HEALTH STATUS MEASUREMENT IN SURGICAL PRACTICE

SIMON CHARLES ALEXANDER FRASER

DOCTOR OF MEDICINE

UNIVERSITY OF EDINBURGH

1992

For Allyson and Hamish

The Doctor next morning was rubbing his hands,
And saying, "There's nobody quite understands
These cases as I do! The cure has begun!
How fresh the chrysanthemums look in the sun!"

The Dormouse lay happy, his eyes were so tight
He could see no chrysanthemums, yellow or white.
And all that he felt at the back of his head
Were Delphiniums (blue) and geraniums (red).

A. A. Milne
from "The Dormouse and the Doctor"

CONTENTS

APPENDICES

PUBLISHED PAPERS

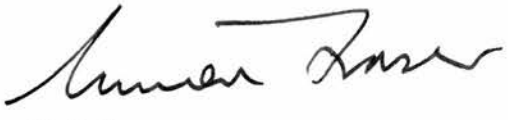
DECLARATION

All of the work described in the following thesis has been performed exclusively by myself, except for the following:

The advanced breast cancer study at Guy's Hospital was conducted by Dr Amanda Ramirez, who collected the Qualitator data.

In the abdominal pain study at the William Harvey Hospital, Dr Kevin Smith initiated the study and, with Dr Meena Agarawal, helped to collect the data on admission of the patients. Mrs Margaret Harrison amassed the returned postal questionnaires.

For the minor surgery study in King's College Hospital, Mr Hamid Khawaja and Mr Nigel Heaton collected the preoperative data.



Simon C A Fraser

LIST OF ABBREVIATIONS

CMF	Cyclophosphamide, Methotrexate, 5-Fluorouracil chemotherapy regimen
HAD	Hospital Anxiety and Depression scale
GHQ	General Health Questionnaire (30 question version referred to in thesis)
LASA	Linear analogue self-assessment
NHP	Nottingham Health Profile
NHS	National Health Service
NSAP	Nos-specific abdominal pain
PMH	Past medical history
QoL	Quality of life
QALY	Quality adjusted life year
SAP	Specific abdominal pain
TNM	Tumour, Node, Metastases staging system
UICC	Union Internationale Contre Cancre
VAS	Visual analogue scale (simalar to LASA)
WBC	White blood cell count
WHO	World Health Organisation

LIST OF TABLES

LIST OF FIGURES

HEALTH STATUS MEASUREMENT IN SURGICAL PRACTICE

Abstract

In the last hundred years, improved social conditions and advances in medical science have rendered previously fatal conditions curable. Modern surgical practice is now too complex to be measured by mortality and morbidity alone. Subjective, patient derived outcome measures are slowly gaining influence in other fields. Health status, or Quality of Life (QoL), measurement has not been widely adopted in surgical practice. To test the hypothesis that **Health status measures, scientifically applied, provide important additional information to the surgeon,** the techniques were applied to three diverse areas of surgical practice as models for broader application.

Chemotherapy for advanced breast cancer patients

Chemotherapy has little effect on survival in patients with advanced breast cancer. UICC response and toxicity criteria are used to measure outcome and QoL measurement is a rarity. Using a diary developed to make QoL measurement simpler, a randomised trial was mounted to compare QoL scores in patients receiving two regimens of differing toxicity. Forty patients received CMF or Epirubicin and were evaluated according to UICC criteria, the Nottingham Health Profile (NHP), Linear Analogue Self-Assessment (LASA) and King's diary. Response rates were better and toxicity worse for patients receiving CMF, but survival was the same for both regimens. Survival and QoL were better for responders than for non-responders, irrespective of therapy.

Good initial QoL scores predicted a response, and longer survival and these findings were repeated in a separate study at Guy's Hospital.

Psychological screening for Non Specific Abdominal Pain

Patients with Non Specific Abdominal Pain (NSAP) are significant consumers of surgical resources but a psychological contributor is often suspected. To determine whether NSAP has a detectable psychological contributor which could be used to predict outcome, 131 patients aged 14-40 admitted with acute abdominal pain were assessed using the General Health Questionnaire and Hospital Anxiety and Depression questionnaires, and a structured interview. In 61 patients with NSAP, more had a psychosocial problem identified by the admitting registrar (x^2=7.28,1df,p<0.01) and marginally more had high questionnaire scores. The risk of having NSAP was high if an abnormality on interview accompanied high questionnaire scores (Relative Risk 1.93, 95%c.i. 1.35-2.77) or if prodromal pain had lasted more than 7 days (relative Risk 2.13: 1.55-2.92). After 2 years, patients with continuing pain had higher HAD (x^2=6.57,1df,p<0.02) and Spielberger anxiety trait (x^2=6.50,1df,p<0.02) scores; NSAP was associated with persisting pain (Relative Risk 2.22, 1.10-4.48). Psychosocial factors are implicated in NSAP and in chronic pain, but the sensitivity and specificity of questionnaire assessment are too low to be useful in diagnosing NSAP. What promotes NSAP still remains largely unknown, but the referral process may be the next direction for productive study.

Health status after minor surgery

To establish whether minor operations cause a perceived and measurable improvement in health and QoL, 57 patients having day-surgery on a Waiting List Initiative were studied prospectively. The NHP, HAD and GHQ were completed before surgery and after 6 months by 81% patients, when an *ad-hoc* questionnaire dwelling on perceived outcome of surgery was also completed. An operative success was reported by 78%, improved health by 64%, improved QoL by 69% and improved work efficiency by 54%. Improvements in HAD *anxiety* (p=0.023), *depression* (p=0.035), NHP *pain* (p=0.001) and *global* NHP (p=0.034) were recorded. In the perceived outcome questionnaire, patients reporting a successful operation had had better preoperative GHQ (p=0.029) and HAD *depression* (p=0.031) scores than those whose operation was not a success. Those reporting an improvement in *health* postoperatively had worse preoperative NHP scores to start with (p=0.027) than those who had no improvement. Minor surgery results in improvements in both perceived and in objectively measured health and QoL. Both are valid outcome measures for minor surgery. Preoperative scores may be related to subsequent perception of outcome.

Health status measurement using validated measures and specially developed measures can yield valuable and unexpected prognostic, prospective and retrospective information on the process and outcome of treatment, on individuals and populations. There are important logistic problems in the gathering of accurate QoL data but the techniques potentially have an important role in allowing the surgeon to evaluate treatment and practice.

CHAPTER ONE

Introduction

1.1 Historical perspective

The teachings of Hippocrates provided the basis on which medicine was taught from the third century before Christ onward. His ethical ideal forms the foundation of medical practice today. He realised the importance of documenting all possible variations in each individual patient and then deducing their relevance to the disease condition:

"We must consider the patient, what food is given to him and who gives it..., the conditions of climate and locality both in general and in particular, the patient's customs, mode of life, pursuits and age. Then we must consider his speech, his mannerisms, his silences, his thoughts, his habits of sleep or wakefulness and his dreams, their nature and time."

He recognised that the outcome of medical treatment was not necessarily beneficial: "Practise two things in your dealings with disease: either help or do not harm the patient".

And even at this early stage in the evolution of professionalised medicine, he warned of the dangers of the quacks:

"If their patient be cured, their reputation for cleverness is enhanced, while, if he dies, they can excuse themselves by explaining that the gods are to blame while they themselves did nothing wrong." (Lloyd, 1978)

In the intervening two millennia, the documentation and categorisation of disease has reached great sophistication. The teaching of *physic* was conducted at the highest academic level in the great teaching centres of the classical world. Surgical treatments were described for many conditions including bowel obstruction and breast cancer. During the Renaissance, a new

era dawned. One of the most prominent surgeons in France, Ambrose Pare (1510-90), echoing Hippocrates, entreated the physician, "First do no harm" (De Moulin, 1989). His teachings on surgical technique reached a wide readership. In England and in Scotland, however, the barber surgeon was reviled by the learned *physicians*, as he was not classically or medically educated. The craft guild of Barber Surgeons of Edinburgh was incorporated by the Town Council in 1505, but surgery did not form a part of university medical education until the 18th century, during the "age of enlightenment". The golden era of hypothetico-inductive reasoning produced advances in understanding and in technique in all branches of medicine, and the art and the science of medicine became inextricable (Baum, 1989a).

The advances which have taken place up until the present day have embodied the early entreaties of Hippocrates. Until relatively recently, the results of treatment could be measured best in terms of death, disability or cure but more usually, simply the former or the latter. Despite the increasing sophistication of new medical and surgical treatments during the latter half of the 20th century, there are now many remedies available to the physician and surgeon which are not measurable in terms of survival. Remarkably little attention has been given to establishing means of measuring the effect of treatment- other than in terms of survival. Baum pointed out the pitfalls in relying upon surrogate outcome measures:

"...Improvements in biochemichal parameters or radiological signs, which may be used because they are quicker or easier to achieve and encourage us to lose sight of the simple issues of length and quality of survival." (Baum, 1989b).

A recurring worry to the profession is that there seems to be increasing dissatisfaction amongst the patient population with conventional medicine, as witnessed by the increasing demand for "alternative" remedies (Baum, 1989a). The old worry, of how to allocate priority in a medical system with finite resources, is becoming ever more pressing. So how are we to judge which treatments are the best to be offered, when, to whom and by whom?

1.2 What is to be measured?

The Scottish physicist Lord Kelvin (1824-1907) stated:

"When you can measure what you are speaking of and express it in terms of numbers you know something about it: when you cannot express it in terms of numbers your knowledge is of a meagre kind" (Duncan, 1985).

The simplest definition of life divides it into two dimensions- quantity and quality (Ware, 1987). Quantity is easy to define and measure. Regarding quality however, the definition alone is a matter worthy of much philosophical discourse (Pirsig, 1974) and defining quality in health care is, similarly, a taxing exercise (Donabedian, 1980a). Measurement cannot therefore be straightforward. Elkington referred to the concept of "Quality of Life" in 1966, as a goal aimed for by every physician for his patient (Elkington, 1966) but it was not until the mid-seventies that the term was to become widespread. The term was first coined by John F Kennedy's presidential commission which set goals for the USA for the year 2000 (Williams, 1991). Until 1966, the term was used with reference to post-war consumer activity (Alexander and Willems, 1981) although the ethos was clearly embodied in the definition of health included in the

constitution of the World Health Organisation in 1947:

"Health is not only the absence of infirmity, but also a state of physical, social and mental wellbeing" (WHO, 1947).

1.3 Quality of Life measurement

In the realms of chronic or incurable disease, it has long been recognised that more subtle factors than simply survival must be taken into account in assessing treatment. It was an oncologist, David Karnofsky, who developed the first measure of "performance" (Karnofsky and Burchanal, 1949). This is a 10-point scale which spans the extremes of physical dependency, as related to nursing burden, and measured by an observer. For many years this scale was used extensively, but not always appropriately, to justify or condemn treatments and not until some 20 years later was it improved upon, with the Activities of Daily Living scale (Katz *et al.*, 1970). In the two decades which have followed, a profusion of scales and questionnaires (called "instruments") have been developed, primarily to measure health status in patients with cancer or chronic disease, and primarily to compare treatments in the context of trials. They have in common the goal of *measurement*. However, the construct being measured differs depending upon the instrument. The terminology which has evolved to describe these is sometimes confusing. The terms "health status" and "quality of life" tend to be used interchangeably. However, Spitzer advises that the former term should be used to describe instruments which are applied primarily to healthy people, whereas the latter should be used for instruments which apply to the sick (Spitzer, 1987). Spitzer goes on to describe a third

category of QoL measures, those designed to test a specific hypothesis. Into these three broad categories, then, fall all known QoL instruments, now in excess of 50.

1.4 Quality of life instruments

Many instruments now exist, to measure QoL in general terms, or health status in specific illnesses. There are certain features common to all. Almost all are now self-assessment instruments, designed to remove the potential for observer bias. Poor observer-respondent correlation in health status and QoL measurement is a well documented phenomenon (Slevin *et al.*, 1988; Brewster and Newman, 1991). Epstein *et al.* found poor correlation between the scores of proxies and subjects compared during structured interviews incorporating various QoL instruments (Epstein *et al.*, 1989). Only a compelling reason such as extreme youth should allow a proxy or observer's QoL assessment to have precedence over that of the subject. All instruments consist of a variable number of *domains*, usually 1-6, which contain or gather information focused on a particular aspect of health and QoL. Ware defined five generic health concepts, or dimensions: physical health, mental health, social functioning, role functioning and general health perceptions (Ware, 1987). The Nottingham Health Profile, perhaps the best known QoL questionnaire, includes six domains: emotional reactions, energy, pain, physical mobility, sleep, and social isolation (Hunt *et al.*, 1985). Each domain contains *items*- which can be complete questions requiring a yes/no answer (dichotomous) or symptoms inviting the

selection of a gradated response on a *scale*. A scale can be categorical (eg. very much, somewhat, a little, not at all), or in the form of a visual analogue scale (VAS), in which a mark is put upon a straight line which spans the extremities of a particular symptom (Bond and Lader, 1974). There is no evidence that a VAS is superior or inferior to a categorical scale, nor evidence that an odd or even number of categories is preferable (Remington *et al.*, 1979), although McQuay favours a categorical scale where simplicity is paramount (McQuay, 1990). More than five points are superfluous on a categorical scale (Lissitz and Green, 1975). The arrival at a final score depends upon the format chosen. A categorical scale is usually given an ascending integer score corresponding to the severity with which the symptom or item is perceived (eg 0-4). Each item of a VAS is usually scored on a 10cm line, divided into 10 portions: the mark on the line is given a score 0-9 (or 1-10) depending again upon severity. Scores for individual items are usually aggregated to produce a simple score for that domain, but in some measures, eg. the Nottingham Health Profile, a weighting system is used, so that a positive response for each (dichotomous) item is multiplied by a factor before aggregation into a score for that domain (Hunt *et al.*, 1985). Aggregation of the scores from separate domains into a global QoL score is a prevalent practice, but in certain circumstances may detract from rather than add to the precision of the data.

The timeframe of a QoL instrument is the period which the respondent is being asked to consider in formulating a response. Many QoL measures leave the timeframe undefined. Alternatively, patients may be asked to consider the

previous day, week, or month before formulating a response (Aaronson, 1988). The periodicity or variability of the condition will dictate the choice of timeframe and for certain instruments, the timeframe is built into the instrument design, for instance a diary card.

1.5 Validity of quality of life instruments

QoL instruments in common usage have all been through a process of development. Briefly, the process involves selection of items, their reduction in number, design of format, pretesting, reliability testing (for repeatability), responsiveness (comparison with the change in other parameters, eg. objective toxicity) and validity. Validation takes the longest time in instrument development and needs time, and patient numbers. Face validity includes the clarity, user-friendliness and how comprehensively the subject is covered; criterion validity uses measurement against an empirical measure (eg a dyspnoea question measured against a treadmill)(Aaronson, 1988); construct validity is the predictability with which items and domains relating to eachother (Guyatt *et al.*, 1986).

As examples of instruments and their developmental processes, some attention will be given to instruments which appear later on in this thesis. The aforementioned Nottingham Health Profile was developed in the latter half of the 1970s and completed in 1981, as a "self-administered questionnaire designed to measure perceived health problems and the extent to which such problems affect normal activities" (Hunt *et al.*, 1981). In the initial stages, 2200

statements relating to ill health were compiled by researchers, eg. "I find it hard to walk upstairs". These were distilled down to 138 which were then used in pilot studies between 1976 and 1978 and reduced further to 82. Further work confirmed that the statements were sensitive to changes over time and to degrees of disability. The statements were refined conform to the yes/no format, exclude negative statements and ambiguity, and to conform to a minimum reading age. Each item was then tested on patients and non-patients for clarity, prior to a series of validation studies. These were on more than ten different groups of patients and healthy subjects, comprising over 5000 altogether. Reliability studies on two groups, one with peripheral vascular disease and the other with osteoarthritis, confirmed a high degree of test-retest reliability in chronic disease. Following this, different weighting was given to each statement using Thurstane's Method of Paired Comparisons, although this approach has since been criticised as inappropriate and leading to logically inconsistent results: eg. "I can only walk about indoors" (weight 11.54) plus "I have trouble getting up and down steps" (weight 10.79) outweigh "I'm unable to walk at all" (weight 21.30) (Jenkinson, 1991). Norms for various differing patient populations were developed, validation studies performed and the instrument has achieved a high level of acceptance in the intervening years. It has been used in many different languages, and validated in such diverse situations as the measurement of psychological disturbance in unemployed males, outcome following coronary artery surgery and results of total knee replacement.

The General Health Questionnaire, developed by Professor David Goldberg of Manchester University, was designed to "discover those features which distinguished psychiatric patients as a class from individuals in the community who considered themselves to be healthy". Items were gathered from interviews conducted by Veroff, Feld and Gurin (1962) with 542 non-hospitalised Americans, covering areas od "adjustment" and "felt distress", and factor analysis produced the following factors: "felt psychological disturbance, unhappiness, social inadequacy and lack of identity", which were then used to select items with high "saturation" of these factors. Items were drawn from other sources and from psychiatric colleagues. An extensive validation process included experimentation with format, which was eventually deecided upon as a four-option response to a statement, but not scored as a Likert-type scale (i.e. 0-1-2-3), but rather as a bimodal response (0-0-1-1). The score for each item is added to produce a total score. Several different lengths of questionnaire were developed in parallel and extensively validated in studies on hundreds of psychiatrically well and unwell subjects. The optimum usage of the instrument has been in using a threshold total score to indicate individuals in a hitherto unselected population likely to represent psychiatric "cases". During the past decade, between 10 and 20 studies every year have added to the data validating versions of the GHQ in various patient populations.

Linear analogue self-assessment (LASA) had been developed in the 1920s (Hayes and Patterson, 1921) and during the late 1960s and early 1970s was used to measure subjective responses in a variety of conditions including rheumatoid

arthritis, mental alertness and sedative effects in drug trials, (Bond and Lader, 1974). The principle is that the subject is presented with opposite extreme options of a symptom or item, spaced by a 10 cm line, on which they are asked to draw a vertical stroke at the point which approximates to how they feel. The length of line at the point of bisection is then measured. Bond and Lader (1974) were able to validate the technique in 500 patients, although were of the opinion that test-retest reproducability may be unstable in their particular patient population. This may, however, be a more general problem, as it is difficult for patients, on retesting, to remember exactly where a mark was made previously on a line. More so than, say, to remember the wording of a response previously chosen in answer to an item. Priestman and Baum (1976) used LASA in their early experiment measuring QoL in breast cancer patients, and as late on as 1988, evidence was still accruing that the LASA compared favourably in terms of validity with other QoL instruments (Boyd *et al.*, 1988).

The Hospital Anxiety and Depression Scale (Zigmond and Snaith, 1983) is a younger instrument than those mentioned above and has, therefore, not undergone the same extent of validation. It was developed as a means of identifying individuals, in a hospital population, suffering from recognisable psychiatric morbidity. It has been through a similar process of validation prior to its release, but up until the late 1980s, only a few studies had employed the instrument. It has the inherent advantage of being simple to use, occupying one side of A4 paper, and being easy to score on a Likert scale of 0-3. It is easily disaggregated into the components of Anxiety and Depression, and is perhaps

more typical of the newer instruments emerging, in that some attention to presentation and "user-friendliness" has made it an attractive instrument to use.

A problem for those considering QoL measurement for the first time is the absence of good information on how to process the enormous amount of data which accrue from QoL measurement. This contrasts with the extensive theoretical information about the relative merits of various constructs and weighting systems. In some of the earlier studies, data were analysed by parametric methods (Priestman and Baum, 1976) or were transformed logarithmically (Bond and Lader, 1974). In large studies, data may be normally distributed, but in recent years, as non-parametric statistical methods have also become more familiar, these have tended to be adopted as the "safest" way of handling data which cannot be regarded as normally distributed, especially if in small numbers. The appropriate statistical tests (Mann-Whitney, Wilcoxon, Kruskal-Wallis, Friedman) are simple and available in many readable books (Siegel, 1956; Swinscow, 1980; Gardner and Altman, 1990; Brown and Beck, 1989) and on many computer programs (Minitab, Systat).

Generally speaking, commonsense should prevail (Cox *et al.*, 1992). Measurements over time should be made against a baseline measurement for that patient. It should be remembered that confounding factors other than the construct under examination may also contrive to affect a change in QoL scores over time (eg a recent bereavement in a patient being tested for perioperative anxiety) so in general, unless group numbers are large, multiple measurements should be made over time. In their seminal paper, Fayers and Jones described

in detail certain principles of QoL data analysis but dwell little upon the statistical methods in use (Fayers and Jones, 1983).

1.6 When to consider quality of life measurement

QoL measurement was originally developed in order to assess treatment strategies which could not be assessed accurately in any other way. It can be applied on an individual level, but is more usually applied to groups of subjects. As in other studies, this has evolved into two main branches: cohort studies and controlled studies. In the former, including the quick dirty experiment of the Baconian variety, to obtain qualitative rather than quantitative information (Medawar, 1979), a pretreatment score is compared to a single post-treatment score, or series of scores. In the latter, the change in scores over time is compared for two or more treatment groups, commonly in the context of randomised trials designed and conducted in the usual manner.

A growing number of trials reported in the literature have incorporated some element of QoL measurement. This probably has more to do with the changing influence of social values on the medical profession than a sudden increase in new treatments or technologies. Between 1966 and 1970, four papers had QoL in the title. In the next five years there were 33 papers (Spitzer, 1987). In 1980, Mosteller *et al.* reported that of 132 trials in cancer, the overwhelming majority reported outcome only in terms of survival or recurrence (Mosteller *et al.*, 1980). Bardelli and Saracci reviewed cancer trials from 1956 to 1976 and found that in fewer than 5% was QoL measured and that

even then, objective functional measures such as the Karnofsky and Zubrod (Zubrod *et al.*, 1960) rather than subjective QoL measures were used (Bardelli and Saracci, 1978). Now, in the early 1990s, QoL has been a separate keyword in the Index Medicus for 15 years. Yet, of all the publications in which QoL is mentioned, only in a small minority is it actually measured.

Amongst the earliest attempts to use patient-derived QoL data was the study by Priestman and Baum who reported the use of linear analogue self-assessment (LASA) to measure QoL in breast cancer chemotherapy patients. They demonstrated that a reduction in tumour size was associated with improved QoL scores (Priestman and Baum, 1976). QoL measurement has developed in a variety of conditions other than cancer during the 1980s, including hypertension (Bulpitt and Fletcher, 1985), inflammatory bowel disease (Guyatt *et al.*, 1989) and rheumatic diseases (Liang and Robb-Nicholson, 1987).

1.7 Quality of life as a measure of utility

Another area of confusion to the casual reader of QoL literature is the use of QoL as a measure of utility. This area accounts, at least in part, for the increased interest in the subject and represents the attempts being made to find a common currency with which to compare, and ultimately cost, different health states and treatment interventions. One of the pioneers in this field is George Torrance, who developed the model of the "quality adjusted life year" (QALY) (Torrance, 1976). He uses a model of health/time, applying to health care the analysis of cost-effectiveness and cost-utility (Torrance, 1987). The

denominator is a Year; the numerator is a number between 0, which represents death and 1, which represents a state of optimum health. Torrance is careful to point out that in this model, QoL is *health related*- that is, without the social or "beyond the skin" considerations which are included in other QoL constructs. Torrance defines utility as "a cardinal measure of the strength of one's preference" and derives his use of the term from the Utility Theory model for decision making in the face of uncertainty (Holloway, 1979). The utilities are measured for various states of disease or health in two ways: as determined by patients actually in that state and hypothetically, by observers, who have experience of dealing with that health state. Torrance proposes that the resulting numbers, or QALYs be used in three ways: in clinical decision making, in group comparisons and in health policy formulation. The principle was seized upon by health economists at an early stage, but application of the utility theory to real life has resulted in some perverse consequences. For example, in the famous, or infamous "Oregon experiment", in response to severe rationing of resources, an attempt has been made to put medical services in some objective order of priority. Using the QALY approach, local citizens were canvassed in a telephone questionnaire using various trade-offs to rank some 2000 conditions in order of priority. Cosmetic breast surgery was ranked higher than treatment for a compound femoral fracture, casting doubt on the whole ranking process (Klein, 1991). Later in the same paper, Klein argues that:

"Technical exercises may be a useful way of starting up the dialogue and providing statistical scaffolding that may subsequently be dismantled, but they cannot resolve conflicts of values or interests".

Harris goes further, arguing that QALYs are unjust because they value time lived instead of individual lives and take an excessively narrow view of what QoL might be; he invites the question "is murdering someone with more QALYs worse than killing someone with fewer?" (Harris, 1988).

1.8 Quality of life measurement in surgery

So what has been happening in surgery and how does the surgeon stand in relation to the QoL research that developed in the last two decades? In 1987 it was reported that only 3% of trials reported in surgical journals mentioned QoL (O'Young J and McPeek B, 1987). Four years later, there has been some progress, although QoL studies have yet to become firmly established in surgical journals. In 1991, no study reported in the British Journal of Surgery contained QoL measurement of any description, even though several would have uncovered clinically relevant information by doing so. Why should this be, especially when the practice of audit has been deeply rooted in many surgical centres and is set to become established in many more?

The assessment of medical intervention, referred to as medical audit, has three components- structure, process and outcome (Donabedian, 1980b). The measurement of surgical interventions dwells almost exclusively upon the last- outcome. Surgical outcome is traditionally evaluated according to two basic, measurable parameters- mortality and morbidity. The latter can be broken down into several components- operation time, re-operation rate, haemorrhage, wound healing, infection and readmission. Most surgeons have found that these

measures, together with the informal and variable feedback on functional results, accorded by the doctor-patient encounter in the outpatient clinic, give all the information needed with which to audit their practice. Bunker and Wennberg drew attention to the inappropriateness of using mortality data to measure certain aspects of surgery:

"It is to the improvement in the quality of life- to the relief of disability, discomfort and disfigurement that elective surgery is primarily directed" (Bunker and Wennberg, 1973).

Ebbs *et al.* cited the example of appendicectomy as an instance where QoL measurement would be superfluous: there is only one treatment: the patient has an appendicectomy and gets better (Ebbs *et al.,* 1989a). However, the advent of laparoscopic surgery has made even this a debatable assertion.

Goligher acknowledged that:

"Surgeons are often accused of adopting too simplistic an approach to the outcome of their handiwork, and of being all too frequently content to judge the results in terms of operative mortality, immediate operative morbidity and, in the case of operations for malignant disease, the length of survival".

He advocated the use of operation specific assessments using a specially formulated and searching questionnaire (Goligher, 1987). One of the first surgeons to attempt *measurement* of the functional results of operation was Visick, who conceded that adverse results may arise from gastric surgery and so devised a simple scale with which to take account of postoperative function (Visick, 1948). It was Goligher *et al.,* in the first randomised studies by surgeons, who used this scale to assess their results (Goligher *et al.*, 1968). In the period since this important study, Goligher was to use this approach to assess *function* as a measure of the outcome of surgery to other operations.

Other surgeons have sought to measure the effect of different treatment modalities using techniques which include *psychological* assessment- by observer and by self-assessment scales. In the study by Priestman and Baum on patients with advanced breast cancer receiving chemotherapy, they developed their own instrument, the LASA (Priestman and Baum, 1976). The adverse psychological consequences of mastectomy were reported by Morris *et al.* (1977) and the following year by Maguire *et al.* (1978). However, in a prospective study of 101 patients randomised to receive mastectomy or breast conserving surgery, detailed psychological QoL measurement by a trained interviewer established that morbidity was equally high in both groups, most probably as a result of the cancer itself (Fallowfield *et al.*, 1986). The same year, the psychological impact of post-mastectomy chemotherapy was reported (Hughson *et al.,* 1986). The same group reported no difference in psychiatric morbidity conferred by postoperative radiotherapy (Hughson *et al.*, 1987). In these three studies, psychological assessment was done using both a trained interviewer and questionnaires- the Hospital Anxiety and Depression scale (or its precursor)(Zigmond and Snaith, 1983) in all 3, the Rotterdam symptom checklist (Pruyn *et al.*, 1981) in the first and the General Health Questionnaire (Goldberg, 1979) in the other two. In 1988, Koivukangas and Koivukangas reported their study into the QoL of patients undergoing surgery for infiltrative brain tumours. They started using simply the Karnofsky scale (Karnofsky and Burchenal, 1949) and the Glasgow Outcome Score (Jennett, 1986) but went on to develop their own scale, using a multidimensional approach based mainly on semi-structured

interviews repeated at intervals from before treatment (Koivukangas and Koivukangas, 1988).

One of the most influential surgeons to promote the role of QoL measurement in surgical practice has been Hans Troidl who said:

"It is a curious paradox that research reports on the effectiveness of surgery focus on mortality, length of hospital stay, major complications and laboratory analyses, when the principle criteria guiding surgeons' clinical decisions are most often the patients' subjective feelings and physical capabilities, on the quality of life."

