them.[6] From these principles, at least two key objectives can be identified: efficiency and equity. Broadly speaking, the former draws from utilitarianism principles of the greatest good for the greatest number of people; technical and allocative efficiency could be realised through the maximisation of health gains across a population. The latter is a more difficult concept to define, drawing from principles of justice and fairness, and concerns the fair distribution of costs and consequences within and across socioeconomic groups. It follows then, that decision-making within the NHS should be consistently aimed at the achievement of both these key objectives. However, one cannot escape the fact that health resources are scarce and humans are increasingly demanding.[7] In practice, the maximisation of health gains may be at the expense of the equitable distribution of the costs and benefits, or alternatively, the achievement of a fair distribution of costs and health gains across various population groups may be at the expense of health gain maximisation.

How then, should decisions be made concerning the allocation of scarce health care resources, whereby health gains are maximised *and* distributed in a fair and just manner? The existence of multiple and potentially conflicting objectives within the constraint of scarce health care resources means that NHS prioritisation decisions are invariably complex. Consequently, the NHS and other health care systems require some rational, transparent and explicitly justified approach to the prioritisation and allocation of health care resources.[8;9] This review explores the factors that may influence how decisions concerning the prioritisation of health technologies are currently made.

2.2 Background to NICE

NICE was established in 1999 to provide patients, health professionals and the public with authoritative, robust and reliable guidance on current best clinical practice.[2] NICE produces national guidance in three areas of health:

(1) Public health. NICE produces guidance on the promotion of good health and the prevention of ill health for those working in the NHS, local authorities and the wider public and voluntary sector;
(2) Clinical practice. NICE produces guidance on the appropriate treatment and care of people with specific diseases and conditions within the NHS;
(3) Health technologies (pharmaceuticals, medical devices, diagnostic techniques, surgical procedures, other therapeutic interventions and health promotion activities). NICE produces guidance on the use of new and existing medicines, treatments and procedures within the NHS.[2] Through the establishment of national guidelines based on current best available evidence, NICE attempts to eliminate the use of ineffective treatment approaches and release resources for the equitable provision of cost-effective health care interventions.[10]

2.3 The Technology Appraisal Process

This review focuses on those decisions concerning the prioritisation of health technologies. NICE's guidance on the use of technologies is based upon the formal appraisal of scientific and non-scientific evidence on the use of health technologies. Technologies to be appraised are formally referred to NICE by the Secretary of State for Health and the Welsh Assembly Government. Following each appraisal, NICE issues guidance on the use of the technology to the NHS in England and Wales. Upon its establishment, the Institute's recommendations were advisory, however, since 2002 the Secretary of State has directed that the NHS is required to provide funding and resources for medicines and treatments recommended by NICE through its technology appraisals work programme. The NHS normally has three months from the date of publication of each guidance document to provide such funding and resources.[11] At the time of writing, NICE had published guidance following 92 health technology appraisals.

Technology appraisals consider evidence relating to the expected health benefits and costs resulting from the use of individual or groups of health technologies in usual clinical practice. Health benefits typically include impacts upon mortality and HRQoL, measured in terms of quality adjusted life years (QALYs). Expected costs of the technology relate to those directly incurred by the NHS and Personal Social Services (PSS).

The Institute considers scientific evidence from an independent assessment of the clinical and cost-effectiveness of the technology, together with submissions from the technology's manufacturer and other consultees such as clinical experts and patient representative groups. Non-scientific evidence such as anecdotal patient accounts is also considered.[12] Following consideration of all relevant available evidence, the Appraisal Committee reaches a judgement as to whether, on balance, the technology can be recommended as a cost-effective use of NHS resources in general, or whether it can be recommended for specific indications or subgroups of patients.[12] The Appraisal Committee then submits an "Appraisal Determination" to the Institute, which forms the basis of NICE's guidance on the use of the technology.

In reaching decisions on whether to recommend health technologies, the Institute and its Appraisal Committees take into account the factors listed in the directions of the Secretary of State for Health and the Welsh Assembly Government,[12] namely:

- the broad clinical priorities of the Secretary of State for Health and the Welsh Assembly Government (e.g. as set out in National Service Frameworks);
- the degree of clinical need of the patients with the condition under consideration;
- the broad balance of benefits and costs;
- any guidance from the Secretary of State for Health and the Welsh Assembly Government on the resources likely to be available and on such other matters as they think fit;
- the effective use of available resources.

The Institute also takes into account the longer-term interests of the NHS in encouraging innovation in technologies that will benefit patients.

2.4 The NICE Appraisal Committees

The Institute currently has three Appraisal Committees, consisting of a total of 81 members. These committees are independent advisory bodies that make recommendations concerning the clinical and cost-effectiveness of treatments for use within the NHS, as well as recommending against those therapies whereby the benefits to patients are either unproven or not cost-effective. The Appraisal Committee's

judgements on clinical effectiveness take account of the nature and quality of several sources of evidence. These include the clinical analysis undertaken by the Assessment Group and the manufacturers of the technology, as well as the views of specialists, patients and their representative groups.[13] The Appraisal Committee also consider uncertainty relating to differences in the effectiveness of the technology between clinical evidence submitted for licensing purposes and usual clinical practice, differential effectiveness or risks of adverse events for specific patient groups, the position of the technology in the overall pathway of care and whether alternative effective treatments are available.[13]

However, the clinical effectiveness of a technology is a necessary but not sufficient condition for the Appraisal Committee to recommend its routine use on the NHS; the Appraisal Committees are also required to consider whether the technology represents good value for money. According to recent methodological guidance,[13] the Committees' judgements on cost-effectiveness include consideration of economic models submitted by both the Assessment Group and the manufacturers of the technology, including their structure, the plausibility of assumptions, the evidence inputs and outputs, and the Committee's preferred modelling approach. During their deliberations, the Appraisal Committee determines the central estimate of cost-effectiveness and whether they consider this to be acceptable, and reviews evidence relating to the uncertainty surrounding mean costs and health effects.[13] The Appraisal Committees do not consider the affordability of the technology, as this lies within the remit of the Department of Health, although consideration is given to how its advice may enable the more efficient use of available health care resources.[13]

A central principle underlying NICE's decision-making approach is that best practice can be determined in a way which is applicable across the whole system.[10] Public documents concerning how NICE conducts health technology appraisals suggest that a broad range of clinical and economic issues are consistently considered when deciding whether or not to recommend technologies for use on the NHS. The central issue does not concern whether the relevant attributes of technologies are considered, but rather *how* such information should be addressed within a consistent, rational framework in order to satisfy the objectives of the NHS. The credibility of NICE's guidance is dependent on the transparency of the Appraisal Committee's decision-making process.

It is crucial that the Appraisal Committee's decisions are consistent across the broad range of health technology appraisals undertaken, and that the views of consultees to the appraisal are taken into account. Only through the application of a coherent and explicit approach can NICE successfully achieve the central objectives of the NHS.

2.5 The use of health economic evidence to prioritise health care resources

2.5.1 The role of economic evaluation in the prioritisation of health care resources

The application of economic evaluation to clinical procedures and health technologies to directly or indirectly inform resource allocation decisions is not new. Economic evaluations of health interventions have been used to inform decision-making since the 1960s.[10] Indeed, NICE's use of economic criterion to inform health resource allocation decisions is not unique; in other countries, such economic evaluations are used for a variety of decision-making purposes. A number of governments including those of Australia, New Zealand, Finland, Norway, and the Canadian provinces of Ontario and British Columbia now require the submission of economic evidence when making recommendations for the purchase of pharmaceuticals.[14] In Portugal, Denmark and France, pharmaceutical companies may be asked to submit health economic evidence when a reimbursement decision is to be made, although such power is used only on a discretionary basis.[14] The role of such economic evidence in government decision-making includes informing *"yes/no"* reimbursement decisions, exerting downward pressures on product prices, and establishing price-volume agreements.[14]

Few would argue against the relevance of economics in health care commissioning decisions. An editorial published in the New England Journal of Medicine argued that a physician who allows cost issues to affect their decision-making has *"embarked on the 'slippery slope' of compromised ethics and waffled priorities."*[15] However, owing to the inevitable scarcity of health care resources, failure to consider cost implications of treatments may lower quality of care and harm the health of patients.[16] It follows then, that resource allocation decisions that are made without the consideration of opportunity cost may be ethically irresponsible.

2.5.2 Theoretical use of economic evidence to prioritise health care resources

Economic evaluation provides one means of informing difficult decisions about which services treatments should receive priority. Budget-holders can use cost-effectiveness analyses to maximise health gains from the allocation of limited resources.[10] Cost-effectiveness evidence has been considered within the vast majority of technology appraisals undertaken by NICE. Of the 92 health technology appraisal guidance documents currently available on the Institute's website, 68 guidance documents report cost-effectiveness estimates or the use of a cost-effectiveness model, 55 of which were undertaken within a cost-utility framework (this excludes the 15 appraisal guidance reports which have become obsolete). Notably, the use of cost-effectiveness analysis and cost-utility analysis to inform NICE technology appraisals appears to be increasing over time. However, there remains ambiguity concerning how cost-effectiveness evidence should be used to inform prioritisation decisions. Whilst the methods for undertaking cost-effectiveness analysis are well developed, efforts expended in ensuring the appropriateness of health care decision-making have been relatively limited.

If the fundamental premise of health technology assessment is to maximise the total aggregate health benefits conferred for any given level of available resources,[17] one is led directly to a mathematical optimisation solution: the maximisation of health benefits subject to the financial and resource constraints of a fixed budget. This optimisation problem highlights three relevant interrelated concepts: (1) the total health budget expenditure, (2) the cost-effectiveness acceptability threshold (or the amount of money society is assumed to be willing to pay for a given level of health benefit); and (3) the overall system efficiency.

Overall system efficiency could be optimised through two alternative approaches, depending on whether the budget constraint is fixed or variable. Assuming a fixed budget constraint, listing all technologies within a "league table",[17;18] and purchasing all interventions in order of their declining cost-effectiveness until funds have been exhausted will maximise the aggregate health benefits achieved and hence maximise the overall system efficiency. Under this system, the cost-effectiveness acceptability threshold is allowed to vary, and is thus defined by the cost-effectiveness of the last intervention purchased. Alternatively, given a fixed cost-effectiveness acceptability

threshold constraint, purchasing only those interventions that demonstrate an incremental cost-effectiveness better than the threshold will minimise the total expenditure and thus maximise the overall system efficiency. The necessary implication is that the partial implementation of both approaches, fixing both the budget and threshold independently and without respect to each other, will inevitably lead to overall system inefficiency, and in turn, the achievement of an inefficient distribution of health care resources.

Whilst the process of priority-setting across a spectrum of alternative programmes using the league table approach[18;19] described above appears intuitively attractive and has some proponents,[20] it is subject to several important problems. Information on the costs and effects of health interventions are imperfect, and the construction of a league table for all health interventions is currently unrealistic.[18;19] In addition, the usefulness of league tables is severely limited by the non-comparability of methods used to derive cost-effectiveness estimates, together with heterogeneity in methods for the valuation of HRQoL and costs. NICE has recently defined a "Reference Case" for cost-effectiveness analyses; this specifies the methods which are currently considered by the Institute to be the most appropriate for decision-making purposes and which are believed to be consistent with the NHS's objective of maximising health gains from limited resources.[13] Whilst the establishment of NICE's Reference Case may help the Institute to make consistent decisions across different technologies and health conditions, at present this does not represent the complete implementation of a formal cost-effectiveness league table approach.

A further problem of the league table approach is that estimates of incremental cost-effectiveness may be influenced by the interests of those undertaking the analysis. A recent review of cost-effectiveness analyses undertaken for the purposes of NICE Appraisals found that incremental cost-effectiveness ratios submitted by the manufacturers of technologies under appraisal were, on average, statistically significantly lower (more favourable) than those submitted by the independent Assessment Groups.[21]

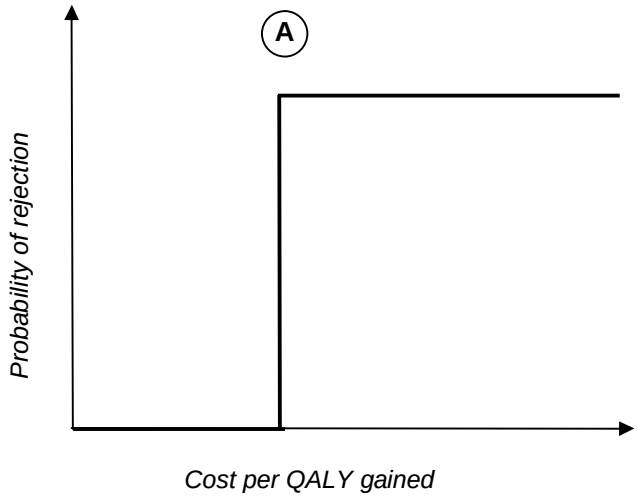
Even if an all-encompassing cost-effectiveness league table were possible, and potential biases and heterogeneities could be suppressed or minimised, the resulting resource

allocation decisions would not take account of who receives the health benefits or how they are distributed across society.

2.5.3 Interpretation of cost-effectiveness evidence – does NICE operate a cost-effectiveness threshold?

NICE's decision-making has been criticised on account of the Institute's ignorance of the principle of opportunity cost.[22] It has further been suggested that *"if economics is not established as the underlying decision making framework for NICE, the credibility of microeconomic evaluation in health care may be seriously disabled."*.[10] If transparency in decision-making is desired, it is crucial for NICE to be explicit about the way in which it uses cost-effectiveness evidence. Indeed, decisions concerning whether technologies represent good value for money require comparisons with some benchmark or threshold, as the notion of "cost-effectiveness" cannot exist in isolation. In line with theoretical approaches to resource allocation on the basis of economic evaluation, it has been suggested that NICE operates a cost-effectiveness threshold, whereby interventions that demonstrate an incremental cost-effectiveness that is better than the threshold are accepted, whilst those interventions that are less cost-effective than the thresholds are rejected, as shown in Figure 1.

Figure 1 *Fixed threshold approach to decision-making (based on Devlin and Parkin[23])*



A recent review of the Institute's technology appraisals by the World Health Organisation (WHO) stated that if a threshold is to be used as the basis for NICE's recommendations, this needs to be specified and justified for reasons of transparency.[24] It is important that NICE is clear about how other objectives should be recognised and traded off against evidence on cost-effectiveness within their decision-making process. Since its inception, much of the discussion surrounding NICE's decision-making criteria has focussed on whether a threshold exists, what value that threshold takes, and what value the threshold should take. Considerably less attention has been placed on how NICE actually interprets cost-effectiveness evidence alongside other considerations in order to satisfy the multiple objectives of the NHS.

2.5.4 Empirical evidence for the existence of a threshold

NICE's formal position is that the Appraisal Committee does not use an "immutable" cost-effectiveness threshold.[3-5;13;25] However, comments from NICE on the existence of a cost-effectiveness threshold appear to be contradictory. Public comments made in 2001 by Sir Michael Rawlins, NICE's Chairman, indicated that a threshold of £30,000 per QALY gained had emerged from the Committee's deliberations, although the Institute subsequently submitted evidence to the House of Commons Health Select Committee claiming that comments relating to the existence of the threshold were misinterpreted.[26] In further support of the operation of a formal cost-effectiveness threshold, recent NICE guidance on the use of Orlistat (Xenical®, Roche Pharmaceuticals) for the treatment of obesity in adults stated that a *"sufficient level of cost-effectiveness"* is *"in the range of a cost per QALY gained of between £20,000 and £30,000."*[27]

In 2001, Raftery undertook review of NICE guidance on 22 technologies according to the three criteria initially outlined in NICE's requirements for submissions of evidence, namely, cost per QALY, clinical benefits, NHS budget impact.[28] Raftery commented on the difficulty in establishing how the balance between clinical benefits and economics had influenced NICE recommendations.[28] A cost per QALY estimate was cited within 50% of guidance documents. Raftery notes that whilst a cost per QALY estimate may have been available from the manufacturers or Assessment Group within the remaining 50% of appraisals, the Appraisal Committee's decision not to cite a cost-effectiveness estimate suggests that they did not find available economic evidence convincing, thus

the basis of their recommendations was unclear.[28] Of those guidance documents in which a cost per QALY was cited, all but one technology (Riluzole, Rilutek®, Aventis Pharma) had a cost per QALY below £30,000 and subsequently received positive recommendations from the Committee (although Riluzole was apparently recommended on the basis of *"the value patients place on tracheostomy free survival time"*[29]). Raftery's review highlighted that NICE have generally said *"yes, but…"*, whereby positive recommendations were accompanied by specified conditions for the use of technologies.

