

Measuring Health-State Utility Values in Clinical Trials: Can We Do Better?

Introduction

Sorrel Wolowacz, PhD, Head, European Health Economics, RTI Health Solutions, Manchester, UK



KEY POINTS

- Assessment of whether and how to measure utility in a trial, and whether alternative or supplementary studies are appropriate.
- Selection of an appropriate measure, optimal timing of assessments, and patient follow-up.
- Design of analyses to provide health state utility values estimates for economic models, and to utilise the power of patient-level utility data.

The quality of health-state utility data used in economic evaluations is critical in health technology assessment (HTA) processes that govern patient access to, and the reimbursement of, new medical technologies. Since HTA agencies expressed a preference for utility data collected from patients [1,2], it has become increasingly common for utility data (most commonly using the EuroQol 5 Dimensions [EQ-5D]) to be collected alongside clinical trials. However, such data have often provided suboptimal utility estimates for economic models for a variety of reasons.

First, the design of clinical trials is driven primarily by regulatory requirements and can present challenges to the optimal collection of utility data for economic models. For example, in many cases, assessments are available for only a small proportion of the health states relevant to the economic model. In cancer trials, for example, patients are often assessed regularly prior to disease progression, but follow-up after progression may be much more limited. Therefore, these trials often provide little opportunity to measure utility in the later stages of disease progression or during terminal illness, which are important parameters in cancer models.

Second, there may be issues related to the generalizability of the trial or study population to the population of interest in the economic model being used, i.e., patients in routine clinical practice. HTA agencies (e.g. NICE [3,4]) has rejected utility weights measured in trials, stating that the values were similar to estimates for the general population of the same age and were therefore implausible for the patient population in question (cancer patients in this case). It has been postulated that trial exclusion criteria (which exclude patients, for example, in certain age groups or with comorbidities or abnormal organ function) may select for a younger, fitter patient group than the population of patients eligible for the new treatment in routine clinical practice and therefore, may generate higher utility estimates.

Third, the timing of utility assessments is usually designed to coincide with other scheduled outcome assessments which tend to occur at regular

intervals. This is often inappropriate for measuring utility for use in economic models. For example, if a utility estimate is needed for events (e.g., a fracture or pulmonary embolism); there may be few or no scheduled assessments during the period that the event affected quality of life. Another common timing problem occurs in cancer trials, where assessments are often made on the first day of each chemotherapy cycle before the chemotherapy is administered. With such a design, the impact of chemotherapy-related toxicity is unlikely to be captured because patients experiencing severe toxicity may discontinue therapy (and will not be assessed after the administration). Nor is the impact of toxicity likely to be captured for patients who do continue therapy because the adverse effects generally will have resolved before the next chemotherapy cycle is received (the next cycle of therapy is commonly delayed if patients are still experiencing toxicity).

The fourth reason for suboptimal utility estimates is that the analyses performed may not be appropriate or optimal to provide estimates for economic

models. Commonly, the mean utility or the mean change from baseline is reported at a series of time points. These data are often inappropriate for use in economic models, which usually require utility estimates for model health states or events.

The quality of health-state utility data used in economic evaluations is critical in health technology assessment (HTA) processes that govern patient access to, and the reimbursement of, new medical technologies.

Finally, in some cases the EQ-5D may not be the optimal utility instrument for the condition of interest (e.g., for reasons of validity or responsiveness) and is included without sufficient analysis of these considerations.

Key considerations for collection of utility data within a planned clinical trial are summarized in Table 1.

Table 1: Considerations for Collection of Health-State Utility Data in a Clinical Trial

Considerations
<ul style="list-style-type: none"> • Is the planned trial appropriate for utility measurement? <ul style="list-style-type: none"> – Is it feasible to observe key model health states or events in the trial? – Can a sufficient number of assessments be included? Could discontinuation of follow-up introduce bias? – Is the trial population representative of the population in routine clinical practice? Could excluded patients who would be eligible for treatment otherwise be followed-up with for utility? – Would an observational study (or a combination of the trial and an observational study) be more appropriate? • Is the EQ-5D is the most appropriate instrument? <ul style="list-style-type: none"> – Is the EQ-5D valid and responsive in this indication? • What is the optimal design of utility assessments in the trial? <ul style="list-style-type: none"> – Number and timing of assessments – Patient follow-up (e.g., after progression, excluded patients) • What analyses should be specified? <ul style="list-style-type: none"> – Align with model health states or events – Optimize sensitivity (e.g., explore the association between change from baseline and continuous, rather than categorical, clinical variables) – Capture any correlations between better and worse health states for probabilistic sensitivity analysis (e.g., using regression modelling)
EQ-5D = EuroQol 5 Dimensions.