He cited the examples of inguinal hernia repair and oesophageal surgery for symptomatic reflux as operations undertaken to improve QoL. Whilst he acknowledged the role of the surgeon's conscious and unconscious estimation of a patient's QoL before and after surgery, he stated the urgent need for better techniques to allow surgeons to measure QoL in a simple and practical way (Troidl *et al.*, 1987). He used this approach to assess outcome in gastric cancer patients having total gastrectomy and either oesophagojejunostomy or construction of a Hunt-Lawrence-Rodino pouch. As well as designing a specific questionnaire, he used a modified version of the Visick scale, as well as general QoL measures such as the Spitzer QoL index (Spitzer *et al.*, 1981). He advocates the use of specifically designed questionnaires to assess the precise outcome of operation, together with more general QoL measures to answer the question "How are you?" This approach has been advocated independently by Aaronson (1988) and Ware (1989).

Perhaps the most significant impediment to the acceptance of QoL measurement as a *bona fide* branch of research was vividly described by

Feinstein: namely that it is regarded as "soft science" by those who are used to measuring phenomena in laboratory conditions (Feinstein, 1977). He describes the science of "clinimetrics": arbitrary ratings, scales, indexes, instruments or other expressions that have been created as "measurements" for clinical phenomena that cannot be measured in the customary dimensions of laboratory data (Feinstein, 1987a). As examples, he cites the Glasgow Coma Scale (Teasdale and Jennett, 1974), the Apgar score (Apgar, 1953), the TNM staging system (American Joint Committee for Cancer, 1977) and the Visick scale (Visick, 1948)(Feinstein, 1987b). Troidl pointed out the value in terms of improved surgical practice that the Visick scale, a simple measure with a weak methodological background, has produced. He later observed: "Clinicians prefer so-called hard data, even when they are totally irrelevant in measuring the outcome of a patient, and they ignore the so-called soft data, like pain, fatigue and quality of life" (Troidl, 1991).

Even traditional gold standards such as Dukes staging for colorectal cancer can be heavily observer-dependent and subject to considerable variation.

1.8 Health status measurement in surgical practice

There is now a great incentive for surgeons to undertake detailed measurement of the effect of their work upon individual patients and upon the population.

The United Kingdom National Health Service reorganisation with the purchaser-provider split and the introduction of medical audit has focused this. Health economists are now scrutinising every aspect of surgical provision. Surgical Audit, hitherto done at the discretion and the pace of the individual

surgeon, is now being established with specific funds allocated from regional health budgets. But, in order to develop this, better measures of *outcome* and *process* must be found. If surgeons need to measure their practice more accurately, they need to know more about the instruments which may help them with this task.

1.10 Hypothesis: Health status measures, scientifically applied, provide important additional information to the surgeon.

QoL measures have been used in four main contexts: *comparing two or more treatment interventions in a clinical trial, measuring the health of populations, determining treatment strategies for the individual patient and assessing the benefit of alternative uses of resources* (Cox *et al.*, 1992). A broadly based enquiry was chosen because of the uncertain role of health status measurement in relation to surgery, and to replicate as nearly as possible the conditions in which a surgeon might wish to conduct research using health status measurement. In order to test the hypothesis it was therefore necessary to apply a variety of techniques to different patient populations, with a variety of objectives for QoL measurement. It was necessary to adapt techniques and instruments which have been established in other disciplines. The goal was to gain wide practical experience of QoL measurement, the endpoint being to formulate guidelines for further use of health status measurement in surgical practice. The models were chosen in common conditions, in three areas in which the author had the most ready access to clinical experience. The first

three of the four broad contexts were studied: the assessment of alternative uses of resources was not addressed, other than in specific areas relevant to each separate study. In each of the studies, the overlap in QoL measurement ojectives is discussed in the final chapter.

1.10.1 Evaluation of treatments in a clinical trial

The first study entailed QoL measurement in patients with advanced breast cancer undergoing chemotherapy. This is an area in which perhaps the most experience of QoL measurement has been reported hitherto. It was considered appropriate to include this subject in a dissertation on health status measurement in *surgical* practice, because there is, as yet, no consensus as to when to offer chemotherapy to patients who develop advanced breast cancer and the decision is often made by surgeons. The view of some surgeons who do not favour chemotherapy for advanced disease is that the distress of chemotherapy is unwarranted where no survival advantage is anticipated. In many centres, the surgeon also administers the chemotherapy; the argument against chemotherapy in advanced breast cancer patients may be reinforced where the anticipated increase in workload involved in giving chemotherapy is not perceived by the surgeon to benefit the patient in any obvious or measurable way. Amongst those who are convinced of its worth, there is still significant doubt as to the optimum type of treatment to be offered to these patients. Yet at the same time there seems to be considerable, though not overtly expressed, resistance to the concept of QoL measurement as judged by the number of trials in which QoL is

measured.

A study was therefore initiated at King's College Hospital, London and the William Harvey Hospital, Ashford, Kent in order to investigate the QoL of patients with advanced breast cancer receiving an aggressive chemotherapy regimen and those receiving a mild regimen, using well validated instruments in addition to the traditional recordings of response, toxicity and survival.

1.10.2 Validation of a new QoL instrument

In addition validation was undertaken of a new QoL instrument designed specifically for such trials in the Department of Surgery at King's College Hospital. This aspect of the study was augmented by data from a trial of different chemotherapy regimens at Guy's Hospital, London to allow comparison of the instrument between studies.

1.10.3 Measuring quality of life to establish a difference in patient populations

The second study is of patients admitted with acute abdominal pain, a subject extensively discussed in the literature. It has been suggested that psychosocial factors are influential in the presentation of patients whose pain will subsequently resolve undiagnosed. This was therefore investigated prospectively at the William Harvey Hospital, Ashford, Kent using validated psychological health status questionnaires and clinical assessment.

1.10.4 Measuring the result of surgical treatment in individual patients

The final study was prompted by the debate surrounding waiting lists for minor surgery, or even removal of minor surgical procedures from NHS waiting lists. A prospective study was undertaken in order to determine whether the results of such treatment can be assessed using QoL measurement prospectively and whether, in individual patients, different outcomes can be detected.

CHAPTER TWO

Aggressive or mild chemotherapy for advanced breast cancer? CMF *versus* Epirubicin measuring quality of life with three different instruments

2.1 Summary

Forty patients with advanced breast cancer, randomised to receive CMF or weekly low dose Epirubicin, were evaluated by UICC criteria of response and WHO toxicity criteria, in addition to three QoL instruments: the "Qualitator" daily diary card, monthly Nottingham Health Profile (NHP) and Linear Analogue Self-Assessment (LASA). Response rates were 58% for CMF and 29% for epirubicin (x^2=3.51,1df,p>0.05). Median time to treatment failure was 24 weeks for CMF, 7 weeks for epirubicin (p<0.05) but survival was similar in both groups. Survival was better for responders than for non-responders (medians 87 and 30 weeks, p=0.02). CMF caused more objective alopecia (p<0.001), nausea and vomiting (p<0.001) and haematological toxicity (p<0.02). However, QoL measures only recorded a significant difference in energy and pain, influenced primarily by the non-responders in each treatment group but with no difference in overall global scores. Scores for responders, irrespective of treatment, were better to start with (LASA p=0.001); at three months scores had improved (Qualitator p=0.021; NHP p=0.041). Scores in non-responders showed no change. In this small study aggressive chemotherapy gave better response and survival without impairing Quality of life overall. Detailed QoL measurement should be integral to all cancer chemotherapy trials.

2.2 Introduction

The treatment of patients with advanced breast cancer using combination chemotherapy can cause significant toxicity without greatly prolonging survival (Powles *et al., 1980*: A'Hern *et al.*, 1988). Recently, studies have been reported in which low-toxicity regimens (single agent or short term) have achieved palliation without affecting survival (Chlebowski *et al.*, 1989; Harris *et al.*, 1990). For example, Jones has reported a response rate of 43% with Epirubicin given with a weekly dose of approximately 20mg. No significant myelosuppression, and minimal nausea and alopecia resulted (Jones, 1988). Further studies have shown no improvement in response rates by doubling the weekly dose from 20 to 40mg. There was, however, a considerable increase in toxicity (Ebbs *et al.*, 1989).

There is a danger that such low toxicity regimens may be accepted without adequate comparison with conventional combination cytotoxics. One of the most widely used regimens in advanced breast cancer is the standard Cyclophosphamide, Methotrexate and 5-Fluorouracil (CMF) treatment which achieves response rates of up to 60% (Bonadonna *et al.*, 1983). This was therefore chosen as the control arm of a direct comparison with low-dose weekly epirubicin. As reduced toxicity was central to the development of the low-dose regimen, the trial was planned around detailed measurement of Quality of Life.

2.3 Objectives

1 Comparison of Epirubicin with CMF using conventional criteria

2 Comparison of Epirubicin with CMF using QoL measures

3 Assessment of QoL measurement as a suitable way of measuring the results of chemotherapy

2.4 Patients and methods

2.4.1 Patients

Between October 1988 and December 1989, forty patients with advanced breast cancer attending the Breast clinics at King's College Hospital and the William Harvey Hospital were randomised to receive CMF or Epirubicin as first line chemotherapy. Criteria for inclusion were: histologically proven locally advanced disease, rapidly progressing primary disease, metastatic disease failing to respond to hormonal measures, a first recurrence which was visceral, or recurrent disease less than 2 years from primary treatment. Excluded, were postmenopausal women with locally advanced disease suitable for a trial of tamoxifen, those with a significant medical condition or known previous or current cardiovascular disease and patients who had received non-adjuvant chemotherapy. The two groups were evenly matched according to the sites of disease, and menopausal status, although there was a difference in their median ages which was not statistically significant (see Table 2.1).

	Epirubicin	CMF
Number	21	19
Median Age	52 (26-80)	63 (39-84)
Premenopausal	9	4
Postmenopausal	12	15
Sites: Soft Tissue	10	10
Nodal	9	12
Lung	6	6
Liver	6	6
Bone	9	9
QoL: NHP	129	91
LASA	53.5	35
Qualitator	64	75

Table 2.1: Characteristics of patients recruited to the CMF/Epirubicin trial, including sites of disease and initial quality of life scores

---

2.4.2 Ethical considerations

The trial was approved by the ethical committees in both participating hospitals. Written informed consent was obtained from the patients prior to randomisation.

2.4.3 Treatment

All therapy was given in the outpatient clinics by one person. The dose schedules were: 1) Epirubicin 20mg intravenously, given into fast-running 0.9% saline every 7 days. 2) Cyclophosphamide 100mg/m^2 orally on days 1-14,

Methotrexate 35mg/m^2 intravenously on days 1 and 8 and 5-fluorouracil 600mg/m^2 intravenously on days 1 and 8, on a 28 day cycle. Anti-emetics were given parenterally or orally as appropriate. In practice, intravenous Metaclopramide 10mg was given prophylactically to every patient receiving CMF at the time of cytotoxic administration and Prochloperazine in tablet or suppository form was given on request to patients to take at home. Dose reductions were made for patients over 65 years old and dose modifications made if the WBC fell below $3000/l^{-6}$ or platelets to below $100/l^{-6}$. One patient on CMF experienced mucositis for which she was given Calcium Folinate 15mg every 6 hours for 24 hours.

2.4.4 Assessment of Disease

The endpoints chosen were: Time to treatment failure, survival, International Union Against Cancer (UICC) response criteria (see Appendix 1), World Health Organisation (WHO) toxicity criteria (see Appendix 2) and Quality of Life. Time to treatment failure was defined as the time to progression of lesions either on measurement or symptomatically requiring addition to or alteration in therapy, or the abandonment of treatment due to toxicity. If treatment failure occurred before completion of a six-month course of treatment, alternative therapy was given as appropriate. After six months, chemotherapy ceased and no treatment was given until or unless recurrence occurred or disease progressed. Clinical and laboratory measurements made at entry to study were a full medical history and examination, weight, height, age,

date of birth, PMH, full blood count, differential WBC, biochemical screen. Photographs were taken of visible lesions and records made of tumour dimensions. All patients had a bone scan, liver ultrasound scan and chest radiograph. CT scan was performed in patients whose lesions were not otherwise measurable. Quality of life assessment was made using the Nottingham Health Profile (Hunt *et al.*, 1885)(see Appendix 3) and Linear Analogue Self Assessment (Priestman and Baum, 1976)(see Appendix 4) at the start of treatment and four weekly thereafter; throughout treatment, patients completed the Qualitator daily diary card (see Appendix 5), a new instrument developed for breast cancer chemotherapy trials (Fraser *et al.*, 1990)(see Appendix 13). Full blood count was measured prior to administration of intravenous cytotoxics. Patient characteristics were compared using the Chi-squared and t-tests.

2.4.5 Survival and response analysis

UICC criteria of response were assessed monthly. The WHO toxicity criteria were recorded every four weeks. UICC response rates were compared using the Chi-squared test and time to treatment failure and survival analyses were done using the Kaplan-Meier life table method (Kaplan and Meier, 1958) and log rank test (Peto *et al.*, 1977). Correlation between initial QoL scores and survival was done using Spearman's rank correlation method.

2.4.6 Quality of life analysis

With all instruments, a high score indicates poor QoL. The NHP scores were analysed as recommended by the authors (Hunt *et al.*, 1976) so that at each completion, a weighted score out of a possible 100 was obtained for each of the six components: emotional state, energy, pain, physical mobility, sleep and social factors. In this study, the components were then added to give a global score range of 0-600. The LASA questionnaire consisted of 26 categories, each scored 0-9 on a visual analogue scale. Two categories, the "open" item and the general statement on QoL were excluded from analysis, as the former was ignored by most patients and the latter was judged to duplicate the rest of the questionnaire. The global range was therefore 0-216. Both NHP and LASA were compared between patient groups at each juncture using the Mann-Whitney-U test. Comparison with subsequent scores was performed using the Wilcoxon rank test. Completion of the Qualitator involves the choice of 5 symptoms from a menu of 23, in 4 domains, scoring on a categorical scale 1-4. The details are described in chapter three; the range of the weekly global score is from 35-140. For comparison, pre-treatment NHP and LASA scores were compared with the first week of the Qualitator and thereafter, the comparison of NHP and LASA monthly scores was with an average of each patient's aggregated Qualitator scores for that month. Analysis was then performed using the same non-parametric methods as for the NHP and LASA. Analysis of individual Qualitator symptoms is also described in chapter three.

2.4.7 Exclusions

Forty patients were entered into the trial. Thirty seven patients completed the NHP and 36 the LASA at the start of the study. Three exclusions were patients who were unable to start treatment following randomisation and subsequently left the study. The other LASA was incorrectly completed by the fourth patient. Thereafter, patients remaining in the study completed the NHP and LASA during each month of treatment. Three CMF patients failed to do so at 1 month and one at 5 months; one patient failed to complete them at 4 months. The Qualitator was commenced by 29 patients. At the start of the study three elderly patients were, in retrospect mistakenly, not offered the Qualitator. One patient, once randomised refused to complete it, one progressed rapidly after 1 month and was unable to return the card. The remaining six patients progressed rapidly within a week of the start of treatment and were also unable to return the diary cards.

2.5 Results

2.5.1 UICC response

The response rates according to UICC criteria were 58% for the CMF group and 29% for the Epirubicin group (x^2=3.51,1df, p>0.05, see Table 2.2).

UICC Response	CMF	Epirubicin
Complete	1	0
Partial	10	6
No change	2	7
Progression	3	5
Rapid progression	3	3

Table 2.2: UICC response of patients according to treatment randomisation

---

If the six patients who relapsed before or within the first week of treatment are excluded as in other studies the difference is significant (x^2=4.30,1df,p<0.05). The time to treatment failure was longer for CMF patients than Epirubicin patients: medians 24 weeks and 7 weeks (X^2=5.17,1df, p<0.05, see figure 2.1).

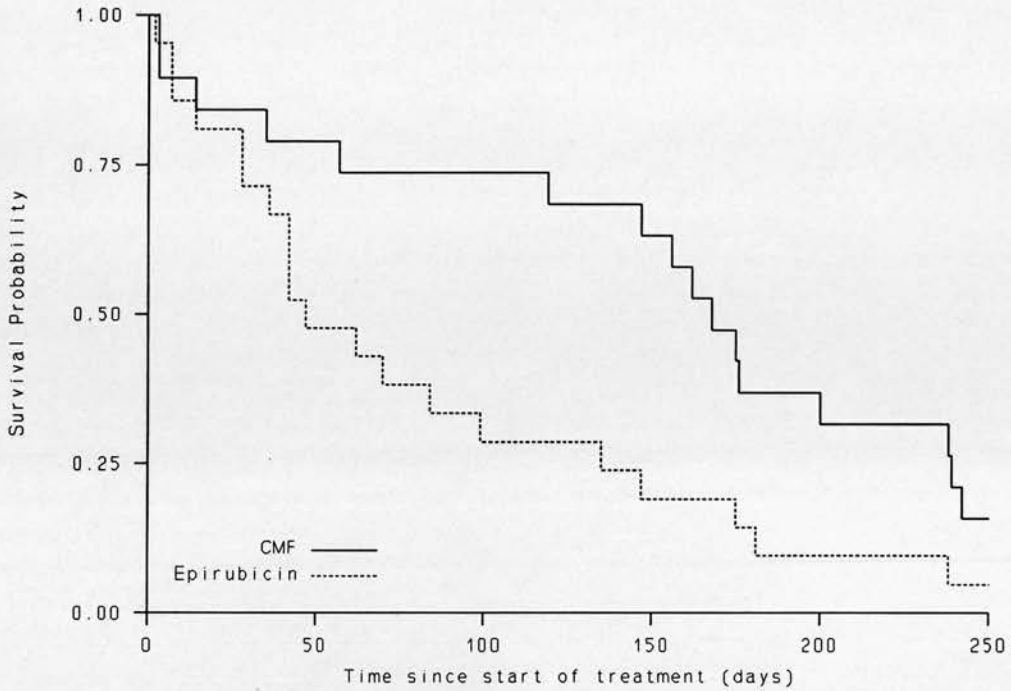


Figure 2.1: Life table indicating time to treatment failure. Patients grouped according to allotted treatment (Epirubicin n=21, CMF n=19)

Survival was similar in both treatment groups: medians 57 weeks and 55 weeks respectively (x^2=1.38,1df,p=0.24) (see figure 2.2).

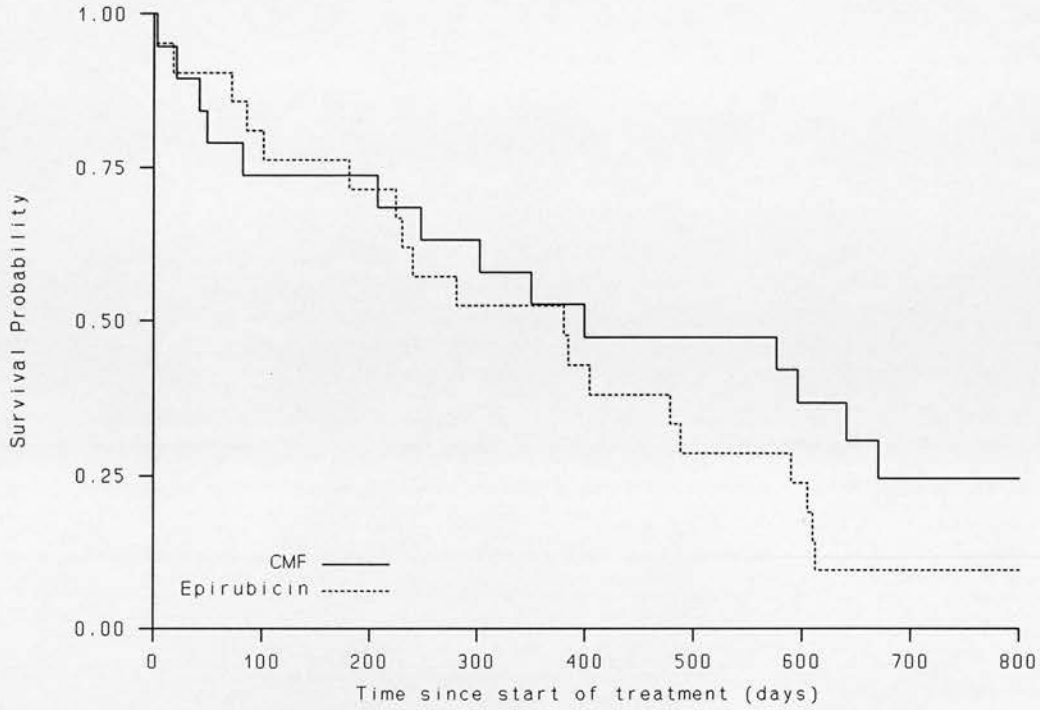


Figure 2.2: Life table indicating overall survival. Patients grouped according to allotted treatment (Epirubicin n=21, CMF n=19)

UICC responders, as expected from many previous studies (A'Hern *et al.*, 1988) survived longer than that non-responders: medians 87 weeks and 30 weeks (X^2=5.42, 1df, p=0.02, see figure 2.3)

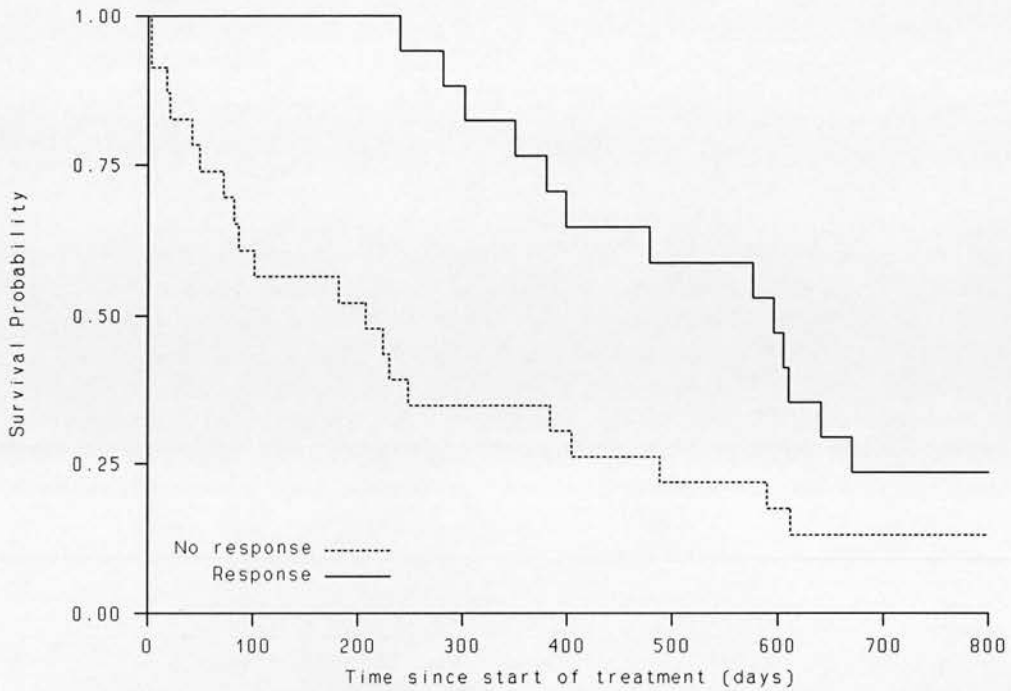


Figure 2.3: Life table indicating overall survival. Patients grouped according to UICC response (No response n=23, Response n=17)

Toxicity was very low for all patients receiving Epirubicin. CMF caused significantly more alopecia ($p<0.001$), nausea and vomiting ($p<0.001$) and haematological toxicity ($p<0.02$) above WHO grades I (see Table 2.3). One CMF patient required hospital admission for treatment of septicaemia. One Epirubicin patient receiving prednisolone for scleroderma developed septicaemia requiring hospital admission. There were no fatalities due to side-effects of treatment.

Rx	Total	WHO grade	Alopecia	Nausea or Vomiting	Haemato-logical
Epirubicin	83	0	75(90)	83(100)	82(99)
		1	8(10)	0	0
		2	0	0	0
		3/4	0	0	1(1)
CMF	106	0	43(41)	64(60)	75(71)
		1	20(19)	22(21)	20(19)
		2	12(11)	10(9)	7(7)
		3/4	31(29)	10(9)	4(4)
			p<0.001	**p<0.001**	**p<0.02**

Table 2.3: Toxicity according to WHO grade: number (%) of each treatment group in each toxicity category, on each month

---

2.5.2 Quality of life at entry to the trial

The respective NHP, LASA and equivalent aggregated weekly Qualitator scores were compared for each month. Patients' QoL scores were analysed according to response and to treatment. Prior to the start of treatment, a poorer QoL was recorded amongst patients who subsequently did not respond, statistically significant only for the LASA, ($p<0.002$). The pre-treatment scores are illustrated in figure 2.4, in which the LASA, NHP and Qualitator scores are standardised to a scale of 0-10.

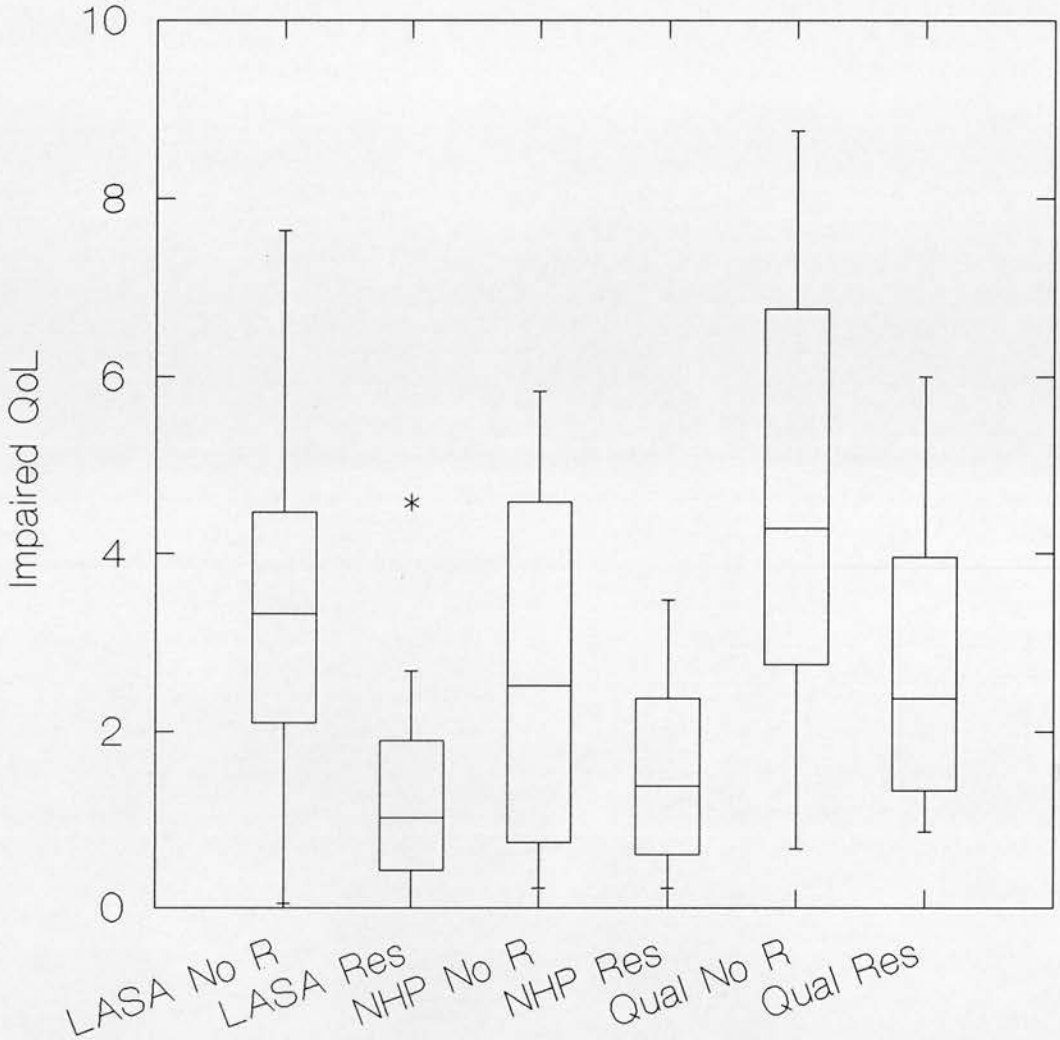


Figure 2.4: Box plot depicting QoL scores at start of treatment according to subsequent response. The data for each instrument are arithmetically adjusted to be comparable on a simplified range of 0-10.

Qual=Qualitator, NHP=Nottingham Health Profile, LASA=Linear Analogue Self-Assessment, Res=Response, No R=No Response.

For key to interpretation of box plots used in this thesis, please refer to Appendix 6.

Patients' Qol scores at the start of the study were correlated by rank with their subsequent survival. The Spearman co-efficients were -0.52 (95% c.i.-0.72,-0.23) for the LASA, -0.35(-0.60,0.04) for the NHP, -0.64(-0.82, -0.36) for the Qualitator.

2.5.3 Quality of life during treatment

Compliance for the 29 patients who started the Qualitator, the 37 who started the NHP and 36 who started the LASA respectively were 88%, 89% and 92%. Figure 2.5 shows the mean global QoL values in each treatment group at each stage for all patients remaining in the study. The means are used purely for graphic representation: statistical comparison between treatment groups was by a rank test at each month.