In 2004, Devlin and Parkin undertook a binary choice analysis using retrospective NICE guidance reports to explore how the characteristics of health technologies operate to influence NICE's decision-making, and to establish the characteristics of NICE's cost-effectiveness threshold.[23] The authors abstracted data from guidance reports on 39 technologies which corresponded to 51 "yes/no" decisions. Data were evaluated using a binary logistic regression model in order to explore how cost-effectiveness, uncertainty, burden of disease, and the existence of alternative therapies influence the decision to recommend for or against individual health technologies. Devlin and Parkin's[23] analysis suggested that NICE's decisions were well explained by incremental cost-effectiveness, with uncertainty and burden of disease explaining some of the rejection decisions where the incremental cost-effectiveness ratio was low, and acceptance decisions where the incremental cost-effectiveness ratio was high.[23] The authors indicated that actual cost-effectiveness threshold used by NICE appears to be somewhat higher than the suggested range of £20,000 to £30,000 per QALY gained.[23]

However, Devlin and Parkin's study[23] was subject to several important limitations. Firstly, the scope of the logistic regression model was narrow and included only a subset of the factors considered by the NICE's Appraisal Committees. Furthermore, missing data items led to the exclusion of 18 of NICE's decisions from the analysis. Further limiting factors included the absence of any distinction between the denominator for the cost-effectiveness ratio (the analysis did not make a distinction between cost per QALY gained and cost per life year gained), and the weak representation of uncertainty surrounding costs and effects (defined as a cost-effectiveness range divided by the central estimate of cost-effectiveness). In addition, where multiple cost-effectiveness estimates were available, the mean was assumed; this

may not reflect the actual central estimate of cost-effectiveness agreed upon by the Appraisal Committee. Most importantly however, the analysis included only factors which appear to relate to concepts of efficiency, and did not include any consideration of equity factors such as the age of the target population or the severity of the disease. Owing to these problems, the interpretation of Devlin and Parkin's results is problematic.

It should be brought to light that if a cost-effectiveness threshold is in operation, the origin of the value of that threshold remains unclear. As noted in Section 2.5.2, the establishment of cost-effectiveness threshold either requires judgements or evidence concerning the intrinsic monetary value of a QALY, and subsequent adjustments to the budget made available for NHS expenditure, or to allow the value of a marginal QALY to emerge from the priority-setting activities of budget-holders based upon a fixed NHS budget.[30] The appropriateness of the £30,000 per QALY gained threshold assumed within the literature, which does not appear to have emerged from either of the approaches described above, is questionable (or at least it would be if the existence of the threshold could be confirmed). An alternative approach suggested by Williams[30] is the adoption of a lower cost-effectiveness threshold of around £18,000 per QALY, which represents the estimated current per capita Gross Domestic Product.[30] However, Rawlins argues that such suggestions rely on judgements which carry no more or less authority than the collective view of the Institute's economic advisors.[5]

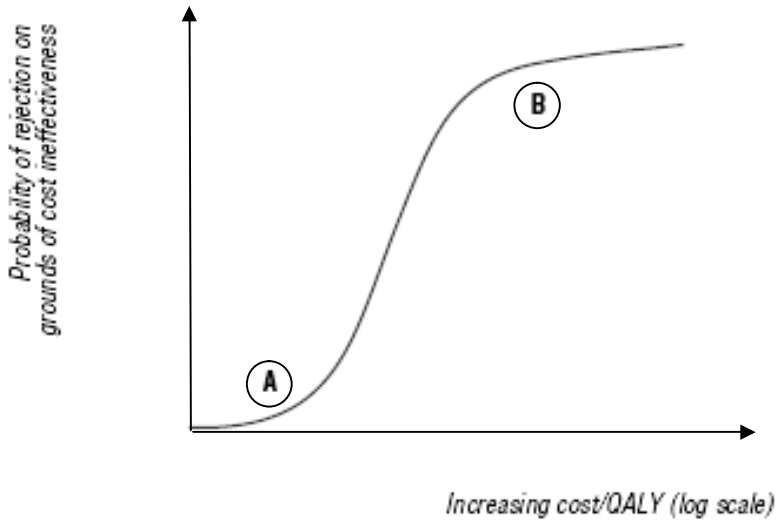
2.5.5 Empirical evidence against the existence of a threshold

Despite some reasonably compelling empirical evidence supporting the existence of a cost-effectiveness threshold, or at least an acceptable range for cost-effectiveness, NICE has repeatedly insisted that they do not operate such a threshold. Rawlins and Culyer[3] recently attempted to clarify the way in which NICE uses cost-effectiveness evidence together with scientific and social value judgements, citing four reasons why NICE rejects the threshold framework:

(1) There is no empirical basis for deciding at what value a threshold should be set;
(2) There may be circumstances in which NICE would want to ignore a threshold;
(3) To set a threshold would imply that efficiency has absolute priority over other objectives such as fairness;
(4) Many of the technology supply industries are monopolies, thus a threshold would discourage price competition.

According to Rawlins and Culyer,[3] the Institute makes decisions concerning the recommendation for or against the use of individual health technologies on a "case-by-case" basis,[3] as described by a probabilistic S-shaped acceptance/rejection curve (See Figure 2). Based upon this curve, as the incremental cost-effectiveness of the technology increases, so too does the probability of rejection.

Figure 2 NICE's empirical acceptance/rejection curve[3]



The key points on the rejection curve are the inflexions at points A and B. According to Rawlins and Culyer, *"NICE and its advisory bodies have taken the view that inflexion A occurs at around £5,000-£15,000/QALY and inflexion B at around £25,000-£35,000/QALY. NICE would be unlikely to reject a technology with a ratio in the range of £5,000-£15,000 solely on the grounds of cost-ineffectiveness, but would need special reasons for accepting technologies with ratios over £25,000-£35,000/QALY as cost effective."*[3]

According to Rawlins and Culyer,[3] the main considerations in making judgements about whether health technologies with incremental cost-effectiveness ratios between £25,000-£35,000 per QALY gained are: the degree of uncertainty; the particular features of the condition and population using the technology (including the availability and clinical effectiveness of other therapies, public health issues such as communicable diseases, and special equity considerations); when appropriate, the wider societal costs and benefits; and when appropriate, reference to previous appraisals.[3]

Further clarity on the relative importance of the characteristics of health technologies is provided within methodological guidance recently published by the Institute,[13] which suggests that for interventions with a "plausible" incremental cost per QALY below £20,000, judgements about the acceptability of the technology as an effective use of NHS resources are based primarily on the cost-effectiveness estimate.[13] For interventions with an incremental cost per QALY greater than £20,000, judgements about the acceptability of the technology as an effective use of NHS resources are more likely to make more explicit reference to factors such as the degree of economic uncertainty, the innovative nature of the technology and the particular features of the condition and population receiving the intervention.[13] Above an incremental cost-effectiveness ratio (ICER) of £30,000 per QALY gained, the case for supporting the technology on these factors is required to be increasingly strong.[13]

It is interesting to note that whilst the Institute has repeatedly dismissed the existence of a cost-effectiveness threshold as an urban myth,[3;13;25;26] the rejection curve shown in Figure 2 indicates that two approximate, rather than absolute, thresholds are in operation; one for acceptance and one for rejection. Whilst Rawlins and Culyer[3] have provided some clarity concerning how NICE interprets cost-effectiveness evidence for interventions which have an incremental cost per QALY gained which is either better than inflexion A, or worse than inflexion B, the way in which other factors may influence adoption decisions between A and B remains unclear.

2.6 Influence of uncertainty surrounding incremental costs and effects

The costs and health gains resulting from the use of health interventions are subject to ubiquitous uncertainty. In recent years, considerable methodological work has been undertaken to resolve methods for handling uncertainty, with an increasing tendency to move away from simplistic one-way sensitivity analyses towards more sophisticated descriptions of uncertainty such as cost-effectiveness planes, net benefit distributions and cost-effectiveness acceptability curves (CEAcc).[31] These approaches provide information on the likelihood that each treatment option is optimal.

The decision-maker's response to uncertainty may affect their decision to accept or reject a health technology. For example, a technology with a low mean cost per QALY

may appear less attractive if it has a high probability of being dominated by an alternative therapy. Claxton[32] argues that only expected (mean) incremental cost-effectiveness estimates should be used within health care commissioning decisions. Claxton argues that information on uncertainty surrounding expected costs and effects should be used to make decisions concerning the prioritisation and planning of further research (to reduce such uncertainty).[32;33] However, the Institute's documentation on how the Appraisal Committees make decisions does not appear to be consistent with this viewpoint.[13] The degree to which uncertainty surrounding expected costs and effects may influence adoption decisions made by the Appraisal Committee is currently unclear.

2.7 The role of social value judgements and equity in health resource prioritisation

The empirical evidence, public conjecture and formal statements from NICE indicate a lack of clarity concerning how the Institute incorporates other factors besides cost-effectiveness into its decision-making processes, particularly with regard to those technologies which are less attractive in terms of cost-effectiveness than other health technologies currently funded by the NHS. Indeed, Sir Michael Rawlins has admitted that the Institute's decisions are *"questions of judgement, difficult to defend, difficult to teach,"*[34] and that *"all decisions in the future will be based on difficult judgements"* which have *"no mathematical quantitative approach."*[34] To date, the Institute has published three papers describing how social value judgements are incorporated into their decision-making processes.[3-5] The broad message within all three papers is that social value judgements must play a critical role if resources are to be distributed equitably as well as efficiently.[3-5] Whilst the objective of distributing health gains in a fair and just manner is unambiguous, the route to achieving this objective and the weight that should be placed upon it are unclear.

The social value judgements made by the Appraisal Committees relate to their beliefs concerning what is good or bad for society.[3] The fundamental value judgement underpinning NICE is that its advice to the NHS should embody values that are generally held by the population that the NHS serves.[3] Undoubtedly, equity lies at the heart of the NHS; one of the principal reasons for the establishment of NICE was to eradicate regional problems of "postcode prescribing."[35] NICE has recently formed a

Citizen's Council to ensure that the values of society are served. However, as Culyer notes, much of the philosophical literature on equity may not be applicable in the real world, and due to conflicts between alternative equity perspectives, there may be no single overriding equity principle to guide the distribution of health care resources.[36;37]

Rawlins and Culyer suggest that an assumption underlying most of the Institute's recommendations is that *"a QALY is a QALY is a QALY"* irrespective of how may QALYs have already been enjoyed, how many are in prospect, age, sex, deservedness, and deprivation.[3] This assumption essentially corresponds to the sum-ranking (utilitarian) approach whereby choices between alternative distributions are determined by the amount of well-being generated by each one. Under this principle, the individual's need for health care is defined by their capacity to benefit from that care i.e. the greater an individual's expected benefit, the greater their need for health care.[38] In this sense, an individual's pre- and post-intervention levels of well-being have no moral relevance, as only the change in well-being is considered to be important.[38] However, there may be instances whereby NICE may wish put greater priority on objectives other than the size of expected health gains,[3] such as how those health gains are distributed. For example, under Rawls' "difference principle" (more commonly referred to as the 'Maximin' principle), *"social and economic inequalities must be to the greatest benefit of the least advantaged."*[39] This differs from the sum-ranking approach, as an individual's need for health care is defined according to their severity of illness, hence resources would be devoted to the most severely ill individual.[38] Other approaches to distributive justice have been suggested, for example "strong egalitarianism", although these are seldom considered reasonable within a health policy context.

The list of potential equity concerns faced by NICE is substantial. For example, greater emphasis may be placed on the aged, the infirm or the vulnerable, who may be discriminated against under the QALY maximisation principle.[40] Conversely, a lesser weight may be placed on the aged population based on the number of QALYs already enjoyed (the so-called "fair innings" argument[41]) and the fewer QALYs they have in prospect. Priority may be given to those interventions which are aimed at individuals with a poor prognosis over those therapies aimed at individuals who enjoy a comparatively high quality of life.[42] Perhaps, differential weightings should be applied depending on the size of the expected treatment effect. Alternatively, the Institute may

place greater weight on those therapies which are life-saving rather than those which merely improve a patient's quality of life, an imperative coined by Jonsen as the "Rule of Rescue."[43] An additional consideration may be whether the patient is in some way responsible for their condition.[44;45] A further practical issue is that the Institute may wish to place differential emphasis on interventions used to treat conditions for which there is no alternative effective therapy.[45]

The list of potential equity concerns presented above is by no means exhaustive. Research concerning methods for weighting QALYs for different groups of individuals is currently underway.[46] The way in which the Institute balances these equity considerations alongside cost-effectiveness evidence is not currently clear, and as such, may be difficult to defend. As Birch and Gafni critically note, "*it implies that the Committee will consider recommending certain technologies be available for identified social groups (those for whom the treatment is 'cost-effective'), but not for others (those for whom the technology is not cost effective and who are not sufficiently deserving in the eyes of the Committee to receive the equity weighting required to make it cost effective). This seems less scientific and more Orwellian than one might expect for a National Institute for Clinical Excellence.*"[22] Consequently, decisions to accept or reject health technologies on the grounds of equity may be considered as a "cop-out." The Institute has gone some way towards outlining the types of social value judgements that its Appraisal Committees make,[3-5] although a coherent framework addressing these alongside cost-effectiveness evidence has not yet been formalised.

2.8 Summary of current evidence on NICE prioritisation decisions

The literature outlined within this review suggests that NICE's Appraisal Committees consider a number of important factors when deciding whether to recommend health technologies for routine use on the NHS. Official documentation from the Institute indicates that decisions are not made on the basis of cost-effectiveness alone, but rather cost-effectiveness evidence is considered alongside evidence relating to clinical need, equity concerns, and decision uncertainty. However, whilst NICE appears to be considering those factors relevant to the realisation of both efficiency and equity objectives, the way in which such evidence is synthesised within a rational, consistent and coherent decision-making framework is not entirely clear. Current evidence suggests the existence of a probabilistic cost-effectiveness threshold approach, whereby

within a range of potentially acceptable cost-effectiveness, numerous alternative considerations may serve to determine whether the technology should be adopted or not. The key uncertainty in NICE's decision-making rationale appears to lie between these inflexion points; how does NICE trade-off alternative programme attributes when evidence on the cost-effectiveness of the technology is not altogether compelling?

Whilst NICE has identified a range of situations whereby special factors may override cost-effectiveness considerations alone, it remains unclear how the presence of these considerations may interact with the interpretation of cost-effectiveness evidence or how much emphasis is placed upon these by the Appraisal Committees. Several key issues remain unresolved: Is it possible to produce a broad framework of how equity concerns for a given technology should be addressed in the light of the technology's cost-effectiveness profile? Can the relative importance of programme attributes be ranked within a hierarchy? Clearly, if the Institute's Appraisal Committees wish to place a lesser weight on the cost-effectiveness of specific health technologies due to the presence of other special factors, it is crucial that such decisions are clearly justified within a consistent and transparent framework.

