FEATURING ARTICLES

ISSPOR Task Force Report


Sorrel E. Wolowacz, PhD1, Andrew Briggs, DPhil2, Vasily Belozeroff, PhD3, Philip Clarke, PhD4, Lynda Doward, MRes1, Ron Goeree, MA5,6, Andrew Lloyd, DPhil7, Richard Norman, PhD8

1RTI Health Solutions, Manchester, UK; 2Health Economics and Health Technology Assessment Research Group, Institute of Health & Wellbeing, Glasgow University, Glasgow, UK; 3Global Health Economics, Amgen, Inc, Thousand Oaks, CA, USA; 4Centre for Health Policy, School of Population and Global Health, University of Melbourne, Melbourne, Australia; 5Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada; 6Goeree Consulting Limited, Hamilton, Ontario, Canada; 7Bladon Associates Ltd., Oxford, UK; 8School of Public Health, Curtin University, Curtin, West Australia, Australia

ABSTRACT

Cost-utility models are increasingly used in many countries to establish whether the cost of a new intervention can be justified in terms of health benefits. Health-state utility (HSU) estimates (the preference for a given state of health on a cardinal scale where 0 represents dead and 1 represents full health) are typically among the most important and uncertain data inputs in cost-utility models. Clinical trials represent an important opportunity for the collection of health-utility data. However, trials designed primarily to evaluate efficacy and safety often present challenges to the optimal collection of HSU estimates for economic models. Careful planning is needed to determine which of the HSU estimates may be measured in planned trials, to establish the optimal methodology, and to plan any additional studies needed. This report aimed to provide a framework for researchers to plan the collection of health-utility data in clinical studies to provide high-quality HSU estimates for economic modeling. Recommendations are made for early planning of health-utility data collection within a research and development program; design of health-utility data collection during protocol development for a planned clinical trial; design of prospective and cross-sectional observational studies and alternative study types; and statistical analyses and reporting.

Keywords: economic model, health-utility, recommendations.

Copyright © 2016, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

Introduction

Health-state utility (HSU) data are estimates of the preference for a given state of health on a cardinal numeric scale, where a value of 1.0 represents full health, 0.0 represents dead, and negative values represent states worse than death [1,2]. Definitions of HSU and other terms used in this report are presented in Table 1. HSU estimates are used in cost-utility analysis, a special case of cost-effectiveness analysis in which health benefits are usually measured in terms of quality-adjusted life-years (QALYs) [3]. QALYs are calculated by multiplying the number of years lived in each state of health by the HSU estimate for each respective state [4]. For example, if an intervention confers 2 extra years of life at an HSU of 0.75, then the intervention would confer an additional 1.5 QALYs (2 × 0.75) to the patient.

The Importance of HSU Data for Economic Models

Cost-utility analyses, most often performed in economic models, are increasingly being used by health technology assessment (HTA), pricing, and reimbursement authorities in many countries to establish whether the cost of a new intervention can be justified in terms of expected health benefits. The decisions made by these authorities affect patients’ access to treatments, physicians’ ability to use them, the product’s price, and, in turn, the return that manufacturers are able to realize on their investment in developing the product. HSU estimates are typically among the most important and uncertain data inputs in cost-utility models; the accuracy and precision of the model results often depend heavily on the quality of the HSU estimates. Some HTA agencies prefer health-utility data to be collected from patients [9,10], and it has become increasingly common for such data to be collected in clinical trials. Clinical trials represent an important opportunity for the collection of health-utility data. However, trials designed primarily to evaluate efficacy and safety for market authorization by regulatory agencies often present challenges to the optimal collection of HSU estimates for economic models. These challenges exist even when health-related quality-of-life (HRQOL) instruments are included as part of the efficacy assessment to meet regulatory authority requirements. Careful planning is needed to define the HSU estimates that will be needed for...
Background to the Task Force

In June 2014, the ISPOR Health Science Policy Council recommended to the ISPOR Board of Directors that the ISPOR Measurement of Health-State Utility Values for Economic Models in Clinical Studies – Good Practices Task Force be established. The Board of Directors approved the task force the same month.