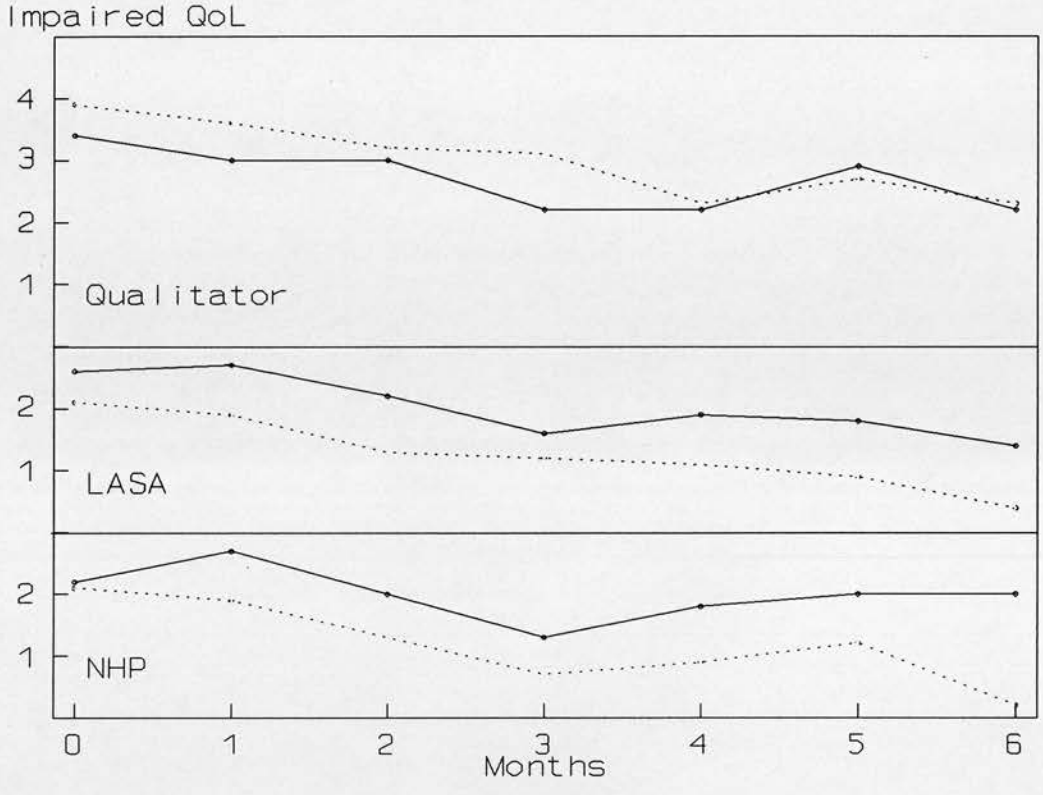
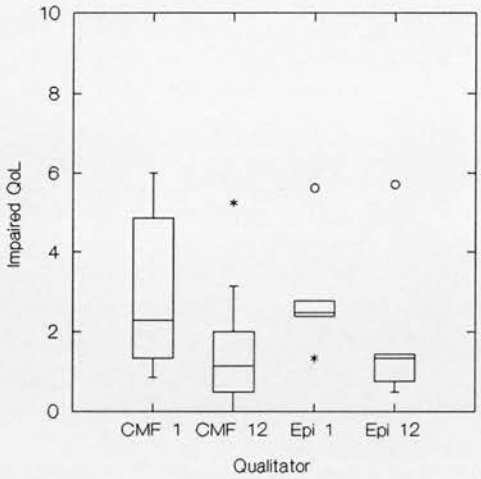


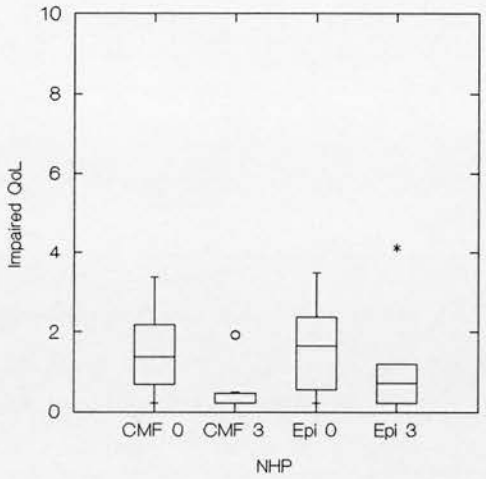
Figure 2.5: Mean QoL scores at each month of treatment for Epirubicin patients(_____) and CMF patients(.....) for each of the three QoL instruments used. The data for each instrument are arithmetically adjusted to be comparable on a simplified range of 0-10.

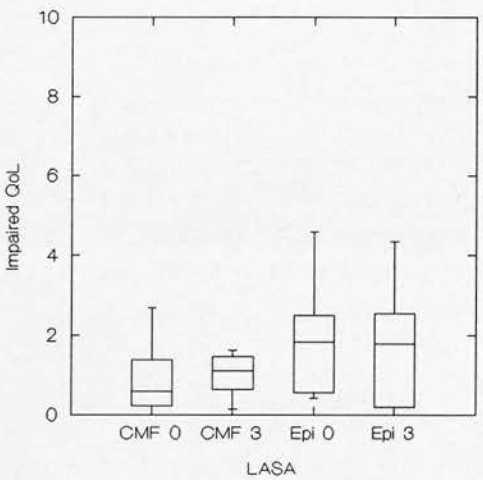
Table 2.4 shows the median improvement in QoL score for the 29 patients who started the Qualitator, the 37 who started the NHP and 36 who started the LASA.

	Quali-tator			NHP			LASA		
Month	1	2	3	1	2	3	1	2	3
Epirubicin	4.35 p<0.05	4.8	7.15 p<0.02	-36.3	-15	17.5	-6	5.5	4.5
CMF	1.6	0.55	8.93 p<0.02	10.5	26.5	53 p<0.02	1	-4	-1
Response	3.75 p<0.06	8.55 p<0.05	12.5 p<0.01	7	14	39.5 p<0.02	-3.5	5	0
No Response	1.95	0.1	12.5	-41.5	-23.8	17	-4	-5	0.5

Table 2.4: Median improvement in QoL score for all patients completing each questionnaire during first 3 months of study (compare Table 2.6)







Figures 2.6,2.7,2.8: The change in QoL scores in patients who responded:

At 1 week and 12 weeks for the Qualitator (15 patients)

Before treatment and at 3 months for the Nottingham Health Profile (17 patients) and Linear Analogue Self-Assessment (17 patients)

The data for each instrument are arithmetically adjusted to be comparable on a simplified range of 0-10.

Non-responders experienced no significant difference in their initial scores and the final scores prior to treatment failure: Qualitator medians 80 to 74 (p=0.5), NHP medians 133 to 182 (p=0.435), LASA medians 64 to 71 (p=0.55). The pre-treatment difference in scores between responders and non-responders

persisted on each monthly comparison: one month (LASA $p<0.02$, NHP $p<0.01$, Qualitator $p<0.05$), 2 months (Qualitator $p<0.05$) 3 months (NHP $p<0.05$, Qualitator $p<0.01$) and 4 months (NHP $p<0.05$).

All of the QoL measures allow sub-analysis in considerable detail. In separate analysis of the six components of the NHP (emotional state, energy, pain, physical mobility, sleep and social factors) and the LASA symptoms in four sub-groups (physical symptoms, social factors, psychological factors and physical performance), non-responders had worse scores, significantly at most stages except for the NHP emotional state and energy. The only significant difference was a better score in CMF than Epirubicin patients in the NHP score for pain at 2 months (median 9.5, $p<0.05$), energy at 3 months (median 24, $p<0.05$) and a worse Qualitator score at 3 months for personal relationships in CMF patients (median 0.65, $p<0.05$). In each case the high scores were amongst the non-responders in each group.

2.6 Discussion

One of the most difficult decisions facing clinicians treating patients with advanced breast cancer is what to do when second line hormone therapy fails. At what point does one advise chemotherapy, to whom and how aggressively? Until recent years, the success of a treatment regimen has been defined almost solely by tumour shrinkage. Although toxic side effects have been measured, there was little evidence of correlation with the patient's experience. The failure of many studies to show a survival advantage to any

regimen caused some clinicians to question the merits of giving chemotherapy at all (Powles *et al.*, 1980). During the last decades, the concept of Quality of Life has become increasingly important in those patients in whom little survival advantage is anticipated through treatment and efforts were made to define and measure it (Fallowfield, 1990). Increasing numbers, but still a minority, of studies measure QoL (Byrne, 1992). The disparate instruments and periods of measurement have made it difficult to interpret how chemotherapy affects QoL for patients with advanced breast cancer. The aim of this study was to compare a standard combination regimen with a single agent regimen in which different toxicity and possibly different response rates could be anticipated, and whether a difference in survival or QoL would result. Detailed intermittent QoL measurement was made with three instruments, two of which were specifically designed for the task. The response data were consistent with previous studies in that the patients who had a measurable response enjoyed longer overall survival. Although survival amongst patients with non-progressive disease was better for CMF patients, the poor survival of CMF non-responders was enough to redress this balance so that survival for the two treatment groups as a whole was equal. Few studies are large enough to show a survival difference between treatment groups, but A'Hern *et al.* showed that a better response rate equated with longer median survival in a statistical overview of 50 chemotherapy trials (A'Hern *et al.*, 1988).

The QoL data were not wholly expected. Although Ebbs *et al.* (1988) had reported that good pre-treatment QoL scores were associated with a subsequent

response, we found that there was a close correlation with subsequent duration of survival too. Morris and Sherwood (1987) described this in terminally ill patients, and Addington-Hall *et al.* (1990) used the Spitzer QoL Index (Spitzer *et al.*, 1981) to predict duration of survival in 230 terminally ill patients. However, it was a surprise that even in this small study, such a consistent trend would emerge. In the context of patients with advanced breast cancer, this may be of significance in deciding on treatment.

Low objective toxicity in patients treated with Epirubicin was reflected in the recording of specific treatment-related symptoms in the Qualitator, but QoL scores overall were unaffected and resembled closely the global scores of the other two instruments. QoL improved for responders in both groups from the start of the study onwards but did not alter for non-responders.

Is a harsher regimen therefore the treatment of choice for advanced breast cancer? The evidence is that it does not impair QoL in non-responders of whom there are fewer anyway and QoL improves for responders. A similar conclusion was drawn by Coates *et al.* (1987) who found that Quality of Life declined significantly in patients on the less aggressive regimen, in which response was poorer. Moreover, Slevin *et al.* (1990) found cancer patients much more willing to contemplate radical chemotherapy than were their doctors for them. However, if pre-treatment QoL scores give not only a guide to response, but to survival as well, then perhaps those patients with clinically advanced disease in whom QoL is poor, who will not respond and whose survival will be poor should not be given chemotherapy at all. A different interpretation of

these findings might be that those patients whose disease is not yet advanced enough to affect their QoL are those most likely to respond to treatment; one could go further, and suggest that a prolongation in survival is independent of any effect afforded by treatment. Therefore, QoL measurement should be used to help define a treatment strategy, early on in the advanced disease.

One way of resolving the difficulty would be to involve the patient more fully in the decision-making process. This is an approach which has recently been advocated by Wennberg in the United States. He and colleagues are conducting pilot experiments into interactive videotapes on early breast cancer treatment in Hanover, New Hampshire (Wall Street Journal, 1992). A study in which the findings also support this approach was reported in the New England Journal of Medicine by Cassileth *et al.* (1991). Patients with metastatic cancer of bowel, lung, pancreas or melanoma, who received conventional therapy, including chemotherapy, had no better survival than matched controls having "alternative" therapy. Chemotherapy was not associated with a *worse* QoL, and although the change in QoL was similar in both groups, the patients treated conventionally started and finished with better QoL measurement. The authors suggested that this may be accounted for by a difference in the social composition of the groups: a higher number of alternative therapy patients had degrees and it may have been that poor QoL contributed to the decision to seek unproven therapy. They concluded that the ideal study in such patients would be randomised, with a no-treatment arm involving only palliative care.

The present study does not provide solutions to these uncertainties.

However, detailed QoL measurement is shown to add valuable and perhaps not wholly expected information in evaluating advanced breast cancer chemotherapy. If nothing else, measurement of QoL would constitute a cheaper alternative trial endpoint to the battery of tests required to fulfil UICC criteria of response. There is a compelling case for inclusion of QoL measurement in all protocols. Only thereby will knowledge of QoL measurement accrue and its precise role in the clinical decision making process become clear.

CHAPTER THREE

Validation of the Qualitator daily diary for quality of life measurement in advanced breast cancer trials

3.1 Summary

The Qualitator daily diary card, mentioned in chapter two, was designed to measure Quality of Life in chemotherapy trials for patients with advanced breast cancer. In addition to the trial at King's College Hospital and the William Harvey Hospital in which 29 patients completed the Qualitator, 31 patients completed the precursor diary card to the qualitator in a separate study at Guy's Hospital. The Qualitator offers accurate prognostic data regarding subsequent UICC response and survival and is simple to use.

3.2 Introduction

The use of combination cytotoxic chemotherapy as palliation for patients with advanced breast cancer became established in the late 1960s (Cooper *et al.*, 1969). Few trials show a survival advantage for a particular regimen and only recently has an overall improvement in survival been associated with treatments giving higher response rates (A'Hern *et al.*, 1988). Although the aim of treatment is to improve the Quality of Life of the patient, regimens are still compared on the basis of their response rate in patients where such measurements can be made. Side effects of chemotherapy such as alopecia, vomiting and lethargy are assumed to affect the QoL of the patients, but their objective measurement is a secondary aspect in most trials. Subjective, patient derived, measurements are seldom made.

The simple technique of QoL measurement using the patients' subjective symptoms using visual analogue scales was adapted for use in breast cancer patients by Priestman and Baum (1976). They reported a significant improvement in QoL scores in patients whose tumour area reduced (Baum *et al.*, 1980). The method has since been well validated (Boyd *et al.*, 1988). Since then, QoL measurement in cancer patients has been advocated widely (Maguire and Selby, 1989). However, in 1986, Macaulay and Smith reported that in a review of over 230 advanced breast cancer trials, in only 2 had overall QoL been measured. They added that assessment of the value of particular treatments should not rest upon response rate alone (Macaulay and Smith, 1986). So it is disappointing that during 1991, 15 years after Priestman and Baum's paper was

written, of 48 studies of chemotherapy in advanced breast cancer listed in the Index Medicus, we found only one which included QoL measurement. Many clinicians still prefer to rely upon their clinical judgement, although Slevin *et al.* found poor correlation between QoL measured by doctor and by patient (Slevin *et al.*, 1988). One problem may have been the QoL instruments on offer. Well-validated instruments did not include items about vomiting, nausea or hair loss and none was specific to breast cancer or chemotherapy. Moreover, QoL measurement is labour-intensive. We therefore addressed these problems.

In QoL measurement, a gold standard does not exist nor is it desirable according to Bergner (1989). Instruments fall into two broad categories: multidimensional, designed to measure specific aspects of disease or treatment, and global, which give a single score for as broad a representation of QoL as possible. The former approach was chosen, to complement existing instruments, with weighting provided by allowing the patient to choose the items of relevance to her. To take account of the fluctuations which may be expected to occur in patients on chemotherapy, a diary format was adopted. Guidelines proposed in 1986 by Guyatt *et al.* (1986) were followed. A six stage process comprises item selection, item reduction, format design, pretesting, construct and test-retest reliability and finally validation. Items were amassed and distilled from all the QoL measures then available and others were added after consultation with a panel which included a psychologist, a surgeon, a GP and a nurse counsellor. The validation of the "King's Diary" in its preliminary format was undertaken by Ebbs *et al.* during a trial comparing Epirubicin in two

different doses and administration systems, in which thirty nine patients completed the initial form of the diary during their treatment (Ebbs *et al.*, 1989b). This development process resulted in the "Qualitator" and has been described previously (Fraser *et al.*, 1990)(see appendix 13). The validation process, continued in two separate trials, is described below.

To test the ability of the Qualitator to measure what it is purporting to measure, it is necessary to consider what is known so far about QoL in advanced breast cancer patients, and in cancer patients in general. Baum *et al.* reported that a response to chemotherapy improved QoL scores, especially for pain and insomnia (Baum *et al.*, 1980); Ebbs *et al.* reported that good pre-treatment QoL scores were associated with a subsequent response (Ebbs *et al.*, 1988). A relationship between poor QoL and poor survival was reported by Morris and Sherwood in a study of terminally ill patients (Morris and Sherwood, 1987). Later, Addington-Hall *et al.* used the Spitzer QoL Index to predict survival in 230 terminal patients (Addington-Hall *et al.*, 1990).

3.3 Objectives

1 Validation of the Qualitator diary by comparison with different QoL measures in the Epirubicin and CMF trial

2 Validation of the Qualitator diary by comparison of data from two trials

3 To find a suitable way of analysing diary data

3.4 Patients and methods

3.4.1 Patients

Data were collected from two different studies, each with two arms. The first is described in chapter two. Forty patients with advanced breast cancer attending King's College Hospital and the William Harvey Hospital were randomised to receive the standard 28 day cycle of CMF (Bonadonna *et al.*, 1983) or weekly Epirubicin 20mg, for six months or until treatment failure between October 1988 and March 1990. The baseline Qualitator was completed by 29 patients who also completed baseline measurements in the Nottingham Health Profile (NHP)(Hunt *et al.*, 1985)(see appendix 3) and the Linear Analogue Self-Assessment (LASA)(Priestman and Baum, 1976)(see appendix 4). Comparisons between the three instruments were made for the 29 patients who completed them all. QoL measurement was continued for six months but stopped if disease progressed first.

In the second study, at Guy's Hospital, thirty nine patients were randomised to receive Adriamycin 25mg/m^2 weekly or 75mg/m^2 three-weekly to examine the influence of treatment schedule on response, survival and quality of life. Thirty one patients completed the diary in its preliminary format between 1986 and 1987, at the commencement of twelve weeks of therapy (Richards *et al.*, 1992) and continued until treatment was complete unless disease had progressed first. Data from 60 patients were therefore available for analysis.

3.4.2 Administration and scoring of quality of life measures

The Qualitator daily diary card, described briefly in chapter two, is administered three-weekly and completed continuously from the first day of treatment (see Appendix 5). From 23 items the patient chooses one she considers the most important from each of four domains: 1) symptoms of disease and side effects of treatment, 2) psychological aspects, 3) personal relationships and 4) physical performance. In addition a weighting variable is chosen from *any* domain. Daily thereafter, a score from 1-4 is given to the five chosen items, corresponding to the severity with which each item is perceived: "Not at all", "A Little", "Somewhat", "Very Much". The opportunity to change items occurs every three weeks, when a new card is exchanged for the old one. This period was chosen to suit the regimens used in the initial study (Ebbs *et al.*, 1988) and was kept for subsequent studies. Each patient's aggregated daily score is added to obtain a weekly total in the range 35-140. In both studies, patient groups (and other QoL measures in the King's study) were compared using a mean diary score taken from the completed weeks during each successive four week period. This allowed inclusion of all the available data, but allowed for any missing weeks. Isolated missing days were given the mean score for the other days that week.

In the King's study the NHP and LASA were administered prior to treatment and every four weeks thereafter, before the administration of chemotherapy and the QoL scores were processed when the study was finished. As previously mentioned, with all three instruments, a high score indicates poor

QoL. The NHP gives a weighted score out of 100 for each of six components: emotional state, energy, pain, physical mobility, sleep and social factors. Adding the components of the NHP was not part of its original design, but allows a global comparison, giving a range of 0-600. The LASA consisted of 24 categories, each scored 0-9, producing a global score range of 0-216. For comparison between instruments, *pre-treatment* NHP and LASA scores were compared with the *first week* of the Qualitator and thereafter, the average four-weekly Qualitator score.

In both studies, the Mann-Whitney-U test was used to compare the QoL scores of responders and non-responders each month and to compare initial and subsequent scores within a patient group. Comparison of QoL scores at each month of treatment with the first week's score was performed using the Wilcoxon signed rank test. Survival according to the Qualitator scores during the first week and the first four weeks of treatment were calculated using the Kaplan-Meier life table method (Kaplan and Meier, 1958) and the log-rank test (Peto *et al.*, 1977).

Patterns of three-weekly item choice were tabulated without statistical analysis. To compare individual items, eg pain, whether chosen in its own domain or as a weighting item, all patients who ever chose that item during the course of treatment had that score processed in the same way as the global scores, giving a range of 7-28. Patients who never chose that item were excluded from the analysis, but those who had not yet chosen the item, or who had stopped choosing it, were given the score 0 for purposes of non-parametric

statistical comparison, making the range for individual items 0-28, the step from "not yet chosen" or "no longer chosen", to "chosen, but given minimum score" being deemed a relevant distinction. The further analysis of individual symptoms is discussed below. Comparison of the global scores between instruments was performed using Pearson's correlation coefficient.

3.5 Results

3.5.1 Compliance

1) The King's study: The NHP and LASA were completed by all 29 patients who completed the Qualitator, 14 in the CMF arm and 15 in the Epirubicin arm. Eleven patients did not complete the Qualitator: three elderly patients were, mistakenly, not asked to do so, one patient refused and the rest either did not start it due to rapid progression of disease, or were unable to return the completed card on early progression of disease. The Qualitator was completed for 419 (88%) out of the total of 474 weeks. The missing weeks were 48(18%) of 262 in the CMF arm compared to 7(3%) of 212 in the Epirubicin arm (x^2=25.8, $p<0.001$). One patient preferred not to indicate the item in each domain which she had chosen so her data were only allowable for numerical analysis of the global scores. One patient failed to choose a weighting question for the fifth domain which was not discovered until the end of the study. Her score was multiplied by 1.25 in order to allow comparison of her global scores. There were eight isolated missing days. The same 29 patients completed the NHP and LASA on 104 of the possible 117 occasions, a compliance of 89%.

2) The Guy's study: The missing weeks were 13(11%) of 123 in the weekly treatment arm and 46(21%) of 220 in the 3-weekly treatment arm (x^2=5.92, p<0.02). Missing weeks were incurred most often as a result of delayed treatment due to haematological toxicity, and omission of the diary during the interim recovery period. There were 15 isolated missing days.

3.5.2 Response to treatment

In the King's study, of the 29 patients who completed the Qualitator, 15 (52%) responded clinically. In the Guy's study, 15 (38%) of the original 39 patients responded clinically, 11 (37%) of the 30 patients who completed the diary.

3.5.3 Quality of life: correlation between instruments

Correlation between the global scores of the individual instruments was 0.78 between the NHP and LASA. Correlation between the Qualitator first week, and subsequent four week aggregate, scores with the corresponding global NHP and LASA scores was 0.68 and 0.67 respectively.

3.5.4 Quality of life at entry to trial

Diary scores for the first week of treatment were taken as the baseline in both the King's and the Guy's trials. Comparison was made between the scores of patients who subsequently had a response to treatment (UICC) and those who did not. Taken separately, the King's responders had a median of 60,

non responders of 80 ($p<0.1$). Guy's responders had a median score of 43, non-responders of 81 ($p<0.05$). Added together, responders from both studies had a median of 59 and non-responders 81 ($p<0.005$). A first-week score of below 52 gives the highest odds ratio of a response to treatment, 6.21 (95%c.i. 1.70-22.8). In the both the King's study and the Guy's study, the first *month's* mean diary scores were significantly better for responders: (King's 56 v 73, $p<0.05$; Guy's 55 v 83, $p<0.05$). The pre-treatment NHP and LASA scores in the King's study gave a similar pattern in predicting responders and non-responders: (LASA responders 22, nonresponders 64, $p<0.005$; NHP responders 88, non-responders 162, $p<0.1$). The initial scores of all instruments in the King's study are illustrated in figure 3.1, standardised to a common scale of 0-10.

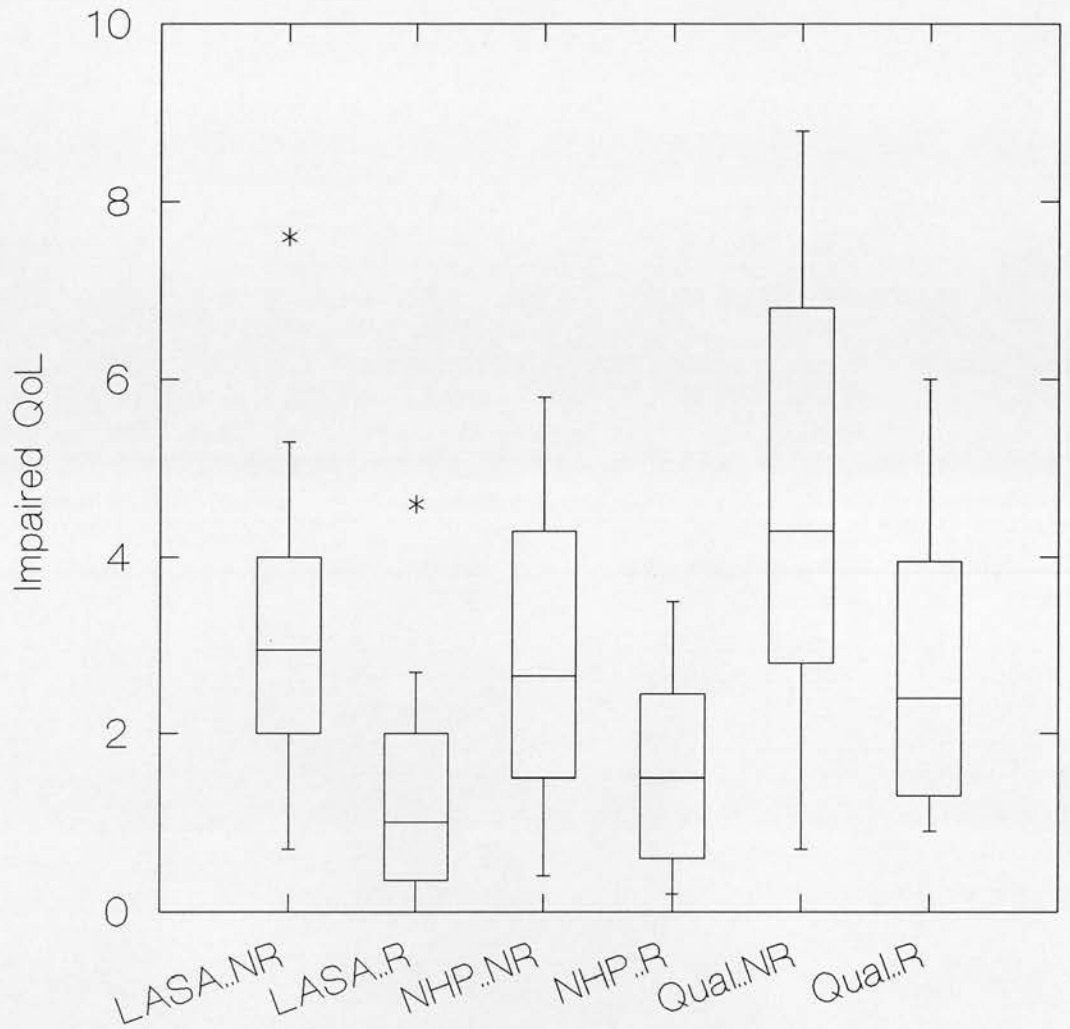


Figure 3.1: Box plot depicting QoL scores at the start of treatment for 29 patients who completed all 3 instruments in the King's study. The data for each instrument are arithmetically adjusted to be comparable on a simplified range of 0-10.

Qual=Qualitator, NHP=Nottingham Health Profile, LASA=Linear Analogue Self-Assessment, R=Response, NR=No Response.

For key to interpretation of box plots used in this thesis, please refer to Appendix 6.

To assess the relationship between early Qualitator scores and subsequent survival, the 60 patients from both studies were divided into high scoring and low scoring groups of nearly equal size using a threshold score of over 65. Survival was significantly better for patients with low scores in both the first week (median survival 57 weeks, 33 weeks; x^2=5.63,1df,p<0.02)(see figure 3.2) and the first 4 weeks (median survival 57 weeks, 30 weeks; x^2=13.14,1df,p<0.001)(see figure 3.3).