This review highlights that NICE has multiple objectives, although specific guidance on how the Institute balances each of these remains unclear. The estimation of equity weights based on the set of relevant characteristics of the beneficiaries is beyond the scope of this study and is currently being undertaken elsewhere.[46] Rather, the purpose of this study is to examine *how* efficiency and equity objectives are currently being weighted and to examine whether NICE's decision-making approach is consistent.

Chapter 3 Methods

3.1 Overview of methods

This chapter details the methods used to examine the decision-making criteria employed by NICE when commissioning health care interventions. The primary research element of this study takes the form of a binary choice experiment. The development of the choice questionnaire, including the identification and selection of attributes and levels included in the experiment are discussed. The generation of the orthogonal sampling design is detailed, and issues surrounding sample sizes are highlighted. The methods for modelling the results of the choice questionnaire within a random effects logit regression model are subsequently reported. Comparisons of stated preference and revealed preference data relating to NICE's decision-making criteria are also explored.

3.2 Alternative methods for examining NICE's decision-making criteria

Numerous alternative approaches could be used to explore how alternative attributes influence NICE's Appraisal Committees in deciding whether or not to recommend health care interventions for routine use on the NHS. These include reviewing current literature, qualitative methods, and quantitative methods.

3.2.1 Literature review

The review of current literature presented in Chapter 2 identified a diverse range of evidence which allows some inference concerning how NICE's Appraisal Committees determine whether health technologies should be recommended or not. However, public comments,[26;34] official guidance from NICE on the methods for Health Technology Appraisal,[13] and published NICE guidance on the use of specific health technologies[27] do not appear to present an entirely consistent approach to making decisions on the basis of cost-effectiveness evidence and other considerations. At the extremes of Rawlins and Culyer's adoption/rejection curve,[3] the Appraisal Committees' adoption decisions may be relatively easy to predict. Indeed, the Institute has afforded considerable effort in clarifying its decision-making approach,[3-5;13] particularly in terms of determining two approximate inflexion points for the respective adoption and rejection of health technologies. The literature however highlights considerable

ambiguity surrounding how equity considerations and uncertainty are balanced alongside cost-effectiveness evidence.

3.2.2 Interview-based methods

Qualitative interviews could potentially provide further information relating to how NICE's Appraisal Committees account for other considerations besides cost-effectiveness when making commissioning decisions. Whilst interview-based methods are a potentially feasible approach, they are unlikely to yield full, clear and accurate information on how alternative programme attributes are traded off with one another in practice. In addition, it is unlikely that interviews would provide data which are appropriate for quantitative analysis. From a practical perspective, it would be particularly difficult to obtain direct access to Appraisal Committee members for in-depth interviews.

3.2.3 Quantitative survey of attitudes

Attitudinal surveys represent an alternative approach to determining whether NICE's Appraisal Committees take factors beside cost-effectiveness into consideration when deciding whether to recommend health technologies. For example, a simple structured questionnaire approach could be administered to each member of the Committees asking questions such as *"Do you take the age of the patient group into consideration in your deliberations concerning whether to recommend health technologies?"* or *"Does a high level of parametric uncertainty influence your interpretation of the cost-effectiveness estimate for a given technology?"* The main problem with this approach is that it provides no information concerning the relative importance of alternative attributes, that is, the respondent's preferences for individual programme attributes. Trade-offs between programme attributes could be indirectly addressed through the use of a scoring system, for example, by asking survey respondents to rate the importance of a given programme attribute on a scale of 1-10.

3.2.4 Preference-based survey methods

Preference data (i.e. data on the choices made by individuals) can yield information concerning how specific programme attributes may be traded off against one another, and thus provide an estimate of their relative importance. The clear benefit of preference data as oppose to simple attitudinal survey data is that they provide a direct,

rather than indirect, means of evaluating preferences. Two types of preference exist: revealed preferences and stated preferences. Revealed preferences refer to the preferences of the respondent as revealed by the choices they make, whilst stated preferences may be inferred from the choices that individuals say they will make within a controlled experiment. Interviews may be effectively used alongside stated preference survey methods in order to appropriately define the attributes and their levels for the experiment.[47]

3.2.4.1 Revealed preference methods

Considerable revealed preference data from NICE's Appraisal Committees has been published within their guidance on the use of specific health technologies. NICE's guidance reports on the use of health technologies typically present the Institute's recommendations to the NHS on the use of the technology, together with details such as the Committee's preferred cost-effectiveness estimate(s), the expected degree of clinical benefit, and any additional considerations that were considered important during the appraisal. NICE's guidance reports represent a potentially rich, reliable and valid data source, however, the use of revealed preference data to estimate the relative importance of individual programme attributes is subject to some important problems:

(1) The factors detailed within guidance documents may not fully reflect all of the individual factors considered by the Committee at the time of the appraisal.

(2) The aggregation of all revealed preference data from guidance reports since the inception of NICE's Appraisal Committees does not allow for organisational learning through experience, and the evolution of the Committee's decision-making criteria over time (although this could be incorporated into a statistical choice model).

(3) The range of factors considered by the Committees is not consistently reported within NICE technology guidance reports. For example, Devlin and Parkin's revealed preference analysis (See Section 2.5.4) was restricted to only four basic attributes, for which full information was reported for only 33 of 51 decisions.[23] Consequently, the attributes included in the revealed preference model, and the number of observations (decisions) available, were constrained by the limited information reported with NICE guidance reports. As equity issues are not consistently reported within guidance documents, Devlin and Parkin's analysis was restricted only to efficiency-related attributes.

3.2.4.2 Stated preference methods

The evaluation of stated preference data using conjoint analysis is typically more flexible than its revealed preference counterpart due to the adoption of a controlled decision-making environment.[48] Stated preference discrete choice methods allow for the estimation of the relative importance of different attributes (that is characteristics or features) in the provision of a good or service.[49] Stated preference techniques have been commonly applied within market research, the environmental sector, the transport sector and health care.[47;49-60]

The main principle which underpins stated preference techniques is the assumption that an individual's preferences are reflected in their respective utility functions. The establishment of a controlled environment and the use of an experimental design enable the utility associated with alternative options to be estimated. Stated preference data may be collected using a simple survey design, whereby potential respondents are presented with a questionnaire consisting of a series of discrete or binary choices. For each choice scenario, individuals are asked to indicate their preferred option. Analysis of the response data using statistical models allows for the delineation of the total utility function such that the statistical contribution of each attribute to the explanation of a choice response, and hence the impact of the attribute on total utility, can be estimated thereby indicating its relative importance.[48]

Each of the respondent's choice can be seen as a comparison of two indirect utility functions and can be analysed within the framework of Random Utility Theory (RUT).[49] Expected Utility Theory[61] suggests that the respondent will choose scenario A over scenario B if the expected utility associated with Scenario B is greater than the expected utility associated with Scenario A.

$$U(B) > U(A) \qquad (3.1)$$

The probability that an individual chooses scenario A is a function of the difference in utility between the two scenarios *{U(B)-U(A)}*. The analysis of discrete or binary choice data within the framework of RUT requires the inclusion of an error term in the utility function to reflect the unobservable and immeasurable factors in the respondent's utility

function. This error term is necessary due to imperfect knowledge concerning the exact form of the utility function and the researcher's inability to recognise and empirically measure all factors which influence a respondent's choice.[47] Consequently, an individual will choose Scenario *B* over Scenario *A* if the measurable component of utility (V_B) plus the unobservable component of utility associated with scenario *B* (E_B) is greater than the measurable component of utility (V_A) plus the unobservable component of utility associated with scenario *A* (E_A).[47]

$$\textit{Choose } B \text{ if } (V_B + E_B) > (V_A + E_A) \qquad (3.2)$$

The measurable components of utility for each scenario (V_B and V_A) can be estimated directly from the response data. The unobservable utility components of utility (E_B and E_A) concerns the effect of those variables which are not included in the choice experiment but which are taken into account by the respondent in making actual choices (*ea*), and the errors specific to the conjoint analysis exercise (*ec*).[47]

$$\textit{Unobservable utility component } E = ea + ec \qquad (3.3)$$

Empirical evidence suggests stated preference surveys can produce data which are consistent with RUT, from which econometric models can be estimated which are indistinguishable from their revealed preference counterparts.[48] One of the key benefits of stated preference techniques is that they may be more flexible than revealed preference data, as the use of a controlled environment allows for the utility function for respondents to be mapped over a large number of scenarios which have not yet arisen. In addition, stated preference techniques can yield multiple observations for each respondent at each observation point.[48]

3.3 Study subjects – The NICE Appraisal Committees

The three NICE Appraisal Committees were identified as the relevant group of participants for the study, as these Committees make recommendations concerning the use of specific health technologies on the NHS. A key benefit of eliciting stated preference data from individual members of the Appraisal Committees is that response data reflect the views of the individual rather than the consensus view of the Committee as a whole; given the multidisciplinary background of the Committee members, this

could lead to important variations in the responses elicited. The main problem associated with eliciting responses from the Appraisal Committees concerns the limited sample size (81 members); previous simulation work suggests that more than 150 respondents may be necessary to protect against biased parameter estimates (*Personal communication, Dr Terry Flynn, Medical Research Council Health Services Research Collaboration, Bristol).* Inevitably, the experiment is likely to be underpowered, thus the results should be considered exploratory.

Two further important problems relating to the use of a stated preference approach to determine NICE's decision-making criteria should be highlighted *a priori*. A generic problem associated with stated preference techniques is that individuals do not always do what they say they will do in reality. Appraisal Committee members may state that they would recommend a technology with a given set of programme attributes within a controlled experiment, but may not make the same decision for a technology with similar characteristics in reality. Secondly, whilst the elicitation of individual preference data from a group of decision-makers may highlight variations in preferences between the individual Committee members, informational influences and normative influences on group decision-making,[62] such as the differential influence of individual Committee members on final recommendations, will not be captured.

3.4 Methods for the design and analysis of a stated preference binary choice experiment to explore NICE's decision-making criteria

3.4.1 Identification of key attributes

The scope of the analysis is crucial to its reliability and validity, as the attributes identified for inclusion in the experiment indirectly determine the quality of the results obtained.[47] Numerous programme attributes may be explicitly considered by health service decision-makers when commissioning health interventions, each of which could be included within a stated choice experiment. Furthermore, it is possible to describe the levels associated with many attributes in several different ways. Table 1 presents a non-exhaustive list of the types of attributes which may influence health care commissioning decisions, as identified within Chapter 2.

Table 1 Potential attributes which may influence health care commissioning decisions

Attribute	Attribute description	Example unit of measurement
1) Clinical effectiveness	Incremental clinical benefits resulting from the technology under consideration compared to the current standard treatment.	Life years gained QALYs gained Disease-specific outcomes
2) Nature of effect	Whether the intervention is life-saving and/or improves HRQoL	Life-saving/HRQoL-gaining
3) Cost-effectiveness/ cost-utility	Central estimate of incremental cost-effectiveness/cost-utility. Calculated as the incremental cost divided by the incremental benefits gained.	Cost per life year gained Cost per QALY gained Cost per event avoided etc.
4) Expected net benefit	Additional benefits associated with the technology under consideration valued in monetary terms after adjusting for cost consequences.	Pounds sterling
5) Baseline HRQoL	Pre-treatment health utility score.	Index ranging from 0 (“dead”) to 1.0 (“perfect health”). Negative scores may be used to represent health states worse than death, depending on Health Status Measure (HSM) used.
6) Cost-effectiveness/ decision uncertainty	Described by illustrative cost-effectiveness plane, cost-effectiveness acceptability curve, probability of being dominated, confidence intervals for cost-effectiveness ratio	Incremental costs and incremental life years/QALYs gained Cost per QALY range or confidence intervals Probability of being dominated Probability of being optimal at a given willingness to pay threshold
7) Age	Age of target population for the technology under consideration	Years/Years of life remaining
8) Responsibility for condition	Whether the individual is in some way responsible for their disease or condition (e.g. lung cancer resulting from smoking)	Patient is partially responsible for condition/Patient is not responsible for condition
9) Availability of other therapies	Whether effective alternative therapies are available for the treatment of the disease or whether the current standard treatment is best supportive care	Yes/No
10) Burden of disease	Prevalence of disease or condition	Number of patients
11) Affordability	Absolute/incremental annual cost to the NHS associated with providing the technology under consideration	Pounds sterling
12) Cost per patient treated	Absolute/incremental annual direct/indirect cost of providing the technology under consideration for an individual patient	Pounds sterling

Clearly, numerous alternative study designs are possible, depending on how many attributes are included within the experiment, how the attributes are described, and how many levels are used to describe each attribute. It should be noted from the outset that design and sampling issues for choice experiments should be considered in concert: good design will not compensate for inadequate sampling and vice versa.[48] Ideally, all independent attributes detailed in Table 1 would be included in the study, each of which would be described over a large number of levels. However, the number of attributes included within this experiment was restricted by a number of practical considerations.

(1) *Cognitive complexity*. Information Processing Theory[63] suggests that people can only process up to around seven pieces of information (plus or minus around two pieces of information). Owing to the potential cognitive complexity of the choice experiment, it was decided that no more than five attributes should be included within each choice scenario.

(2) *Task burden*. It was anticipated *a priori* that the increasing burden of the task, particularly in terms of time required, would reduce the response rate for the survey. The presentation of a large number of choice scenarios within each questionnaire may cause respondents to become bored and lead to respondent fatigue biases which may undermine the validity of the study. A full factorial design across the five attributes would require 108 scenarios, which is clearly infeasible (See Section 3.5). The study group decided that no more than 20 choice scenarios should be included within the experiment.

(3) I*ndependence of attributes*. In order to allow for the estimation of the utility associated with individual attributes, all attributes included within the utility model must be independent of one another, thus avoiding problems of multicollinearity. Interactions were excluded from the analysis, as their inclusion would increase the burden of the task. It should be noted that several of the attributes are simply alternative descriptions of the same dimension (for example expected net benefit and cost-effectiveness are closely related concepts), and some dimensions incorporate information on other dimensions (for example, age, the clinical effectiveness of the intervention, and the nature of the treatment effect should be captured within the cost per QALY estimate). Consequently, the number of independent programme attributes is reduced considerably.

The initial set of attributes and levels to be included in the analysis were selected by the study group (See Acknowledgements) using official NICE documentation describing the factors considered by the Appraisal Committees,[13] and through correspondence with a representative from NICE (*Dr Carole Longson, Appraisals Programme Director, NICE, London*). The attributes included in the experiment are detailed in Table 2.

Table 2 List of attributes and descriptions included in the choice experiment

Attribute	Attribute description
1) Incremental cost per QALY gained	The central estimate of cost-effectiveness for the intervention as compared to the current standard treatment. This is calculated as the additional cost of the intervention compared to the current standard treatment divided by the additional benefits of the intervention compared to the current standard treatment.
2) Uncertainty	The degree of uncertainty surrounding incremental costs and effects.
3) Age	The mean age of the population who will benefit from the intervention.
4) Baseline HRQoL	The mean health-related quality of life score of patients prior to receiving the intervention, whereby "1" represents a state of perfect health and "0" represents dead.
5) Availability of other therapies	Whether alternative effective therapies are available to manage the condition or not.

