

UTILITY MEASUREMENT OF HEALTH-RELATED QUALITY OF LIFE: PAST, PRESENT AND FUTURE

George W. Torrance, Centre for Health Economics and Policy Analysis, and Dept. of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Canada; Innovus Research Inc. and Health Utilities Inc.

Utility measurement has gained a prominent role in the field of health care. It is used as a measure of health-related quality of life; as a quantitative measure of process and outcome in clinical decision analysis; as an outcome measure in technology assessment and economic evaluation, particularly cost-effectiveness and cost-utility analyses; as a measure of population health; as a method of tracking the progress of patients through time; and as input into policy decision making. Its appeal lies in the fact that it measures preferences for process/outcomes/health status and summarizes these into a single cardinal number, and it does this based on a precise, well-defined, theoretical foundation. But exactly what does this number mean, and how can it legitimately be used?

Initially the methods were complex and time consuming (e.g., interviewer-administered Standard Gamble complete with complicated visual props). The history of the field is a continual movement towards increasingly simple, more user-friendly methods (e.g., time trade-off, visual analog scales, computer interactive measurement, telephone administration, questionnaire-based multi-attribute utility-scored systems like the Health Utilities Index, and most recently attempts to convert Short-Form 36 scores into utilities). But at what point do the costs outweigh the benefits? At what point do the approximations inherent in any simplification outweigh the advantages of the simplification itself?

I have been deeply involved in utility measurement for more than three decades. I will share my personal reflections based on this experience. What was (von Neumann-Morgenstern) utility theory designed to do? What have we been using it for in health applications? How have we used/misused it? What have we learned in three decades? Is it all worthwhile? Where do we go from here?

WHEN DOES A PROFILE BEAT A SINGLE NUMBER, OR ARE TWO ALWAYS BETTER THAN ONE?

Ron D. Hays, UCLA Department of Medicine, Los Angeles, CA, USA, RAND Health Program; Santa Monica, CA, USA

Profile measures such as the SF-36 provide multiple bits of information about health-related quality of life (HRQOL) whereas preference-based measures are designed to yield a single score that can be used for cost utility analyses. Previous research has consistently identified two underlying dimensions of HRQOL: physical and mental health. Although physical health and mental health are significantly correlated, there is more unique than shared variance between these two dimensions. This presentation examines various strategies for aggregating to a single number in HRQOL studies. The implications of different approaches and the strengths and weaknesses of one versus multiple scores are discussed.

HEALTH-RELATED QUALITY OF LIFE: WHERE ARE WE AND WHERE ARE WE GOING?

Ian F Tannock, Princess Margaret Hospital and University of Toronto, Toronto, Canada

My interest in Quality of Life (QL) stemmed from the reality that as a practicing medical oncologist, I was involved primarily in trying to provide palliation to patients with metastatic cancer, but the clinical trials on which I was basing decisions did not measure palliation. Rather they were using some measure such as tumor shrinkage, with the implicit and potentially false assumption that this was a valid surrogate for palliation. Palliation is a composite of improving the duration and or quality of survival, and in many fields of medicine, meaningful improvements in the duration of survival remain elusive. Research has provided tools that allow QL to be assessed in such trials, with reasonable psychometric properties, so we have at least arrived at the point where palliation can be measured and compared between different management options.

Unfortunately QL is rarely regarded as a primary endpoint of clinical trials, and it is seldom applied with the same rigor as the more traditional endpoints of duration of survival or tumor response. The methods that have been used to assess QL in clinical trials relating to cancer will be reviewed briefly. Few of them have adhered to the following principles:

1. It is essential to define a primary endpoint relating to QL that is relevant to the patient. This might be a global QL score or a measure of the dominant symptom such as pain. Changes in other aspects of QL can be regarded as exploratory and hypothesis-generating.
2. It is essential to define a primary hypothesis about the magnitude of change that is meaningful to a patient. This meaningful change needs to be based on evidence.
3. It is essential to recognize that QL is a property of an individual patient and that it will change at different rates and in different directions among a population of patients.

A common way of evaluating QL in a clinical trial is to measure a mean score for multiple QL endpoints as a baseline for the groups at randomization, and then at a fixed time such as 3 months later. The groups are then compared for changes in each endpoint and for differences at the 3-month time point. I do not know how to interpret such data. Mean QL has little meaning, enough comparisons are done that some will inevitably appear "significant", and much mental torture is spent on dealing with drop-outs and lack of compliance. It is like measuring tumor response by measuring the mean size of all metastatic tumors across 100 patients and determining whether this mean size decreases by a given amount (in those patients who remain on study!). Instead we need to record changes in the major endpoint of QL in the individual patient and determine if he or she has had a palliative response, and for how long it lasts; here we can learn by analogy from evaluation of tumor response. These principles will be illustrated by reference to the Canadian trial of simple chemotherapy versus prednisone for patients with symptomatic hormone-resistant prostate cancer.