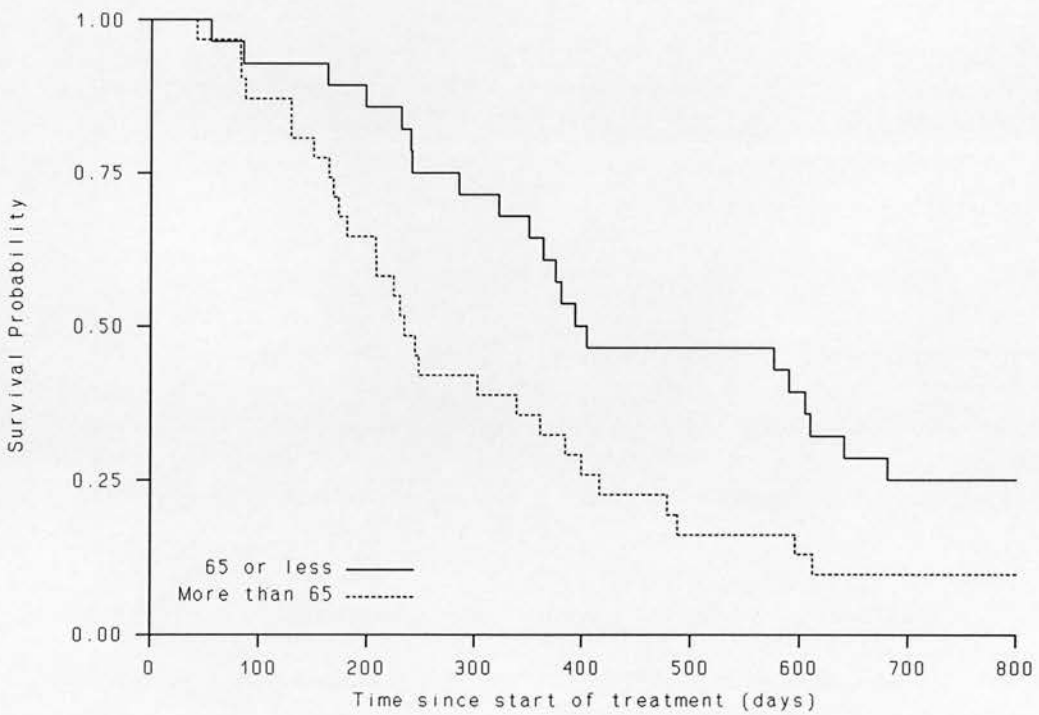


Figure 3.2: Life table indicating survival: patients grouped according to first week Qualitator scores (>65 n=31, 65 or less n=28)

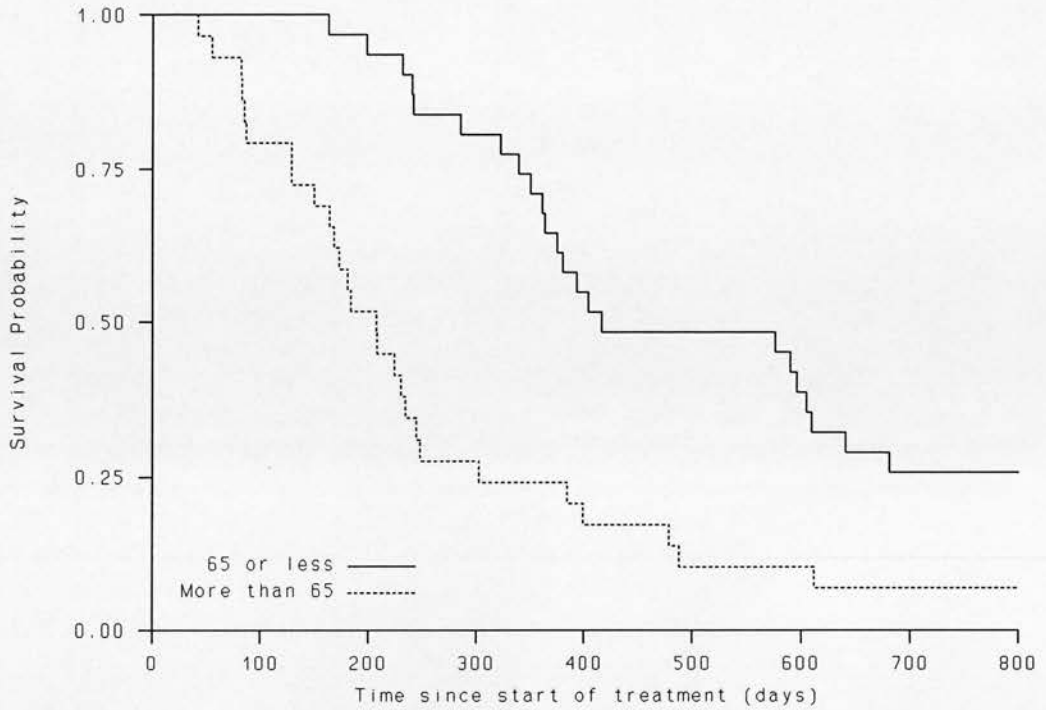
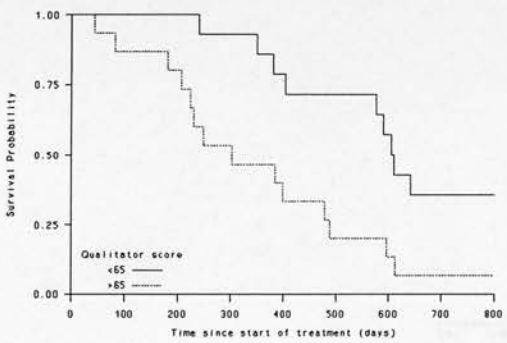


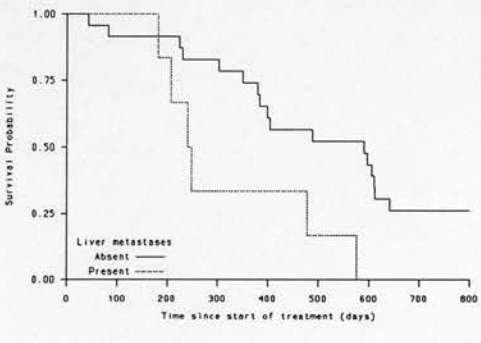
Figure 3.3: Life table indicating survival: patients grouped according to first 4 weeks' Qualitator scores (>65 n=29, 65 or less n=31)

In the King's study, in order to assess the prognostic significance of a high qualitator score at the start of treatment, comparison was made with other data documented at the start of treatment in the 29 patients in whom all data had been recorded (see Table 3.1). This indicates that a Qualitator score over 65 was a more potent indicator of survival than all other factors (see figures 3.4 and 3.5). The LASA also demonstrated a significant difference in the survival of high scorers and low scorers (see Table 3.1).

	Total	x^2	p
LASA > 45	14	4.98	0.03
NHP > 100	16	0.70	0.40
Qualitator > 65	15	6.90	0.01
Menopausal	19	0.004	0.95
Site of disease:			
Soft tissue	15	3.37	0.07
Nodal	14	0.03	0.86
Bone	13	0.28	0.60
Lung	6	0.13	0.72
Liver	6	4.85	0.03

Table 3.1: Survival differences according to factors recorded in 29 patients at entry to the King's study. Each factor is used, in turn, to calculate an overall survival difference using the log-rank method.





Figures 3.4 and 3.5: Life tables indicating survival in 29 patients in the King's study, grouped according to their first week's Qualitator score, or presence of liver metastases at start of treatment

However, if all 40 patients who entered the trial, but did not necessarily complete QoL measurement, are considered, then the most significant indicator of poor prognosis was the known presence of liver metastases at entry to the trial (x^2=10.69,1df,p=0.001).

3.5.5 Quality of life during treatment according to response

The initial difference between the diary scores of responders and non-responders persisted for three months in the King's study (p<0.05, p<0.05, p<0.02) and four months in the Guy's study (p<0.05, p<0.05, p<0.05, p<0.05). The corresponding differences in global scores for the NHP and LASA for the 29 patients in the King's study were not significant after 1 month.

Comparing patients' first week's Qualitator score with the corresponding aggregated score for one, two and three months, there were significant improvements for responders in the King's study at two months (median 8.55, p<0.05) and three months (median 12.5, p<0.01). There was no difference in the scores of non-responders. In the same patients, a similar pattern was observed in the NHP responders at three months (median 67.8, p<0.06) though less so in the LASA (median 1.75, p<0.8). The same trend of improvements in QoL score were not significant in Guy's responders. In order to illustrate the weekly trend in QoL scores amongst patients, figure 3.6 shows the mean diary scores in each group (although non-parametric methods were used for statistical analysis).

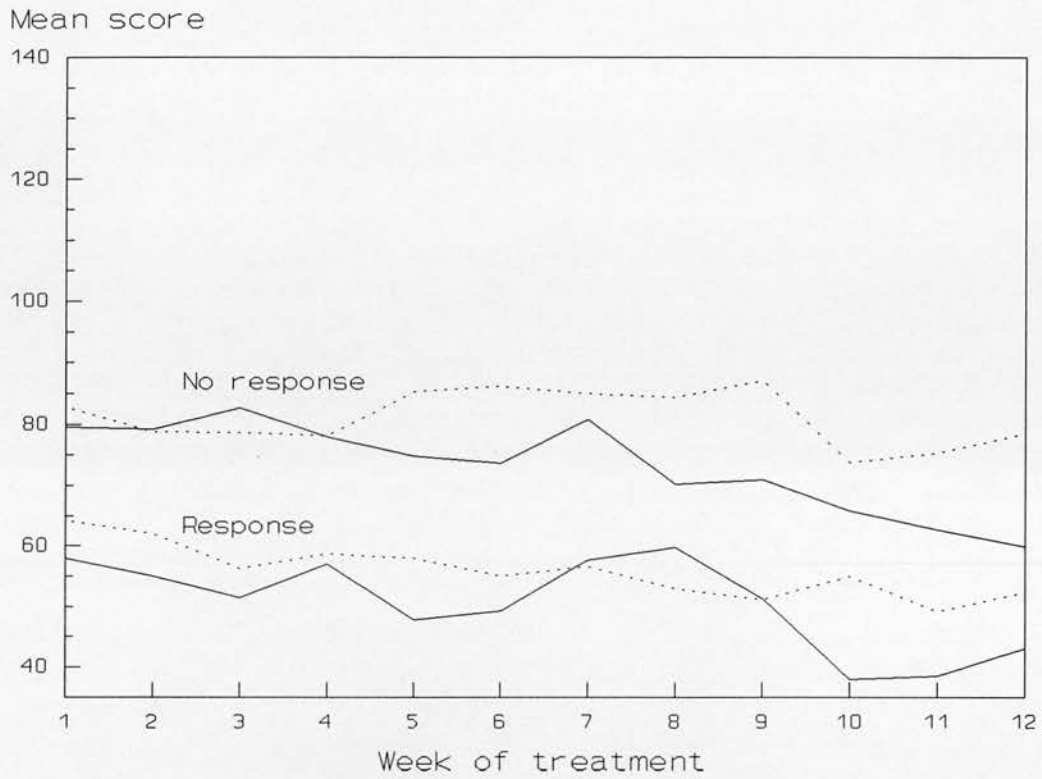


Figure 3.6: Mean Qualitator scores during the first 12 weeks of the King's (.....) and Guy's (_____) studies, grouped according to UICC response

---

3.5.6 Quality of life during treatment according to therapy

Comparing the change in Qualitator scores, between the first week and the subsequent aggregated score for one, two and three months, an improvement was recorded in the King's study for Epirubicin (median 7.2, $p<0.02$) and CMF (median 8.93, $p<0.02$) patients remaining at three months. This pattern was seen in the NHP score at three months for the same patients on CMF (median 67.75,

p<0.02). In the Guy's study, patients on the three weekly regimen had improved Qualitator scores at three months (median 9.4, p<0.05). The King's scores are shown in Table 3.2.

	Quali-tator			NHP			LASA		
Month	1	2	3	1	2	3	1	2	3
Epi-rubicin	4.35 p<0.05	4.8	7.15 p<0.02	-38.5	-7	17.25	10.75	5	9.25
CMF	1.6	0.55	8.93 p<0.02	13.5	32.5	67.75 p<0.02	2.5	-2.75	0
Resp.	3.75 p<0.06	8.55 p<0.05	12.5 p<0.01	6	9	36.75 p<0.06	-2	-1	1.75
No Resp.	1.95	0.1	12.5	-35.5	24	37.25	0	5.5	7

Table 3.2: Median improvement in QoL score, from initial measurement, during first 3 months of study for 29 patients in whom data were available (compare Table 2.4)

3.5.7 Analysis of separate items and domains

The Qualitator can be sub-analysed in detail but caution has been exercised to avoid producing spurious results. In Table 3.3, the total number of 3-weekly item choices in each study has been compared. Domain 1 receives

most of the weighting scores. Pain, tiredness, hair loss, activity and overall condition are chosen frequently in both studies. One patient out of the 60 in both studies did not indicate her item choices and 11 progressed or died on treatment after 3 weeks. Eight who did not change items from the first week onwards had a total of 42 opportunities to do so and the remaining 40 who did change had a total of 144 opportunities on which to do so. This opportunity was exercised, respectively, in groups 1,2,3,4 and the weighting group on 48, 46, 21, 41 and 75 occasions.

Analysis of the separate domains, 1-4, in the King's study demonstrated no significant improvement in score for any patient group in any domain. Differences between the scores of responders and non-responders are illustrated in Table 3.4. In the separate items, in the King's study, the only significant change was that at 3 months the scores for pain had improved for responders (medians 15 to 8.25, $p<0.02$) but not for non-responders (15.25 to 16). A similar, though non-significant trend was observed in the Guy's study.

Table 3.3 (over page): Relative proportions of Qualitator items chosen throughout study. In the column marked (%), the figure represents the total number of three weekly choices, including the weighting choice, made for that item, expressed as a percentage of the figure that would be obtained by distributing all choices evenly between each item.

	Guys				K'gs				Resp	No Resp
No. patients	31				28				27	32
	Smp	Wt	Tot	%	Smp	Wt	Tot	%	%	%
Pain	27	4	31	**187**	57	14	71	**292**	266	267
Breathing	22	1	23	**135**	9	5	14	**58**	89	107
Tired	13	13	26	**153**	24	32	56	**230**	155	179
Appetite	11	7	18	**106**	13	5	18	**74**	44	107
Feel sick	10	5	15	**88**	19	10	29	**119**	111	107
Vomiting	4	4	8	**47**	0	1	1	**4**	67	36
Bowel upset	7	11	18	**106**	1	5	6	**25**	44	71
Hair loss	4	16	20	**118**	23	20	43	**177**	155	89
Total	99	61	160	**940**	146	92	238	**979**	931	963
Anxiety	23	5	28	**115**	34	1	35	**99**	123	100
Depress.	28	3	31	**127**	11	5	16	**45**	77	87
Sleep	19	3	22	**90**	74	5	79	**224**	107	174
Future	17	6	23	**94**	22	8	30	**85**	107	112
Life	12	1	13	**53**	5	0	5	**14**	61	25
Total	98	18	117	**480**	146	19	165	**467**	475	498
Partner	27	1	28	**115**	41	0	41	**116**	107	124
Family	39	1	40	**164**	53	0	53	**150**	184	149
Friends	6	0	6	**25**	8	0	8	**23**	31	12
Sexual	1	0	1	**4**	1	0	1	**3**	15	12
Social	23	0	23	**107**	43	0	43	**122**	92	112
Total	98	2	101	**414**	146	0	146	**414**	429	409
Work	8	1	9	**37**	14	4	18	**51**	77	50
Hobbies	2	0	2	**8**	5	2	7	**20**	15	0
Activity	40	3	43	**176**	54	9	63	**178**	168	149
Overall	25	11	36	**148**	49	11	60	**170**	168	124
Self care	23	3	26	**107**	23	4	27	**76**	77	149
Total	98	18	116	**476**	145	30	175	**496**	505	472

	Start	Month 1	2	3	4
NHP					
Emotional	NS	NS	NS	NS	NS
Energy	NS	NS	NS	NS	NS
Pain	NS	p<0.05	p<0.01	NS	NS
Performance	<0.05	p<0.05	p<0.05	NS	NS
Sleep	NS	NS	NS	NS	NS
Social isolation	p<0.02	p<0.001	NS	NS	NS
LASA					
Physical Symptoms	p<0.01	p<0.05	NS	NS	p<0.05
Emotional	p<0.001	p<0.05	NS	NS	NS
Relationships	NS	NS	NS	NS	NS
Performance	p<0.05	p<0.05	NS	NS	NS
Qualitator					
Physical Symptoms	NS	NS	NS	NS	NS
Emotional	NS	NS	NS	NS	NS
Relationships	p<0.05	p<0.05	p<0.05	NS	NS
Performance	p<0.05	p<0.05	p<0.05	p<0.05	p<0.05

Table 3.4: Levels of significance, of the difference in scores, between responders and non-responders at each point in the King's study, in separate domains of the NHP and LASA and Qualitator

3.6 Discussion

There is no common currency of QoL measurement. In advanced breast cancer, QoL comprises many facets and as in other cancers, symptoms change

in importance between patients and over time (Clement Jones, 1985); chemotherapy adds to this complexity. The recently developed Rotterdam Symptom Checklist is a multidimensional instrument specifically designed for advanced cancer patients on chemotherapy which measures many facets but at intermittent timepoints (de Haes *et al.*, 1990). The development of the Qualitator represents a different response to the same perceived problem, rather than an attempt to "reinvent the wheel" (Aaronson, 1988).

The frequency of diary completion allows few items. Geddes *et al.* used a diary comprising eight obligatory items to measure QoL in lung cancer patients receiving chemotherapy (Geddes *et al.*, 1990). Fallowfield pointed out that most QoL questionnaires have fixed components that might not be relevant to an individual (Fallowfield, 1990). The Qualitator only measures five items on any three week cycle. However, permitting the patient to define areas of her life contributing most to its overall quality was the most novel and important departure from more traditional instruments. Lumping symptoms altogether in a global measurement is regarded by some as unscientific, akin to "trying to compare apples and oranges". This, however, was the intention: the sum of the parts was of overall interest. The number of changes of item made by patients in both studies supports this view.

In spite of small numbers in both the King's and Guy's studies, the Qualitator can predict patients likely to respond, supporting the findings of Baum *et al.* (1980) and Ebbs *et al.* (1988a). Moreover, patients with high initial scores had poorer survival, supporting the findings of Morris and Sherwood (1987)

and Addington-Hall *et al.* (1990). The qualitator score was the most potent indicator of prognosis of any of the factors recorded at the start of the King's trial.

Compliance for diary completion overall was good in both the King's study and Guy's study, comparing favourably with Geddes *et al.* who obtained 85%. In both studies, more weeks of diary completion were omitted by patients on intermittent regimens who ran out of diaries during the treatment delay due to neutropenic episodes. Aaronson advocates that QoL measures should be capable of disaggregation (Aaronson, 1988). This can be done with the Qualitator but, as with other instruments, subscales may not necessarily reflect the paramount concerns of the patient. In any case it may be more appropriate to apply a specific instrument to a specific area of interest (Ware, 1987). In the original processing of the Qualitator data, it was found that analysis of individual symptoms or domains can result in error. Following the method used by Geddes *et al.* (1990), the only other group to have published details about how they processed their data, the number of days all patients in one treatment group spent with significant severity for a particular symptom were compared using a chi-squared test. In the present study, it was possible to obtain a statistically significant difference between Epirubicin and CMF for every symptom, in one direction or other. After further scrutiny of this method, it was realised that each individual patient's baseline level was not taken into account, and that patients present in the study for differing durations were exerting disproportional influence on the direction of the result. In all trials

where QoL is measured, the QoL of patients who have left the study may continue to be affected by the treatment they received, irrespective of response. By measuring QoL only in patients still receiving treatment, a bias is incurred which will tend to exclude non-responding patients, who have a poorer QoL. This function of study design rather than instrument design may favour the use of an intermittent QoL measure beyond the intended treatment period.

The Qualitator is not presented as the long-awaited gold-standard and modification may be desirable with experience. However, it does offer a simple alternative to relying on clinical judgement alone. Collecting symptoms together and measuring an overall score is feasible. It provides an alternative or an adjunct to multi-dimensional measures with the aim of encouraging more clinicians to incorporate QoL measurement into trial design.

CHAPTER FOUR

Psychological screening for non specific abdominal pain

4.1 Summary

To determine whether Non Specific Abdominal Pain (NSAP) has a detectable psychological contributor which could be used to predict outcome, 131 patients aged 14-40 admitted with acute abdominal pain were assessed using the General Health Questionnaire and Hospital Anxiety and Depression questionnaires, and a structured interview. In 61 patients with NSAP, more had a psychosocial problem identified by the admitting registrar (x^2=7.28,1df,p<0.01) and marginally more had high questionnaire scores. The risk of having NSAP was high if an abnormality on interview accompanied high questionnaire scores (Relative Risk 1.93, 95%c.i. 1.35-2.77) or if prodromal pain had lasted more than 7 days (relative Risk 2.13: 1.55-2.92). After 2 years, patients with continuing pain had higher HAD (x^2=6.57,1df,p<0.02) and Spielberger anxiety trait (x^2=6.50,1df,p<0.02) scores; NSAP was associated with persisting pain (Relative Risk 2.22, 1.10-4.48). Psychosocial factors are implicated in NSAP and in chronic pain, but the sensitivity and specificity of questionnaire assessment are too low to be useful in diagnosing NSAP. What promotes NSAP still remains largely unknown, but the referral process may be the next direction for productive study.

4.2 Introduction

Previous studies of patients admitted with acute abdominal pain have shown that 25-45% will remain undiagnosed, representing a large and unwelcome burden on surgical resources. The term non-specific abdominal pain (NSAP) is applied to this group but no common pathology has been demonstrated. In their review of the subject, Gray and Collin concluded that a variety of possible causes may include a psychological component in patients with otherwise benign self-limiting conditions (Gray and Collin, 1987). No common psychological pathology has ever been identified and in the only prospective study to date, in which questionnaires were the only assessment tool, the findings were negative (Raheja *et al.*, 1990). A prospective study with long term follow up was therefore undertaken in order to determine whether psychosocial factors were implicated in patients with NSAP and whether, by screening with surgical examination, structured psychological interview or questionnaires, outcome could be predicted and admissions policy modified.

4.3 Objectives

1 To assess the likelihood of a psychological component to NSAP

2 To compare clinical methods with questionnaires in screening for NSAP on admission

3 To assess the ability of questionnaires to identify patients whose problems will not resolve following admission to hospital

4.4 Patients and Methods

4.4.1 Patients

131 consecutive patients between the ages 14-40 were admitted with abdominal pain to one surgical firm between October 1987 and December 1988. They were asked to complete two psychological questionnaires, the Hospital Anxiety and Depression Scale (HAD)(Zigmond and Snaith, 1983)(see Appendix 7) and the 30 item General Health Questionnaire (GHQ)(Goldberg, 1972)(see Appendix 8) . These are global psychological measures designed to detect people at high risk of having psychiatric illness. After completion, the forms were filed unseen. In a structured interview enquiring about recent marital problems, life events or other emotional problems, the admitting surgical registrar recorded the duration of symptoms, preliminary diagnosis and whether an underlying psychosocial problem was suspected (see appendix 9). The process of surgical management continued as normal and the discharge diagnosis was recorded. The questionnaires were readministered to 63 patients who attended the outpatient clinic at a median of 2 months. The hospital records of all patients were examined after 1 year for details of histology, clinical course and reattendance. A final diagnosis of NSAP was accepted if subsequent outpatient attendance or investigation had resulted in a discharge from clinic without a specific diagnosis. At a median of 2 years all 131 patients were contacted by post to complete further identical questionnaires and another *ad hoc* questionnaire to establish whether patients continued to have pain, whether pain had been adequately dealt with at the time of admission, and whether further

medical attention had been required. Patients were also offered a further appointment (see Appendix 10). The Spielberger Anxiety Trait questionnaire was introduced at this point (Spielberger *et al.*, 1983)(see Appendix 11). Sixty eight (52%) replied.

4.4.2 Scoring of questionnaires

After first follow-up, each admission questionnaire was scored. A high score was defined by a total score above 10 for either the anxiety or depression component of the HAD and above 5 for the GHQ (Zigmond and Snaith, 1983; Goldberg, 1972). The Spielberger Anxiety Trait questionnaire scores at 2 years were compared using a threshold of above 44.5, the mean score found in those with psychiatric complications amongst general medical and surgical patients (Spielberger *et al.*, 1983). For all questionnaires and for the registrar's assessment, the numbers of high and low scorers between the groups with specific abdominal pain (SAP) and NSAP were compared at each stage using the Chi-squared test. The scores of 63 patients who completed follow-up questionnaires in the outpatient clinic and 68 who completed postal questionnaires were compared with their respective admission scores using Wilcoxon's signed rank test. The patients who responded to the question at two years concerning continued abdominal pain were then considered separately: questionnaire scores at each stage were compared between those *with* and those *without* abdominal pain using the Mann-Whitney-U test and scores were compared with the admission score using the Wilcoxon signed rank test.

4.5 Results

4.5.1 Missing Data

Patients were excluded from analysis only for the specific question or questionnaire on which data were missing. On admission, all patients completed the HAD correctly, but 8 (2 NSAP) failed to complete the second page of the GHQ. For 9 patients, the record of the admitting registrar's psychosocial assessment and for 14 the admission diagnosis were incomplete. At 2 months the GHQ and HAD were incomplete for 4 and one respectively of the total of 63 patients. At 2 years, 2 NSAP and 2 SAP patients failed to answer the questions about further pain.

4.5.2 Diagnoses

Sixty one (47%) had a final diagnosis of NSAP and 70 (53%) of SAP. The final diagnoses and the sex distribution are listed in Table 4.I. The NSAP group included five whose diagnosis was unconfirmed beyond hospital discharge. A diagnosis of NSAP was made in 44% of males and 48% of females. There were 36 operations: of 33 appendicectomies, histology confirmed the diagnosis of appendicitis in 24 patients. Normal appendices were removed from three patients with no demonstrable pathology, one each with a diagnosis of Meckel's diverticulum, pelvic inflammatory disease, enteritis, ruptured fimbrial cyst, and cholecystitis. One patient had a pinworm infection and in the absence of acute inflammation was classified in the NSAP group. Two patients with the diagnosis of NSAP had normal appendices removed at later dates.

	Total n=131	Males n=50(100%)	Females n=81(100%)
NSAP	61	22(44)	39(48)
Appendicitis	24	13(26)	11(14)
Pancreatitis	5	1(2)	4(5)
Gynaecological	8	0	8(10)
Upper GI	16	8(16)	8(10)
Renal colic/UTI	10	3(6)	7(9)
Miscellaneous	7	3(6)	4(5)
Appendicectomy (NSAP)	4	1(1)	3(6)
Appendicectomy (Other diagnosis)	5	2(4)	3(4)

Table 4.1: Number and percentage of each sex in each diagnostic group and those undergoing unnecessary appendicectomy

---

4.5.3 Scores on admission and at follow-up

The percentage of patients with NSAP and with SAP having a high score or a psychosocial problem at each stage are represented in figure 4.1. After the interview, 38 were thought to have a psychosocial problem: 24 (44%) out of 55 with an eventual final diagnosis of NSAP and 14 (21%) out of 67 with SAP (X^2=7.28,1df,p<0.01).

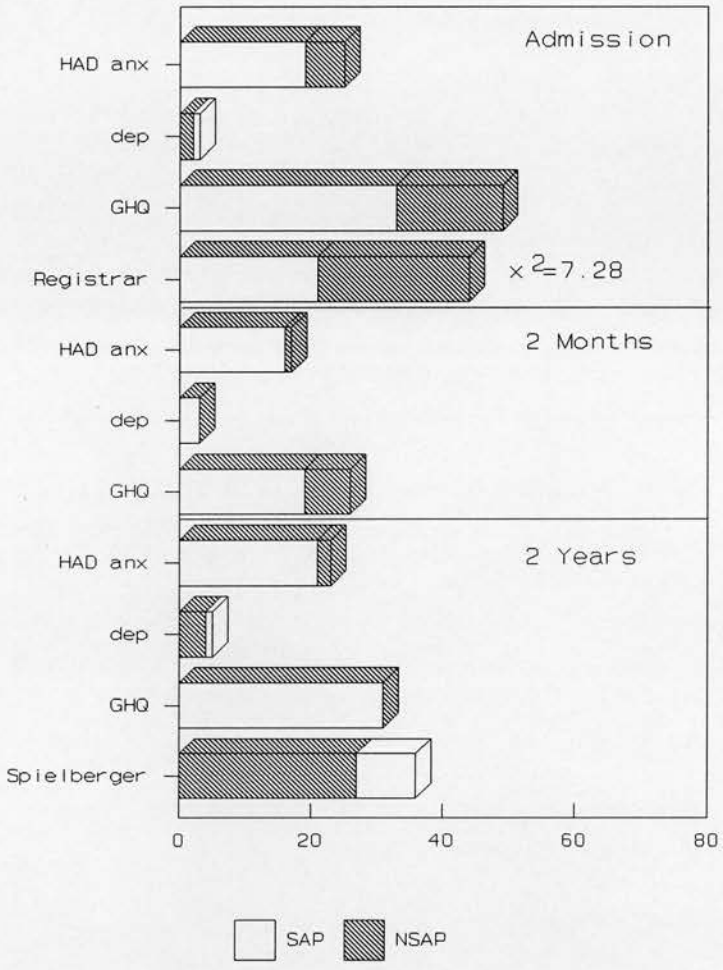


Figure 4.1: Percentage of "cases" in each group according to each test used. The two blocks in the right end of each bar indicate the extent to which the number of cases in one diagnostic group exceeds that in the other.

Although there appear to be more high scores in those with NSAP for the HAD (anxiety) and the GHQ on admission, this was not statistically significant. At 2 months and at 2 years the scores were even. This trend was observed in the Mann-Whitney-U tests. Again this was not statistically significant (see Table 4.2). Comparison of the admission score with the follow-up score of each patient using Wilcoxon's paired rank test demonstrate no statistically significant change for individual patients in either the SAP or the NSAP groups at 2 months or 2 years (see Table 4.3).