3.4.2 Discrete versus binary choices

For this particular experiment, the standard pairwise discrete choice design is inappropriate due to the inclusion of the cost-effectiveness attribute. The description of incremental cost-effectiveness using the pairwise discrete choice model means that the respondent is actually asked to compare four technologies: Intervention A, Intervention B as well as the relevant comparator for Intervention A and the relevant comparator for Intervention B. Not only would such an approach be cognitively difficult, but comparing cost-effectiveness ratios between interventions with different baselines is theoretically flawed. For example, consider Intervention A, which costs an additional £5,000 and provides an extra 2 QALYs compared to its comparator, and Intervention B which costs an additional £20,000 and provides an extra 4 QALYs over its comparator. On the basis of cost-effectiveness information alone, one would prefer Intervention A (A=£2,500 per QALY gained, B=£5,000 per QALY gained). However, if the value of each QALY is greater than £7,500, intervention B may be considered preferable on the basis of incremental net benefit. The net benefit approach would avoid the lack of sensitivity of the cost per QALY ratio, however, the Appraisals Committees do not

work in these terms (*Personal communication, Dr Carole Longson, Appraisals Programme Director, NICE, London*).

Perhaps more importantly, whilst the Institute's Appraisal Committees may draw from their experience during previous technology appraisals, the traditional discrete model is not representative of the Institute's "case by case" decision-making approach.[3] Consequently, the standard pairwise discrete choice model would have poor face validity, and would not reflect the way in which the Appraisal Committees actually make decisions in practice. In order to resolve these problems, a binary response scenario approach was adopted. Using the binary response format, Committee members were presented with a series of scenarios which presented information relating to the five attributes, and were asked to indicate whether they would recommend the intervention under consideration or not (See Appendix 1).

3.4.3 Identification of attribute levels

The levels assigned to each attribute, that is the values that each attribute may take, must be plausible to respondents and realisable in practice.[49] In addition, the levels assigned to each attribute should be capable of being traded off against one another; if changes in the levels of an individual attribute are considered by the respondent to be too small to be of importance, the attribute is likely to be ignored.[49] Whilst assigning a larger number of levels to each attribute would provide additional information, in practice, the desire for more information must be traded off with the complexity that arises from the inclusion of a large number of attributes and levels.[47] The levels assigned to each attribute within this choice experiment were initially selected by the study group with the input of Dr Carole Longson (*Appraisals Programme Director, NICE, London*).

In order to assess the validity of the questionnaire and to determine whether the selected attributes and levels were capable of being traded off against one another within a stated preference framework, a limited sample pilot study was undertaken within the Health Economics and Decision Science (HEDS) Section within the School of Health and Related Research (ScHARR). The HEDS Section was selected as the study population for the pilot, as the majority of staff members are familiar with the concepts commonly faced by NICE's Appraisal Committees. Two minor adjustments were made to the

levels assigned to the attributes based on the feedback from the pilot study. Firstly, the cost per QALY gained levels were changed to better reflect Rawlins and Culyer's[3] adoption/rejection curve (previously described as £20,000, £30,000 and £40,000 per QALY gained). Secondly, the description of uncertainty was changed due to problems in interpreting this attribute, and due to the presence of multicollinearity between the incremental cost per QALY gained attribute and the uncertainty attribute (previously described as the probability of being optimal at £30,000 per QALY gained, as represented by a single point on a CEAcc). The final set of attributes and levels used within the main study is detailed in Table 3.

Table 3 Attributes and levels included in the choice experiment

Attribute	Levels included
1) Incremental cost per QALY gained	0) £15,000 per QALY gained 1) £25,000 per QALY gained 2) £35,000 per QALY gained
2) Uncertainty	0) Low degree of uncertainty 1) High degree of uncertainty
3) Age	0) Children (<18 years) 1) Working (18-64 years) 2) Retired (>64 years)
4) Baseline HRQoL	0) 0.25 1) 0.50 2) 0.75
5) Are other therapies available?	0) No 1) Yes

As noted in Section 3.4.1, the selection of attributes and levels included within the choice experiment are central to its validity and reliability. The selection of levels used to describe each attribute therefore warrants further justification. It should be noted that the choice questionnaire included a brief key which described the meaning of each attribute and outlined their possible levels (See Appendix 1).

3.4.3.1 Incremental cost per QALY gained

The inclusion of the cost-effectiveness attribute was led by the objective of the study (See Section 1.2). Three levels were assigned to the cost-effectiveness attribute: (0) £15,000, (1) £25,000 and (2) £35,000 per QALY gained. These three levels were selected in order to broadly reflect the inflexion points for adoption and rejection suggested by Rawlins and Culyer.[3]

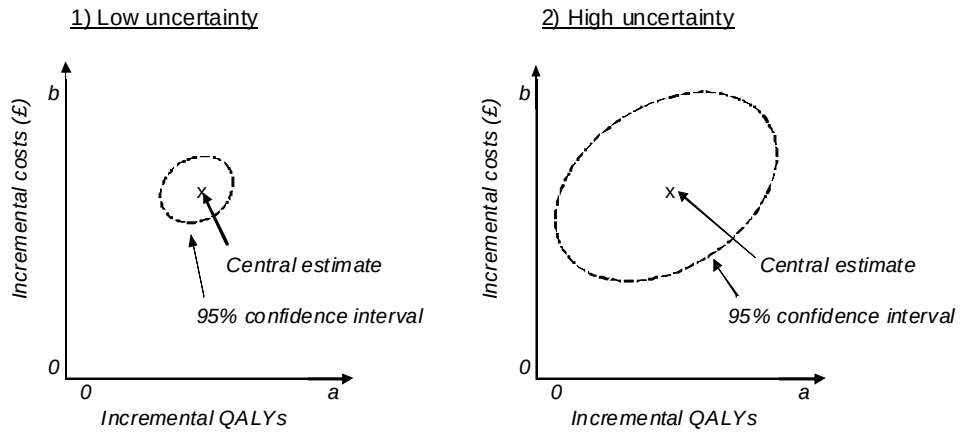
3.4.3.2 Uncertainty

The decisions made by NICE's Appraisal Committees may be influenced by the degree of uncertainty surrounding the incremental costs and effects of a health technology, and the likelihood that the technology under consideration is optimal compared to the current standard treatment. However, describing the level of uncertainty surrounding incremental costs and benefits without introducing correlations with the cost-effectiveness attribute is problematic. Initially, it was intended that the uncertainty attribute would be described as a single point on a CEAcc, i.e. the probability of being optimal at £30,000 per QALY gained. However, this approach would imply the existence of a fixed cost-effectiveness threshold, and leads to unrealistic combinations of the cost-effectiveness and uncertainty attribute levels, e.g. mean cost-effectiveness of £15,000 per QALY gained combined with a 30% probability of having a cost-effectiveness that is better than £20,000 per QALY gained. As noted in Section 3.4.3, this problem was highlighted through feedback from the pilot study.

Alternatively, using a range of costs and benefits (e.g. +/-£10,000 per QALY gained) is problematic as the cost per QALY scale is not fully continuous; as incremental costs and effects cross the axes of the cost-effectiveness plane, the resulting cost-effectiveness ratio tends towards plus infinity or minus infinity. A range of uncertainty around the expected incremental net benefit would be more appropriate, however, as noted in Section 3.4.2, the Appraisal Committee does not directly make decisions using this metric (although they do consider the results of probabilistic sensitivity analyses presented using acceptability curves).

As a result of these practical problems, a more subjective approach was adopted. Two levels were assigned to the uncertainty attribute: (0) low level of uncertainty, and (1) high level of uncertainty. The main benefit of this approach is that the uncertainty and cost-effectiveness attributes remain independent. However, the limitation of describing the uncertainty surrounding incremental costs and effects in this manner is that the interpretation of the degree of uncertainty may differ between respondents e.g. one respondent may interpret a high degree of uncertainty as the therapy being potentially dominated, whilst another respondent may not. Figure 3 shows the description of the two levels of uncertainty used within the experiment; this diagram was displayed in the attribute key presented at the beginning of the choice questionnaire (See Appendix 1).

Figure 3 Description of uncertainty used in choice questionnaire



3.4.3.3 Age

The Institute's draft consultation document on social value judgements suggests that whilst health should not be valued more highly for some age groups rather than others, where age is an indicator of benefit or risk, age discrimination is appropriate.[5;64] Whilst the element of age is incorporated in the number of QALYs gained, and as such directly affects the cost-effectiveness of an intervention, the age of the patient population may be given additional equity weight; it is this potential additional weighting that this attribute was intended to capture. For the purpose of simplicity, three levels were assigned to this attribute: (0) children under 18 years of age, (1) working adults between 18-64 years of age, and (2) retired adults over 64 years of age.

3.4.3.4 Baseline HRQoL

The baseline HRQoL of the population who would receive the intervention was taken to represent a proxy for clinical need. This attribute describes the baseline health status of the target population without treatment with the therapy under consideration, as described by a health utility index scale ranging from 0 ("dead") to 1 ("perfect health). Three levels were assigned to describe this attribute: (0) 0.25, (1) 0.50, and (2) 0.75; these levels were intended to broadly capture patients with high, moderate and low HRQoL respectively.

3.4.3.5 Availability of other therapies

The degree to which the availability of established effective therapies to manage or treat diseases influences the Appraisal Committees' decision-making is not clear within the

literature. Needless to say however, the Institute does undertake appraisals of technologies used to treat conditions whereby the current standard treatment is best supportive care (for example, the use of disease-modifying therapies for the management of multiple sclerosis). The relative importance of other programme attributes may be dependent on whether effective therapies are established for the treatment of the disease under consideration. Two levels were assigned to this attribute: (0) no other effective therapies are available, thus the comparator is best supportive care, and (1) other therapies are available.

It should be noted that the inclusion of the "other therapies" attribute essentially changes the baseline option if respondents do not recommend the intervention under consideration, therefore the baseline choice is different for scenarios in which other therapies are available and scenarios in which the comparator is best supportive care (*Personal communication, Dr Terry Flynn, Medical Research Council Health Services Research Collaboration, Bristol*). In order to explore the impact of this potential confounding, a separate regression model was estimated for those scenarios in which other therapies were available, and the results of this analysis were then compared to the results of the main analysis.

3.5 Design of experiments

As noted in Section 3.4.1, the design of the experiment was considered alongside decisions concerning the number of attributes and levels to be included in the choice experiment. Whilst a full factorial design ($3^3 \times 2^2$=108 scenarios) would allow for more precise estimates of utility through the inclusion of all possible effects, it is unlikely that any of the Appraisal Committee members would have responded to such a lengthy questionnaire. Instead, a fractional factorial main effects design (excluding interactions) was sampled from the complete factorial. It should be noted that currently there remains no consensus on the number of profiles that respondents need to complete to ensure reliable and valid parameter estimates. Louviere et al[48] note that many experiments that have implemented 32 or more profiles successfully. However, it was felt that presenting potential respondents with more than 20 scenarios would dramatically reduce response rates. Consequently, 18 scenarios were sampled using an orthogonal main effects only design. ORTHOPLAN®, an orthogonal random sampling routine within the SPSS®

statistical program (SPSS Inc, Illnois), was used to generate the final set of 18 choice scenarios.

The sampling design used within the choice questionnaire is presented in Table 4. It should be noted that given the use of three three-level attributes and two two-level attributes, it was not possible to generate a fractional factorial design which maintained both balance and orthogonality (*Personal communication, Professor David Grey, Department of Probability and Statistics, University of Sheffield*). As a result, it was decided that orthogonality should be preserved at the cost of balance. The implication of this trade-off is that the magnitude of standard errors estimated from the regression model would differ between attribute levels (*Personal communication, Dr Terry Flynn, Medical Research Council Health Services Research Collaboration, Bristol)*.

The same set of scenarios was presented in each choice questionnaire. Whilst there may be some potential benefit in randomly sampling a set of possible choices for each potential participant, as more choices could be presented, this approach would have assumed homogeneity of preferences between Appraisal Committee members, and as such may have biased the results obtained from such a small sample.

Table 4 Main effects only design codes for binary choice experiment

Scenario	Incremental cost per QALY gained	Degree of uncertainty around mean cost-effectiveness	Age of target population	Baseline HRQoL	Are other therapies available?
1	1	0	1	2	1
2	0	1	1	1	1
3	1	1	1	1	0
4	2	1	1	2	1
5	2	1	0	2	1
6	0	1	2	2	0
7	0	0	2	1	1
8	1	1	0	1	0
9	1	0	0	0	1
10	1	1	2	0	1
11	0	1	1	0	1
12	1	1	2	2	1
13	0	0	0	2	0
14	2	0	1	0	0
15	0	1	0	0	1
16	2	0	2	1	1
17	2	1	2	0	0
18	2	1	0	1	1

Table 5 shows the corresponding quantitative and qualitative attribute level descriptions for the choice scenarios included in the questionnaire.

Table 5 Main effects only choice scenarios for binary choice experiment

Scenario	Incremental cost per QALY gained	Degree of uncertainty around mean cost-effectiveness	Age of target population	Baseline HRQoL	Are other therapies available?
1	£25,000	Low	Working	0.75	Yes
2	£15,000	High	Working	0.50	Yes
3	£25,000	High	Working	0.50	No
4	£35,000	High	Working	0.75	Yes
5	£35,000	High	Children	0.75	Yes
6	£15,000	High	Retired	0.75	No
7	£15,000	Low	Retired	0.50	Yes
8	£25,000	High	Children	0.50	No
9	£25,000	Low	Children	0.25	Yes
10	£25,000	High	Retired	0.25	Yes
11	£15,000	High	Working	0.25	Yes
12	£25,000	High	Retired	0.75	Yes
13	£15,000	Low	Children	0.75	No
14	£35,000	Low	Working	0.25	No
15	£15,000	High	Children	0.25	Yes
16	£35,000	Low	Retired	0.50	Yes
17	£35,000	High	Retired	0.25	No
18	£35,000	High	Children	0.50	Yes

3.6 Questionnaire format and administration

The complete questionnaire is presented in Appendix 1. Each sampled choice scenario was presented as an individual table of characteristics, analogous to a simplified version of the range of evidence considered during each technology appraisal. For each scenario, participants were asked to indicate whether they would be likely to recommend the intervention or not. A key describing each of the attributes and the levels was presented at the beginning of the questionnaire. An information sheet detailing the purpose of the study, instructions for participants, contact details of the study group, and the proposed dissemination strategy was also presented at the beginning of the questionnaire (See Appendix 1). Whilst it would have been useful to collect demographic information on study respondents, the study group felt that absolute anonymity would maximise response rates.

The objectives, methods of the study and the requirements of study participants were briefly presented to each of the three Appraisal Committees at the end of routine

Appraisal Committee meetings in May and June 2005 by Professor David Barnett (*Chair of NICE Appraisals Committees, National Institute for Health and Clinical Excellence, London*) on behalf of the study group (See Appendix 2).

3.7 Statistical analysis

3.7.1 Assessment of logical consistency

The internal validity of the binary response data elicited from Appraisal Committee members was explored using a test of logical consistency. The logical consistency of the response data was tested by comparing individual-level responses between pairs of scenarios in which all but one of the attribute levels were identical. For example, choice scenarios 7 and 16 differ (See Tables 4 and 5) only in terms of the cost per QALY attribute level (£15,000 per QALY gained vs. £35,000 per QALY gained for scenarios 7 and 16 respectively). Logically, an individual who recommends the intervention under consideration described by scenario 16 should also recommend the intervention under consideration described by scenario 7, as it describes the same population of patients but the intervention has a lower incremental cost-effectiveness ratio. Table 6 shows the scenarios comparisons used to test the logical consistency of the response data.