The task force members and primary reviewers were selected on the basis of their expertise in fundamental or applied utility research, or economic modeling, or their understanding of quality appraisal of utility estimates during health technology assessments. Considerable effort was made to ensure international representation of health care systems in selecting task force members and primary reviewers. A list of leadership group members is available via the http://www.ispor.org/Estimating-Health-State-Utality-Economic-Models-Clinical-Studies-guidelines.asp

The task force identified five areas of guidance to enable researchers to design an optimum utility measurement program for a planned trial program or clinical study. They developed the outline, drafted sections and the full report, reviewed drafts via e-mail, ASANA, teleconference, and in person at two ISPOR international meetings and two European congresses. All task force members, as well as primary reviewers, provided feedback either as oral comments or as written comments.

The draft task force report was reviewed several times: once by the primary reviewer group of experts and twice by the Measurement of Health-State Utility Values for Economic Models in Clinical Studies Review Group. Comments were also received during two forum presentations: at the ISPOR European Congress in 2014 and again at the Annual International Meeting in 2015. All comments received during the review processes and presentations were considered, discussed, and addressed as appropriate in revised drafts of the report. We gratefully acknowledge our reviewers for their contribution to the task force consensus development process and to the quality of this ISPOR task force report.

All written comments are available on request by contacting taskforce@ispor.org. The task force report and Web page may be accessed from the ISPOR home page (http://www.ispor.org) via the purple Research Tools menu, ISPOR Good Practices for Outcomes Research, heading: Preference-Based Methods; Estimating Health-State Utility for Economic Models in Clinical Studies.

cost-utility analyses; to determine which of the HSU estimates may be measured in the planned trial or trials; to establish the optimal health-utility instrument, mode of administration, timing and frequency of assessments, period of follow-up, and data analyses; and to plan additional studies to collect HSU estimates that may not be adequately estimated within the trial.

Aim and Scope

This guideline aims to provide a framework for researchers to plan the collection of health-utility data in clinical studies to provide high-quality HSU estimates appropriate for economic modeling. High quality in this context means HSU estimates that are aligned with the definitions of the economic model health states, are free from known sources of bias, and were measured using a validated method appropriate to the condition and population of interest and the perspective of the decision maker for whom the economic model is being developed. Various study designs (Table 2) and measures (Table 3) have been used to estimate HSU. The primary focus of this guideline is collecting data within clinical trials because clinical trials often represent the best opportunity to collect high-quality data; other study types are considered briefly.

This guideline includes considerations for the following:

1. Early planning of health-utility data collection within the research and development program;
2. Design of health-utility data collection during protocol development for a planned clinical trial;
3. Design of prospective and cross-sectional observational studies for estimating health utility;
4. Use of alternative study types for HSU estimation;
5. Statistical analyses and reporting to maximize the value of patient-level health-utility data for economic models.

Recommendations are discussed in the following sections and are summarized in Tables 4 through 7 and in Figure 1. The guideline focuses on approaches for collecting HSU estimates for cost-utility models. It does not include longitudinal measurement and analysis to perform statistical comparisons of health utility between treatment arms or QALY estimation by treatment arm for cost-effectiveness analysis alongside clinical trials. A separate guideline from ISPOR is available for cost-effectiveness analysis alongside trials [28]. We focus on collection of HSU estimates via preference-based measures (PBMs) in patients (particularly validated multiattribute utility instruments with preference-based value sets) because these methods are favored by many HTA agencies at the time of writing (e.g., Canadian Agency for Drugs and Technologies in Health [11], Haute Autorité de Santé [12], National Institute for Health and Care Excellence [13], Pharmaceutical Benefits Advisory Committee [10], and agencies in other countries who recommend economic evaluations [29–34]). Other methods of estimation, including direct valuation of patients’ own health state and health-state descriptions (vignettes) valued by members of the general population, using methods such as time trade-off or standard gamble, are discussed briefly. Discussion of utility theory, development of instruments and measures, and details of how to conduct utility elicitation (e.g., time trade-off and standard gamble interviews) are beyond the scope of this guideline.