		NSAP	SAP	p
Admission	**HAD Anxiety**	6 (n=61)	7 (n=70)	0.663
	HAD Depression	3 (n=61)	3 (n=70)	0.241
	GHQ	5 (n=59)	3 (n=64)	0.814
2 Months	**HAD Anxiety**	5 (n=30)	5.5 (n=32)	0.740
	HAD Depression	2 (n=30)	1 (n=32)	0.611
	GHQ	3 (n=26)	1.5 (n=32)	0.392
2 Years	**HAD Anxiety**	7 (n=26)	6 (n=42)	0.495
	HAD Depression	2 (n=26)	2.5 (n=42)	1
	GHQ	2 (n=26)	1 (n=42)	0.507
	Spielberger	36.5 (n=26)	38 (n=42)	0.925

Table 4.2: Median questionnaire scores of SAP and NSAP patients at each juncture

		NSAP	p	SAP	p
2 Months	**HAD Anxiety**	0.5 (n=30)	0.443	0 (n=32)	0.742
	HAD Depression	0 (n=30)	0.867	0.5 (n=32)	0.276
	GHQ	1.5 (n=29)	0.158	1 (n=32)	0.061
2 Years	**HAD Anxiety**	-0.5 (n=26)	0.332	0 (n=42)	0.868
	HAD Depression	-0.5 (n=26)	0.313	0 (n=42)	0.694
	GHQ	0 (n=26)	0.888	0.5 (n=40)	0.539

Table 4.3: The median fall in paired questionnaire scores at each stage (statistical comparison using Mann-Whitney test)

Few patients had high scores in the depression component of the HAD at any stage. Of the nine patients who had a non-inflamed appendix removed, including the four with NSAP, eight had normal questionnaire scores on admission. The numbers of high questionnaire scores were similar amongst males and females: 24% and 20% respectively for the HAD, 45% and 39% respectively for the GHQ. Against this trend, a psychosocial problem was diagnosed in 24% of males and 34% of females but this was not statistically significant.

4.5.4 Assessment of admission data as screening tools

The relative risks and predictive values for a diagnosis of NSAP were calculated for the variables recorded on admission and are shown in Table 4.4.

	HAD	GHQ	Psycho-social Problem	Initial Diagnosis NSAP	Pain >7 days	HAD+ GHQ	HAD + GHQ + Psych. problem
Sensit-ivity %	26	49	44	47	15	24	21
Specif-icity %	81	67	79	86	99	86	97
Relative Risk (95% ci)	1.25 0.84-1.85	1.41 0.98-2.03	1.71 1.18-2.48	2.24 1.57-3.21	2.13 1.55-2.92	1.35 0.92-2.00	1.93 1.35-2.77
Positive Predict-ive Value	55%	58%	63%	76%	89%	61%	82%

Table 4.4: Recordings on admission as screening tools for NSAP

In the 117 patients with a recorded admission diagnosis, the *initial* diagnosis of NSAP was the single factor most strongly associated with a subsequent *final* diagnosis of NSAP. The registrar's diagnosis of a psychosocial problem was also significantly associated with a final diagnosis of NSAP; a high score in the HAD and GHQ questionnaires, alone or combined, was not. An admission diagnosis of NSAP was not associated with the diagnosis of a psychosocial problem: (odds ratio 1.39, 95% c.i. 0.62-3.14) and there was no significant association between the registrar's diagnosis of a psychosocial problem and with the HAD or GHQ scores: odds ratios 2.31 (0.95-5.64) for the

HAD, 1.82 (0.82-4.05) for the GHQ.

Of 11 patients with high scores in both questionnaires *and* deemed to have a psychosocial problem on interview, 9 had NSAP (Relative Risk 1.93: 1.35-2.77). Of 9 patients with symptoms for more than one week before admission, 8 had NSAP (Relative Risk 2.13, 1.55-2.92).

4.5.5 Assessment of outcome at 2 years

In the 2 year postal questionnaire, four questions were asked (see appendix 10). The replies to these questions are shown in Table 4.5.

	n=64	NSAP, n=24
1) Still suffer from pain	YES 21 (33%)	12 (x^2=5.15, p<0.05)
2) Pain adequately dealt with	YES 45 (70%)	16
3) Further medical attention	YES 21 (33%)	10
4) Another appointment sought	YES 4 (6%)	2

Table 4.5: The answers to questions asked in 2 year questionnaire

In answer to the first question, about continuing pain, more with an anxiety trait were found among those who replied "yes" (x^2=6.50,1df,p<0.02); more had a high HAD at 2 years (x^2=6.57,1df,p<0.02). Those with continuing pain included more originally diagnosed as having a psychosocial problem (x^2=10.94,1df,p<0.001) and more with a final diagnosis of NSAP

(x^2=5.15,1df,p<0.05). More patients with continuing pain also thought their pain had not been adequately dealt with on admission (x^2= 4.82, p<0.05) and more had sought further medical attention in the interim (x^2=5.43, p<0.02). There were no significant differences in questionnaire scores according to responses to the other questions. In Table 4.6, the risk of having continuing pain is compared for each of the variables recorded on admission and at 2 years. The corresponding non-parametric data are displayed in Tables 4.7 and 4.8.

	NSAP	Admis-sion			2 Years		
		HAD	GHQ	Psych. Problem	HAD	GHQ	Spiel-berger
Pain n=21	12	5	11	10/19	9	9	11
No Pain n=43	12	4	13	5/40	6	11	9
Rel.Risk Pain at 2 Years (95% ci)	1.91 0.93-3.91	1.83 0.92-3.66	3.26 1.64-6.46	2.22 1.10-4.48	2.45 1.29-4.66	1.65 0.83-3.27	2.42 1.23-4.75
Positive Predictive value	57%	24%	52%	53%	43%	43%	52%

Table 4.6: Factors associated with continuing pain at 2 years

		Pain	No Pain	p
Admission	**HAD Anxiety**	8 (n=21)	5.5 (n=42)	0.213
	HAD Depression	3 (n=21)	2 (n=42)	0.789
	GHQ	7 (n=21)	2 (n=42)	0.239
2 Months	**HAD Anxiety**	6 (n=10)	5 (n=25)	0.442
	HAD Depression	3 (n=10)	2 (n=25)	0.163
	GHQ	5 (n=10)	1 (n=25)	0.001
2 Years	**HAD Anxiety**	8 (n=21)	5 (n=43)	0.008
	HAD Depression	4 (n=21)	1 (n=43)	0.009
	GHQ	3 (n=21)	1 (n=43)	0.127
	Spielberger	46 (n=21)	35.5 (n=43)	0.046

Table 4.7: Median scores for patients with and without abdominal pain at 2 years

		Pain	p	No Pain	p
2 Months	**HAD Anxiety**	0.5 (n=11)	0.878	0 (n=25)	0.948
	HAD Depression	-0.5 (n=11)	0.732	0.5 (n=25)	0.36
	GHQ	-3 (n=11)	0.114	1.5 (n=25)	0.07
2 Years	**HAD Anxiety**	-1 (n=21)	0.365	0 (n=43)	0.844
	HAD Depression	-2 (n=21)	0.078	0.5 (n=43)	0.137
	GHQ	0.5 (n=21)	1	0 (n=43)	0.746

Table 4.8: The median fall in questionnaire score at each stage for patients with and without abdominal pain

4.6 Discussion

Many organic causes of NSAP have been suggested (Gray and Collin, 1987). NSAP is a loose concept binding disparate conditions but a significant consumer of surgical resources. In this study, 47% of patients had NSAP. In a study of 1190 patients with acute abdominal pain, 50.8% of patients aged 10-29 had NSAP (Irvin, 1989) and in a multi-centre study of 6097 patients the rate was 43% (De Dombal, 1979a). Because psychosocial aspects are seldom addressed routinely, a genuine psychological or social problem may be overlooked. Authors of previous studies assessing patients after operation or diagnosis have reported that patients who had normal appendices removed were emotionally distressed (Blanton and Kirk, 1947; Barraclough, 1967) or had an increased incidence of adverse life events (Creed, 1981) but in the only prospective study low scores were reported in NSAP patients and controls alike (Raheja *et al.*, 1990). In another study, of 105 patients, most of the 18 with NSAP were female and scored higher in anxiety state-trait and illness behaviour questionnaires (Joyce *et al.*, 1986). In this study, NSAP and higher psychological scores were no more common in women. Most of the patients undergoing unnecessary appendicectomy had other surgical or gynaecological pathology and had if anything, lower psychological scores. However, abdominal pain has been claimed to be ameliorating in depression (Gomez and Dally, 1977).

A *preliminary* diagnosis of NSAP was the best predictor of a *final* diagnosis of NSAP. The registrar's diagnosis of a psychosocial problem was better than the HAD and GHQ but may have been encouraged where a surgical

diagnosis was uncertain. However the preliminary diagnosis of NSAP and the diagnosis of a psychosocial problem were not significantly associated. Some feature, perhaps social rather than psychological, and independent of clinical findings, was being detected less accurately by the HAD and GHQ, which could not distinguish the anxiety surrounding acute hospital admission in patients with SAP. A larger study may have established a significant difference. Even using a positive score in *both* GHQ and HAD, 10 patients out of 25 would have been wrongly diagnosed as NSAP, including three with appendicitis and two with pancreatitis. The combination of psychosocial assessment, HAD and GHQ identified a small but specific group at high risk of having NSAP, but similar specificity was obtained simply by using a cut-off of more than seven days of prodromal pain. In a study of 158 patients, the simple measure of recording a "closed eyes sign" on abdominal palpation had a predictive value of 79% for NSAP (Gray *et al.*, 1988). De Dombal reported that computer assisted diagnosis could reduce unnecessary admissions by 25% (De Dombal, 1979b) but this has not been widely adopted, a probable drawback of any approach needing extensive data collation.

The ability of a test to diagnose a patient having NSAP was expressed by the relative risk. However, a more pragmatic approach is to calculate the positive predictive value, which expresses as a percentage the chance of a positive test being accurate in predicting NSAP. This is of direct value, as for screening purposes, the only patients we are interested in identifying are those who have a high probability of NSAP, and therefore may not need to be

admitted to hospital.

Higher HAD anxiety and Spielberger trait scores were found in patients with persistent pain at 2 years, but volunteer bias cannot be discounted since compliance was low despite assiduous pursuit by mail and telephone. The registrar's psychosocial diagnosis was the closest predictor of persistent pain but it remains debatable as to whether psychosocial factors or pain come first. Similar uncertainty as to the direction of causality arose from a study of the irritable bowel syndrome, in which patients with bowel dysfunction had higher HAD scores than controls (Heaton *et al.*, 1991). Other studies have made use of the concept of abnormal illness behaviour to study how bowel symptoms are acted upon (Drossman *et al.*, 1988); a similar approach might reveal the way in which a person in the community becomes a patient with NSAP.

Amongst the unmeasured influences on the patient population in this study are the referral patterns, which may change as GP fundholding and purchasing become established. At present, the momentum acquired by a patient already in the reception area and the fear of a missed diagnosis increase the pressure to admit a patient with a questionable surgical diagnosis. To reduce non-surgical admissions may require study of the referral process and in future may be driven by reduced bed availability.

An adverse psychosocial history with positive questionnaires can identify a small group of patients at high risk of having NSAP. However, the patient with NSAP still remains easier to admit than to diagnose.

CHAPTER FIVE

Health status measurement after minor surgery: a prospective study

5.1 Summary

To establish whether minor operations cause a perceived and measurable improvement in health and QoL, 57 patients having day-surgery on a Waiting List Initiative were studied prospectively. The NHP, HAD and GHQ were completed before surgery and after 6 months by 81% patients, when an *ad-hoc* questionnaire dwelling on perceived outcome of surgery was also completed. An operative success was reported by 78%, improved health by 64%, improved QoL by 69% and improved work efficiency by 54%. Improvements in HAD *anxiety* (p=0.023), *depression* (p=0.035), NHP *pain* (p=0.001) and *global* NHP (p=0.034) were recorded. In the perceived outcome questionnaire, patients reporting a successful operation had had better preoperative GHQ (p=0.029) and HAD *depression* (p=0.031) scores than those whose operation was not a success. Those reporting an improvement in *health* postoperatively had worse preoperative NHP scores to start with (p=0.027) than those who had no improvement. Minor surgery results in improvements in both perceived and in objectively measured health and QoL. Both are valid outcome measures for minor surgery. Preoperative scores may be related to subsequent perception of outcome.

5.2 Introduction

Evidence has emerged that at least one health authority has considered excluding minor operations for non-life threatening conditions from its health care provisions, pleading financial stringency (Godlee, 1991). This may have been encouraged by evidence from health utility measures, which are based on hypothetical trade-offs and tend to award low priority to such operations (Cochrane *et al.*, 1991).

The measurement of health status *per se* has gained momentum in branches of medicine mainly concerned with chronic diseases and cancer (Maguire and Selby, 1989). But more recently the techniques have also been used to advance the cause of new or expensive technologies such as liver transplantation or coronary bypass grafting (Tarter *et al.*, 1991; Caine *et al.*, 1991). Where the outcome of treatment *is*, or has been assumed to be, obvious, surprising results can occur, such as the failure to demonstrate any superiority of extracorporeal shock wave lithotripsy over more invasive percutaneous therapy for renal calculi (Mays *et al.*, 1990). Only in one study has QoL been measured during minor surgery for non-life threatening conditions and this reported no change (Hunt *et al.*, 1984).

To find out what, if any, benefit accrues to patients having these operations, a prospective study was undertaken in which QoL was measured before and after surgery using a variety of well validated instruments, to allow wider comparison of results. In addition, patients were asked to comment on the results of operation.

5.3 Objectives

1 To determine if, and which, conventional instruments can be used to measure the outcome of minor surgery

2 To compare prospective use of QoL questionnaires with a simple retrospective questionnaire

3 To determine whether minor surgery does actually improve quality of life

5.4 Patients and methods

Fifty seven patients underwent day surgery in March 1991 as a result of a King's College Hospital waiting list initiative. Patients were asked to complete three Health Status questionnaires before surgery. The Nottingham Health Profile (NHP)(Hunt *et al.*, 1985) was chosen as a well validated multidimensional instrument to indicate patients' overall health status. The Hospital Anxiety and Depression Scale (HAD)(Zigmond and Snaith, 1983) and the 30-Question General Health Questionnaire (GHQ)(Goldberg, 1972) were chosen in order to indicate whether a psychological improvement could be measured in patients generally judged to be fit by conventional standards. In addition patients were asked to record duration of symptoms and time on waiting list. Between 6 months and 10 months (median 7.5) after the operation, the NHP, HAD and GHQ were sent out by post together with a short questionnaire devised to assess the perceived outcome of surgery (see Appendix 12). This included the dichotomous questions: "Has the operation been as successful as you had hoped?" and "Do you wish you had been offered the operation much sooner?" together

with seven questions on general postoperative Health, QoL and function scored on a five point categorical scale ("much worse", "worse", "same", "better" and "much better"). Two more postal sweeps were made to non-responders, the final one accompanied by a telephone call. Each component of the HAD, *anxiety* and *depression*, gives a range 0-21, the GHQ gives a global score of 0-30 and the NHP components of *Emotional reactions, Energy, Pain, Physical mobility, Sleep and Social isolation* give a range 0-100 or added, give a global score of 0-600.

In keeping with standard practice in health status measurement, all comparisons were made using non-parametric statistical methods. Paired preoperative and postoperative scores were compared using the Wilcoxon signed rank test. HAD, GHQ and NHP scores were then classified according to the patients' responses to the postoperative perceived outcome questions: on operation success (Yes or No), and health, QoL and function ("much worse, worse and the same", or "better and much better"). Comparison was by the Mann-Whitney-U test.

5.5 Results

5.5.1 Operations

Patients had been awaiting surgery for a mean of 1.9 years (range 0-4.7). The operations were for varicose veins 30 (53%), hernia 8 (14%), anal lesions 5 (9%), skin lesions 7 (12%), ingrowing toenail 3 (5%) and 4 (7%) others.

5.5.2 Missing data

The overall compliance rates for questionnaire completion are shown in Table 5.1. The *health* and *QoL* items were completed by all patients who answered the perceived outcome questionnaire but there was a small attrition rate for the other items which individual patients may have considered inapplicable to their situation.

	Preoperatively	Postoperatively
GHQ	50 (88)	43 (75)
HAD	53 (93)	44 (77)
NHP	55 (96)	45 (79)
Successful operation?	-	41 (72)
Preferred sooner?	-	40 (70)
Health	-	46 (81)
QoL	-	46 (81)
Work efficiency	-	38 (67)
Social life	-	45 (79)
Sex life	-	42 (74)
Interests & Hobbies	-	45 (79)
Holidays	-	45 (79)
Total number of patients	57	46

Table 5.1: Number (%) of patients completing each questionnaire

Of the 11 patients who did not return questionnaires despite three postal sweeps, seven who were contacted by telephone expressed satisfaction with the operation and agreed to return questionnaires but none was received. Four patients were untraceable.

5.5.3 Outcome measured by paired HAD, GHQ and NHP scores

In patients who completed both sets of questionnaires, there was a significant improvement in the scores for HAD *anxiety* (median improvement 1, range -4 to 6, p=0.023), *depression* (median 1, -8 to 9, p=035), NHP *pain* (median 8.74, -11.2 to 85.2, p=0.001) and *global* NHP (median 13.73, -75.8 to 167, p=0.034). The medians and ranges of these scores, standardised to a scale of 1-10, are illustrated in Figure 5.1. The ranges of change in these scores are illustrated in Figure 5.2.

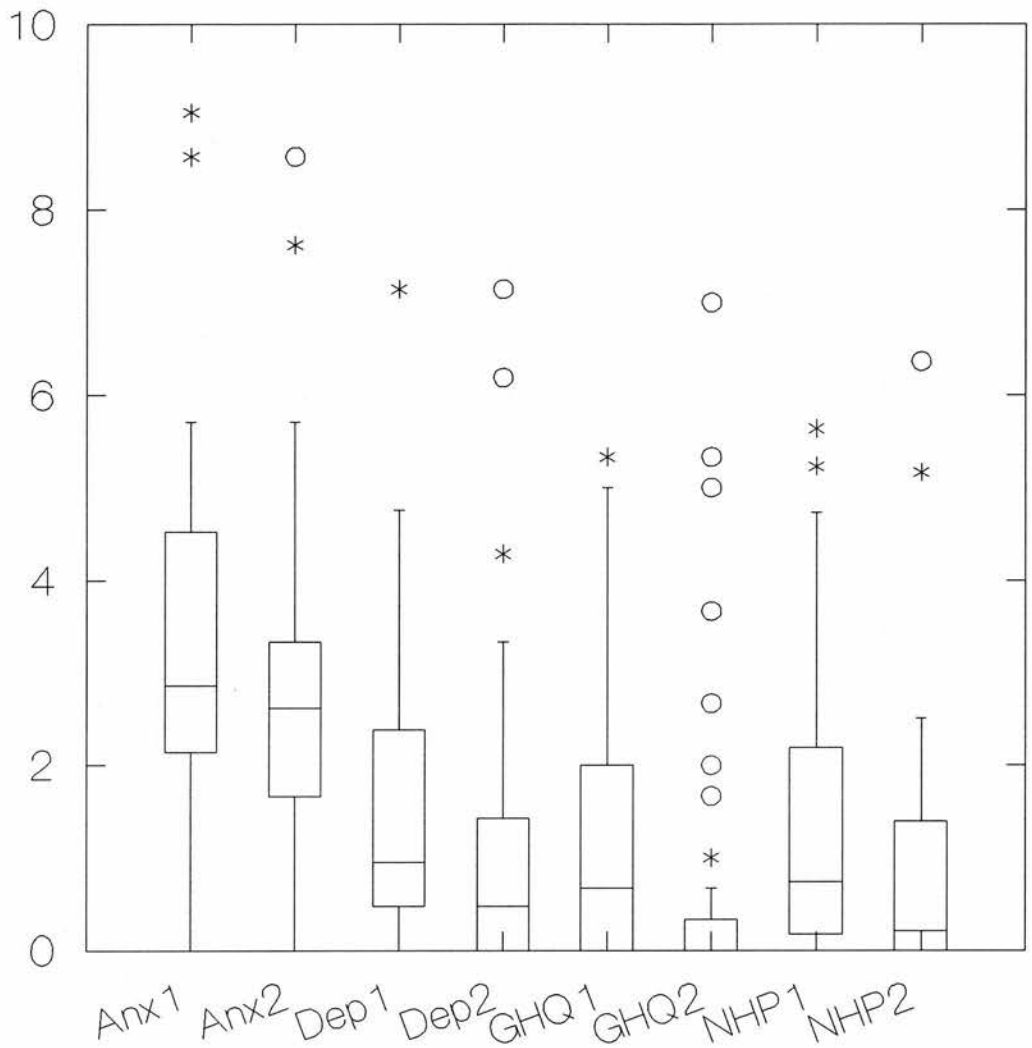


Figure 5.1: Box plots indicating medians and ranges of preoperative (1) scores and postoperative (2) scores; Anx = HAD anxiety, Dep = HAD Depression. Scores have been arithmetically adjusted to be comparable on a scale of 0-10

For key to interpretation of box plots used in this thesis, please refer to Appendix 6.

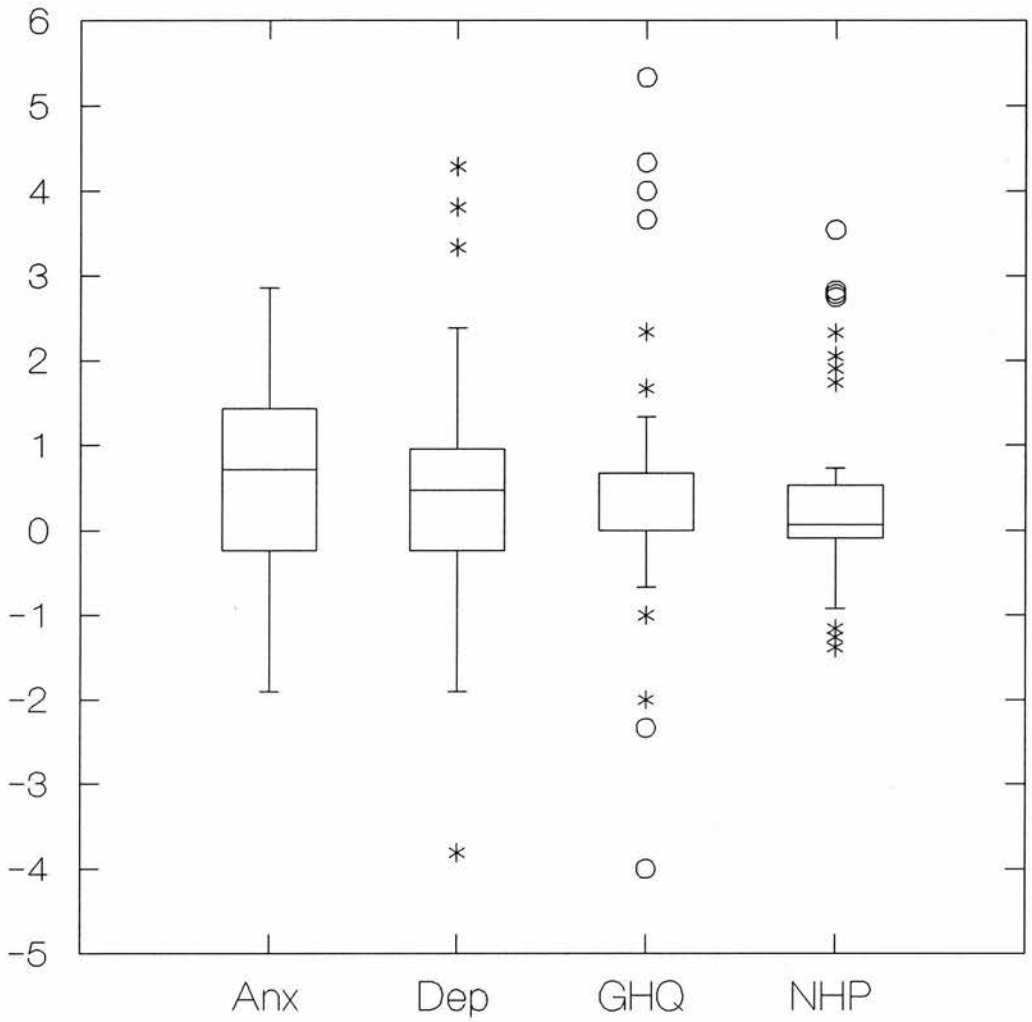
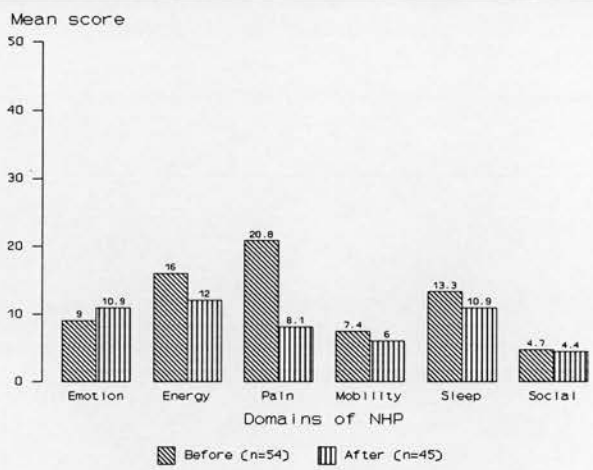
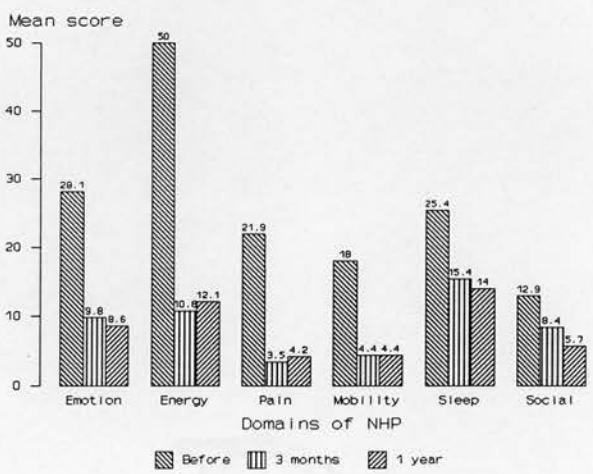


Figure 5.2: Box plots indicating medians and ranges of the change in score following operation. The data for each instrument are arithmetically adjusted to be comparable on a simplified range of 0-10 (a <u>plus</u> value indicates an improvement)

To give an idea of how these scores compare with NHP scores in another study, the mean scores before and after surgery in the NHP domains have been represented in graphical form in Figure 5.3 and are juxtaposed with the scores from a study of 100 patients in whom QoL was measured before and after coronary bypass grafting in Figure 5.4 (Caine *et al.*, 1991).





Figures 5.3 and 5.4: Bar graphs indicating mean NHP scores in the minor surgery patients compared to those in a study of patients undergoing coronary artery surgery (Caine et al., 1991).

5.5.4 Perceived outcome

The responses to the perceived outcome questionnaire are shown in Table 5.2. Thirty two of forty one (78%) patients said the operation had been as successful as they had hoped and 9 said not; 38/40 (95%) wished they had had their operation sooner and 2 did not. The mean time off work was 2.1 weeks.

	Much worse	Worse	Same	Better	Much Better	Total
Health	0	1 (2)	15 (33)	22 (48)	8 (17)	46
QoL	0	1 (2)	12 (26)	23 (50)	10 (22)	46
Work efficiency	0	2 (5)	16 (42)	16 (42)	4 (11)	38
Social life	0	1 (2)	22 (49)	19 (42)	3 (7)	45
Sex life	0	1 (2)	29 (69)	9 (21)	3 (7)	42
Interests & Hobbies	0	1 (2)	30 (67)	11 (24)	3 (7)	45
Holidays	0	1 (2)	23 (51)	18 (40)	3 (7)	45

Table 5.2: Responses (%) in each category in perceived outcome questionnaire

5.5.5 Change in paired HAD, GHQ and NHP scores according to perceived outcome

The responses to the perceived outcome questionnaire were used to distinguish between two groups of patients for each item: those who reported

an operative *success*, or not; an improvement *QoL*, or not; and similarly for *health*, *work efficiency*, *social life*, *sex life*, *hobbies and interests*, and *holidays*. The change in HAD, GHQ and NHP scores was then compared between the two groups for each item. There was a tendency, in patients who perceived no improvement, for HAD, GHQ and NHP scores to show a lesser improvement. The median change in score was similar for each item between patients who perceived a success, or improvement, and those who did not. Nevertheless, ranking the changes in score demonstrated a significantly greater improvement in NHP *energy* scale for those who perceived an improvement in *health* (p=0.018) or *QoL* (p=0.026) and those who did not. The *global* NHP also showed a greater improvement in those who perceived an improvement in *health* (p=0.043). There was no statistical difference in other NHP scales, nor between patients who perceived an improvement in the other items and those who did not.

5.5.6 Preoperative indicators of perceived outcome

When *preoperative* NHP, HAD and GHQ scores were compared, patients who perceived the operation a *success* turned out to have had better scores than those whose operation was not a *success*, in the HAD *depression* (medians 5 *versus* 2, p=0.031) and GHQ (medians 6.5 *versus* 1, p=0.029). Moreover, those who perceived improved *health* as a result of surgery had recorded significantly poorer *global* NHP scores preoperatively than those whose *health* was no better after surgery (medians 64.9 *versus* 17, p=0.027). Patients who perceived an

improvement in *work efficiency* as a result of surgery had recorded significantly poorer *global NHP* scores preoperatively (75.8 *versus* 24.1, p=0.02). There was no difference for the other items.