Issues with Utility Data: Industry Perspective

Jennifer Petrillo, PhD, Associate Director, Global Market Access, Biogen Idec, Cambridge MA USA



INTRODUCTION

While typical clinical outcome assessments for clinical trials have been thoroughly researched for reliability and validity to a particular condition and detailed analysis plans have been derived for the assessments' evaluations, preference-based measures (PBMs) tend to be less so. Utilities are used to reflect the incremental cost-effectiveness ratio (= cost ÷ effectiveness) of an intervention. The more cost-effective the treatment, the greater the chance it will be reimbursed by the local HTA authorities. Unfortunately, many researchers tend to include a PBM as a tag-along endpoint without much consideration for the measurement's concepts, its design implications, or the analysis plan.

EXAMPLE FROM RESEARCH

In one recent phase 2 intervention study of patients with Parkinson's disease with levodopa-induced dyskinesia, the EQ-5D Dimensions were included at baseline and week 9 to coincide with the clinical endpoints [5]. While the primary endpoint, the modified Abnormal Involuntary Movement Scale

(mAIMs), demonstrated clinically meaningful change (mean, 4.8; standard deviation, 5.2), the EQ-5D did not (mean, 0.024; standard deviation, 0.255). Additionally, there was little to no correlation of change in mAIMs score with the change in the EQ-5D. The lack of concept correlation, together with little utility change, suggested that the concepts measured in the EQ-5D might not be consistent with relevant concepts for this population. Furthermore, change from baseline tended to overlook any minor improvements experienced in the duration of the study.

CONCLUSIONS

Several considerations must be kept in mind for pharmaceutical researchers during collection and analyses of utilities.

First, timing of the utility assessment is crucial. Severity ranges, comorbidities, and acute events all could be missed if the timing does not coincide with important patient events. There is also a need to ensure the study design is focused on clinically

meaningful time points. This may be short term or longer term focused, depending on the timing of events and the ability to measure desired outcomes.

Next, treatment interventions geared toward symptom and/or biomarker reduction may not allow improvement in functioning and quality of life within a single trial. Components of the treatment response (and measurement timing) may not correlate with changes on PBMs, which could lead to recommendations of another utility measure being included.

Finally the trial population may not be representative of patients found in routine clinical practice. Specifically, the trial population may be healthier and have a narrower range of severity. This could limit a trial's findings and the generalizability of the utilities. >>

- ADVERTISEMENT -

More Than Data. **Answers.**

TRUVEN
HEALTH ANALYTICS™

**25 Years of Big Data and Innovation:
New Oncology EMR Linked to Claims**

Demonstrate the value of oncology treatments using our new data asset — linked claims data with detailed clinical data. The robust Truven Health MarketScan® Oncology EMR-Claims Dataset enables researchers to conduct **high-quality, patient-level analyses** linking total healthcare utilization and costs from all providers with associated clinical details from EMR records. It's the critical link to your next research project.

Email us at lifesciences@truvenhealth.com to learn more.

Stephen S. Johnston
Director, Outcomes Research

© 2014 Truven Health Analytics Inc. All rights reserved.