*Table 6 Scenario comparisons used for tests of logical consistency**

Scenario comparison	Different attribute level	Identical attribute levels	Logical consistency rule
Scenario 2 versus Scenario 11	Scenario 2) HRQoL=0.50 Scenario 11) HRQoL=0.25	£15,000 per QALY gained High level of uncertainty Working adults Other therapies available	Respondents who recommend the intervention for scenario 2 should recommend the intervention for scenario 11
Scenario 5 versus Scenario 18	Scenario 5) HRQoL=0.75 Scenario 18) HRQoL=0.50	£35,000 per QALY gained High level of uncertainty Children Other therapies available	Respondents who recommend the intervention for scenario 5 should recommend the intervention for scenario 18
Scenario 10 versus Scenario 12	Scenario 10) HRQoL=0.25 Scenario 12) HRQoL=0.75	£25,000 per QALY gained High level of uncertainty Retired adults Other therapies available	Respondents who recommend the intervention for scenario 12 should recommend the intervention for scenario 12
Scenario 7 versus Scenario 16	Scenario 7) £15,000 per QALY gained Scenario 16) £35,000 per QALY gained	Low level of uncertainty Retired adults HRQoL=0.50 Other therapies available	Respondents who recommend the intervention for scenario 16 should recommend the intervention for scenario 7

**Three pairs of scenarios differed only in terms of age, but were excluded from the internal consistency assessment due to absence of a logical order of preference*

3.7.2 Logit regression modelling

The use of a simple binary response format means that a simple binary logit model (BLM) is appropriate. The mathematical formulation for the logit model has been previously detailed by Louviere et al,[48] and is presented below for purposes of transparency (equations 3.4–3.8). The functional form of the binary logit model may be expressed as:

$$P(yes \mid yes,no) = \exp(V_{yes})/[\exp(V_{yes}) + \exp(V_{no})] \tag{3.4}$$

where *V*s are the systematic utility components. The value of V_{no} can be set to zero with no loss of generality, which gives:

$$P(yes \mid yes,no) = \exp(V_{yes})/[\exp(V_{yes}) + 1 \tag{3.5}$$

Through considering the odds of responding "yes", which in this case corresponds to "recommend intervention under consideration over the current treatment", relative to "no", which corresponds to "do not recommend intervention under consideration over the current treatment", this gives:

$$\frac{P(yes \mid yes,no)}{P(no \mid yes,no)} = \frac{\dfrac{\exp(V_{yes})}{\exp(V_{yes})+1}}{\dfrac{\exp(V_{no})}{\exp(V_{yes})+1}} = \frac{\exp(V_{yes})}{\exp(V_{no})} \tag{3.6}$$

If the $\exp(V_{no})$ is specified as 1, taking natural logarithms of both sides gives:

$$\log_e\left[\frac{P(yes \mid yes,no)}{P(no \mid yes,no)}\right] = V_{yes} \tag{3.7}$$

When V_{yes} is specified as a linear-in-the-parameters expression, this gives:

$$V_{yes} = \sum_k b_k X_k + \sum_m a_m Z_m \tag{3.8}$$

where β_k is a vector of taste weights associated with *K* attribute vectors, X_k; and α_m is a vector of effects associated with *M* individual characteristics interacted with either the

"yes" intercept or elements of the X vector, Z_m.[48] A simple additive logit regression function, which gives the odds of a "yes" response relative to a "no" response, can thus be expressed as:

$$Log\frac{(yes \mid yes, no)}{no \mid yes, no} = b^0 + b^1X^1 + b^2X^2 .....b^LX^L + e \quad (3.9)$$

$where\ X^1, X^{2\cdots}X^L$ = specified levels of L attributes included in the model

$\beta^0, \beta^1, \beta^{2\cdots\cdots}\beta^L$ = coefficients which are estimated from the regression equation

e = error term which is included to allow for factors that influence utility which have not been controlled for in the model

The corresponding utility function was estimated by transforming the logit function as shown below.

$$P(yes) = \frac{\exp(b^0 + b^1X^1 + b^2X^2 .....b^LX^L + e)}{1 + \exp(b^0 + b^1X^1 + b^2X^2 .....b^LX^L + e)} \quad (3.10)$$

All statistical analysis was undertaken using STATA® statistical software (STATA Corporation, Texas). Logit regression analysis was undertaken whereby "Response" was the dependent variable and the five attributes were the independent variables (See Table 7). Robust standard errors were calculated by allowing for clustering within individual responses. Both log odds ratios and standardised odds ratios were calculated for β coefficients. For each of the attributes, the base level for the regression analysis was defined as the level which is most likely to lead to a positive recommendation for the therapy under consideration, thus the resulting odds ratios describe the reduction in the odds of adoption for increasingly less preferable attribute levels relative to the base level. Categorical variables were used to describe the levels for each of the five attributes, as the cost per QALY scale is not continuous (See Section 3.4.3.2), and the baseline HRQoL attribute was intended to describe three groups of individuals (healthy, sick, very sick) rather than a continuous scale.

Table 7 Description of variables included in the logit regression model

Variable	Variable type	Coded attribute levels*
Response	Binary	1=recommend intervention under consideration 0=do not recommend intervention under consideration
Cost per QALY	Categorical	0=£15,000 per QALY gained 1=£25,000 per QALY gained 2=£35,000 per QALY gained
Uncertainty	Categorical	0=low 1=high
Age	Categorical	0=Children 1=Working 2=Retired
Baseline HRQoL	Categorical	0=0.25 1=0.50 2=0.75
Other Therapies	Categorical	0=no other therapies available 1=other therapies available

**Attribute level 0 held as base value for logit regression*

The explanatory power of the model was assessed by comparing the observed adoption probabilities from the empirical response data for each scenario with the expected probability of adoption from the logit regression model.

It was intended that marginal rates of substitution, that is, the rate at which individuals are willing to give up one attribute in order to receive another, would be calculated by dividing the logit coefficients describing the change in the level for one attribute by the change in the level for the incremental cost-effectiveness ratio. However, as the cost per QALY scale is not continuous, the resulting marginal rates of substitution would be misleading and therefore are not presented here. Despite these problems in estimating marginal rates of substitution, the relative importance of the alternative attributes may still be observed by examining the coefficients estimated within the regression model.

3.7.3 Assessing the external validity of the stated preference model: comparison against revealed preferences

Clearly, the utility of the model is largely dependent on its criterion validity, that is, the model's ability to predict the Appraisal Committees' actual recommendations concerning the use of specific health technologies. In order to test the predictive ability of the model, choice scenarios were generated based upon published guidance on the use of seven technologies (or groups of technologies) recently appraised by NICE, as

detailed in Table 8. Information relating to incremental cost-effectiveness, uncertainty, age, baseline HRQoL, and the availability of other therapies was extracted and inputted into the regression model. Where data were unavailable, evidence from the corresponding Assessment Group reports were used instead. The estimated probability that an individual technology would be recommended was then compared against the recommendation made by the Institute. As the model estimates odds ratios or probabilities of acceptance rather than binary "yes" or "no" outcomes, it was assumed that a predicted probability greater than 0.5 represents a "yes", whilst a predicted probability less than 0.5 represents a "no."

Table 8 Case study appraisals used to assess criterion validity

Appraisal number	Technology appraisal	Date published
80	Clopidogrel in the treatment of non-ST-segment-elevation acute coronary syndrome	July 2004
41	Routine antenatal anti-D prophylaxis for RhD-negative women	May 2002
72	Anakinra for rheumatoid arthritis	November 2003
65	Rituximab for aggressive non-Hodgkin's lymphoma	September 2003
69	Liquid-based cytology for cervical screening	October 2003
32	Beta interferon and glatiramer acetate for the treatment of Multiple Sclerosis	January 2002
14	Ribavirin and interferon alpha for Hepatitis C	October 2003

Two limitations of the assessment of criterion validity should be noted:

(1) The stated preference model uses five categorical variables to describe the attribute levels, thus dummy variables are used to describe whether the selected attribute level is present or not. In order to replicate revealed preference scenarios for use in the stated preference model, it was necessary to assume each actual attribute level took the value of its nearest equivalent in the choice model, e.g. an actual cost-effectiveness ratio of £28,700 would be inputted as "level 1" (£25,000 per QALY gained) within the choice model.

(2) Complete information on all attributes was not always available within either the NICE guidance reports or the corresponding Technology Assessment Reports. For example, baseline HRQoL and the degree of uncertainty surrounding incremental costs and effects were not consistently reported within the NICE guidance documents. Where quality of life evidence was not available within these reports, utility scores were obtained from the Harvard Catalogue of

Preference Scores.[65] Where available, the degree of economic uncertainty (simply categorised in terms of high or low uncertainty) was assumed from further investigation of uncertainty analysis presented within the Technology Assessment Reports. Where multiple cost-effectiveness ratios were reported, the mean value was assumed.

The necessary use of these assumptions means that the assessment of criterion validity should be considered to be exploratory.

3.8 Research ethics

Ethical approval for this study was obtained from the Sheffield University Research Ethics Committee (UREC). The letter of ethical approval is presented in Appendix 3. Permission to distribute the choice questionnaire was kindly granted by Professor David Barnett (*Chair of Appraisal Committee, NICE, London*), Professor Andrew Stevens (*Chair of Appraisal Committee, NICE, London*), and Dr Carole Longson (*Dr Carole Longson, Appraisals Programme Director, NICE, London*).

Chapter 4 Results

4.1 Overview of results

This chapter details the results of the binary choice experiment, beginning with an assessment of the internal validity of the elicited response data. The results of the logit regression analysis are presented and discussed. Estimated odds ratios for the adoption of technologies obtained from the regression analysis are compared to the empirical response data to demonstrate the explanatory power of the model. Sensitivity analyses are presented using probit regression analysis, and separate results for different comparators are reported and their validity discussed. The external validity of the logit regression model is explored by testing its ability to predict adoption/rejection decisions made by NICE's Appraisal Committees during recent technology appraisals. Finally, the limitations of the model and analysis are discussed.

4.2 Empirical response data

4.2.1 Response rate

A total of 37 responses were obtained from the three NICE Appraisal Committees (approximately 46%), which corresponded to 664 choice responses. Whilst the number of responses obtained represents a usable sample of the Committee members, further responses could have helped to provide a more representative sample and additional statistical power.

4.2.2 Empirical response data

Table 9 presents a summary of the response data obtained from members of the NICE Appraisal Committees.

Table 9 Summary of response data

Scenario	Attributes					Adopt the intervention?		P(adopt)
	Cost per QALY gained	Uncertainty	Age	Baseline HRQoL	Other therapies?	Yes	No	
1	£25,000	Low	Working	0.75	Yes	24	13	0.65
2*	£15,000	High	Working	0.50	Yes	30	6	0.83
3*	£25,000	High	Working	0.50	No	27	9	0.75
4	£35,000	High	Working	0.75	Yes	0	37	0.00
5	£35,000	High	Children	0.75	Yes	0	37	0.00
6	£15,000	High	Retired	0.75	No	33	4	0.89
7	£15,000	Low	Retired	0.50	Yes	35	2	0.95
8	£25,000	High	Children	0.50	No	29	8	0.78
9	£25,000	Low	Children	0.25	Yes	25	12	0.68
10	£25,000	High	Retired	0.25	Yes	15	22	0.41
11	£15,000	High	Working	0.25	Yes	32	5	0.86
12	£25,000	High	Retired	0.75	Yes	10	27	0.27
13	£15,000	Low	Children	0.75	No	37	0	1.00
14	£35,000	Low	Working	0.25	No	19	18	0.51
15	£15,000	High	Children	0.25	Yes	29	8	0.78
16	£35,000	Low	Retired	0.50	Yes	6	31	0.16
17	£35,000	High	Retired	0.25	No	13	24	0.35
18	£35,000	High	Children	0.50	Yes	5	32	0.14

**One respondent did not state a preference for scenarios 2 and 3*

Table 9 demonstrates that for some of the scenarios, the responses were entirely consistent across all participating Appraisal Committee members. For example, all respondents stated that they would not recommend the interventions under consideration described by scenario 4 and scenario 5, whereby the incremental cost per QALY gained appears to be unfavourable and incremental costs and effects are highly uncertain. Similarly, all study respondents stated that they would recommend the intervention under consideration described by scenario 13, whereby the intervention under consideration has a cost-effectiveness profile which is better than many interventions currently funded by the NHS, incremental costs and effects are subject to only a low degree of uncertainty and no other therapies are available.

Interestingly, less consistency is observed for some scenarios, for example scenario 10 and scenario 14, for which around half of the respondents stated that they would recommend the intervention under consideration, whilst the remainder stated that they would not. For these scenarios, either the incremental cost per QALY gained is less favourable, or incremental costs and effects are subject to a high degree of uncertainty.

4.2.3 Assessment of internal consistency

Table 10 presents the results of the assessment of logical consistency of the Appraisal Committee responses, as described in Section 3.7.1.

Table 10 Logical consistency of response data

Scenario comparison	Number logically consistent (%)	Number inconsistent (%)
Scenario 2 versus 11	35 (97.22%)	1 (2.78%)
Scenario 7 versus 16	37 (100%)	0 (0%)
Scenario 5 versus 18	37 (100%)	0 (0%)
Scenario 10 versus 12	36 (97.3%)	1 (2.70%)

For the four scenario comparisons presented in Table 10, the data suggests almost complete logical consistency for individual responses. Only 2 of the 147 respondent-level comparisons which could be assessed were found to be logically inconsistent.

4.3 Results of logit regression modelling

Table 11 presents the results of the logit regression model; parameter values are expressed both in terms of model coefficients (the natural log of the odds ratio) and standardised odds ratios. The odds ratios describe the reduction in the odds of adoption associated with moving from the base attribute level to the selected attribute level.

Table 11 Results of logit regression model

Model parameter	Coefficient (log odds)	Odds ratio	Robust standard error	Significance (p)	Lower 95% c.i.	Upper 95% c.i.
Cost per QALY gained level 1 (£25,000)	-1.7823	0.1683	0.3434	0.0000	-2.4554	-1.1092
Cost per QALY gained level 2 (£35,000)	-4.0771	0.0170	0.4237	0.0000	-4.9076	-3.2467
Uncertainty level 1 (high)	-1.1624	0.3127	0.2278	0.0000	-1.6089	-0.7160
Age level 1 (working)	0.1391	1.1492	0.1830	0.4470	-0.2197	0.4978
Age level 2 (retired)	0.0752	1.0781	0.2609	0.7730	-0.4362	0.5865
HRQoL level 1 (0.50)	0.0183	1.0185	0.2205	0.9340	-0.4139	0.4506
HRQoL level 2 (0.75)	-0.7327	0.4806	0.2480	0.0030	-1.2187	-0.2467
Other therapies level 1 (yes)	-1.5292	0.2167	0.2688	0.0000	-2.0561	-1.0024
Constant	4.1471	-	0.4305	0.0000	3.3033	4.9908

The logit model suggests that increases in cost-effectiveness, uncertainty, and the availability of other therapies are associated with a statistically significant reduction in the odds of adoption at the 5% level. In addition, the transition from very low HRQoL (0.25) to a comparatively high level of HRQoL (0.75) is also statistically significant. The model suggests that small changes in HRQoL (utility of 0.25-0.50), and the age of the target population are not associated with a statistically significant reduction in the odds of a positive recommendation ($p>0.05$). Table 11 suggests that an increase in incremental cost-effectiveness from £15,000 to £25,000 per QALY gained, the odds of a positive recommendation are multiplied by 0.1683 (a dramatic 83% decrease). Switching from an incremental cost-effectiveness of £15,000 to £35,000 per QALY further reduces the odds of a positive recommendation by a factor of 0.0170 (a 98% decrease). The analysis suggests that switching from a low level of uncertainty to a high level of uncertainty reduces the odds of a positive recommendation by a factor of 0.3127 (a 69% decrease). Moving from interventions which are aimed at patients with low baseline HRQoL (0.25) to high baseline HRQoL (0.75) is associated with a reduction in the odds of a positive recommendation of 52%. The model also suggests that switching from interventions targeted at diseases and conditions where other effective therapies are already available to best supportive care is associated with a reduction in the odds of adoption of 78%.