5.6 Discussion

The majority of patients report an improvement in general health, quality of life and function following minor surgery. This is supported by small, but significant, improvements in psychological, pain and global health questionnaire scores. These scores are proportionately better in patients who do report improvements resulting from surgery, indicating that established general QoL instruments may be used to measure outcome of minor surgery. As might be expected, the improvements in QoL are proportionately less than those to be obtained by coronary artery bypass grafting (Caine *et al.*, 1991). The vast majority wished that their operations had been done sooner and a majority perceived an improvement in function.

Patients who reported an improvement in health and work efficiency had poorer preoperative NHP global health scores, indicating that these patients had genuinely impaired health-related QoL which was then improved by surgery. Moreover, patients who considered their operation not to be a success had had poorer preoperative depression scores and GHQ scores, suggesting an underlying psychological modifier to their perception of outcome. An alternative approach to the study design may have been to focus more precisely in the postoperative perceived outcome questionnaire on the symptoms associated with each

particular condition. However, as the conditions were heterogenous, the data thus derived may not have been easily interpreted and the NHP proved a suitable outcome measure for these conditions.

Ideally, this study should have had a control group which would go through the same preoperative work-up and QoL measurement as the study population, but not undergo surgery. However, there are reasons why this approach was not adopted. The optimum control group would be treated identically, except to be denied treatment. This is ethically unacceptable. An alternative would be to measure QoL on patients in whom conservative management or deferred surgery had been decided upon, but this group represents a different patient population. This was the approach used by Hunt *et al.* in patients undergoing minor surgery. They found no improvement nor any difference in QoL between this and a control group, but conceded that the control group could not be assumed to be closely matched (Hunt *et al.,* 1984).

A compromise may have been to measure QoL of all patients on the waiting list, and then to measure controls who remained on the waiting list. The logistic problem that made this difficult was that on the waiting list initiative, the objective had been to operate on all patients waiting more than a few months. Another approach would be to compare a population of patients having a different procedure, for instance liver transplant. The problem with this approach is that again, this would represent a different patient population. Compliance in the postoperative questionnaires was good and although volunteer bias cannot be discounted, the direction which such bias might take is not

readily surmised.

Conventional QoL measures and a simple unvalidated questionnaire can be used to assess outcome in minor surgical procedures. Both demonstrate that minor operations improve quality of life, but with the caveat that there was no control population to compare with. To defer indefinitely such operations or exclude them from NHS waiting lists is to deny measurable health gain to these patients. This supports the inclusion of such operations in NHS provision.

CHAPTER SIX

Discussion: The application of health status measurement to surgical practice

"As all scientific knowledge is only an approximation to the truth, with sufficiently rigorous experimentation, the hypothesis will in time be falsified. As an inevitable consequence of this act of falsification, new data will be derived that will have to be incorporated as a new set of observations into a new hypothesis that better fits all the data. This, in turn, will have to be subjected to the most rigorous tests until it is once again found wanting. A cascade of hypotheses, together with their refutations, develops, so that we constantly approximate closer and closer to the truth, without ever having the arrogance to believe that we have arrived at the ultimate goal." (Baum, 1989b).

The hypothesis is: **Health status measures, scientifically applied, provide important additional information to the surgeon.**

In considering the hypothesis, it is first necessary to review what additional information has been provided by the measurement of QoL in the studies described. It is necessary to review the methodology employed, and the potential drawbacks and finally, the recent literature concerning QoL measurement in relation to surgical practice. The role and relevance of such methods in surgical practice is discussed.

6.1 Additional information provided by QoL measurement

6.1.1 Comparison of treatment interventions

The chemotherapy study established, using conventional measures of outcome, the UICC response criteria, that the aggressive chemotherapy regimen was the more successful. It was anticipated that QoL measurement may challenge this finding, but the results of the study confirm the findings of other

studies that response *per se* is associated with improved QoL. In common with other studies, the design of the chemotherapy study did not allow firm conclusions to be drawn as to how each regimen affected the QoL of patients not enjoying a response. They left the study on progression of disease. This, together with the finding that nonresponders started off with a poorer QoL may suggest chemotherapy had little opportunity to affect an already poor QoL, but supportive care, such as pain control and hospice care could have been more appropriate at this stage in their disease. The principle of QoL measurement in this context is, however, vindicated. The use of frequent QoL measurement and the innovation of a diary can be translated into other areas where, for specific surgical interventions, a survival difference is not anticipated and alternative interventions can only be assessed in terms of ongoing QoL. An example might be pouch *versus* ileostomy following total colectomy, where isolated postoperative QoL measurement may not give a complete account of QoL changes over time.

A potential advantage of using a universal QoL instrument, the NHP, in the minor surgery study was that direct comparison was possible with, for example, the results of the study by Caine *et al.* on coronary artery bypass surgery (Caine *et al.*, 1991). Although purists rightly caution against comparison between unlike populations, it may be helpful for surgeons to compare treatment interventions between studies using a "gold standard", especially in the context of the "quick, dirty experiment".

6.1.2 Measuring the health of populations- predicting outcome

Information regarding the anticipated outcome of treatment has a long tradition in the surgical literature, and usually concerns survival and operative morbidity. This depends upon the careful documentation of clinical signs and clinical and laboratory staging of disease. However, such detail can be largely absent from surgical practice outside specific studies, although computerised audit may change this. A most significant finding from each of these studies is that QoL scores *prior* to treatment could predict outcome.

The Qualitator score predicted response and survival in advanced breast cancer patients undergoing chemotherapy in both the King's and the Guy's studies. In the King's study, the LASA score was able to predict survival, the NHP score to predict response and the qualitator to predict both. No single factor pre-treatment predicted survival more accurately. The common-sense view is that people in their terminal months know they are unwell because they are weak, anorexic, in pain, or depressed and that these factors inter-relate. That the QoL measurements appear to support this confirms their validity. If this study is considered in isolation, it could be argued, however, that a large amount of work has been done, in order merely to establish that QoL measurement is as good a prognostic indicator as a simple liver scan. However peripheral this finding may seem at present, it may be of singular value in future, when difficult decisions are forced by renewed stringency in NHS budgets. The finding that better response, and therefore better QoL were obtained with the combination regimen (CMF) will comfort clinicians convinced

of the benefits of aggressive chemotherapy. However, the finding that response rates are predictable using QoL measurement offers the opportunity for a deeper study into a gearing treatment strategies to individual patients.

In the NSAP study, health related QoL measurements alone were used to try to identify the population at risk of having NSAP. However, the instruments used were inferior to clinical judgement in predicting NSAP. The simple expedient of recording the number of days' prodromal pain was better than questionnaires at predicting which patients would have NSAP and on the strength of this, QoL measurement would appear to have little to offer. However, in the self-selected subgroup who responded to postal questionnaires, the association between continued perception of pain and significantly poorer anxiety scores indicates that expectations of treatment outcome amongst individual patients differ, and can be measured numerically, and this may offer guidance as to a worthwhile direction for future study. A measure more appropriate to QoL and general rather than psychiatric health could perhaps have increased the accuracy of the QoL data. Nevertheless, the patients in this study with a degree of chronicity to their pain had consistently high scores at all stages, suggesting that for some patients, their perception of physical health may be less susceptible to physical intervention than might be supposed.

The small study of minor surgery patients demonstrated this phenomenon quite clearly: patients claiming a poor outcome had significantly poorer psychological scores before treatment. In this study, psychological and pain scores were of use in predicting outcome, but weight was added to their results

by the use of a general QoL instrument. Again, the common-sense view would be that a practised clinician can tell which patients will benefit from treatment. That this can be measured, however, may be of more than academic interest in future.

There is a case, therefore, to be made for conducting more detailed research upon predictors of outcome using QoL measurement in other areas to see if such findings are replicated and to explore the potential applications.

6.1.3 Measuring the health of populations- measuring outcome

In the chemotherapy study, outcome was measured using established methods and QoL measures. QoL measures supported the response data, but predicted patients who would respond. In abdominal pain and in minor surgery, outcome measurement is a poorly developed science. Mortality was zero in both studies and wound infection rates (nil in the minor surgery patients and inappropriate in NSAP patients) were irrelevant. The usual outcome measure after surgical admission for non-life threatening conditions is the answer to the question "How are you feeling now?" and in a busy outpatient clinic, this may be of lesser interest than the presence of a healed wound or the absence of surgical complications. It is possible that in each of these rather dissimilar studies, QoL measurement would have given full and pertinent outcome data even if uncomplemented by any other recordings. It is even possible to envisage a time when the majority of routine surgical follow-up could be conducted by postal questionnaire with the option of face to face encounter if desired.

6.1.4 Making treatment decisions on individual patients

In each of the studies, it was demonstrated that an individual patient's QoL measurement may be pertinent to their subsequent outcome, or perception of outcome. However, the use of QoL measurement alone to decide on a treatment strategy for an individual is frought with uncertainty. In the NSAP study, QoL measurement was singularly unsuccessful in predicting which patients would need surgical inpatient treatment, and in the other studies, although QoL was a good indicator of *groups*, QoL was too nonspecific to be a sole determinant of treatment in individuals. However, it could be envisaged that, in the context of other information, QoL data may be of use. For instance, in patients being assessed for routine interventions for non life-threatening conditions, knowledge of an individual's psychological or pain score may indicate the appropriate time scale for, or type of treatment. It is not unreasonable, however, to expect that an attentive and sympathetic clinician could evince such information without recourse to QoL measurement.

6.2 Methodological considerations

6.2.1 The choice of instrument

"Research into quality of life in medical care is dominated by sociologists and epidemiologists. Due to the methodologists' lack of clinical practice, the instruments available to them are far removed from clinical reality and are inadequate or unacceptable for clinical purposes." (Troidl, 1991).

At the time of commencing the studies in this thesis, the instruments available to measure QoL were either not specific for the conditions under study or

deemed to be lacking in important respects. Independently, Ware and Aaronson have suggested that in studies where QoL is measured, there should be a well validated general questionnaire and that this should be augmented by a specific questionnaire (validated, or *ad hoc*) focused on the condition being examined (Ware, 1987; Aaronson, 1988). Both state that where a specific instrument seems to be lacking for a particular angle of enquiry, existing instruments should be adapted, rather than attempting to "reinvent the wheel". In the studies comprising this thesis, this advice was followed.

In the chemotherapy study, the NHP and the LASA were chosen as the best available validated instruments. As judged by their ability to support the known facts on QoL in advanced cancer patients, this decision was correct. The development of the Qualitator was to fulfil a perceived gap among the then currently available instruments in capturing day-to-day variations in cancer chemotherapy patients and so a degree of *de novo* development was necessary. The finding that, using basic concepts which were not original, it was possible to derive a reasonably competent measure of QoL would support the adaptation of other instruments to surgical uses without repeating the whole development process. For instance, changing a few of the items of the Qualitator may make it an acceptable instrument for other conditions. Many possible applications unfold.

In the abdominal pain study, no generic QoL measure was used and this is a flaw because as a result, no data were available concerning any change in general health status. Thus the psychological factors had to be studied in

isolation, without any reference to, say, the pain experienced by these patients. This was addressed in the issuing of the postal questionnaires at 2 years, with interesting results concerning the perception of continuing pain- despite the poor response.

In the minor surgery study, this omission was rectified, with the inclusion of the NHP, which yielded important additional information regarding pain and general health.

In the latter two studies, the simple *ad-hoc* questionnaires yielded additional significant information. They were kept simple, and designed along lines recommended by other authors (Fitzpatrick, 1990). They were, however, unvalidated. Although several authors caution against using unvalidated measures, citing as a particular problem their face and content validity, Troidl points out the profound influence on peptic ulcer surgery which resulted from the use of an early, unvalidated measure, the Visick scale (Troidl, 1987). On a more practical level, all new measures are unvalidated when new. A more subtle drawback is the suspicion with which untried measures may be treated by other specialists.

A glance at the appendices to this thesis will confirm that QoL instruments are relatively uncomplicated. Moreover, having noted the problems associated with using an unvalidated instrument, for internal comparison in the same study using randomised controls or baseline control scores, there is no reason why such measures should not yield pertinent and useful information. However, in order to allow a more general appreciation or comparison in the

context of other studies, it is wise to use a validated measure as well. It is necessary that if QoL measures are to be widely used by surgeons, their simplicity be made known and some general instruments be adopted for comparison between studies.

6.2.2 Frequency of testing

In the vast majority of studies where QoL is measured, measurement is made once, before a treatment intervention and on a limited number of occasions afterwards. Little specific advice exists in the literature as to how often measurements should be made. Cox *et al.* (1992) advise of the need for base-line observations, avoidance of unnecessary assessments to the detriment of doctor or patient compliance and targeting of assessments a) to distinguish early from late treatment effects, b) to reflect the pattern of treatment administration and c) to concentrate measurements when maximum treatment response is expected. In the studies described, this pattern was generally followed, although it could be argued that more frequent measurement would have elicited useful information about the status of the acute abdominal pain patients during and immediately after their hospital stay, and similarly for the minor surgery patients. However, the logistics prohibited this. The use of the Qualitator, a diary, would seem to have been an appropriate solution to the logistical problem in the chemotherapy patients, and may be a useful line of investigation in the future in studies involving perioperative patients.

6.2.3 Compliance

A methodological problem common to all the studies is the rate of non-response to questionnaires. This is especially relevant to postal questionnaires and to studies of patients who are basically physically well. Poor patient compliance is blamed for deficiencies in design or conduct of many studies, not just those involving QoL measurement. However, in QoL studies, it is particularly important that the completion rate is good and this is the responsibility of the clinician. The studies in this thesis suffer from a poor completion rate in a variety of ways. The most notable is in the abdominal pain study, in which the two year questionnaires were returned by only 52% of patients- despite two postal sweeps and telephoning by a research assistant. In the minor surgery study, in which follow-up was shorter, and in which three postal sweeps were sent, together with telephoning of the intransigent non-responders by the principle investigator, the fall-out rate was still 19%. In the chemotherapy study, data were not gathered when the routine was upset by treatment delay or progression of disease. In every study, at every stage, therefore there was a loss of data, introducing the possibility of volunteer bias. The effect of this on the validity of each study's findings is not known. However, if it is accepted that even with committed investigators such gaps could occur, it is logical that a less committed approach may produce larger gaps. This is probably one of the major flaws in any QoL study, and one for which there is no single solution; the general rule that it only takes one committed person to make a success of such studies still applies, however. This

view is emphasised by Williams in a discussion of QoL strategies in Surgery (Williams, 1991). It does also affect the potential of QoL measurement as a routine tool in surgical audit.

An ethical problem arises in the pursuit of non-responders. To do so too assiduously may be to breach the ethical line which should prevent the researcher from interfering with the QoL of the patient. Patients have, after all, a right not to comply. This calls for tact, discretion and persuasiveness in addition to empathy on the part of any investigator in the field of QoL research, a factor often omitted in the discussion of methodological issues in QoL research. In this respect, QoL research has much in common with the older science of market research, already a valued tool in the manufacturer's armamentarium, but subject to the vagaries of public opinion and individual mood.

6.2.4 Statistical considerations

In spite of the amount of work done in recent years to validate instruments which measure QoL, relatively little has been written on how to process the data which accrue from QoL measurement. This has been discussed previously (section 1.4). Although non-parametric tests are relatively easy to perform, there arose several problems in the conduct of these studies which have not been addressed in the literature and for which, moreover, three statisticialns consulted by the author, and attendance at a forum on the very subject (Cox *et al.*, 1992), offered no clear answers.

First, is the question of when do patient numbers in a study become large enough to justify logarithmic transformation of the data? This issue was avoided altogether because of the ample justification in the literature for using non-parametric tests, but there may have been a considerable loss of sensitivity as a result.

Second, is how to cope with missing data. This was handled in a variety of ways in this thesis, but no help was gleaned from published literature, which is presumably heavily biased in favour of studies with near-complete databases. In the Qualitator studies, the odd day, week or even month of missing data was simply given the mean score for the measurements on either side. Looked at closely, this method is rather unscientific, as it cannot be assumed that quality of life carried on in a continuum during the missing period. Anything could have been happening- in fact, in those with marrow toxicity, QoL could have reached a low. In the chemotherapy patients who had missing NHP or LASA data at any point during the study, these were omitted from the analysis. However, the effect of having these "holes" in the data, is to reduce the validity of the overall analysis. The only solution which appeared honest was to perform a separate comparison each month, with the pre-treatment scores, with the data available, rather than to use a method requiring complete data, such as the Friedman analysis.

The other problem is the fall-out rate of patients from the chemotherapy study. This was tackled in two ways: again, by analysing data every month; then by comparing responders in both groups (as it is the responders who generally

remained in study the longest). Only one paper addressed this problem directly- in response to a similar fall-out in animal tumour immunotherapy experiments (Koziol *et al., 1981).* This method, which involves calculation of a multivariate rank statistic and its permutation distribution, would probably have been entirely appropriate, but was not used in the end, because, on a pragmatic level, the data analysis (intended to be repeatable by others) was already becoming unduly complicated.

In the NSAP and minor surgery studies, the absence of data at entry, or follow-up effectively removed the patient from that particular analysis, reducing the power of the studies at each point of data collection. As mentioned before, there is only one solution to this- better data collection at the time. The alternative is to repeat *all* questions again at a later timepoint and include this as a separate analysis.

Finally, threshold scores, which were used to denote "caseness" in the NSAP study, but avoided in the minor surgery study, may distort the data. When comparing two groups, and not necessarily wishing to quantify them for out-and-out psychological pathology, is is better to use comparative non-parametric tests. In fact, both methods *were* used in this study, and no differenc found in the overall results. However, the raising or lowering of the threshold score in the Spielberger data would have *removed* any statistical difference between the "Pain" and "No pain" groups, when the Mann-Whitney confirmed that there was such a difference (Table 4.7).

Amongst other impediments to the easy statistical processing of Qol

data, the one most consistently encountered during the background reading for this thesis was how, even in papers in reputable journals, so little space was devoted to the details of statistical methods.

6.2.5 Allocating resources

One of the biggest obstacles to the introduction of new research methods into clinical practice is the time which is required to gather the data. Very often the data gathering is left to the most junior member of the medical or nursing staff. If, as is usually the case, this person has little or no personal investment in the project, then data collection will be deficient. In all the studies described in this thesis, the vast majority of data were collected by the author. Where this was not the case, co-investigators had considerable commitment to the study in question. In the chemotherapy study, the time involved in administration of treatment, counselling and QoL data collection alone was two days per week for the 18 month duration of the study, followed by several months of data processing. In the other two studies, administration of the control questionnaires was conducted by a variety of staff, at a total cost in time of about 20 minutes per patient. The postal questionnaires took considerable time to organise. Finally, the processing of data from the latter two studies took approximately two months. So the time consequences of setting up a study in which QoL is to be measured properly are considerable and should be planned for at the start. Ideally, one person should be responsible for data collection. Where this is not the case, compliance inevitably deteriorates,

threatening the validity of the eventual findings of the study.

6.3 The current state of QoL measurement in surgical practice

So what has been happening in surgery and how does the surgeon stand in relation the QoL research that developed in the last two decades? In 1989, Troidl suggested several areas in which QoL could be measured in Surgery (see Table 6.1).

Clinical situation	Surgical examples	Ranked importance of outcome variables
Impaired QoL but not life threatening	Hernia Gallstones Lacerated meniscus Preipheral arterial occlusion	Discomfort> Disability> Disease> Dissatisfaction> Death
Impaired QoL and life threatening: surgery prolongs life but increases morbidity	Stoma for colitis Amputation for vascular disease Transplantion	Discomfort> Disability> Disease> Dissatisfaction> Death
Different therapeutic approaches: similar operative mortality, complications and survival	Arterial occlusion: bypass vs. profundoplasty Stomach cancer: different reconstructions	Discomfort> Disability> Dissatisfaction> Disease> Death
Trade-off between better QoL and increased therapeutic risk	Colitis: stoma vs. pelvic pouch Osteoarthrosis: conservative vs. endoprosthesis	Discomfort> Death> Disability> Dissatisfaction> Disease
Palliation	Oesophageal carcinoma Pancreatic carcinoma	Discomfort> Disability> Dissatisfaction> Disease> Death

Table 6.1: Example of ranking the importance of outcome variables in surgical conditions (Troidl, 1991)

In 1987 it was reported that only 3% of trials reported in surgical journals mentioned QoL (O'Young J and McPeek B, 1987). Four years later, there has been progress. Studies measuring QoL have been reported in neurosurgery (Schulte *et al.*, 1987; Bach *et al.*, 1988; Stewart-Amedei and Penckofer, 1988; McKenna *et al.*, 1989; Trojanowski *et al.*, 1989), carotid endarterectomy (De Leo *et al.*, 1987), head and neck cancer (Strauss, 1989; Rathmell *et al.* 1991), congenital heart disease (Shimada and Tsunemoto, 1990; Torii *et al.*, 1990; Aigueperse and Marechal, 1991), coronary artery and valve surgery (Mayou and Bryant, 1987; Langeluddecke *et al.*, 1989; Rogers *et al.*, 1990; Jenkins *et al.*, 1990; Caine *et al.*, 1991; Booth *et al.*, 1991), cardiac dysrhythmia surgery (de Carvalho *et al.*, 1989), heart transplantation (Lawrence and Fricker, 1987; Mai *et al.*, 1990), aortic aneurysm (Rohrer *et al.*, 1988), oesophageal and gastrointestinal surgery (Habu *et al.*, 1988; Sakamoto *et al.*, 1989; Roder *et al.*, 1990; Buhl *et al.*, 1990; Nogughi *et al.*, 1991), hepatobiliary surgery (Spina *et al.*, 1988; Little and Wong, 1991), liver transplantation (Kober *et al.*, 1990; Tarter *et al.*, 1991), colorectal surgery (Kennedy, 1988; Yasutomi *et al.*, 1988; Drossman *et al.*, 1989; Oresland *et al.*, 1989; Pemberton *et al.*, 1989; Walsh *et al.*, 1990; Anseline, 1990), benign and malignant prostatic surgery (Fowler *et al.*, 1988; Singer *et al.*, 1991), renal tract stones (Mays *et al.*, 1990), gynaecological cancer (Sichel, 1990) and early breast cancer (Ganz *et al.*, 1990), However, in 1991, no study reported in the British Journal of Surgery contained QoL measurement of any description, even though several would have uncovered

clinically relevant information by doing so. Why should this be, especially when the practice of audit has been deeply rooted in many surgical centres and is set to become established in many more?

6.4 The Meran consensus conference

Many areas of contention in surgical practice come down to straight Quality of Life issues. This was recognised by a group of surgeons and QoL methodologists who met in Meran, Italy, in October 1989 to discuss the issues and to formulate objectives and guidelines for QoL measurement in surgical practice (Neugebauer *et al.*, 1991a). It had been recognised that in a changing world, traditional technical measures of surgical success were no longer sufficient justification for surgical treatment. The conference met in small multidisciplinary groups in the specialties of transplantation, thoracic, cardiovascular, trauma/orthopaedic and abdominal surgery to select related diseases where formal QoL assessment was a priority, to consider how useful such information was in deciding whether to operate or not, and to monitor or evaluate patient status. The important QoL domains in each disease state were defined (see Table 6.2) and finally, the QoL instruments were reviewed. The results were then presented by each group to a plenary session and are summarised in an issue of *Theoretical Surgery* given over to the reporting of the conference (Neugebauer *et al.*, 1991b). Each specialty group provided a comprehensive list of specific situations in which QoL measurement would be appropriate.

There was broad agreement as to the obstacles which prevent QoL measurement being taken up widely at present. First was the perception amongst clinicians used to handling laboratory or pathological data that QoL measurement somehow represents "soft", or less respectable data. Second, was the perception amongst surgeons that their clinical judgement of a given situation was likely to be more reliable than QoL data derived from such a situation. Third, was the complaint that QoL measurement would consume significant resources from those whose workloads were already considerable. More specifically, the sheer number of QoL instruments, without a definite "gold standard", may act as a deterrent to their use and finally, evidence is still lacking that, even in specialties where QoL research has been most prolific, the results of studies make any actual difference to clinical practice.

Priority rating	Transplant		Thoracic		Cardio-vascular		Trauma/ orthopaedic		Abdominal	
High	Renal failure	bdgh	Meso-thelioma		Coronary artery disease	efgd (ab)	Brain Spinal cord	ahif hi	Peptic ulcer	bcag
	Liver failure	cdfh	Metastases		Congenital	gf(b)	Burns reconstr.	gib	Ch pancreatitis	g
	Pancreatic failure	bcg	Breast Ca: -reconstructon -Adjuvant	i^2	Cardio-myopathy	d	Fractures	afbie	Portal HT/ varices	bg
	Heart failure	cdfh	Oesophagus -Benign -Malignant	g g	Peripheral vascular disease	abg	Spinal low back pain	abe hi	Crohn's/ UC	gbc
	GIfailure	bcg	Bronchial Ca -advanced	bg			Hand	abegni f	GI malignancy	
							Joint		Curative	ba
								abcefg hi	Palliative	gb
									Adjuvant	b
Mediu m			Breast Ca -"curable"	bg	Valve replacemen t	bef(g)	Craniofacial Amputation	gfh i^1h	Polyposis coli	b
					Arrhythmi as	abdg				
					Carotid surgery	af	Visceral	i		
					Var. veins	ag				
Low			Bronchial Ca -"curable" mediastinal tumours	g	Aneurysm	h	Burns-acute	j	Appendicit is	c
									Hernia	
									Piles	c
									Gallstones	

Table 6.2: Reasons for choice: a = prevalence of disease; b = no. of treatment options; c = newness of intervention; d = scarcity of resources; e = patient characteristics; f = timepoint of intervention; g = QoL is goal of intervention; h = others; i^1 = impact of capacity; i^2 = treatment associated morbidity; j = survival vs QoL

6.5 The future of health status measurement in surgical practice

Evidence is lacking, as to whether or not treatment or management policies of people who do not conduct QoL research have been substantially altered by the results of those who do. This is illustrated by the abandonment of amputation in favour of limb-sparing surgery for extremity sarcoma despite evidence that QoL was no better in patients having the conservative procedure (Sugarbaker *et al.*, 1982). This may in part be due to ignorance amongst clinicians of work on QoL (Troidl, 1991). It is noteworthy that in plastic surgery, a specialty where QoL must be the major, if not the only endpoint, in no study in the decade up to 1990 was QoL measured as an endpoint (Spilker and Stark, 1991).

In the reorganised NHS and with the advent of the Patients' Charter, and competition for the internal market, QoL considerations may require more complete documentation and may, in future, be scrutinised more closely than was previously considered necessary. That the technology now exists, as indicated by these studies, offers cold comfort to surgeons already steeling themselves to deliver more accurate audit. However, if surgeons became familiar with the potential advantages and pitfalls of QoL measurement, they could rely less heavily upon the skills of the methodologists, who are less familiar with the complexities of the delivery of surgical care. The incorporation of QoL measurement into clinical audit is technically feasible, and is one of the few ways in which treatment strategies can be evaluated other than by measures of survival, complications and activity.

6.5.1 The changing climate in health care delivery- non-surgical considerations

The two biggest influences on treatment strategy are still the amount of available resources on one hand, and the interacting goals of provider and consumer on the other. The latter factor is influenced largely by fashion, which includes the provider-driven introduction of new technologies. If QoL research is to have any bearing upon the practice of surgery, then neither in the reorganised NHS nor in health systems similarly constrained financially, will it still be a matter left to the discretion of the surgeon as in days of old. The purchaser of surgical services is now all-powerful.

The danger is that QoL data may not be used to make impartial judgements on scrupulously gathered and processed data, leading to changes in treatment practices where appropriate, but to support dogma:

"It is a common failing- and one that I have myself suffered from- to fall in love with a hypothesis and to be unwilling to take no for an answer. A love affair with a pet hypothesis can waste years of precious time." (Medawar, 1979).

It therefore becomes all the more important that a scientific approach evolves to the measurement of QoL and health status in surgical practice, with well validated instruments and methodology.

From the three studies described, it can be seen that a large amount of data accrue. Depending upon the views of the author, such data can be adapted or omitted to suit the goals of the study. It is important, therefore, that a study involving QoL measurement starts out with a clear *a priori* hypothesis which is not adapted or forgotten under the deluge of data which accumulate. This will become even more relevant as there is considerable commercial

advantage to be gained by drug companies which can demonstrate a QoL advantage to their particular product: many more pharmaceutical companies are using QoL measurement in their trials (Luce *et al.*, 1989). This early in the use of QoL measurement in surgery, no consensus has yet formed as to the approaches to be adopted.

6.5.2 Surgical trends

The delivery of health care in Europe and in the USA is becoming much more carefully costed and, consistent with social trends, much more consumer-sensitive. It is inevitable that the delivery of surgical services will bend to the same influences. There is recognition amongst influential surgeons in the UK not only that closer scrutiny of outcome is warranted, but that the structure and process of the delivery of surgical services be examined (Bates, 1990; Kettlewell, 1990; Ellis, 1991). The Royal college of Surgeons of England has an Audit Unit, and although its main efforts are directed towards the standardisation of information gathering regarding process and mortality, amongst other projects underway is the Patient Satisfaction Study, designed to assess the extent to which the clinical treatment and outcome meets the patients' expectations of care and recovery (Emberton *et al.*, 1991). This involves validation of questionnaires which clinicians can use routinely to investigate whether services meet patients' expectations and requirements and it will examine closely the whole process of surgical care delivery, as it is experienced by the patient (Meredith, 1991). Health status measurement has

been sanctioned by the English Royal College of Surgeons in assessing the outcome in prostatectomy patients (Emberton, 1992). A study in the Freeman Hospital, Newcastle has now gathered a large number of patients in whom the NHP has been used to measure outcome (Coles, 1990). This study has bridged the period during which laparoscopic cholecystectomy has all but taken over as the orthodox treatment for gallstones so the results will be of great clinical relevance during the tidal wave towards the laparoscopic operations.