Issues with Utility Data: Patient Perspective

Lynda Doward, MRes, European Head, Patient-Reported Outcomes, RTI Health Solutions, Manchester, UK



The collection of optimal utility data in the context of clinical trials presents a number of challenges to pharmaceutical researchers. Data are typically collected via standardized patient-reported outcome questionnaires designed to capture health-related quality of life (HRQOL), thus providing the HRQOL utility for inclusion in quality-adjusted life-year calculations. However, the quality of utility data generated via clinical trials rests on three key issues:

- Representativeness of the trial population
- Choice of instrument
- Timing of assessments

From the researcher perspective, getting it “right” in these three areas may affect reimbursement decisions by HTA authorities. From the patient perspective, getting it “wrong” presents ethical issues surrounding the collection of suboptimal data; further, if those data contribute to an HTA agency deciding not to reimburse a drug, it is ultimately the patient who suffers. The importance of these three areas is described in more detail in the following sections.

REPRESENTATIVENESS OF CLINICAL TRIALS

The generalizability of the clinical trial population to the broader population seen in clinical practice will have a bearing on the plausibility of utility weights derived from the trial. Patients enrolled in clinical trials may not reflect the general patient population in a number of areas, notably age; diagnostic subgroup; disease severity; presence of co-morbidities; and additional sociodemographic factors, including socioeconomic status, geographic location, race/culture, or language. A review of trials funded by the National Cancer Institute found that only 2.5% of cancer patients in the United States are enrolled. Among the patients underrepresented

were adults aged 65 and older, adolescents, African-American men, Latinos/Hispanics, Asian and Pacific Islanders, American Indians/Alaska Natives, individuals residing in rural areas, and individuals of low socioeconomic status [6,7]. A recent study by Chiò and colleagues [8] found that patients enrolled in clinical trials for amyotrophic lateral sclerosis were not representative of the broader patient population. Specifically, they tended to be younger, to have longer diagnostic delays, were more likely to have spinal involvement, and were more likely to be male.

The lack of diversity in randomized study populations may reduce the opportunity to discover health effects that may be particularly relevant to underrepresented populations. In terms of utilities, the values generated may be skewed in favor of a fitter; younger; and less socially, economically, and racially diverse population, leading ultimately to potential flaws in cost-effectiveness arguments generated on the basis of these values.

CHOICE OF INSTRUMENT

Selecting the most appropriate instrument for collecting utility data in clinical trials requires a judgment between the preferences of the target audience (i.e., the HTA authorities) and the relevance, suitability, and appropriateness of the instrument for the therapeutic area under study. Instruments may include preference-based measures (PBMs) or HRQOL measures that can be used to generate preferences. Understandably, HTA agencies prefer consistency in utility measures, because it allows comparison across appraisals; thus, there is a preference for utility data based on generic PBMs. However, generic measures, by definition, are designed to be broadly acceptable to a wide range of conditions. Thus, the content validity of the instrument for the therapeutic area under study may need to be proven. Many widely

used PBMs are most relevant for therapeutic areas where physical limitations are a key feature of the disease. They may be less valuable for therapeutic areas where social, relationship, or emotional issues are important. In such cases, an HRQOL measure that can be used to generate preferences may be a better choice. Ultimately, the use of an instrument with poor content validity will have a detrimental effect on the quality of the data produced.

TIMING OF ASSESSMENTS

The collection of utility data in clinical trials is usually timed to coincide with scheduled assessments associated with primary and secondary endpoints or is scheduled at regular intervals for the study duration. This timing poses a particular challenge for the collection of utility data for chronic, episodic conditions. For example, a review of clinical trials in relapsing-remitting multiple sclerosis, where utility data were collected using a PBM, revealed that most trials are conducted over a 12-month or 18-month period, with only two scheduled assessments. However, as the rate of flare-ups for relapsing-remitting multiple sclerosis is reported to be 2.1 ± 1.2 per year, there is limited likelihood of collecting utility data associated with a disease flare-up in a trial of this design [9]. Thus, the design of trials may not allow for the assessment of utility (or disutility) associated with events of interest.

The collection of optimum utility data in the context of clinical trials requires careful consideration of key issues related to the representativeness of the trial population, instrument selection, and assessment timing. These factors are key to ensuring that the utility values used in economic modeling are the most realistic and the most representative for the therapeutic area under study.