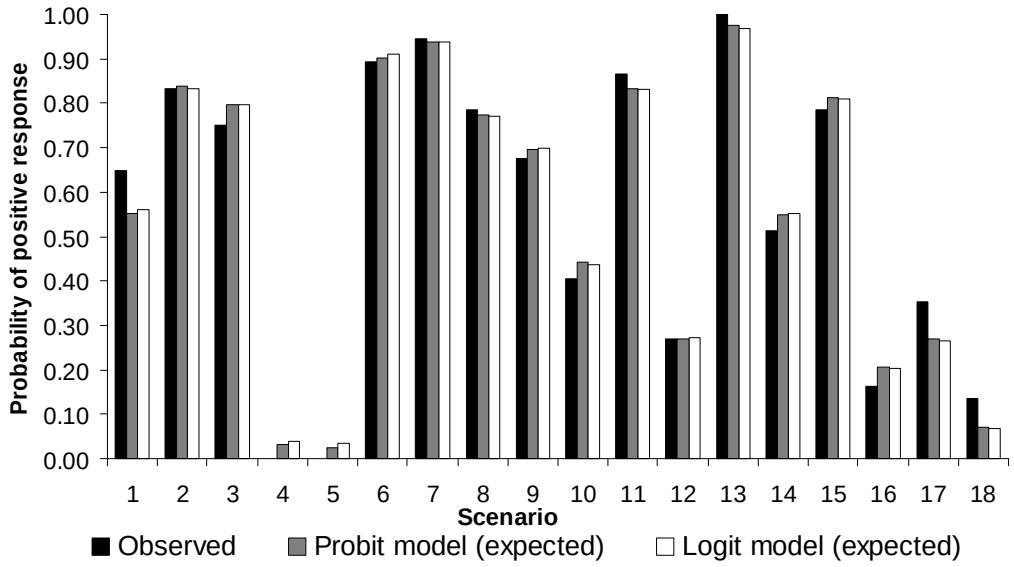
The overall goodness of fit indicated by the pseudo r-squared value was 0.35; Louviere et al suggests that for nonlinear models, a pseudo r-squared value of between 0.2 and 0.4 indicates an adequate model fit.[48] Whilst it would be beneficial to test the goodness of fit of the regression model using a statistical method such as the chi-squared test, this is inappropriate due to the inclusion of adjustments for within-respondent clustering effects (*Personal communication, Dr Stephen Walters, Senior Lecturer in Medical Statistics, Health Economics and Decision Science Section, University of Sheffield*).

4.4 Assessment of explanatory power

In order to test the explanatory power of the model, the probability of adoption for each of the eighteen scenarios was estimated using both the random effects logit model (as described in Section 4.3) and its equivalent probit counterpart. The results of these two regression models were then compared to the observed probabilities of adoption

observed within the response data. Figure 4 suggests that both the logit and probit models closely predict the observed response data. The comparison suggests very minor predictive differences between the probit and logit estimations.

Figure 4 Observed versus expected probabilities of "yes" response



4.5 Sensitivity analysis

As noted in Section 3.4.3.5, interpretation of the Appraisal Committee members' preferences may be problematic due to the use of the "other therapies" attribute; those scenarios in which the comparator is described as best supportive care suggest a different baseline to those scenarios in which other therapies are already available to manage the disease or condition. In other words, the therapy the respondent is assumed to prefer if they reject the intervention under consideration is different depending on the nature of the comparator. To explore the impact of this potential confounding, a separate logit model was estimated for those choice scenarios in which other therapies were available, as shown in Table 12.

Table 12 Logit regression results for scenarios where comparator is "other therapies available" compared to logit results for all scenarios

Model parameter	Other therapies available scenarios (n=12)			All scenarios (n=18)		
	Coefficient	Odds ratio	P (sig)	Coefficient	Odds ratio	P (sig)
Cost per QALY gained level 1 (£25,000)	-1.4092	0.2443	0.0010	-1.7823	0.1683	0.0000
Cost per QALY gained level 2 (£35,000)	-4.3335	0.0131	0.0000	-4.0771	0.0170	0.0000
Uncertainty level 1 (high)	-1.0636	0.3452	0.0010	-1.1624	0.3127	0.0000
Age level 1 (working)	0.2440	1.2764	0.2850	0.1391	1.1492	0.4470
Age level 2 (retired)	-0.1903	0.8267	0.6100	0.0752	1.0781	0.7730
HRQoL level 1 (0.50)	0.6026	1.8268	0.2090	0.0183	1.0185	0.9340
HRQoL level 2 (0.75)	-0.7159	0.4888	0.0060	-0.7327	0.4806	0.0030
Constant	2.3271	-	0.0000	4.1471	-	0.0000

Table 12 suggests only minor differences in the impact of individual attribute levels on the odds of adoption. For those scenarios in which other therapies were available, the incremental cost per QALY gained, uncertainty, and age attributes appear to remain largely unaffected. Notably, the odds ratio for the change in baseline HRQoL from level 0 (0.25) to level 1 (0.50) is nearly double the estimate from the main analysis (1.8268 vs. 1.0185). This appears to be a slightly perverse result, as one would expect the odds ratio to be below 1.0, however, this change in odds is not statistically significant in either analysis ($p>0.05$). This sensitivity analysis suggests that only negligible interactions exist between the "other therapies" attribute and the other attributes included in the experiment. Unfortunately, due to the design of the experiment and the decision to maintain orthogonality at the cost of balance, only six choices scenarios described interventions for which the comparator was best supportive care. Consequently, insufficient data were available to estimate the parameters to a regression model describing these scenarios alone.

4.6 External validity: Comparison of stated preference and revealed preference data

Table 13 shows extracted data on the five attributes included in the model, together with the resulting appraisal determination for each of the seven case study technology appraisals. Data sources used to inform the parameter assumptions used in the model are shown in brackets.

Table 13 Revealed preference attribute levels and adoption decisions for the seven case study appraisals

Appraisal	Cost per QALY	Uncertainty	Age	Baseline HRQoL	Other therapies available?	Adoption decision
(1) Clopidogrel for non-ST-segment-elevation acute coronary syndrome (No. 80)[66]	All cost per QALY estimates reported in NICE guidance <£15,000 per QALY gained.[66] Level 0 (CPQ=£15,000) assumed.	0.72 probability of being optimal at £10,000 per QALY gained.[66] Level 0 (low uncertainty) assumed.	Clinical evidence obtained from individuals aged 66-68 years of age.[66] Level 2 (Retired) used in model.	Utility of 0.80 used in health economic model.[67] Level 2 (0.75) used in logit model.	Yes	Yes
(2) Routine antenatal anti-D prophylaxis for RhD-negative women (No. 41)[68]	Incremental cost per QALY =£7,600 for primigravidae vs. post-partum, and £28,000 for All rhesus-negative vs. primigravidae reported in NICE guidance.[68] Level 0 (CPQ=£15,000) assumed.	High uncertainty due to absence of information on QALYs lost due to fetal/neonatal loss.[68] Level 1 (high uncertainty) assumed.	Working age assumed.	Level 2 (0.75) assumed.	No	Yes
(3) Anakinra for rheumatoid arthritis (No.72)[69]	Using the Appraisal Committee's preferred biologics, Assessment model estimates > £67,000 per QALY gained.[69] Level 2 (£35,000 per QALY gained) assumed in model.	NICE guidance notes considerable uncertainty surrounding parameter values.[69] Level 1 assumed (high uncertainty)	Working age assumed from evidence in Assessment Report.[70]	No information available. Utility of 0.25 assumed.	Yes	No
(4) Rituximab for aggressive non-Hodgkin's lymphoma (No. 65)[71]	All cost per QALY estimates reported in guidance <£15,000 per QALY gained.[71] Level 0 assumed (£15,000 per QALY gained).	PSA indicated 5% chance that the cost per QALY would exceed £23,400 in people aged > 60, or £19,000 in people < 60 years.[71] Level 0 assumed.	Retired age assumed as clinical evidence relates to individuals aged 60-80.[71]	Utility scores reported in Assessment Report >0.75. Level 2 assumed (0.75).[72]	Yes	Yes

(5) Liquid-based cytology for cervical screening (No. 69)[73]	Base case analysis suggested CPQ=<£10,000 per QALY gained.[73] Level 0 assumed (£15,000 per QALY).	High uncertainty assumed from Assessment Report; notably due to lack of good quality sensitivity and specificity studies and importance of these parameters in the model.[74]	Working age assumed.	General population utility assumed. Level 2 assumed (0.75).	No	Yes
(6) Beta Interferon and Glatiramer Acetate for the treatment of Multiple Sclerosis (No. 32)[75]	Base case cost-effectiveness >£40,000 per QALY gained.[75] Level 2 assumed in model (£35,000 per QALY gained).	PSA suggested considerable parametric uncertainty. Time horizon identified as central determinant of cost-effectiveness.[76]	Working age assumed.	Utility of 0.5 assumed.[76]	No	No
(7) Ribavirin and Interferon Alpha for Hepatitis C (No. 41)[77]	Cost/QALY gained from treatment with combination therapy for 6 months in comparison with monotherapy estimated to be £7,000. Following relapse after previous course of interferon alpha, CPQ from 6 months of combination therapy compared with monotherapy estimated to be £3,050.[77] Level 0 assumed (£15,000per QALY gained).	High uncertainty assumed.	Working age assumed	0.75 assumed from Harvard Preference Scores.[65]	No	Yes

A series of scenarios were generated based upon the information detailed in Table 13; these data were then inputted into the logit regression model. Table 14 shows the resulting estimated probability of adoption for each of the scenarios.

Table 14 Results of the assessment of criterion validity

Technology appraisal	Incremental cost per QALY gained	Uncertainty	Age	HRQoL	Other therapies	NICE decision	P(adopt)
Clopidogrel for non-ST-segment-elevation acute coronary syndrome	£15,000	Low	Retired	0.75	Yes	Yes	0.88 (yes)
Routine antenatal anti-D prophylaxis for RhD-negative women	£15,000	High	Working	0.25	No	Yes	0.96 (yes)
Anakinra for rheumatoid arthritis	£35,000	High	Working	0.25	Yes	No	0.08 (no)
Rituximab for aggressive non-Hodgkin's lymphoma	£15,000	Low	Retired	0.25	Yes	Yes	0.94 (yes)
Liquid-based cytology for cervical screening	£15,000	Low	Working	0.25	No	Yes	0.99 (yes)
Beta interferon and glatiramer acetate for multiple sclerosis	£35,000	High	Working	0.50	No	No	0.28 (no)
Ribavirin and interferon alpha for Hepatitis C	£15,000	High	Working	0.25	Yes	Yes	0.83 (yes)

Table 14 suggests that the model is capable of distinguishing between those technologies which the Appraisal Committees would be highly likely to recommend, and those technologies which appear to be less attractive. It should be noted however, that the two rejections were for interventions which had less attractive cost-effectiveness profiles (£35,000 per QALY gained). An important question therefore is *"what other factors must be present for a less cost-effective intervention to become viable for recommendation?"* For interventions with an incremental cost-effectiveness of £35,000 per QALY gained, the highest probability of adoption that could be obtained from the

model was 0.56 (assuming low uncertainty, working age, no therapies available, low baseline quality of life). For interventions which have an incremental cost-effectiveness of £25,000 per QALY gained, the equivalent scenario suggests a much higher potential probability of adoption (0.93). Thus, the model suggests that it is possible for interventions which have weaker cost-effectiveness profiles to be viable for adoption; between £25,000 and £35,000 per QALY gained, the differential impact of these other factors on the probability of adoption may be substantial.

Chapter 5 Discussion

5.1 Discussion of study findings

The results of the stated preference modelling suggest that incremental cost-effectiveness, the degree of uncertainty surrounding incremental costs and effects, the baseline HRQoL of the target patient group and the availability of other therapies are significant factors which influence the Appraisal Committees' decisions whether to recommend health technologies. Interestingly, the results of the stated preference modelling appear to support the Institute's recent dissemination of draft guidelines on social value judgements,[5;64] as the modelling suggests that the age of the target population is not associated with an additional equity weighting.

The modelling exercise presents compelling evidence in support of Rawlins and Culyer's[3] probabilistic adoption/rejection curve, rather than the operation of a single threshold value. The analysis suggests for interventions which have a less attractive cost-effectiveness profile of £25,000 per QALY gained and £35,000 per QALY gained, the odds of a positive recommendation are substantially reduced by a factor of 0.8317 (p=0.000) and 0.9830 (p=0.000) respectively. Interestingly, the "mythical" £30,000 per QALY gained cost-effectiveness threshold is not supported by this stated preference modelling analysis. Rather, the stated preference modelling suggests that the Institute's Appraisal Committees would be unlikely to adopt technologies with this level of cost-effectiveness unless other factors are present, for example those diseases in which patients have a very low baseline quality of life, or diseases for which no established effective therapies are available. The results of the modelling exercise appear to contrast with the findings of Devlin and Parkin's analysis of revealed preference data from NICE guidance,[23] which suggested that the Institute's Appraisal Committees operated a substantially higher cost-effectiveness threshold of between £35,000-£57,000.†

The modelling analysis suggests that the presence of a high degree of uncertainty also significantly influences the Committees' decision to adopt health technologies. The model suggests that switching from a low level of uncertainty to a high level of uncertainty around incremental costs and effects reduces the odds of adoption by around 69% (p=0.000). Switching from interventions which are targeted at individuals who have a low level of HRQoL prior to receiving treatment (utility score=0.25) to those

† *Mean values from reported probabilistic threshold models*

interventions which are targeted at individuals who are comparatively healthy (utility score=0.75) leads to a reduction in the odds of adoption of 52% (p=0.003). The existence of other therapies to treat diseases is associated with a reduction in the odds of adoption of 78% compared to interventions used to manage diseases for which the only available therapy is best supportive care.

The evidence of trading between attributes within the response data suggests that the attributes and levels selected for use in the study were broadly appropriate. Interestingly, one of the study respondents commented that the use of a higher cost-effectiveness level of £40,000 per QALY gained would represent more difficult choices, however, the dramatic reduction in the odds of adoption associated with switching from £15,000 to £35,000 per QALY gained suggests that this would have little impact upon the model.

Both the logit and probit regression models provided a good fit to the elicited response data. Exploratory sensitivity analysis using only those choice scenarios which compared the intervention under consideration to some established therapy (Scenarios numbered 1, 2, 4, 5, 7, 9, 10, 11, 12, 15, 16 and 18) did not indicate the presence of confounding due to differences in the comparator treatment option the respondent is assumed to prefer if they reject the intervention under consideration. The analysis of logical consistency suggested that the choice responses elicited from individual Committees members were on the whole, internally consistent, although it should be noted that this test only covered two of the attributes (HRQoL and incremental cost-effectiveness). The comparison of the stated preference model against the revealed preference data reported within the seven NICE guidance reports and Assessment Reports indicates a potentially high degree of external validity, although further comparisons between the stated preference and revealed preference data are warranted.

5.2 Limitations of the study

The results of this choice experiment should be considered exploratory rather than definitive, as the decisions faced by the Appraisal Committees are invariably more complex than the scope of the choice experiment presented here. Indeed, the utility of the preference model is restricted by the limited number of attributes included in the experiment, and the limited number of levels used to describe each attribute. As

suggested in Chapter 3, it could have been preferable to include additional attributes in order to explain more variation in the elicited responses. For example, one respondent suggested that the expected size of the clinical benefit would have been an appropriate attribute for inclusion in the experiment (*Personal communication, Dr Richard Cookson, Senior Lecturer in Health Economics / member of NICE Appraisals Committee, University of York*), or perhaps an attribute could have been included to describe whether the intervention under consideration saves lives, improves HRQoL, or both. However, these factors should already be incorporated into the incremental cost per QALY gained attribute.