As well as being an important endpoint in clinical trials measures of QL are now being used to define psychosocial aspects of disease and its treatment that may be as important as the physical effects. There is evidence that patients receiving adjuvant chemotherapy for breast cancer now perceive side effects such as fatigue and the symptoms due to an accelerated menopause induced by treatment as more disabling than symptoms such as nausea or vomiting. New medications have done much to control the latter but not the former. Also there is evidence from our studies and those of others that substantial cognitive changes can occur with chemotherapy and that they are not rapidly reversible. The results of our pilot study and the use of questionnaires and psychological tests, which are being used to evaluate these effects will be described.

There are many challenges in the further application of QL research to clinical practice. As a clinician involved in trying to improve the care of patients, and in developing new treatments through clinical trials, perhaps those most important to me are:

1. Convincing doctors whose major role is palliation that they need to measure palliation as a primary endpoint in clinical trials that seek to improve patient benefit, and making it easy for them to do so.
2. Extending assessment of QL to routine clinical practice so that it can be incorporated into every-day clinical decision making.

THE ADDED VALUE OF HEALTH-RELATED QUALITY OF LIFE EVIDENCE: IS SAFETY AND EFFICACY ENOUGH?

Dennis A. Revicki, Center for Health Outcomes Research, MEDTAP International, Bethesda, MD, US.

Health-related quality of life (HRQL) outcomes are frequently incorporated into clinical trials comparing medication and surgical interventions. In this context, the HRQL endpoints provide the patients' perspective on the impact of both the treatment and the underlying disease on their functioning and well-being. It is thought that HRQL is the ultimate endpoint for evaluating medical treatments (assuming survival); therefore HRQL outcomes extend measures of effectiveness beyond safety and efficacy. Skeptics counter that HRQL measures add little to the existing safety and clinical efficacy data in comparing medical treatments for regulatory or clinical decision making. HRQL outcomes, in most clinical trials, provide additional support for the effectiveness of treatments, taking into account the adverse effects of treatments. The added value of HRQL outcomes is demonstrated in several ways: (1) HRQL outcomes provide added data on effectiveness not captured by clinical efficacy endpoints; (2) HRQL outcomes document the trade-off, from the patients' perspective, of the risks and benefits of treatments; (3) HRQL outcomes provide more meaningful endpoints to physicians and patients for evaluating many chronic diseases; (4) HRQL outcomes can be used to differentiate treatments with comparable clinical endpoints or survival; and (5) HRQL outcomes best capture treatment effectiveness for chronic diseases with no objective clinical endpoints. These concepts are illustrated by examples from clinical trials in oncology, HIV disease, psychiatry, gastrointestinal disease, and respiratory disease. It is important for researchers to document and demonstrate the value of HRQL endpoints for evaluating medical treatments. Future research must focus on providing a stronger theoretical framework and rationale for measuring HRQL within the context of the medical care setting and in providing psychometric evidence on HRQL instruments for different patient populations and applications. Patients, their families, physicians, and regulatory agencies need HRQL information to make decisions about the benefit and risks of new therapies. The relevancy of HRQL data for regulatory and clinical decision making depends on the strength of the research evidence of the added value of HRQL outcomes.

METHODS FOR ASSESSING CLINICAL SIGNIFICANCE IN QUALITY OF LIFE MEASUREMENT.

Jeff A. Sloan, Quality of Life Research, Mayo Clinic, Rochester, MN, USA

The integration of quality of life (QOL) assessment is increasingly a routine part of oncology clinical research. The challenge for the QOL research community that remains unmet is to provide clinical researchers with cogent, coordinated, and accessible descriptions of how to tell when a clinically significant change in QOL has been achieved.

This presentation will summarize the results of a meeting held October 6-7, 2000 at the Mayo Clinic which attempted to take the first step in meeting this challenge. Thirty international experts on QOL assessment gathered for a two-day "think tank" with an objective to produce working drafts of six papers targeted at assessing clinical significance in QOL measurement. These papers, which will be published in a major journal and/or as a separate monograph, deal with the following topics:

- 1) Methods used to date for clinical significance
- 2) Group versus individual clinical significance differences
- 3) Single item versus summated scale scores
- 4) Patient versus clinician versus population perspectives of clinical significance
- 5) Assessing changes over time
- 6) Incorporating clinical significance into clinical practice/industry

Synopses of each paper will be presented and discussed. The goal of this collection of papers is to provide those who might wish to assess QOL with a set of practical implementation guidelines balancing the theoretical framework with practical limitations. The meeting was focused specifically on oncology patient QOL assessment, but the results are at least partially generalizable to other populations. This work will continue through an open registration meeting to be held October 18-19, 2001 at the Mayo Clinic so that refinements can be made, and further input/discussion can be sought.