Methods for estimation of HSU by mapping from condition-specific measures are not considered in detail; separate guidelines for these methods are available [35], and are under development [23]. Measurement of HRQOL using scales other than the health-utility scale (i.e., using non-preference-based HRQOL measures such as the 36-item short form health survey [SF-36] for which scores are not anchored by a value of 1.00 for full health and 0.00 for dead) is a broad topic and is not covered here. Several guidelines for the measurement of patient-reported outcomes have been developed by ISPOR’s task forces [36–42].

This article states the consensus position of the ISPOR Task Force on Good Practices for Outcome Research – Measurement of Health-State Utility Values for Economic Models in Clinical Studies and provides recommendations for best practice developed by discussion and based on collective experience. Recognizing that best practices may not always be feasible because of practical constraints, pragmatic approaches are also discussed. This position represents the best judgment of the task force at the time of writing and is subject to change as new methods for health-utility estimation and economic modeling emerge and as the preferences of HTA, pricing, and reimbursement authorities change.
Table 1 – Definition of terms.

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical study</td>
<td>For the purposes of this guideline, a clinical study is defined as a clinical trial investigating one or more interventions, or an observational study (prospective or cross-sectional) in a routine clinical practice setting.</td>
</tr>
<tr>
<td>HRQOL</td>
<td>Health-related quality of life.</td>
</tr>
<tr>
<td>HSU</td>
<td>Health-state utility. The estimate of the health utility for a given health state. For the purpose of this guideline, this refers to the health-utility estimate for an economic model health state.</td>
</tr>
<tr>
<td>Health utility</td>
<td>A representation of strength of preference for a given health-related outcome on a cardinal numeric scale, where a value of 1.0 represents full health, 0.0 represents dead, and negative values represent states worse than dead.</td>
</tr>
<tr>
<td>PBM</td>
<td>Preference-based measure. A preference-based measure of health utility is a measurement system that allows patients to describe the impact of ill health and assigns a utility score to those descriptions on the basis of peoples’ preferences for health states. These measurement systems consist of two components: 1. A standardized descriptive system for health or its impact on HRQOL (sometimes referred to as a multiattribute utility instrument), composed of a number of multilevel dimensions that together describe a universe of health states, and 2. An algorithm for assigning utilities to each health state described by the system (often described as a value set). Algorithms have been based on various valuation methods, e.g., time trade-off, standard gamble, and discrete-choice experiments.</td>
</tr>
</tbody>
</table>

Table 2 – Options for collection of HSU estimates for economic models.

- **Review of published estimates.** A review of existing estimates available in the published literature is an important starting point to determine whether high-quality estimates relevant to the economic model are available and whether new research is needed. Such a review can also establish the range of available estimates that should be explored in sensitivity analyses. A formal systematic review maximizes the generalizability and representativeness of the estimates used in any economic model and is required by some HTA agencies as part of the evidence submission (e.g., Canadian Agency for Drugs and Technologies in Health, Haute Autorité de Santé, and NICE). Critical appraisal of studies should examine three broad issues: the relevance of the health-utility data to the health states and population in the economic model, the potential sources of bias in a study design, and the extent to which data meet the needs of decision makers. For example, NICE prefers data estimated using the EQ-5D instrument that reflects the preferences of the population of the United Kingdom. Pooling of HSU estimates may be performed to improve the precision of the estimates (and estimates of uncertainty), provided the populations are sufficiently homogeneous and the same method of elicitation and instrument was used. If these conditions are not met, it may be appropriate to explore alternative relevant estimates in the economic model.

- **Prospective data collection in clinical trials.** In most cases, with careful consideration of the optimum methodology, clinical trials and open-label extension studies represent an important opportunity and an efficient way to collect HSU data. However, careful design of the health-utility data collection is critical to the quality and usefulness of the data for the economic model. In some cases, it may not be possible to collect all HSU data needed for a model within the registration trial, and additional work may be needed to provide a complete set of estimates for all the modeled health states. In particular, it is important to time the collection of data so that it captures profiles of change around the relevant health states. These issues are discussed in more detail in Recommendations for the design of health-utility data collection during the protocol development for a planned clinical trial. The incremental cost of adding a utility measure to a regulatory trial is expected to be substantially less than the cost of performing a separate utility study.