Whilst the analysis suggests that the selected attributes and levels were capable of being traded off against one another, greater precision in the elicited preferences could be realised through introducing a larger number of levels used to describe the attributes. For example, using a range from £10,000 to £40,000 per QALY gained, perhaps increasing in increments of £5,000 per QALY gained, could provide more sensitive estimates of the Committees' preferences for incremental cost-effectiveness.

Despite the simplicity of the experiment, as acknowledged above, it was felt that the attributes and levels selected for inclusion would provide a parsimonious model which included the key factors which influence decisions concerning the adoption of health technologies. However, it should also be noted that due to the limited number of potential responders (81 NICE Appraisal Committee members) and the disproportionate increase in the cognitive burden resulting from the inclusion of additional attributes and levels, it was necessary to trade-off certain aspects of the ideal experiment design against a design which would maximise response rates whilst maintaining a reasonable degree of precision. Despite these practical considerations, the experiment successfully managed to explore the key trade-offs between incremental cost-effectiveness, uncertainty and relevant equity concerns.

It should also be noted that the number of choice scenarios presented to each respondent within the questionnaire was less than the number currently recommended within the literature.[48] Whilst the inclusion of a larger number of scenarios within the choice questionnaire could potentially improve the precision of the coefficients estimated by the regression model, it is likely that adherence to current recommendations for stated

preference designs would have substantially reduced the already limited response rate. Consequently, the preferences presented here should be interpreted as exploratory rather than precise estimates.

A further limitation of the study concerns the description of the scenarios used within the choice experiment. The interpretation of the uncertainty attribute within the study was purposefully subjective; different interpretations of this attribute and its associated levels may have led to some bias in the results. For example, uncertainty surrounding the incremental costs and effects of an intervention may be large, yet the probability that the intervention is optimal at a given willingness to pay threshold may be high. It is therefore important to note that the potential difference between parametric uncertainty and decision uncertainty is not captured by this experiment. Furthermore, the description of the uncertainty attribute does not include information concerning whether the intervention is potentially dominated by the current standard treatment. Consequently, some respondents may have assumed that incremental costs and effects always lie in the North-West quadrant of the cost-effectiveness plane (i.e. the intervention is more effective but more costly than the current treatment), whilst others may have assumed that the intervention under consideration may be potentially dominated by the current standard treatment. It is likely that these two alternative interpretations would affect respondents' preferences differently. As noted in Chapter 3, the use of a range to describe the uncertainty in incremental net benefits would have been more appropriate from a theoretical perspective, although this would have reduced the face validity of the experiment

It should also be noted that the description of the uncertainty attribute used within this experiment focuses only on parametric uncertainty, and assumes that all relevant parameter uncertainty is appropriately modelled. However, the Appraisal Committees may also consider the impact of structural and methodological uncertainties on cost-effectiveness estimates (*Personal communication, Louise Longworth, Technical Analyst, Centre for Health Technology Evaluation, NICE, London)*.

It should noted once more that the stated preference data elicited from the Appraisal Committee members represents the preferences of individuals, and therefore cannot capture the dynamics of organisational decision-making. In addition, the choice data

used within the experiment represent the choices that the Appraisal Committee members say they would make, and as such may not reflect the actual decisions that the Committees would make in reality.

The principal limitation of the analysis of the preference data concerns the relatively low response rate (46%). Whilst sufficient observations were obtained for use in the regression modelling, there is a chance that the preferences elicited may not be representative of the Appraisal Committees as a whole. Whilst the questionnaire did not request personal information from study participants, personal correspondence from some of the study respondents suggested that participating Committee members varied in terms of their professional background (including statisticians, health economists, and clinicians).

A final limitation is that the use of a fractional factorial main effects design and a simple additive regression model is likely to account for only around 70-90% of the explained variance.[48] The inclusion of two-way or higher-order interactions between attribute levels could potentially account for further explained variance, albeit at the cost of a greater number of degrees of freedom and the potential under-determination of the model. However, the literature suggests that models derived from main effects only designs often predict well in attribute regions of greatest interest even if their parameters are biased.[48]

5.3 Areas for further research

A number of areas for further research are merited:

- The case study technology appraisals used to assess the external validity of the preference model represent only a small proportion of the total number of appraisals undertaken by NICE to date. Systematic analysis of all existing revealed preference data using the choice model developed within this study could be used to further assess the model's external validity. In addition, discrepancies between the predicted outcomes from the model and the actual recommendations published within NICE guidance may provide some indication of the impact of group decision-making phenomena on the recommendations for the use of technologies issued by the Institute.

- The inclusion of other potential equity concerns, for example, distinctions between those technologies which are life-saving and those which are HRQoL-gaining, could be explored within a similar experiment.
- It would be interesting to explore the consistency between the preferences of the members of the Institute's Centre for Health Technology Evaluation and its Appraisal Committees. An extension study to address this research question is currently underway; the results of this analysis will be forthcoming.
- Finally, it would be particularly interesting to understand how NICE's health care decision-making criteria differ from those used by other health care commissioning groups such as the Scottish Medicines Consortium (SMC), the Australian Pharmaceutical Benefits Advisory Committee (PBAC), and the Canadian Co-ordinating Office for Health Technology Assessment (CCOHTA). For the purposes of comparability, this extension study could be undertaken using the same design as the NICE study.

6. Conclusion

The results of the stated preference modelling suggests that incremental cost-effectiveness, the degree of uncertainty surrounding incremental costs and effects, the baseline HRQoL of the target patient group and the availability of other therapies are significant factors which influence the NICE Appraisal Committees' decisions concerning the recommendation for or against the use of health technologies.

References

1. Department of Health. The New NHS. Modern. Dependable. White Paper. 1997. London, The Stationary Office.

2. National Institute for Clinical Excellence. A guide to NICE. 1-34. 2005. London, NICE.

3. Rawlins MD, Culyer AJ. National Institute for Clinical Excellence and its value judgements. British Medical Journal 2004;329:224-7.

4. Rawlins MD. Scientific and social value judgements. 2004: 1-20. London, NICE.

5. Rawlins MD. Social value judgements: Guidelines for the Institute and its advisory bodies. Draft for consultation. 2005:1-31. London, NICE.

6. A National Health Service. 1944. London, HMSO.

7. Fuchs VR. Who shall live? Health, Economics and Social Choice. London: World Scientific, 1998.

8. Sculpher M, Drummond M, O'Brien B. Effectiveness, efficiency, and NICE. British Medical Journal 2001;322:943-4.

9. Smith R. The failings of NICE. British Medical Journal 2000;321:1363-4.

10. Hutton J,.Maynard A. A NICE challenge for health economics. Health Economics 2000;9:89-93.

11. National Institute for Clinical Excellence. Legal context of NICE guidance. 2005. London, NICE.

12. National Institute for Clinical Excellence. Guide to the Technology Appraisals Process. 1-31. 2004. London, NICE.

13. National Institute for Clinical Excellence. Guide to the methods for technology appraisal. 2004:1-71. London, NICE.

14. Pritchard C. Overseas approaches to decision making. *In* Towse A, Pritchard C, Devlin N, eds. Cost-effectiveness thresholds, pp 56-68. London: King's Fund / Office for Health Economics, 2002.

15. Loewry EH. Cost Should Not Be a Factor in Medical Care [editorial]. New England Journal of Medicine 1980;302:697.

16. Eddy DM. A manual for assessing health pratices and designing practice policies: an explicit approach. Philadelphia: Oxford University Press, 1992.

17. Weinstein MC,.Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. New England Journal of Medicine 1977;296:716-21.

18. Drummond M, Torrance G, Mason J. Cost-effectiveness league tables: More harm than good? Social Science and Medicine 1993;37:33-40.

19. Gold M, Siegel JE, Russell LB, Weinstein MC (eds). Cost-Effectiveness in Health and Medicine. New York: Oxford University Press , 1996.

20. Cookson R, McDaid D, Maynard A. Wrong SIGN, NICE mess: is national guidance distorting allocation of resources? British Medical Journal 2001;323:743-5.

21. Miners AH, Garau M, Fidan D, and Fischer AJ. Comparing estimates of cost effectiveness submitted to the National Institute for Clinical Excellence (NICE) by different organisations: retrospective study. British Medical Journal 2004:330.

22. Birch S, Gafni A. On being NICE in the UK: guidelines for technology appraisal for the NHS in England and Wales. Health Economics 2002;11:185-91.

23. Devlin N, Parkin D. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. Health Economics 2004;13:437-52.

24. Hill S, Garattini S, van Loenhout J, O'Brien BJ, and de Joncheere K. Technology Appraisal Programme of the National Institute for Clinical Excellence: A review by WHO. 2003. World Health Organisation.

25. Littlejohns P. Does NICE have a threshold? A response. In Towse A, Pritchard C, Devlin N, eds. Cost-effectiveness thresholds, pp 31-7. London: King's Fund / Office for Health Economics, 2002.

26. National Institute for Clinical Excellence. Supplementary evidence to the House of Commons Health Select Committee. Paper 2. 2002:1-9. London, NICE.

27. National Institute for Clinical Excellence. Guidance on the use of orlistat for the treatment of obesity in adults. Guidance no.22. 2001:1-12. London, NICE.

28. Raftery J. NICE: Faster access to modern treatments? Analysis of guidance on health technologies. British Medical Journal 2001;323:1300-3.

29. National Institute for Clinical Excellence and Nation. Guidance on the use of riluzole (Rilutek) for the treatment of motor neurone disease. Guidance no. 20. 2001:1-10. London, NICE.

30. Williams A. What could be NICER than NICE? Office for Health Economics Annual Lecture 2004.

31. Briggs AH,.Gray AM. Handling uncertainty when performing economic evaluations of healthcare interventions. Health Technology Assessment 1999;3:i-134.

32. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. Journal of Health Economics 1999;18:341-64.

33. Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute for Clinical Excellence (NICE). The Lancet 2002;360:711-5.

34. Yamey G. Chairman of NICE admits that its judgements are hard to defend. British Medical Journal 1999;319:1222.

35. Dobson F. Secretary of State's speech launching the National Institute for Clinical Excellence. 31st March 1999. Available from:

http://www.nice.org.uk/embcat.aspx?page=oldsite/back/frank_dobson.htm [Accessed 02/04/2005].

36. Culyer AJ. Equity—some theory and its policy implications. Journal of Medical Ethics 2001;27:275-83.

37. Culyer AJ. Economics and ethics in health care. Journal of Medical Ethics 2001;27:217-22.

38. Dolan P, Olsen JA. Distributing health care: Economic and ethical issues. New York: Oxford University Press, 2002.

39. Rawls J. A theory of justice. Cambridge, Massachusetts: Harvard University Press, 1971.

40. Harris J. Unprincipled QALYs - a response to Cubbin. Journal of Medical Ethics 1991;17:185-8.

41. Williams A. Intergenerational equity: an exploration of the 'fair innings' argument. Health Economics 1997;6:117-32.

42. Nord E. Cost-value analysis in health care: Making sense out of QALYs. Cambridge: Cambridge University Press, 1999.

43. Jonsen AR. Bentham in a box: technology assessment and health care allocation. Law, Medicine and Health Care 1986;14:172-4.

44. LeGrand J. Equity, health and health care. Social Justice Research 1987;1:257-74.

45. Dolan P, Shaw R, Tsuchiya A, Williams A. QALY maximisation and people's preferences: a methodological review of the literature. Health Economics 2005;14:197-208.

46. Dolan P, Tsuchiya A, Brazier J, and Ratcliffe J. Social QALY Project: Summary. 2005. Available from http://www.shef.ac.uk [Accessed 23/05/2005].

47. Ratcliffe J. Valuing the benefits of health care technologies: a case study of liver transplantation. PhD Thesis. 2000. London, Health Economics Research Group, Brunel University.

48. Louviere J, Hensher D, Swait J. Stated choice methods: Analysis and application. Cambridge, UK: Cambridge University Press, 2000.

49. Ryan M. Using Consumer Preferences in Health Care Decision Making: the Application of Conjoint Analysis. 1996. London, OHE Publications.

50. Kocur G, Alder T, , Hyman W, and Aunet B. A guide to forecasting travel demand with direct utility assessment. 1982. Washington DC, US Department of Transportation.

51. Bates J. Sensitivity to level of service: Evidence from stated preference work. Research for Transport Policy 1986;13:289-305.

52. Wardman M. Route choice and the value of motorists travel time: empirical findings. Institute for Transport Studies Working Paper 224. 1986. University of Leeds.

53. Magat WA, Viscusi WK, Huber J. Paired comparison and contingent valuation approaches to morbidity risk valuation. Journal of Environmental Economics and Management 1988;15:396-411.

54. Yoo DI,.Ohta H. Optimal pricing and product planning for new multiattribute products based on conjoint analysis. International Journal of Production Economics 1995;38:245-53.

55. Dennis DF. Analysing public inputs to multiple objective decisions on national forests using conjoint analysis. Forest Science 1998;44:421-9.

56. Sculpher M, Bryan S, Fry P, de Winter P, Payne H, Emberton M. Patients' preferences for the management of non-metastatic prostate cancer: discrete choice experiment. British Medical Journal 2004;328.

57. Ryan M, Farrar S. Using conjoint analysis to elicit preferences for health care. British Medical Journal 2000;320:1530-3.

58. Propper C. Contingent valuation of time spent on NHS waiting lists. The Economic Journal 1991;100:193-9.

59. Ubach C, Scott A, French F, Awramenko, M, Needham G. What do hospital consultants value about their jobs? A discrete choice experiment. British Medical Journal 2003;326.

60. Ryan M. Using conjoint analysis to take account of patient's preferences and go beyond health outcomes. An application to in-vitro fertilisation. Social Science and Medicine 1999;48:535-46.

61. Von Neumann J, Morgenstern O. Theory of Games and Economic Behavior. Princeton University Press, New Jersey, 1944.

62. Beach LR. The Psychology of Decision-Making. London: Sage, 1997.

63. Miller GA. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. The Psychological Review 1956;63:81-97.

64. Mayor S. NICE says that patients' age should affect treatment. British Medical Journal 2005;330:1102.

65. Harvard Center for Risk Analysis. Catalog of Preference Scores. 2005.

66. National Institute for Clinical Excellence. Clopidogrel in the treatment of non-ST-segment-elevation acute coronary syndrome. Technology Appraisal Guidance 80. 2004. London, NICE.

67. Main C, Palmer S, Griffin S, Jones L, Orton V, Sculpher M, et al. A rapid and systematic review of the clinical effectiveness and costeffectiveness of clopidogrel used in combination with aspirin compared to aspirin alone in the treatment of non-ST-segment-elevation acute coronary syndromes (ACS). Technology assessment report prepared for the National Institute for Clinical Excellence, 2003:1-180.

68. National Institute for Clinical Excellence. Guidance on the use of routine antenatal anti-D prophylaxis for RhD-negative women. Technology Appraisal Guidance No. 41. 2002. London, NICE.

69. National Institute for Clinical Excellence. Anakinra for rheumatoid arthritis. Technology Appraisal Guidance 72. 2003. London, NICE.

70. Clark W, Jobanputra P, Barton P, and Burls A. The clinical and cost-effectiveness of anakinra for the treatment of rheumatoid arthritis in adults. Technology Assessment Report prepared for the National Institute for Clinical Excellence, 2003:1-117.

71. National Institute for Clinical Excellence. Rituximab for aggressive non-Hodgkin's lymphoma. Technology Appraisal 65. 2003. London, NICE.