Baum has argued eloquently for the randomised trial, and quotes Maimonides, of 12th century Alexandria: "Teach thy tongue to say I do not know, and thou shall progress". David Hume, the Scottish philosopher averred:

> "No testimony is sufficient to establish a miracle unless the testimony be of such a kind that its falsehood would be even more miraculous than the fact which it endeavours to establish" (Baum, 1989a).

The trouble is, that the basic instinct of surgeons and patients alike is to search for and seize upon the miracle cure.

Minimally invasive surgery (MIS) is the most obvious example of a "quantum leap" in contemporary surgical practice. This area seems to be well suited to QoL research, as the main reason for performing MIS is to reduce the incidence of wound pain and the length of hospital stay, so improving the quality of life of the patient undergoing such surgery. Amongst the minimally invasive operations, laparoscopic cholecystectomy does seem truly miraculous and has now become the "norm", so that the opportunity to conduct a randomised trial of this, with conventional cholecystectomy, measuring QoL, has almost certainly been lost.

The rapid introduction of laparoscopic cholecystectomy has presented enquiring surgeons with a dilemma. On one hand no clinician would wish to deny their patient the option of a less invasive procedure; on the other, there is a learning curve, and there is anecdotal evidence that considerable morbidity may result from laparoscopic procedures. The "obviously" better treatment may not be so much better after all, as was found by Mays *et al.* who reported no difference in QoL between patients having extracorporeal shock wave lithotripsy compared to the more invasive percutaneous therapy (Mays *et al.*, 1990). However, even if a measured strategy were acceptable to clinicians, then it would probably not be acceptable to patients. Neugebauer *et al.* believe they have done the next best thing, which is to measure as many outcome indicators as they can, including QoL, as their laparoscopic cholecystectomy programme accelerates (Neugebauer *et al.*, 1991c). What will be the conclusions in five years' time? The indications at present are that laparoscopic cholecystectomy may have cost some patients a vastly impaired QoL or even their lives in order to give the rest a marginally improved QoL. Under such circumstances, measurement of QoL may have singular importance in establishing the correct approach to gallbladder disease. The relative merits of conservative or operative treatments would only have been quantifiable if, in addition to the other measures of process and outcome, QoL measurement had also been undertaken.

Although examples of important QoL issues in surgery are many, QoL measurement will not suddenly transform surgical research and practice.

Moreover, there is not an obvious single method of quantifying QoL for all situations, however attractive the simplistic approach may seem: "The answer to the question of the meaning of life is.. 42" (Adams, 1979). However, momentum is gaining and certain key studies are likely to appear in the coming years. They will command wider attention which may encourage others to undertake further research (Goligher, 1987). The principles and practice of QoL measurement in surgery will only become clearly delineated thereafter. From the studies in this thesis, some pointers emerge: Reliable data can be obtained by including a general, well validated measure, and a specific, simple measure directed towards the condition under study. In trials, controls should, where possible, be used. Statistical analysis should be by simple, nonparametric methods. The pre-intervention QoL of study patients may have considerable bearing upon the post-intervention QoL measurements. Common sense is one of the most important ingredients in QoL research.

In the area of surgical audit, which as mentioned previously is becoming standardised, it may be that if computer software should allow later inclusion of outcome measurement using QoL, *built in* to the data capture systems, self-assessment questionnaires could form a key part of the process of surgical admission, discharge and follow up, and contribute much information, not just about the achievement of treatment objectives, but about the whole "surgical experience" seen from the patient's viewpoint- as yet uncharted territory. Centuries after the dawning of the "Age of Enlightenment" there is still a paucity of objective information on how surgery touches the lives of those

undergoing it.

The role of health status measurement in surgical practice is not yet defined. Its measurement in specific studies seems set to increase, but such studies will only gain credence if they are properly conducted. This will only occur if surgeons themselves devote some energy to this new and exciting area of enquiry.

REFERENCES

A'Hern R P, Ebbs S R, Baum M. Does chemotherapy improve survival in advanced breast cancer? A statistical overview. British Journal of Cancer 1988, 57:615-618.

Aaronson N K. Quantitative issues in health-related quality of life assessment. Health Policy 1988, 10:217-230.

Adams D. The Hitch Hiker's Guide to the Galaxy. BBC Radio 4, 1974. London: Pan Books 1979.

Addington-Hall J M, MacDonald L D, Anderson H R. Can the Spitzer Quality of Life Index help to reduce prognostic uncertainty in terminal care? British Journal of Cancer 1990, 62:695-699.

Aigueperse J, Marechal MC. Evaluation of the quality of life in adulthood of 158 patients surgically-treated for tetralogy of Fallot. Archives Des Maladies Du Coeur Et Des Vaisseaux 1991, 84:685-90.

Alexander JL, Willems EP. Quality of Life: some measurement requirements. Archives of Physiotherapy and Medical Rehabilitation 1981, 62:261-265.

American Joint Committee for Cancer. Staging and end results reporting: Manual for staging of cancer. Chicago, Illinois: American Cancer Society 1977.

Anseline PF. Quality of life after restorative proctocolectomy. Australian and New Zealand Journal of Surgery 1990, 60:683-8.

Apgar V. A proposal for a new method of evaluation of the newborn infant. Anaesthesiology and Analgesia 1953, 32:260-267.

Bach M, Schramm J, Brock M. Quality of life and mortality following major neurosurgical interventions in advanced age. Zeitschrift Fur Gerontologie 1988, 21:134-7.

Bardelli D, Saracci R. Measuring the quality of life in cancer clinical trials: a sample survey of published trials. UICC (International Union Against Cancer) Technical Report 1978, 36:75-94.

Barraclough BM. Appendicectomy in men. Journal of Psychosomatic Research 1967, 2:203-205.

Bates T. Avoiding inappropriate surgery: a discussion paper. Journal of the Royal Society of Medicine 1990, 83:176-178.

Baum M, Priestman T, West R R, Jones E M. A Comparison of subjective responses in a trial comparing endocrine with cytotoxic treatment in advanced carcinoma of the breast. European Journal of Cancer 1980 (supplement 1):223-226.

Baum M. Rationalism versus irrationalism in the care of the sick: science versus the absurd. The Medical Journal of Australia 1989, 151:607-8.

Baum M. Limitations of nonscience in surgical epistemology. Journal of Technology Assessment in Health Care 1989, 5:381-389.

Bergner M. Quality of Life, Health Status and Clinical Research. Medical Care 1989, 27;3(suppl):148-156.

Blanton s, Kirk V. A psychiatric study of 61 appendicectomy cases. Ann Surg 1947, 126:305-314.

Bonadonna G, van Oosterom A. Treatment of advanced breast cancer; workshop report. European journal of Cancer & Clinical Oncology 1983, 19;12:1779-1781.

Bond A, Lader M. The use of analogue scales in rating subjective feelings. British Journal of Medical Psychology 1974, 47:211-218.

Booth DC, Deupree RH, Hultgren HN, DeMaria AN, Scott SM, Luchi RJ. Quality of life after bypass surgery for unstable angina. 5-year follow-up results of a Veterans Affairs Cooperative Study. Circulation 1991, 83:87-95.

Boyd N F, Selby P J, Sutherland H J, Hogg S. Measurement of the clinical status of patients with breast cancer: evidence for the validity of self-assessment with linear analogue scales. Journal of Clinical Epidemiology 1988, 41;3:243-250.

Brewster S, Newman J. Can knee replacements be assessed by post? Health Trends 1991, 23;3:113-114.

Brown RA, Beck JS. Medical Statistics on Microcomputers. London: British Medical Journal 1990.

Buhl K, Schlag P, Herfarth C. Quality of life and functional results following different types of resection for gastric carcinoma. European Journal of Surgical Oncology 1990, 16:404-9.

Bulpitt CJ, Fletcher AE. Quality of life in hypertensive patients on different anti-hypertensive treatments: rationale for methods employed in a multi-centre randomised controlled trial. Journal of Cardiovascular Pharmacology 1985, 7(suppl):137-145.

Bunker JP, Wennberg JE. Operation rates, mortality statistics and the quality of life. New England Journal of Medicine 1973, 289:1249-1251.

Byrne M. Cancer chemotherapy and quality of life. British Medical Journal 1992, 992:1523-1524.

Caine N, Harrison SC, Sharples LD, Wallwork J. Prospective study of quality of life before and after coronary artery bypass grafting. British Medical Journal 1991 302:511-6.

Cassileth BR, Lusk EJ, Guerry DuP, Blake AD, Walsh WP, Kascius L, Schultz DJ. Survival and quality of life among patients receiving unproven as compared with conventional cancer therapy. New England Journal of Medicine 1991, 324;17:1180-1185.

Chlebowski R T, Smalley R V, Weiner J M, Irwin L E, Bartolucci A A, Bateman J R. Combination *versus* sequential single agent chemotherapy in advanced breast cancer: associations with metastatic sites and long-term survival. British Journal of Cancer 1989, 59:227-230.

Coates M D, Gebsky v, Stat M and 13 others. Improving the Quality of Life during chemotherapy for advanced breast cancer. A comparison of Continuous and intermittent treatment strategies. New England Journal of Medicine 1987, 317:1490.

Clement-Jones V. Cancer and Beyond: the formation of BACUP. British Medical Journal 1985, 291:1021-23.

Cochrane M, Heginbotham C, Smith R. Rationing: at the cutting edge. British Medical Journal 1991, 303:1039-1042.

Coles J. Outcomes management and performance indicators. In: Measuring the Outcomes of Medical Care. Eds. Hopkins A and Costain D. London: Royal College of Physicians of London 1990.

Cooper R G. Combination chemotherapy in hormone resistant breast cancer. Proceedings of the American Association for Cancer Research 1969, 10:15.

Cox DR, Fitzpatrick R, Fletcher AE, Gore SM, Spiegelhalter DJ, Jones DR. Quality of life assessment: Can we keep it simple?. Journal of the Royal Statistical Society 1992, 155;3 (in press).

Creed F. Life events and appendicectomy. Lancet 1981, 1:1381-1385.

de Carvalho TA, Sosa E, Marcial MB, Scanavacca M, Bonetti D, Bellotti G, Pileggi F. The quality of life of patients with cardiac arrhythmia undergoing surgical treatment. Arquivos Brasileiros De Cardiologia 1989, 53:147-50.

De Dombal FT. Acute abdominal pain- an OMGE survey. Scandanavian Journal of Gastroenterology 1979, 14(suppl):29-43

De Dombal FT. Computers and the surgeon- a matter of decision. Surgical Annals 1979, 11:33-57.

De Haes JC, Van Knippenberg FC, Niejt JP. Measuring psychological and physical distress in cancer patients: structure and applications of the Rotterdam Symptom Checklist. British Journal of Cancer 1990, 62:1034-1038.

De Leo D, Serraiotto L, Pellegrini C, Magni G, Franceschi L, Deriu GP. Outcome from carotid endarterectomy. Neuropsychological performances, depressive symptoms and quality of life: 8-month follow-up. International Journal of Psychiatry In Medicine 1987, 17:317-25.

De Moulin D. A short history of breast cancer. Dordrecht: Kluwer Academic Publishers 1989.

Donabedian A. The definition of quality and approaches to its assessment. Ann Arbor, Michigan: Health Administration Press 1980, 79-128.

Donabedian A. The definition of quality and approaches to its assessment. Ann Arbor, Michigan: Health Administration Press 1980, 3-76.

Drossman DA, McKee DC, Sandler RS, Mitchell CM, Cramer EM, Lowman BC, Burger AL. Psychosocial factors in the irritable bowel syndrome. Gastroenterology 1988, 95:701-708.

Drossman DA, Patrick DL, Mitchell CM, Zagami EA, Appelbaum MI. Health-related quality of life in inflammatory bowel disease. Functional status and patient worries and concerns. Digestive Diseases & Sciences 1989, 34:1379-86.

Duncan W. Caring or curing: conflicts of choice. Journal of the Royal Society of Medicine 1985, 78:526-535.

Ebbs S R, A'Hern R P, Graham H, Baum M. Subjective measurements of quality of life predict response to chemotherapy for advanced breast cancer (abstract). British Journal of Surgery 1988, 75;6:601.

Ebbs SR, Fallowfield LJ, Fraser SCA, Baum M. Treatment outcomes and quality of life. International Journal of Technology Assessment in Health Care 1989, 5:391-400.

Ebbs S R, Graham H, A'Hern R P et al. Advanced breast cancer: A randomised trial of epirubicin at two different dosages and two administration systems. Acta Oncologica 1989, 28:fasc.6.

Elkington JR. Medicine and the Quality of Life. Annals of Internal Medicine 1966, 64:711-714.

Ellis BW. Surgical workload: variation in throughput. Annals of the Royal College of Surgeons of England 1991, 73;4(suppl):72-74.

Emberton M, Rivett R, Ellis BW. Comparative Audit: a new method of delivering audit. Annals of the Royal College of Surgeons of England 1991, 73;6(suppl):117-120.

Emberton M. Personal communication. 1992.

Epstein AM, Hall JA, Tognetti J, Son LH, Conant L. Using proxies to evaluate quality of life. Medical Care 1989, 27;3(suppl):91-98.

Fallowfield LJ, Baum M, Maguire GP. Effects of breast conservation on psychological morbidity associated with diagnosis and treatment of early breast cancer. British Medical Journal 1986, 293:1331-1334.

Fallowfield LJ. The Quality of Life- The missing measurement in health care. 1990 Souvenir Press, London.

Fayers PM, Jones DR. Measuring and analysing quality of life in cancer clinical trials: a review. Statistic in Medicine 1983, 2:429-446.

Feinstein AR. Clinical Biostatistics XLI. Hard science, soft data, and the challenge of choosing clinical variables in research. Clinical Pharmacology and Therapeutics 1977, 22:485-498.

Feinstein AR. Clinimetric perspectives. Journal of Chronic Disease 1987, 6:635-640.

Feinstein AR. Clinimetrics. Newhaven, Connecticut: Yale University Press 1987.

Fitzpatrick R. Measurement of patient satisfaction. In: Measuring the Outcomes of Medical Care. Eds. Hopkins A and Costain D. London: Royal College of Physicians of London 1990.

Fowler FJ Jr, Wennberg JE, Timothy RP, Barry MJ, Mulley AG Jr, Hanley D. Symptom status and quality of life following prostatectomy. Journal of the American Medical Association 1988, 259:3018-22.

Fraser SCA, Ebbs SR, Dobbs HJ, Fallowfield LJ, Baum M. The design of advanced breast cancer trials - new approaches. Acta Oncologica 1990, 29:397-400.

Ganz PA, Schag CA, Cheng HL. Assessing the quality of life--a study in newly-diagnosed breast cancer patients. Journal of Clinical Epidemiology 1990, 43:75-86.

Gardner MJ, Altman DG. Statistics With Confidence. London: British Medical Journal 1989.

Geddes D M, Dones L, Hill E, Law K, Harper P, Spiro S G, Tobias J S, Souhami R L. Quality of Life during chemotherapy for small cell lung cancer: Assessment and use of a daily diary card in a randomised trial. European Journal of Cancer 1990, 26;4:484-492.

Godlee F. Minor procedures off the list. British Medical Journal 1991, 302:311.

Goldberg DP. The detection of psychiatric illness by questionnaire. London: OUP 1972 (Maudsley Monograph 21).

Goldberg DP. Manual of the General Health Questionnaire. London: NFER-Nelson 1979

Goligher JC, Pulvertaft CN, De Dombal FT. Five to eight year results of Leeds/York controlled trial of elective surgery for duodenal ulcer. British Medical Journal 1968, 2:781.

Goligher JC. Judging the quality of life after surgical operations. Journal of Chronic Disease 1987, 40:631-633.

Gomez J, Dally P. Psychologically mediated abdominal pain in surgical and medical outpatient clinics. British Medical Journal 1977, 1:1451-1453.

Gray DWR, Collin J. Non-specific abdominal pain as a cause of acute admission to hospital. British Journal of Surgery 1987, 74:239-242.

Gray DWR, Dixon JM, Collin J. The closed eyes sign: an aid to diagnosing non-specific abdominal pain. British Medical Journal 1988, 297:837.

Guyatt G, Mitchell A, Irvine EJ, Singer J, Williams N, Goodacre R, Tompkins C. A new measure of health status for clinical trials in inflammatory bowel disease. Gastroenterology 1989, 96:804-810.

Guyatt G H, Bombardier C, Tugwell P X. Measuring disease-specific quality of life in clinical trials. Canadian Medical Association Journal 1986, 134:889-95.

Habu H, Saito N, Sato Y, Takeshita K, Sunagawa M, Endo M. Quality of postoperative life in gastric cancer patients seventy years of age and over. International Surgery 1988, 73:82-6.

Harris J. Life: quality, value and justice. Health Policy 1988, 10:259-266.

Harris A L, Cantwell M J, Carmichael J, Wilson R, Farndon J, Dawes P, Ghani S, Evans R G B. Comparison of short-term and continuous chemotherapy (mitozantrone) for advanced breast cancer. Lancet 1990, 335:186-90.

Hayes MHS, Patterson DG. The graphic rating scale. Psychological Bulletins 1921, 18:98-99.

Heaton KW, Ghosh S, Braddon FEM. How bad are the symptoms and bowel dysfunction of patients with the irritable bowel syndrome? A prospective, controlled study with emphasis on stool form. Gut 1991, 32:73-79.

Holloway CA. Decision making under uncertainty: Models and choices. Englewood Cliffs 1979, NJ: Prentice-Hall.

Hughson AVM, Cooper AF, McArdle CS, Smith DS. Psychological impact of adjuvant chemotherapy in the first two years after mastectomy. British Medical Journal 1986, 293:1268-1271.

Hughson AVM, Cooper AF, McArdle CS, Smith DS. Psychological effects of radiotherapy after mastectomy. British Medical Journal 1987, 294:1515-1517.

Hunt S M, McEwen J, McKenna S P. Manual of the Nottingham Health Profile 1980, Appendix: 240-283 (copyright Hunt, McEwen, McKenna 1980).

Hunt SM, McEwen J, McKenna SP, Backett EM, Pope C. Subjective health assessments and the perceived outcome of minor surgery. Journal of Psychosomatic Research 1984, 28: 105-14.

Hunt S M, McEwen J, McKenna S P. Measuring health status: A new tool for clinicians and epidemiologists. Journal of the Royal College of General Practitioners 1985, 35:185-8.

Irvin TT. Abdominal pain: a surgical audit of 1190 emergency admissions. British Journal of Surgery 1989, 76:1121-5.

Jenkins CD, Jono RT, Stanton BA, Stroup-Benham CA. The measurement of health-related quality of life: major dimensions identified by factor analysis. Social Science & Medicine 1990, 31:925-31.

Jenkinson C. Why are we weighting? A critical examination of the use of item weights in a health status measure. Social Science in Medicine 1991, 32(12): 1413-6.

Jennett B. High technology medicine: benefits and burdens. Oxford: Oxford Medical Publications 1986, 144.

Jones W G. Weekly low-dose epirubicin: effective single agent therapy for advanced breast cancer without significant toxicity. Unpublished study 1988.

Joyce PR, Bushell JA, Walshe JW et al. Abnormal illness behaviour and anxiety in non-organic abdominal pain. British Journal of Psychiatry 1986, 149:57-62.

Kaplan E L, Meier P. Non-parametric oestimation from incomplete observation. Journal of the American Statistical Association 1958, 53:451.

Karnofsky D and Burchenal J. The clinical evaluation of chemotherapeutic agents in cancer. In: Evaluation od Chemotherapeutic Agents (ed. C Macleod) 1949, 191-205. New York: Columbia University Press.

Katz S, Downs TD, Cash HR, Grotz RC: Progress in development of the index of ADL. Gerontologist 1970, 10:20-30.

Kennedy HJ. Quality of life in patients with an ileostomy. International Disability Studies 1988, 10:175-6.

Kettlewell MGW. Surgical Audit. Annals of the Royal College of Surgeons of England 1990, 72;5(suppl):70-71.

Klein R. On the Oregon trail: rationing health care. British Medical Journal 1991, 302:1-2.

Kober B, Kuchler T, Broelsch C, Kremer B, Henne-Bruns D. A psychological support concept and quality of life research in a liver transplantation program: an interdisciplinary multicenter study. Psychotherapy and Psychosomatics 1990, 54:117-31.

Koivukangas P, Koivukangas J. Role of quality of life in therapeutic strategies in brain tumours. Health Policy 1988, 10:241-257.

Koziol JA, Maxwell DA, Fukushima M, Colmeraurer MEM, Pilch YH. A distribution-free test fro tumour-growth curve analysis with application to an animal tumour immunotherapy experiment. Biometrics 1981, 37:383-90.

Langeluddecke P, Fulcher G, Baird D, Hughes C, Tennant C. A prospective evaluation of the psychosocial effects of coronary artery bypass surgery. Journal of Psychosomatic Research 1989, 33:37-45.

Lawrence KS, Fricker FJ. Pediatric heart transplantation: quality of life. Journal of Heart Transplantation 1987, 6:329-33.

Liang MH, Robb-Nicholson C. Health status and utility measurement viewed from the right brain: experience from the rheumatic diseases. Journal of Chronic Disease 1987, 40;6:579-583.

Lissitz RW, Green SB. Effect of the number of scale points on reliability: a Monte Carlo approach. Journal of Applied Psychology 1975, 60:10-13.

Little JM, Wong KP. Palliation of malignant jaundice. Australian & New Zealand Journal of Surgery 1991, 61:501-4.

Lloyd GER (Ed). Hippocratic Writings. London: Pelican 1978.

Luce BR, Weschler JM, Underwood C. The use of quality of life measures in the private sector. In: Quality of Life and Technology Assessment. Wachington DC: National Academy Press 1989.

Macaulay V, Smith I E. Advanced breast cancer. In: Randomised Trials in Cancer- A critical review by sites (Slevin M, Staquet M L, Eds) New York, Raven Press, 1986, 273-357.

McKenna P, Willison JR, Lowe D, Neil-Dwyer G. Recovery after subarachnoid haemorrhage. British Medical Journal 1989, 299:485-7.

McQuay HJ. Assessment of pain and effectiveness of treatment. In: Measuring the Outcomes of Medical Care. Eds. Hopkins A and Costain D. London: Royal College of Physicians of London 1990.

Maguire P Selby P. Assessing quality of life in cancer patients. British Journal of Cancer 1989, 60:437-440.

Maguire GP, Lee EG, Bevington DJ *et al.*. Psychiatric problems in the year after mastectomy. British Medical Journal 1978, i:963-965.

Mai FM, McKenzie FN, Kostuk WJ. Psychosocial adjustment and quality of life following heart transplantation. Canadian Journal of Psychiatry 1990, 35:223-7.

Mayou R, Bryant B. Quality of life after coronary artery surgery. Quarterly Journal of Medicine 1987, 62:239-48.

Mays NB, Petruckevitch A, Snowdon C. Patients' quality of life following extracorporeal shock-wave lithotripsy and percutaneous nephrolithotomy for renal calculi. International Journal of Technology Assessment In Health Care 1990, 6(4):633-642.

Medawar PB. Advice to a Young Scientist. New York: Harper and Row 1979, 69-77.

Meredith P. Audit and the quality of clinical care: patient satisfaction. Annals of the Royal College of Surgeons of England 1991, 73;3(suppl):51-54.

Milne AA. When we were very young. London: Methuen 1927.

Minitab Statistical Software, PA, USA.

Morris T, Greer HS, White P. Psychological and social adjustment to mastectomy (a 2-year follow-up study). Cancer 1977, 40:2281-7.

Morris J N, Sherwood S. Quality of Life of cancer patients at different stages of the disease trajectory. Journal of Chronic Disease 1987, 40:545-553.

Mosteller F, Gilbert JP, McPeek B. Reporting standards and research strategies for controlled trials: Agenda for the editor. Controlled Clinical Trials 1980, 1:37-58.

Neugebauer E, Troidl H, Wood-Dauphinee S, Bullinger M. Quality of life assessment in surgery: Editorial. Theoretical Surgery 1991, 6:121-122.

Neugebauer E, Troidl H, Wood-Dauphinee S, Eypasch E, Bullinger M. Quality of life assessment in surgery: results of the Meran Consensus Development Conference. Theoretical Surgery 1991, 6:123-137.

Neugebauer E, Troidl H, Spangenberger W, Dietrich A, Lefering R and The Cholecystecomy Study Group. Conventional *versus* laparoscopic cholecystectomy and the randomised controlled trial. British Journal of Surgery 1991, 78:150-154.

Noguchi Y, Baba M, Nakano S, Fukumoto T, Yoshinaka H, Aikou T, Shimazu H. Quality of life of patients after esophagectomy for esophageal cancer. Journal of the Japanese Surgical Society 1991, 92:281-7.

O'Young J, McPeek B. Quality of life variables in clinical trials. Journal of Chronic Disease 1987, 40:513-522.

Oresland T, Fasth S, Nordgren S, Hulten L. The clinical and functional outcome after restorative proctocolectomy. International Journal of Colorectal Disease 1989, 4:50.

Pemberton JH, Philips SF, Ready RL, Zinmeister AR, Beahrs OH. Quality of life after Brooke ileostomy and ileal pouch anal anastomosis. Annals of Surgery 1989, 209:620-618.

Peto R, Mike MC, Armitage P and 7 others. Design and analysis of randomised clinical trials requiring prolonged observation of each patient. Part 2. Analysis and examinations. British Journal of Cancer 1977, 35:1.

Pirsig R. Zen and the art of motorcycle maintenance. London: The Bodley Head, 1974.

Powles T J, Smith I E, Ford H T *et al.* Failure of chemotherapy to prolong survival in a group of patients with metastatic breast cancer. Lancet 1980, i:580-582.

Priestman T J, Baum M. Evaluation of quality of life in patients receiving treatment for advanced breast cancer. Lancet 1976:899-901.

Pruyn JFA, Maguire P, De Haes JCJM. Two methods of measuring some aspects of quality of life. In Quality of life methods of measurement and related areas. Proceedings of the second EORTC quality of life workshop. Copenhagen: European Organisation for the Research and Treatment of Cancer 1981, 52-55.

Raheja SK, McDonald PJ, Taylor I. Non-specific abdominal pain-an expensive mystery. Journal of the Royal Society of Medicine 1990 Jan, 83:10-11.

Rathmell AJ, Ash DV, Howes M, Nicholls J. Assessing quality of life in patients treated for advanced head and neck cancer. Clinical Oncology 1991, 3:10-6.

Remington M, Tyrer PJ, Newson-Smith J, Cicchetti DV. Comparative reliability of categorical and analogue rating scales in the assessment of psychiatric symptomatology. Psychological Medicine 1979, 9:765-770.

Richards MA, Hopwood P, Ramirez AJ, Twelves CJ, Fergusson J, Gregory WM, Swindell R, Scrivener W, Miller J, Howell A, Rubens RD. Doxorubicin in advanced breast cancer: Influence of schedule on response, survival and quality of life. European Journal of Cancer 1992 (in press).

Roder JD, Herschbach P, Ritter M, Kohn MM, Sellschopp A, Siewert JR. The quality of life following esophagectomy. The results of a psychosocial study of 80 cancer patients. Deutsche Medizinische Wochenschrift 1990, 115:570-4.

Rogers WJ, Coggin CJ, Gersh BJ, Fisher LD, Myers WO, Oberman A, Sheffield LT. Ten-year follow-up of quality of life in patients randomized to receive medical therapy or coronary artery bypass graft surgery. The Coronary Artery Surgery Study (CASS). Circulation 1990, 82:1647-58.

Rohrer MJ, Cutler BS, Wheeler HB. Long-term survival and quality of life following ruptured abdominal aortic aneurysm. Archives of Surgery 1988, 123:1213-7.

Sakamoto K, Nakano G, Hashimoto I, Nagamachi Y. Potential usefulness of a physical activity index in the assessment of quality of life of patients after gastrointestinal surgery. Acta Chirurgica Scandinavica 1989 155:305-12.

Schulte FJ, Matthes-Martin S, Zarbock G, Herrmann HD. Prognosis and quality of life following tumors in the pineal region in childhood. Klinische Padiatrie 1987, 199:429-39.

Shimada M, Tsunemoto M. Long-term (10 to 23 years) clinical evaluation of quality of life in patients following open heart repair of tetralogy of Fallot. Japanese Journal of Thoracic Surgery 1990, 43:634-9.

Sichel MP. Quality of life and gynecological cancers. European Journal of Gynaecological Oncology 1990, 11:485-8.

Siegel S. Nonparemetric Statistics for the Behavioural Sciences. Tokyo: McGraw-Hill Kogakusha 1956.

Singer PA, Tasch ES, Stocking C, Rubin S, Siegler M. Sex or survival: trade-offs between quality and quantity of life. Journal of Clinical Oncology 1991, 9:328-34.

Slevin M L, Plant H, Lynch D, Drinkwater J, Gregory W M. Who should measure quality of life, the doctor or the Patient? British Journal of Cancer 1988, 57:109-112.