- **Prospective longitudinal or cross-sectional observational studies.** These studies may offer the greatest flexibility in terms of the data that can be collected; and, for many health states, it may be more appropriate to collect HSU data in observational or routine data sets rather than in trials. However, observational studies may be time consuming and costly when performed in addition to registration trials and may not produce measures that are good proxies for measures of change associated with the occurrence of key health states. These studies are discussed in Recommendations for the design of prospective or cross-sectional observational studies for health-utility estimation.

- **Early-access or compassionate-use-type programs, phase 4 studies, registries, and other postlicensing commitments.** These types of studies can be an efficient way to capture HSU data and may include patients eligible for the treatment in routine practice who would be excluded from registration trials. Although such studies may be performed too late in the product development program to provide HSU estimates for HTA submissions of new technologies, the data may be useful for reassessments and ongoing cost-utility research for products.

- **Vignette studies.** In this approach, detailed descriptions of each health state are developed from different sources of information (e.g., patient and physician interviews, trial data, and published literature). Members of the public are asked to rate these states in a stated-preference experiment (such as time trade-off or standard gamble). These methods are limited because the resulting estimates are entirely dependent on the validity of the vignette descriptions, vignettes are not able to fully reflect the varied HRQOL experience among patients within a given vignette health state, and do not offer the opportunity for patients in the health state to describe their own health (as is the case if a multiattribute utility instrument is used).
Table 3 – Measures for the estimation of HSU estimates for economic models.

- **Generic PBMs.** Many HTA authorities require or prefer HSU to be estimated using a generic PBM and a value set algorithm developed using data from the general population in the authority’s country. Examples of generic (condition-nonspecific) PBMs include EQ-5D, HUI, 15D, AQoL, and SF-6D. PBMs differ considerably in the content and size of their descriptive system, the methods of valuation, and the populations used to value the health states [20].

- **Condition-specific PBMs.** These measures are similar in concept to the generic PBMs but have been developed for a specific disease or condition. Examples include the AQL-5D for asthma [21] and the EORTC-8D for cancer [22]; other examples are summarized by Brazier and Rowen [20]. Many condition-specific measures describe symptoms or symptom impact rather than HRQOL [20].

- **Mapping of a patient-reported outcome measure onto a PBM.** Mapping or “cross-walking” involves estimating a relationship between a (often condition-specific) patient-reported outcome measure and a PBM, using a statistical model, and making predictions from the estimates [23]. These methods can be valuable in those instances in which there is no health-utility measure available but some form of patient-reported outcome or clinical end point can be used as a basis for predicting HSUs. Mapping may also be performed from one PBM to another, e.g., from the EQ-5D five-level version to the EQ-5D three-level version.

- **Direct valuation of patients’ own health state using PBMs such as time trade-off or standard gamble.** This may be appropriate if no multiattribute utility instrument is available that is valid and responsive in the condition of interest and that meets the needs of decision makers and no means of mapping to a PBM is available. For example, it has been reported that generic PBMs may not effectively capture the impact of eye diseases [24,25]. Guidance on performing time trade-off studies is available in the published literature [26,27]. Although this approach provides direct observations of health utility, there are technical, ethical, and practical obstacles to performing time trade-off and standard gamble experiments with patients. Furthermore, the data represent patients’ values rather than those of the general population, which is the preference of many HTA authorities [20].

Note: The ordering of methods in this box is not intended to convey any order of preference; preferences for alternative methodologies vary among individual HTA authorities and other audiences.

AQL-5D, Asthma Quality of Life Utility Index; AQoL, Assessment of Quality of Life; EORTC-8D, European Organization for Research and Treatment of Cancer-specific instrument; HRQOL, health-related quality of life; HSU, health-state utility; HTA, health technology assessment; HUI, Health Utilities Index; PBM, preference-based measure; SF-6D, six-dimensional health state short form (derived from 36-item short form health survey).