Analyzing Utility Data Collected Alongside Clinical Trials

Andrew Briggs, DPhil, MSc, Professor of Health Economics, University of Glasgow, Glasgow, UK



Increasingly, clinical trials are being used as vehicles for collecting data that are relevant for reimbursement submissions. However, when these data are analyzed, the approach to the analysis often is strongly influenced by the between-arm comparisons that typically are used for regulatory submissions. This is particularly true of preference-based utility data used to estimate quality adjusted life-years. These measures are a form of patient-reported outcome, yet their primary use is to estimate quality-adjusted life-years (QALY), not to make label claims. If reimbursement submissions are to be made using an economic model, then it

is appropriate that the utility data in a clinical trial be analyzed to inform that model, rather than be analyzed by treatment arm.

Consider the following example of utility data collected as part of the UK Prospective Diabetes Study (UKPDS). When the study began in 1977, economic evaluation was barely used and utility measures had not been invented. However, toward the end of this study, there was much interest in performing economic evaluation alongside clinical trials. Therefore, in 1996, as the study was drawing to a close, the EuroQol 5 Dimensions (EQ-5D) was

administered cross-sectionally to all 3,667 patients remaining in the study. When the data were analyzed by treatment arm, no difference in EQ-5D utility could be detected, despite the fact that this landmark trial demonstrated that intensive blood glucose control significantly reduced the long-term sequela of diabetes. A subsequent analysis regressed EQ-5D against the long-term clinical outcomes that formed the primary endpoints of the UKPDS study. As expected, the results showed very clearly that these long-term outcomes had a significant detrimental effect on health-related >>

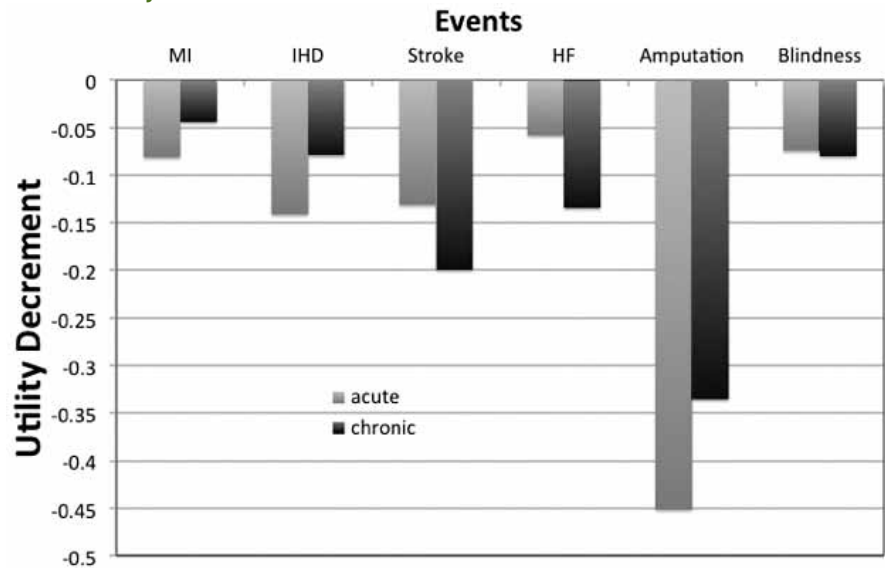
quality-of-life (HRQOL) as measured by the EQ-5D. Figure 1 shows the estimated utility decrements associated with the different clinical events and differentiates the “acute” impact (an event occurring in the year of EQ-5D measurement) from the “chronic” impact (an event occurring more than 1 year before EQ-5D measurement). [10]

The UKPDS example is compelling. The “story” of the economic model—that treatment impacts the long-term risk of events and those events impact HRQOL—is confirmed by the analysis of the data. Yet, with all the noise in the observed data and given the relative insensitivity of the EQ-5D measure, a between-arms difference in HRQOL could not be detected at conventional significance levels. Of course, there are problems with using cross-sectional data in this way. One issue is that patients who experienced an event could have had a lower starting utility than patients who did not experience an event. However, and in contrast to the UKPDS, most studies collect utility at randomization and at intervals throughout the study. With longitudinal data, more sophisticated analyses are possible that control for baseline utility. The lesson is that the approach to analyzing data for reimbursement outcomes should reflect the needs of the economic analysis, not be constrained by the traditional regulatory approach to analyzing data.