Invited Speakers - 2

OVER-USE AND ABUSE OF PSYCHOMETRICS – IS THERE A ROLE FOR CLINIMETRICS?

Peter Fayers, Institute of Applied Health Sciences, University of Aberdeen, UK and Unit of Applied Clinical Research, Norwegian University of Science & Technology (NTNU), Trondheim, Norway

Quantitative psychometric techniques mainly originated from 1900 onwards, and were initially developed for applications such as intelligence testing, educational examinations or personality testing. This has led to extensive literature on the development and validation of multi-item measurement scales. However, much psychometric theory is based on the assumption that individual items on the questionnaires form "parallel tests", so that simple weighted or unweighted summation is an appropriate method of aggregation. More recent psychometric work either continues to promulgate these traditional methods or places emphasis on modern techniques centred upon item response theory (IRT). All too frequently these methods – including factor analysis, Cronbach's α and IRT – are employed uncritically to HRQL scales.

However, the underlying principles of many clinical measuring instruments are crucially different from those of psychological tests, making such psychometric approaches inappropriate; this has led to the use of "clinimetrics" in some branches of medicine. We show that HRQL instruments also frequently contain variables that do not behave like traditional psychometric items, and that their fundamentally different relationship to patients' HRQL levels results in misleading analyses. We illustrate the errors of interpretation that arise if psychometric methods are used (abused) on items that have a causal component, and suggest that the concept of "causal items" underpins the ideas of clinimetrics. Thus we conclude there is a need to consider both clinimetric and psychometric approaches when developing or evaluating HRQL scales, especially when instruments contain disease-related symptoms or treatment side-effects.

EVALUATION OF HRQL IN SPECIAL POPULATIONS: CHILDREN

David Feeny, Institute of Health Economics and Univ. of Alberta, Edmonton, AB, Canada

Six key issues are listed, stated as assertions. Panelists need not agree with the assertions!

1. Measures are Currently Available. There are reliable, valid, and responsive measures of health-related quality of life (HRQL) suitable for administration (self assessment) to children and adolescents. The existing set of measures include specific (for instance there are a number of pediatric oncology and pediatric asthma measures), generic profile measures (for instance, the Child Health Questionnaire), and generic preference-based measures (for instance, the Health Utilities Index Mark 2 and Mark 3 systems). Although clearly the number of measures is quite limited, especially when compared to the availability of measures of HRQL in adult populations, the lack of availability of measures is not as serious a barrier as it was a decade ago.
2. Responses based on Self Assessment by Children are not, in general, interchangeable with Proxy Assessments by Parents. A reasonable body of evidence has been accumulated indicating that responses from children do not agree completely with responses from parents (or clinicians). The degree of disagreement varies by dimension (domain) of health status and may vary with the age of the child and situation. There is some evidence that in situations that are highly stressful to parents (for instance, when the child is undergoing bone marrow transplantation), proxy assessments offered by parents may be affected by their own emotional health. Even though there is evidence on agreement between a self report by children and proxy assessments by parents, accumulating additional evidence on when and for what dimensions we might expect to observe agreement and when and for what dimensions we might expect to observe disagreement is important. It is also important to examine agreement between parents and clinicians and patients and clinicians.
3. Challenge to Accumulate Evidence/Identify Gaps. We need to accumulate evidence on reliability, validity, and responsiveness in new settings. This process will also help to identify gaps in the set of measures, especially diseases or problems for which specific measures have not yet been developed and are needed to address important questions.
4. Challenge to Develop Measures for Preschool Children. Most of the existing evidence on the ability of children to complete questionnaires or act as respondents is for school-aged children. A key challenge is to develop innovative approaches for preschool children and assess the reliability, validity, and responsiveness of their responses.
5. Challenge to Develop Skill in the Interpretation of HRQL Scores. Clinicians have had years of experience with standard clinical measures (for instance, white blood counts) and have learned how to attach meaning to these measures. There is much less familiarity with HRQL scores. Are there strategies to speed learning and assist users in interpreting HRQL scores? Are there strategies to identify clinically (or policy) important differences?
6. Challenge to Demonstrate the Usefulness of Assessing HRQL in Children and Adolescents. Do we have any evidence that the use of HRQL measures matters? Do we have evidence that results have affected clinical policy, the research agenda, or resource allocation? Do we have evidence that these measures have played an important role in managing patients and families? How would we accumulate such evidence?

Summary: The session will examine available measures and the use of proxy respondents, identify gaps in the availability of measures, and discuss key challenges including self report by preschool children, the interpretation of scores, and evidence on the usefulness of assessing health-related quality of life in children and adolescents.