General Considerations for the Collection of Health-Utility Data for Economic Models

There are several key issues to consider when designing studies to collect HSU estimates for economic models.

1. **The HSU estimates must be fit for the purpose of the anticipated economic evaluation.** The health-utility data collection should be designed (as far as is possible) to capture HSU estimates for each health state anticipated to be in the economic model. The criteria used to define the health states in the economic model should be used to label individual patient utility assessments for analysis for calculation of HSU estimates. The health-utility measure should be selected on the basis of its acceptability to the economic model’s audience (e.g., HTA authorities), appropriateness for the health condition of interest, and (where appropriate) consistency with previous economic evaluations.

2. **The participants in the study should reflect the indication and the population that will be likely considered in the economic model.** This is usually the population that would receive the investigational intervention if it is adopted in routine clinical practice. Ideally, studies should recruit a sample of participants that will produce HSU estimates representative of the entire population under consideration in the economic model. Regarding representativeness more generally, any sampling approach should be free from sources of bias. It is better to recruit from multiple centers and, if appropriate, to include patients on different types of treatment. Trial inclusion and exclusion criteria may result in populations that are younger and fitter than populations in routine practice or may exclude specific groups of patients. A further potential source of bias is that participants in more severe health states can be less likely to complete assessments [43]. Similarly, the participants recruited into any study should be representative of the types of patients being treated in the health care system where the cost-effectiveness analysis will be used. For example, if the HSU estimates are being collected for an economic model for the National Institute for Health and Care Excellence, then it is most appropriate to include participants from England and Wales. If it is not possible to recruit a representative patient sample, it is important to ensure adequate sampling of the model population and, if appropriate, to adjust HSU estimates in the analysis. This is discussed further in the “Recommendations for Data Analysis and Reporting” section of this task force report.

3. **Longitudinal collection of data is often valuable.** Many economic models describe longitudinal changes in patients’ health utility as their disease progresses through different stages. Sampling needs to consider variability among patients within a health state (i.e., the study population should adequately reflect variability within the model population) and variability over time within a single health state (e.g., a decline in HRQOL over time within a disease progression category). The order in which health states are experienced can also be important; for example, health utility in patients with low body mass index (BMI) who previously had high BMI might be expected to differ from health utility in patients who always had low BMI because the latter may be less likely to have complications associated with high BMI. The best way to understand how health utility changes over time is to capture data from patients at multiple time points as they progress through health states. There is emerging evidence that changes in health utility associated with progressing through health states over time may differ significantly from differences inferred from cross-sectional data [44]. Collection of longitudinal data is not always practical, due to time and cost constraints. Where cross-sectional studies are performed instead, these should be designed to ensure that the data sample for each health state represents variability as much as possible within the patient population and over time within the health state.
4. **Collection of real-world utility data is an important addition in many instances.** Trials contain many protocol-driven procedures and have very tight entry criteria, both of which may reduce the generalizability of the data. Collecting subjective data in a trial may also promote placebo-type effects that could inflate health-utility scores (e.g., regular monitoring or the possibility of receiving an investigational new treatment may inflate utility scores). Data collection from routine clinical practice may produce more representative data for guiding decision making. For example, the Patient Reported Outcome Measures program in the United Kingdom attempts to capture data from all patients undergoing hip and knee replacement surgery, and part of the assessment includes the EuroQol five-dimensional questionnaire (EQ-5D) [45]. It should be noted that data collected in routine clinical practice for patients receiving existing therapies may not fully reflect the HRQOL of patients receiving a new treatment. Inevitably, there will be instances in which real-world utility data cannot be collected; in those instances, the economic evaluation should consider whether, and in which direction, trial-derived data are likely to be biased when estimating incremental QALYs.