REFERENCES

- [1] National Institute for Health and Care Excellence (NICE). Guide to the methods of technology appraisal. June 2008. Available at: <http://www.nice.org.uk/about/nice/howwe/work/devnicetech/GuideToMethodsTechnologyAppraisal2008.jsp?domedia=1&mid=B52851A3-19B9-E0B5-D48284D172BD8459>. [Accessed February 26, 2014].
- [2] Pharmaceutical Benefits Advisory Committee (PBAC). Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee. 2013. Available at: <http://www.pbac.pbs.gov.au/information/table-of-contents.html>. [Accessed March 6, 2014].
- [3] National Institute for Health and Care Excellence (NICE). Technology Appraisal No. 118: bevacizumab and cetuximab for the treatment of metastatic colorectal cancer. January 2007. Available at: <http://guidance.nice.org.uk/TA118>. [Accessed January 14, 2014].
- [4] National Institute for Health and Care Excellence (NICE). Technology Appraisal No. 150: cetuximab for the treatment of metastatic colorectal cancer following failure of oxaliplatin-containing chemotherapy. June 2008. Available at: <http://guidance.nice.org.uk/TA150>. [Accessed January 14, 2014].
- [5] Petrillo J, Pendergraft T, Sacco P, et al. Sensitivity of EQ-5D to clinical change in patients with Parkinson’s disease L-dopa induced dyskinesia. Abstract presented at the 18th Annual International Meeting of the International Society for Pharmacoeconomics and Outcomes Research; May 18-22, 2013. New Orleans, LA.
- [6] Sateren WB, Trimble EL, Abrams J, et al. How sociodemographics, presence of oncology specialists, and hospital cancer programs affect accrual to cancer treatment trials. *J Clin Oncol* 2002;20:2109-17.
- [7] Ford JG, Howerton MW, Lai GY, et al. Barriers to recruiting underrepresented populations to cancer clinical trials: A systematic review. *Cancer* 2008;112:228-42.
- [8] Chiò A, Canosa A, Gallo S, et al; PARALS group. ALS clinical trials: Do enrolled patients accurately represent the ALS population? *Neurology* 2011;77:1432-7.
- [9] Fernández-Megía MJ, Casanova B, Magraner MJ, et al. Assessment of the effectiveness and safety of natalizumab for treating relapsing-remitting multiple sclerosis. *Fam Hosp* 2011;35:75-9.
- [10] Clark P, Gray A, Holman L. Estimating utility values for health states of Type 2 diabetic patients using the EQ-5D (UKPDS 62). *Med Decis Making* 2002;22:340-9. **IC**

Figure 1: Utility Decrements Associated With Clinical Endpoints in the UK Prospective Diabetes Study. [10]



HF = heart failure; IHD = ischemic heart disease; MI = myocardial infarction; UK = United Kingdom



ISPOR Good Practices for Measurement of Health State Utility Values for Economic Models in Clinical Studies Task Force

An ISPOR Task Force on Measurement of Health State Utility Values for Economic Models in Clinical Studies was recently formed. The goal of this task force is to develop good practices on utility measurement within the clinical trial program for a new intervention.

Task Force Chairs: Andrew Briggs, DPhil, MSc, Professor of Health Economics, University of Glasgow, Glasgow, UK and Sorrel E. Wolowacz, PhD, Head European Health Economics, RTI Health Solutions, Manchester, UK. **Task Force members:** Lynda Doward, MRes, RTI Health Solutions, Manchester, UK, Jennifer Petrillo, PhD, Novartis AG, Basel, Switzerland; Andrew Lloyd, PhD, ICON, plc., England; Vasily Belozeroff, PhD, Amgen Inc., USA; Philip Clarke, PhD, University of Melbourne, Australia; and Ron Goeree, MA, McMaster University, Canada.

Information on the Task Force is available at the ISPOR website www.ispor.org under the Task Force tab or <http://www.ispor.org/taskForces/TFindex.asp>. If you are interested in joining the reviewer group for this ISPOR Task Force, please complete the Special Interest Group/Task Forces Form (available at the bottom of the task force index page).