Slevin M L, Stubbs L, Plant H, Wilson P, Gregory W M, Armes P J, Downer S M. Attitudes to chemotherapy: comparing views of patients with cancer with those of doctors, nurses and general public. British Medical Journal 1990, 300:1458-60.

Spielberger CD, Gorsuch RL, Lusherie RE. STAI Manual (Form Y). Palo Alto California, Consulting Psych. Press, 1983.

Spilker G, Stark GB. Quality of life considerations in plastic and reconstructive surgery. Theoretical Surgery 1991, 6:216-220.

Spina G, Santambrogio R, Opocher E, Galeotti F, Cucchiaro G, Strinna M, Pezzuoli G. Improved quality of life after distal splenorenal shunt. A prospective comparison with side-to-side portacaval shunt. Annals of Surgery 1988, 208:104-9.

Spitzer WO. State of Science 1986: Quality of life and functional status as target variables for research. Journal of Chronic disease 1987, 40;6:465-471.

Spitzer WO. Dobson AJ, Hall J. *et al.* Measuring the quality of life of cancer patients. A concise QL-Index for use by physicians. Journal of Chronic Disease 1981 34:585-597.

Stewart-Amidei C, Penckofer S. Quality of life following cerebral bypass surgery. Journal of Neuroscience Nursing 1988, 20:50-7.

Strauss RP. Psychosocial responses to oral and maxillofacial surgery for head and neck cancer. Journal of Oral and Maxillofacial Surgery 1989, 47:343-8.

Sugarbaker PH, Barofsky I, Rosenberg SA, Gianola FJ. Quality of life assessment of patients in extremity sarcoma clinical trial. Surgery 1982, 91:17-23.

Swinscow TDV. Statistics at Square One. London: British Medical Journal 1980.

Systat, Inc., USA.

Tarter RE, Switala J, Arria A, Plail J, Van Thiel D. Quality of life before and after orthotopic hepatic transplantation. Archives of Internal Medicine 1991 51(8):1521-6.

Teasdale G, Jennett B. Assessment of coma and impaired consciousness: A practical scale. Lancet 1974, 2:81-84.

Torii S, Suma K, Miyawaki F, Shiroma K, Inoue K. Postoperative long-term quality of life at present after total correction of tetralogy of Fallot. Japanese Journal of Thoracic Surgery 1990, 43:620-4.

Torrance GW. Health status index models: A unified mathematical view. Management Science 1976, 22:990-1001.

Torrance GW. Utility approach to measuring quality of life. Journal of Chronic Disease 1987, 40:593-600.

Troidl H. Quality of life:definition, conceptualisation and implications - a surgeon's view. Theoretical Surgery 1991, 6:138-142.

Troidl H, Kusche J, Vestweber K-H, Eypasch E, Koeppen L, Bouillon B. Quality of life: an important endpoint both in surgical practice and research. Journal of Chronic Disease 1987, 40;6:523-528.

Trojanowski T, Peszynski J, Turowski K, Markiewicz P, Goscinski I, Bielawski A, Bendarzewska B, Szymona J, Dabrowska A, Lopatkiewicz J et al. Quality of survival of patients with brain gliomas treated with postoperative CCNU and radiation therapy. Journal of Neurosurgery 1989 70:18-23.

Veroff J, Feld S, Gurin G. Dimensions of subjective adjustment. Journal of Abnormal Social Psychology 1962, 64:192.

Visick AH. A study of the failures after gastrectomy. Annals of the Royal College of Surgeons of England 1948, 3:266.

Wall Street Journal 1990, February 1992, b1-b2

Walsh RM, Aranha GV, Freeark RJ. Mortality and quality of life after total abdominal colectomy. Archives of Surgery 1990, 125:1564-6.

Ware J E. Standards for validating health measures: Definition and content. Journal of Chronic Disease 1987, 40;6:473-480.

Williams JI. Strategies for quality of life assessment- a methodologists view. Theoretical Surgery 1991, 6:152-157.

World Health Organisation. The constitution of the WHO. WHO Chronicle 1947, 1:29.

Yasutomi M, Shindo K, Matsuda T, Maruyama J, Morikawa E, Kubo R, Hida J. Rectal cancer therapy compatible with preservation of function. Gan To Kagaku Ryoho 1988, 15:2681-5.

Zigmond AS, Snaith RP. The Hospital Anxiety and Depression scale. Acta Psychiatrica Scandanavia 1983, 67:361-70.

Zubrod CG, Schneiderman M, Frei E, Brindley C, Gold GL, Shnider B, Oriedo R, Gorman J, Jones R, Jonsson V, Colsky J, Chalmers T, Fergusson B, Dederick M, Holland J, Selawry O, Regelson W, Lasagna L, Owens AH. Appraisal of methods for the study of chemotherapy of cancer in man: comparative therapeutic trial of nitrogen mustard and triethylene phosphoramide. Journal of Chronic Diseases 1960, 11:7-33.

ACKNOWLEDGEMENTS

I am indebted to the following people and organisations without whom the studies contributing to this thesis would not have possible:

Professor Michael Baum, Mr Stephen Ebbs and Dr Lesley Fallowfield, whose ideas and advice were freely available and freely used during every stage of the planning, research and writing up of this thesis.

The advanced breast cancer study at King's College Hospital:
Farmitalia Carlo Erba (UK) plc, who contributed substantial financial and technical assistance.
Dr Jane Dobbs, for her considerable expertise and advice on the running of the study and who allowed me to study her patients.
The Staff of the Department of Radiotherapy, whose support and assistance were gratefully received during the advanced breast cancer study.

The advanced breast cancer study at Guy's Hospital:
Dr Amanda Ramirez and Dr Michael Richards, who collected the Qualitator data and allowed me to study their patients.

The abdominal pain study at the William Harvey Hospital:
Dr Kevin Smith who initiated the abdominal pain study.
Dr Meena Agarawal, who helped to collect the QoL data on admission of the patients.
Mrs Margaret Harrison, who amassed the returned postal questionnaires.
Mr Tom Bates, who encouraged the studies, and allowed me to study his patients in the abdominal pain and advanced breast cancer studies.

The minor surgery study in King's College Hospital
Mr Hamid Khawaja, whose idea led to the minor surgery study and who with **Mr Nigel Heaton,** collected the preoperative QoL data.
The Surgeons of King's College Hospital, who allowed me to study their patients.

My wife, **Dr Allyson Pollock,** who supported the family during the months it took to process the data and write this thesis and then objectively criticised the end result.

APPENDICES

PUBLISHED PAPERS

APPENDIX 1

CRITERIA OF RESPONSE IN SOLID TUMOURS (UICC/WHO)

Complete Response (CR)

Complete disappearance of all clinically detectable malignant disease. No new lesions. Patients with bone metastases should have reversion to normal of all X-rays

Partial Response (PR)

50% or greater decrease in tumour size, without increase in size of any area of known malignant disease. No new lesions.

(i) Measurable bidimensional: 50% or greater decrease in tumour area (multiplication of two greatest diameters) or 50% decrease in sum of products of perpendicular diameters of multiple lesions.

(ii) Measurable unidimensional: 50% or greater decrease in linear tumour measurement.

(iii) Non-measurable: Appreciable change confirmed by photography or radiography, agreed upon *independently* and objectively.

No Change (NC)

No significant change in measurable lesions (<50% decrease or <25% increase in size) and no new lesions.

Progressive Disease (PD)

Significant increase (>25%) in size of some or all lesions present at start of therapy or appearance of new metastatic lesions known not to be present at start of therapy.

APPENDIX 2

TOXICITY CRITERIA

	Grade 0	Grade 1	Grade 2	Grade 3	Grade 4
Hb (g/100ml)	11 or more	9.5-10.9	8.0-9.4	6.5-7.9	less than 6.5
WBC ($10^3/mm^3$)	4 or less	3.0-3.9	2.0-2.9	1.0-1.9	less than 1
Platelets ($10^3/mm^3$)	100 or more	75-99	50-74	25-49	less than 25
Nausea/ Vomiting	none	nausea	transient vomiting	vomiting requiring therapy	intractable vomiting
Diarrhoea	none	transient (<2 days)	tolerable (<2 days)	Intolerable requiring therapy	Haemorr-hagic dehydration
Alopecia	none	minimal	moderate patchy	complete reversible	non-reversible

APPENDIX 3 The Nottingham Health Profile

Below are some problems people may have in their daily life. Look down the list and put a tick in the box under YES for any problem you have at the moment. Tick the box under NO for any problem you do not have. Please answer every question. If you are not sure whether to answer YES or NO, tick whatever answer you think is more true at the moment.

	YES	NO
I'm tired all the time		
I have pain at night		
Things are getting me down		
I have unbearable pain		
I take tablets to help me to sleep		
I've forgotten what it's like to enjoy myself		
I'm feeling on edge		
I find it painful to change position		
I feel lonely		
I can only walk about indoors		
I find it hard to bend		
Everything is an effort		
I'm waking up in the early hours of the morning		
I'm unable to walk at all		
I'm finding it hard to make contact with people		
The days seem to drag		
I have trouble getting up and down stairs/steps		
I find it hard to reach for things		

Remember, if you are not sure whether to answer YES or NO to a problem, tick whichever answer is more true at the moment

	YES	NO
I'm in pain when I walk		
I lose my temper easily these days		
I feel there is nobody I am close to		
I lie awake for most of the night		
I feel as if I'm losing control		
I'm in pain when I'm standing		
I find it hard to dress myself		
I soon run out of energy		
I find it hard to stand for long (eg at the kitchen sink, waiting for a bus)		
I'm in constant pain		
It takes me a long time to get to sleep		
I feel I am a burden to people		
Worry is keeping me awake at night		
I feel that life is not worth living		
I sleep badly at night		
I'm finding it hard to get on with people		
I need help to walk about outside (eg a walking aid or someone to support me)		
I'm in pain when going up and down stairs/steps		
I wake up feeling depressed		
I'm in pain when I'm sitting		

APPENDIX 4. LASA (Priestman and Baum)

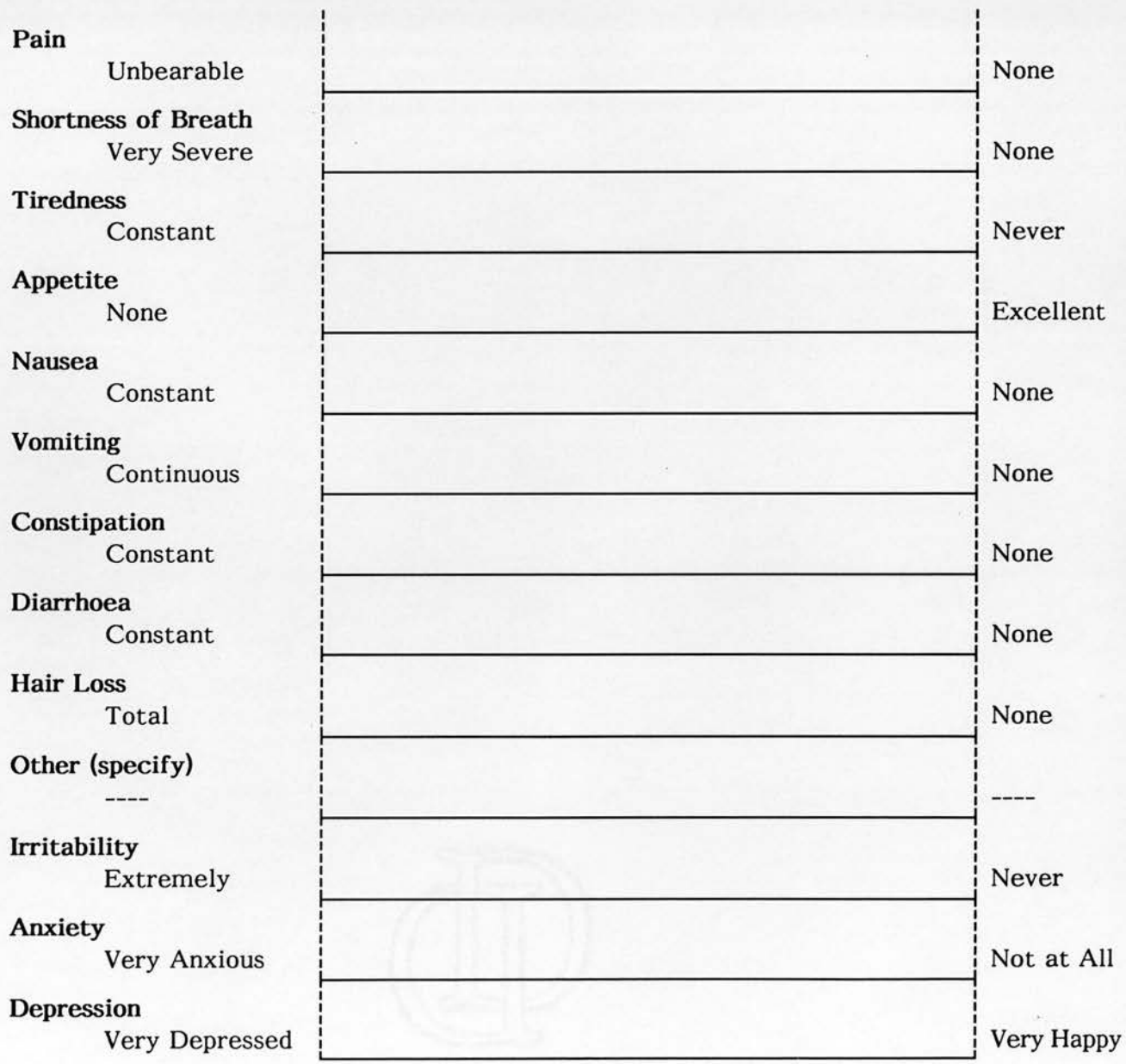
Instruction Sheet

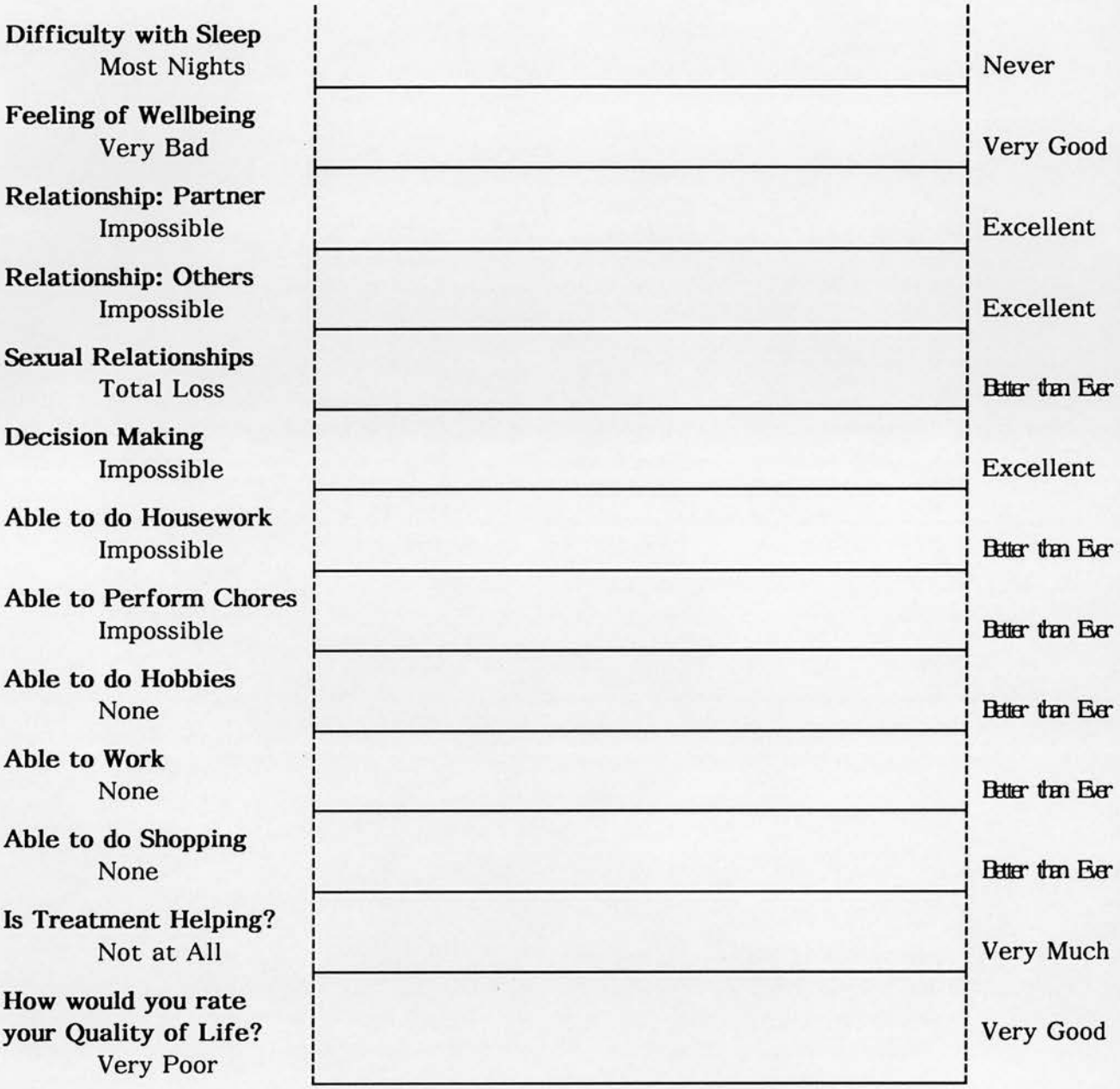
The doctors and nurses looking after you are concerned about your quality of life. We would like you to tell us how you have been feeling recently.

The first time you fill in this form, someone will explain to you how to mark it. If you need help, please ask the doctors or nurses.

There is a list of items that people with your disease undergoing treatment are often concerned about.

One of the spaces (marked "Other") is left free for you to tell us about any item that you consider important but that is not on the list.




Difficulty with Sleep
Most Nights
Never
Feeling of Wellbeing
Very Bad
Very Good
Relationship: Partner
Impossible
Excellent
Relationship: Others
Impossible
Excellent
Sexual Relationships
Total Loss
Better than Ever
Decision Making
Impossible
Excellent
Able to do Housework
Impossible
Better than Ever
Able to Perform Chores
Impossible
Better than Ever
Able to do Hobbies
None
Better than Ever
Able to Work
None
Better than Ever
Able to do Shopping
None
Better than Ever
Is Treatment Helping?
Not at All
Very Much
How would you rate
your Quality of Life?
Very Poor
Very Good

APPENDIX 5

The Qualitator

APPENDIX 6

Key to interpretation of box plots (SYSTAT, Inc)

The box plot is a means of representing non-normally distributed data diagramatically.

The median of the batch is marked by the centre horizontal line and splits the batch in two. The lower and upper boundaries of the enclosed box (the "hinges") split the remaining halves in half again.

Hspread is the term used to denote the interquartile range, i.e. the absolute value of the difference between the values of the two hinges.

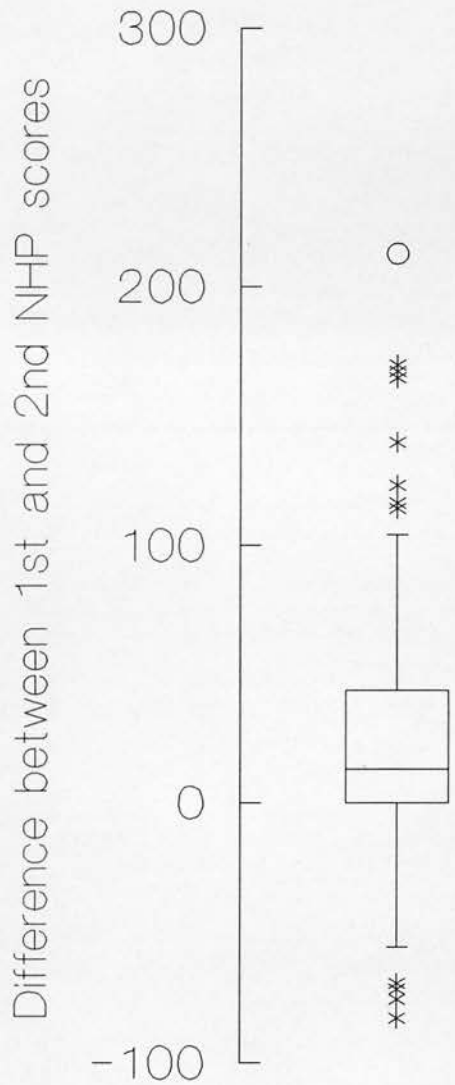
The *inner fences* are defined as follows:
lower fence = lower hinge - (1.5Hspread)
upper fence = upper hinge + (1.5Hspread)

The *outer fences* are defined as follows:
lower fence = lower hinge - (3Hspread)
upper fence = upper hinge + (3Hspread)

The *whiskers* show the range of values which fall within 1.5 Hspreads of the hinges. They do *not* necessarily extend to the inner fences.

Values outside the inner fences are plotted with asterisks. Values outside the outer fences are plotted with empty circles.

The adjacent hypothetical example represents the difference in absolute values of total NHP score between preoperative and postoperative measurements, for each individual. The raw data are as follows:



36.8, 58.56, 19.87, 30.67, 0, -75.81, -25.21, 1.61, 139.58, 21.04, -55.5, 36.57, 0, 0, -71.74, 167.04, 0, 3.75, 23.41, 169.51, -11.07, 0, 11.43, 0, -12.57, 212.75, 0, 116.26, 47.18, 49.67, 91.45, 12.91, 49.71, 24.13, 123.01, -43.47, 114.28, -82.98, 43.58, 9.76, 164.69, -11.22, 31.65, 30.4, -69.63, 27.47, 11.2, 31.38, 104.1, 0, 30.67, 0, -23.55, -13.52, 12.91, 0, 1.51

APPENDIX 8

GENERAL HEALTH QUESTIONNAIRE

Please read this carefully:

We should like to know if you have had any medical complaints, and how your health has been in general, *over the past few weeks.* Please answer ALL the questions on the following pages simply by underlining the answer which you think most nearly applies to you. Remember that we want to know about present and recent complaints, not those that you had in the past.

It is important that you try to answer ALL the questions.

Thank you very much for your cooperation.

HAVE YOU RECENTLY:

1	been able to concentrate on whatever you're doing?	Better than usual	Same as usual	Less than usual	Much less than usual
2	lost much sleep over worry?	Not at all	No more than usual	Rather more than usual	Much more than usual
3	been having restless, disturbed nights?	Not at all	No more than usual	Rather more than usual	Much more than usual
4	been managing to keep yourself busy and occupied?	More so than usual	Same as usual	Rather less than usual	Much less than usual
5	been getting out of the house as much as usual?	More so than usual	Same as usual	Less than usual	Much less than usual
6	been managing as well as most people would in your shoes?	Better than most	About the same	Rather less well	Much less well
7	felt on the whole you were doing things well?	Better than usual	About the same	Less well than usual	Much less well
8	been satisfied with the way you've carried out your task?	More satisfied	About same as usual	Less satisfied than usual	Much less satisfied
9	been able to feel warmth and affection for those near to you?	Better than usual	About same as usual	Less well than usual	Much less well
10	been finding it easy to get on with other people?	Better than usual	About same as usual	Less well than usual	Much less well
11	spent much time chatting with people?	More time than usual	About same as usual	Less time than usual	Much less than usual
12	felt that you are playing a useful part in things?	More so than usual	Same as usual	Less useful than usual	Much less useful
13	felt capable of making decisions about things?	More so than usual	Same as usual	Less so than usual	Much less capable

HAVE YOU RECENTLY:

14	felt constantly under strain?	Not at all	No more than usual	Rather more than usual	Much more than usual
15	felt you couldn't overcome your difficulties?	Not at all	No more than usual	Rather more than usual	Much more than usual
16	been finding life a struggle all the time?	Not at all	No more than usual	Rather more than usual	Much more than usual
17	been able to enjoy your normal day-to-day activities?	More so than usual	Same as usual	Less so than usual	Much less than usual
18	been taking things hard?	Not at all	No more than usual	Rather more than usual	Much more than usual
19	been getting scared or panicky for no good reason?	Not at all	No more than usual	Rather more than usual	Much more than usual
20	been able to face up to your problems?	More so than usual	Same as usual	Less able than usual	Much less able
21	found everything getting on top of you?	Not at all	No more than usual	Rather more than usual	Much more than usual
22	been feeling unhappy and depressed?	Not at all	No more than usual	Rather more than usual	Much more than usual
23	been losing confidence in yourself?	Not at all	No more than usual	Rather more than usual	Much more than usual
24	been thinking of yourself as a worthless person?	Not at all	No more than usual	Rather more than usual	Much more than usual
25	felt that life is entirele hopeless?	Not at all	No more than usual	Rather more than usual	Much more than usual
26	been feeling hopeful about your own future?	More so than usual	About same as usual	Less so than usual	Much less hopeful
27	been feeling reasonably happy, all things considered?	More so than usual	About same as usual	Less so than usual	Much less than usual
28	been feeling nervous and strung-up all the time?	Not at all	No more than usual	Rather more than usual	Much more than usual
29	felt that life isn't worth living?	Not at all	No more than usual	Rather more than usual	Much more than usual
30	found at times you couldn't do anything because your nerves were too bad?	Not at all	No more than usual	Rather more than usual	Much more than usual

APPENDIX 9

QUESTIONNAIRE 1

To be completed on Surgical Admission

Name

Number

Address

D o B

Tel No

Occupation

Psychological assessment of patients admitted with abdominal pain aged 16–40 years

Date

History

Duration of present episode of pain

Number of previous attacks

Total duration of history

Working diagnosis

(after assessment of initial investigations)

Emergency Surgery (please ring one)

Yes Review Unlikely No

Psychosocial Problems

Is there a problem? (please ring one) Yes No

If yes, is the nature of the problem obvious?

APPENDIX 10

Consultant: Mr Tom Bates
Research assistant: Mrs M Harrison
Tel. Ashford 633331, ext 326

Department of Surgery,
William Harvey Hospital,
Ashford,
Kent.
January 1990

<Address 1>
<Address 2>
<Address 3>
<Address 4>
<Address 5>

Dear <Dearname>,

You will remember that you were admitted to the William Harvey Hospital in <date> with abdominal pain. You may recall that when you were admitted, you filled some questionnaires.

Because we are interested in the after-effects of hospital admission for this type of abdominal problem, we would be very grateful if you would kindly answer the following questions, together with the enclosed questionnaires (on both sides) and return them in the envelopes provided. There quite a few questions but they are meant to be answered fairly quickly. If you have any queries, then please phone.

Your answers will be treated in the strictest confidence. Thank you for your co-operation.

Yours sincerely,

Simon Fraser, Honorary Registrar.

Tom Bates, Consultant

Questions

1) Do you still suffer from the pain which caused your first admission to hospital? YES/NO

2) Was your pain adequately dealt with at the time of hospital admission? YES/NO

3) Did you require further medical attention, hospital attendance or admission? YES/NO

If so, please state which doctor or which hospital and when:

4) Would you like another appointment to be seen at the William Harvey Hospital? YES/NO

APPENDIX 11 Spielberger anxiety trait questionnaire

e_____Date_____

CTIONS: A number of statements which people have used to ribe themselves are given below. Read each statement and then ken the appropriate number to the right of the statement to ate how you *generally* feel. There are no right or wrong ers. Do not spend too much time on any one statement but give nswer which seems to describe how you generally feel.

1 = Almost never
2 = Sometimes
3 = Often
4 = Almost always

eel pleasant	1	2	3	4
eel nervous and restless	1	2	3	4
eel satisfied with myself	1	2	3	4
vish I could be as happy as others seem to be	1	2	3	4
eel like a failure	1	2	3	4
eel rested	1	2	3	4
m "calm, cool, collected"	1	2	3	4
eel that difficulties are piling up so that I cannot overcome them	1	2	3	4
vorry too much over something that really doesn't matter	1	2	3	4
m happy	1	2	3	4
ave disturbing thoughts	1	2	3	4
ack self-confidence	1	2	3	4
eel secure	1	2	3	4
nake decisions easily	1	2	3	4
eel inadequate	1	2	3	4
m content	1	2	3	4
me unimportant thought runs through my mind and bothers me	1	2	3	4
ake disappointments so keenly that I can't put them out of my nd	1	2	3	4
m a steady person	1	2	3	4
et in a state of tension or turmoil as I think over my recent ncerns and interests.....	1	2	3	4

uestions 1,3,6,7,10,13,14,16,19 subtract from 5; add to rest)

PPENDIX 12

1. Has the operation been as successful as you had hoped? **YES NO**

If not, please say in what way.

2. Do you wish you had been offered the operation much sooner?

YES NO

For the following questions, please circle the response which most closely agrees with your present state.

3. Following your operation, how do you grade your health?

Worst possible Worse Same Better Best possible

4. How has the operation affected your quality of life?

Much worse Worse Same Better Much better

5. How is your efficiency at work affected?

Much worse Worse Same Better Much better

6. How is your social life affected?

Much worse Worse Same Better Much better

7. How is your sex life affected?

Much worse Worse Same Better Much better

8. How are your interests and hobbies affected?

Much worse Worse Same Better Much better

9. How have your holidays been affected?

Much worse Worse Same Better Much better

10. How much time did you have to take off work or normal activity after the operation?