5. **Collection of utility data relating to rare but important events.** Many economic evaluations will include health states that occur infrequently but nevertheless are important to capture because of their impact on patients’ health (e.g., a severe adverse event or a complication related to a disease). If health states are rare, then it is often hard to capture enough data from patients in that health state. When making decisions regarding how data should be captured, it is worth reflecting on how important these states are for the outcome of the economic evaluation. If they are important for determining cost-effectiveness, then this may provide justification for investing in a prospective study. One way to address this problem is by purposively recruiting patients who are at increased risk for developing the infrequent health state. So, rather than recruiting a large number of patients and waiting for the health state to emerge, it may be better to specifically target patients at increased risk. However, this approach may not work for some disease areas. For some health states, it is extremely difficult to capture data outside of a clinical trial. For example, there are substantial difficulties in collecting health-utility data from patients experiencing severe adverse events. Similarly, published quantitative data regarding health utility in patients approaching the end of life in palliative care are rare. However, such information can be important. Data collection after the end of the active treatment phase of the study is informative to illustrate how HRQOL changes as people move through subsequent care and treatment (although may be more prone to missing data due to the difficulty of retention beyond active treatment; see Section Recommendations for the design of health-utility data collection during the protocol development for a planned clinical trial-Missing data for further discussion).

6. **Mode of administration should be considered.** Regarding the mode of administration of the assessment (e.g., paper instruments, electronic data capture systems via computer, dedicated device, or platform mobile apps), the mode can affect response rates, task comprehension, response strategies, and representativeness of the sample. Standardization across data collection and devices is preferred for all respondents.
the highest category or the lowest category (i.e., excellent and poor) when asked an oral rather than a written self-reported health question. Furthermore, where electronic data capture is used, care must be taken to ensure that the questionnaires can be standardized across all potential devices. For example, the length of a visual analogue scale should be consistent for all participants on all devices at all time points [49]. In addition, it is possible that certain population groups (e.g., the elderly or the severely ill) will have more difficulty in using electronic data collection tools; the choice of mode of administration should consider whether this is an issue.

7. **Good clinical practice should be followed.** The application of any research solution should be within the appropriate guidance for good clinical practice, patient consent, safety reporting protocols, and data transparency obligations appropriate to the study.

**Recommendations for Early Planning of Health-Utility Data Collection Within a Product’s Research and Development Program**

To successfully plan the collection of health-utility data for an economic model, it is important to perform early research to establish the following:

1. The HSU estimates that will be required for the model;
2. The data that are already available from the published literature;
3. Whether any important differences exist among countries or cultures;
4. If there is any uncertainty, the appropriateness of available instruments for the health condition and population of interest, and availability of versions suitable for use in the countries in which utility measurement will be made.

This information will inform the selection of an appropriate instrument for inclusion in clinical studies. Alternatively, if no existing instrument is appropriate, plans can be made for the adaptation of an existing instrument or the development of a new instrument and value sets. The planned trial program then may be evaluated for opportunities to collect the required HSU estimates; and, if necessary, additional research can be performed to explore the appropriateness of a particular health-utility instrument in the condition of interest. Any expected gaps in the data available from the clinical trial program should be identified, and additional studies to address these gaps should be planned. Engaging the teams responsible for economic modeling, health-utility assessment, patient-reported outcome assessment, and clinical trial design in strategic planning at an early stage within the product development lifecycle will maximize the potential for producing robust data that support the economic model. **Figure 1** summarizes recommendations for activities to be performed within, or in parallel with, the clinical studies in the product development program. The figure uses the pharmaceutical development program as a framework; however, a similar set of activities during product development may be appropriate for vaccines, diagnostics, and medical devices for which cost-utility analyses are anticipated.

It is important to first describe the impact of the health condition or disease on HRQOL and the benefits that the interventions are expected to provide in terms of HRQOL (as well as any harms, e.g., from side effects of treatment), so that the research may focus on capturing these aspects. As part of the planning process, it may be helpful to develop a preliminary economic model or review existing models of the condition of interest, to identify and define the health states for which HSU estimates are likely to be required, and then to perform exploratory analyses to determine the influence that the estimates may have on the results. This can be helpful in determining how much resource to invest in the collection of specific estimates. It may be appropriate to consider whether it will be important to measure the impact of the patient’s condition on the health utility of caregivers and/or family members or dependents, especially for chronic conditions. A literature review should be performed to establish the availability of existing data, as well as the quality of the available data, the relevance of the data to the economic model health states, and the acceptability of the data to the model’s audience (e.g., specified HTA authorities).