72. Knight C, Hind D, Brewer N, and Abbott V. Rituximab (MabThera) for aggressive non Hodgkin's lymphoma: systematic review. Technology Assessment Report prepared for the National Institute for Clinical Excellence, 2003:i-81.

73. National Institute for Clinical Excellence. Guidance on the use of liquid-based cytology for cervical screening. Technology Appraisal Guidance 69. 2003. London, NICE.

74. Karnon J, Peters J, Platt J, Chilcott J, and McGoogan E. Liquid-based cytology in cervical screening: an updated rapid and systematic review. Technology Assessment Report prepared for the National Institute for Clinical Excellence, 2003:i-67.

75. National Institute for Clinical Excellence. Beta interferon and glatiramer acetate for the treatment of multiple sclerosis. Technology Appraisal Guidance 32. 2002. London, NICE.

76. Tappenden P, Chilcott J, O'Hagan A, McCabe C, Cooper N, Abrams K, and et al. Cost effectiveness of beta interferons and glatiramer acetate in the management of multiple sclerosis. Technology Assessment Report prepared for the National Institute for Clinical Excellence, 2001:i-51.

77. National Institute for Clinical Excellence. Guidance on the use of Ribavirin and Interferon Alpha for Hepatitis C. 2003. London, NICE.

Appendix 1 Binary choice questionnaire to examine NICE's decision-making criteria

ScHARR
School of Health and Related Research

The University Of Sheffield.

A discrete choice experiment to examine the criteria used in health services commissioning decisions

Choice questionnaire

Information for study participants

A research group at the School of Health and Related Research (ScHARR) within the University of Sheffield is undertaking a study on the importance that health services decision-makers place on the various attributes of health interventions when deciding whether to commission them. We are interested in the importance that you place on central estimates of cost-effectiveness, uncertainty, age, disease severity and the availability of alternative therapies. This project is being internally funded by ScHARR as part of a postgraduate qualification.

Enclosed is a questionnaire on this issue, which will take ten to fifteen minutes to complete. For each the eighteen hypothetical scenarios, we would like you to decide whether or not you would be likely to recommend the intervention. Please indicate your selection by placing a tick in the appropriate box. We are interested in what you think; there are no right or wrong answers. Please note that this is a simple experiment and only a limited amount of information can be contained within each scenario. Your decision whether to recommend the intervention or not should be based only on the information available.

We would be grateful if you would complete the questionnaire and consent form, and return these to Paul Tappenden in the stamp addressed envelope by 30th June 2005. If you have any problems completing the questionnaire, please do not hesitate to contact us (Tel: 0114 2220855, email: P.Tappenden@Sheffield.ac.uk).

Once completed questionnaires have been returned to the project team, these will be analysed using a form of regression analysis to estimate the importance of each decision attribute. The methods and results of the analysis will be submitted as part of an academic thesis, and results will be made available to all study participants at this stage. Subject to the approval of the National Institute for Health and Clinical Excellence (NICE), it is anticipated that the study will be published in a high quality peer reviewed journal. Study participants will be acknowledged within any forthcoming publications of this study.

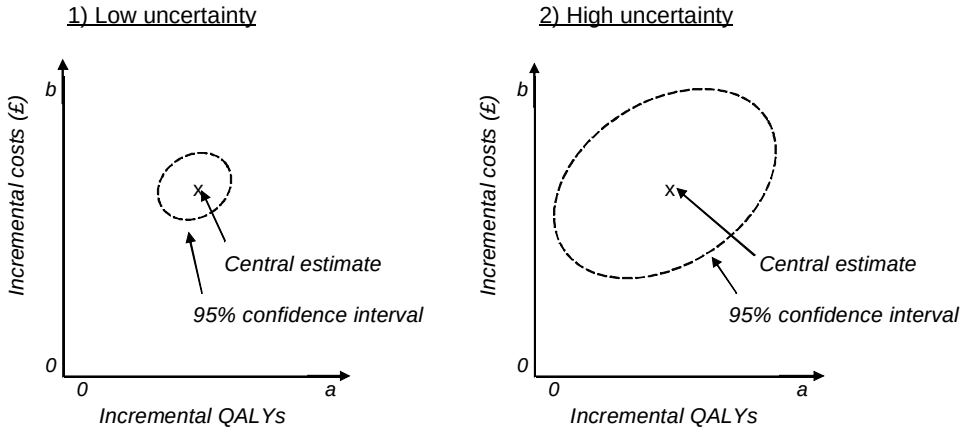
Key to choice attributes

Mean cost per quality adjusted life year (QALY) gained

This is the central estimate of cost-effectiveness for the intervention as compared to the current standard treatment. This is calculated as the additional cost of the intervention compared to the current standard treatment divided by the additional benefits of the intervention compared to the current standard treatment. Three levels are specified: (1) £15,000 per QALY gained; (2) £25,000 per QALY gained; (3) £35,000 per QALY gained.

Degree of uncertainty around mean cost-effectiveness

This relates to how uncertain the central estimate of cost-effectiveness is. Two levels are specified: (1) Low level of uncertainty – limited dispersion of incremental costs and benefits; (2) High level of uncertainty – large dispersion of incremental costs and QALYs. Illustrative examples of these are shown below.



Age of target population

The mean age of the population who will benefit from the intervention. Three levels are specified: (1) Children, <18 years; (2) Working, 18-64 years; (3) Retired, >64 years.

Baseline health related quality of life

The mean health-related quality of life score of patients prior to receiving the intervention, whereby "1" represents a state of perfect health and "0" represents dead. Three levels are specified: (1) 0.25; (2) 0.50; (3) 0.75.

Are other effective therapies available?

Whether alternative effective therapies are available to manage the condition or not. Two levels are specified: (1) Yes (i.e. comparator is an alternative effective therapy); (2) No (i.e. comparator is best supportive care).

CHOICE 1

Attribute	Intervention versus current standard treatment
Mean cost per QALY gained	£25,000 per QALY gained
Degree of uncertainty around mean cost-effectiveness	Low level of uncertainty
Age of target population	Working (18-64 years)
Baseline health-related quality of life	0.75
Are other effective therapies available?	Yes

Would you recommend the intervention? *(please tick)* Yes ☐ No ☐

CHOICE 2

Attribute	Intervention versus current standard treatment
Mean cost per QALY gained	£15,000 per QALY gained
Degree of uncertainty around mean cost-effectiveness	High level of uncertainty
Age of target population	Working (18-64 years)
Baseline health-related quality of life	0.50
Are other effective therapies available?	Yes

Would you recommend the intervention? *(please tick)* Yes ☐ No ☐

CHOICE 3

Attribute	Intervention versus current standard treatment
Mean cost per QALY gained	£25,000 per QALY gained
Degree of uncertainty around mean cost-effectiveness	High level of uncertainty
Age of target population	Working (18-64 years)
Baseline health-related quality of life	0.50
Are other effective therapies available?	No

Would you recommend the intervention? *(please tick)* Yes ☐ No ☐

CHOICE 4

Attribute	Intervention versus current standard treatment
Mean cost per QALY gained	£35,000 per QALY gained
Degree of uncertainty around mean cost-effectiveness	High level of uncertainty
Age of target population	Working (18-64 years)
Baseline health-related quality of life	0.75
Are other effective therapies available?	Yes

Would you recommend the intervention? *(please tick)* Yes ☐ No ☐

CHOICE 5

Attribute	Intervention versus current standard treatment
Mean cost per QALY gained	£35,000 per QALY gained
Degree of uncertainty around mean cost-effectiveness	High level of uncertainty
Age of target population	Children (<18 years)
Baseline health-related quality of life	0.75
Are other effective therapies available?	Yes

Would you recommend the intervention? *(please tick)*

Yes ☐
No ☐

CHOICE 6

Attribute	Intervention versus current standard treatment
Mean cost per QALY gained	£15,000 per QALY gained
Degree of uncertainty around mean cost-effectiveness	High level of uncertainty
Age of target population	Retired (>64 years)
Baseline health-related quality of life	0.75
Are other effective therapies available?	No

Would you recommend the intervention? *(please tick)*

Yes ☐
No ☐

CHOICE 7

Attribute	Intervention versus current standard treatment
Mean cost per QALY gained	£15,000 per QALY gained
Degree of uncertainty around mean cost-effectiveness	Low level of uncertainty
Age of target population	Retired (>64 years)
Baseline health-related quality of life	0.50
Are other effective therapies available?	Yes

Would you recommend the intervention? *(please tick)*

Yes ☐
No ☐

CHOICE 8

Attribute	Intervention versus current standard treatment
Mean cost per QALY gained	£25,000 per QALY gained
Degree of uncertainty around mean cost-effectiveness	High level of uncertainty
Age of target population	Children (<18 years)
Baseline health-related quality of life	0.50
Are other effective therapies available?	No

Would you recommend the intervention? *(please tick)*

Yes ☐
No ☐

CHOICE 9

Attribute	Intervention versus current standard treatment
Mean cost per QALY gained	£25,000 per QALY gained
Degree of uncertainty around mean cost-effectiveness	Low level of uncertainty
Age of target population	Children (<18 years)
Baseline health-related quality of life	0.25
Are other effective therapies available?	Yes

Would you recommend the intervention? *(please tick)*

Yes ☐
No ☐

CHOICE 10

Attribute	Intervention versus current standard treatment
Mean cost per QALY gained	£25,000 per QALY gained
Degree of uncertainty around mean cost-effectiveness	High level of uncertainty
Age of target population	Retired (>64 years)
Baseline health-related quality of life	0.25
Are other effective therapies available?	Yes

Would you recommend the intervention? *(please tick)*

Yes ☐
No ☐

CHOICE 11

Attribute	Intervention versus current standard treatment
Mean cost per QALY gained	£15,000 per QALY gained
Degree of uncertainty around mean cost-effectiveness	High level of uncertainty
Age of target population	Working (18-64 years)
Baseline health-related quality of life	0.25
Are other effective therapies available?	Yes

Would you recommend the intervention? *(please tick)*

Yes ☐
No ☐

CHOICE 12

Attribute	Intervention versus current standard treatment
Mean cost per QALY gained	£25,000 per QALY gained
Degree of uncertainty around mean cost-effectiveness	High level of uncertainty
Age of target population	Retired (>64 years)
Baseline health-related quality of life	0.75
Are other effective therapies available?	Yes

Would you recommend the intervention? *(please tick)*

Yes ☐
No ☐

CHOICE 13

Attribute	Intervention versus current standard treatment
Mean cost per QALY gained	£15,000 per QALY gained
Degree of uncertainty around mean cost-effectiveness	Low level of uncertainty
Age of target population	Children (<18 years)
Baseline health-related quality of life	0.75
Are other effective therapies available?	No

Would you recommend the intervention? *(please tick)*

Yes ☐
No ☐

CHOICE 14

Attribute	Intervention versus current standard treatment
Mean cost per QALY gained	£35,000 per QALY gained
Degree of uncertainty around mean cost-effectiveness	Low level of uncertainty
Age of target population	Working (18-64 years)
Baseline health-related quality of life	0.25
Are other effective therapies available?	No

Would you recommend the intervention? *(please tick)*

Yes ☐
No ☐

CHOICE 15

Attribute	Intervention versus current standard treatment
Mean cost per QALY gained	£15,000 per QALY gained
Degree of uncertainty around mean cost-effectiveness	High level of uncertainty
Age of target population	Children (<18 years)
Baseline health-related quality of life	0.25
Are other effective therapies available?	Yes

Would you recommend the intervention? *(please tick)*

Yes ☐
No ☐

CHOICE 16

Attribute	Intervention versus current standard treatment
Mean cost per QALY gained	£35,000 per QALY gained
Degree of uncertainty around mean cost-effectiveness	Low level of uncertainty
Age of target population	Retired (>64 years)
Baseline health-related quality of life	0.50
Are other effective therapies available?	Yes

Would you recommend the intervention? *(please tick)*

Yes ☐
No ☐

CHOICE 17

Attribute	Intervention versus current standard treatment
Mean cost per QALY gained	£35,000 per QALY gained
Degree of uncertainty around mean cost-effectiveness	High level of uncertainty
Age of target population	Retired (>64 years)
Baseline health-related quality of life	0.25
Are other effective therapies available?	No

Would you recommend the intervention? *(please tick)*

Yes ☐
No ☐

CHOICE 18

Attribute	Intervention versus current standard treatment
Mean cost per QALY gained	£35,000 per QALY gained
Degree of uncertainty around mean cost-effectiveness	High level of uncertainty
Age of target population	Children (<18 years)
Baseline health-related quality of life	0.50
Are other effective therapies available?	Yes

Would you recommend the intervention? *(please tick)*

Yes ☐
No ☐

Thank you for your co-operation.

Appendix 2 Study outline presented to NICE's Appraisal Committees

The University Of Sheffield.

A discrete choice experiment to examine the criteria used in health services commissioning decisions

Mr Paul Tappenden

Prof. John Brazier

Dr Julie Ratcliffe

Study outline

- **Study question**
  - Do decision-makers take into account programme attributes such as uncertainty, equity and budget impact rather than incremental cost-effectiveness alone in commissioning healthcare services?
- **Study methods**
  - The study takes the form of a discrete choice experiment (conjoint analysis).
  - Participants are asked to complete a choice questionnaire consisting of 18 scenarios
  - For each scenario, participants are asked *"would you recommend the intervention or the current standard treatment?"*

Attributes and levels

Attribute	Level 1	Level 2	Level 3
Mean cost per QALY gained	£15,000 per QALY gained	£25,000 per QALY gained	£35,000 per QALY gained
Degree of economic uncertainty	Low level of uncertainty	High level of uncertainty	N/a
Age of target population	Children (<18 years)	Working (18-64 years)	Retired (>64 years)
Baseline health related quality of life	0.25	0.50	0.75
Are other effective therapies available?	Yes (i.e. current treatment is other therapy)	No (i.e. current treatment is best supportive care)	N/a

Analysis of results

- Regression analysis will be used to estimate coefficients for each attribute
- The coefficients for each attribute represent the participant's strength of preference

Dissemination of results

- Presentation of early study results to each Appraisal Committee
- The study will form part of a postgraduate thesis at University of Sheffield (available online subject to NICE's approval)
- Publication in high quality peer reviewed journal (subject to NICE's approval)

Participation

- Please complete the questionnaire and consent form and return these in the stamp addressed envelope.
- There are no right or wrong answers – we want to know what you think.

And finally…..

- Please participate! Without a high response rate the results will be very difficult to interpret.
- *Thank you from the research team.*

Appendix 3 Letter of ethical approval from University of Sheffield Research Ethics Committee

Academic Division
Research Office

Sarah A Fulton
Deputy Academic Secretary
Head of Research Office

85 Wilkinson Street
Sheffield S10 2GJ
Direct line: 0114 222 1448
International: +44 114 222 1448
Fax: 0114 222 1452
Email: r.j.hudson@sheffield.ac.uk

29 March 2005

Mr Paul Tappenden
ScHARR

Dear Paul

PROJECT TITLE: 'A Discrete Choice Experiment to examine the criteria used in health services commissioning decisions'

I am pleased to inform you that on 29 March 2005 the above-named project was unconditionally **approved** on ethics grounds, on the basis that you will adhere to the following documents that you submitted for ethics review:

- University research ethics application form (dated 1 March 2005)
- Related participant consent form (dated 1 March 2005)
- Related participant information sheet (dated 1 March 2005)

If during the course of the project you need to deviate from the above-approved documents please inform me. Written approval will be required for significant deviations from or significant changes to the above-approved documents. Please also inform me should you decide to terminate the project prematurely.

Yours sincerely

Richard Hudson
Research Development Manager
(& Interim Ethics Administrator)