Health-utility data may be generated using one of the following categories of instruments (see also **Table 3**):

1. A generic PBM (e.g., EQ-5D, Health Utilities Index [HUI], Short-Form Six-Dimensions [SF-6D, derived from SF-36], 15D, Assessment of Quality of Life [AQoL], or Quality of Well-being [QWB]);
2. A condition-specific PBM (e.g., the Asthma Quality of Life Utility Index [AQL-5D] or the European Organization for Research and Treatment of Cancer eight dimensions [EORTC-8D]);
3. A non–preference-based, condition-specific patient-reported outcome measure mapped onto a generic, preference-based measure [23,35].

An instrument or instruments should be selected on the basis of suitability for the disease or condition of interest, suitability for the population of respondents (including availability of translated and culturally adapted versions suitable for use in all study countries), and acceptability to the model’s audience (e.g., the HTA authorities to which the model is expected to be submitted). If there is any doubt about the appropriateness of utility instruments for the condition of interest, this should be evaluated in terms of practicality, reliability, validity, and responsiveness on the basis of empirical evidence [17] and using methods that take into consideration any requirements of the audience for the economic model (e.g., HTA agencies) for such evaluations. Some examples of this type of evaluation are available in the published literature (e.g., Brazier et al. [49]). The requirements and preferences of HTA authorities are evolving over time; researchers should refer to guidelines issued by individual authorities (links to many of these guidelines are available on the ISPOR Web site) and consult with the HTA authorities, if needed, regarding the appropriateness of instruments in the condition of interest. If additional research is required to validate an instrument or develop an appropriate instrument, identification of this requirement early in the product’s development (e.g., before initiation of phase 2 studies) may make it possible to perform such research in parallel with the clinical research program. If new patient-reported outcomes instruments are being developed, researchers should consider including a generic preference-based health-utility measure in the research studies. This may have benefits in evaluating the suitability of the PBM for the condition and population of interest and/or in providing data that could be used to develop a mapping algorithm.

These steps should establish what HSU estimates are needed and the appropriate instrument for their measurement, as well as help to guide selection of the optimal timing and frequency of administrations of the instrument to collect data for model HSU estimates. Researchers should determine whether the planned trial program has the capacity to provide all HSU estimates required for the economic model. The benefits of collecting all key health-utility estimates for an economic model within a single study, or at least using a single methodology, should be recognized (e.g., consistency of population characteristics and methodology among all HSU estimates used in the economic model). Often researchers are faced with limitations on the number of measures that can be included in a clinical study
Begin planning for HSU estimation early in the product development process.

Prepare a description of the HRQOL impact of the disease and the potential HRQOL benefits of the new intervention.*

Prepare a list of HSU estimates required for the model, their definitions, appropriate measurement methods, and data already available; define new data needs.

• Identify and clearly define the patient population of interest and the modeled health states and events for which HSU estimates will be required. It may be appropriate to consider the need for health-utility data for patients’ caregivers and/or family members or dependents. Development of an early economic model can be helpful in this process.

• Identify the requirements and preferences of key audiences (e.g., HTA authorities) for HSU data.*

• Identify the most appropriate health-utility instrument, based on appropriateness for the condition of interest and acceptability to the audience(s) for the economic model(s). If there is any uncertainty about the appropriateness of HSU instruments in the disease condition, evaluate the appropriateness of possible instruments (e.g., by review of published qualitative and quantitative studies).†

• Review the published literature to identify HSU estimates that are already available; assess their quality, relevance to the model health states, and acceptability to the audience for the model (e.g., HTA authorities).†

• It can be helpful to determine the sensitivity of cost-effectiveness estimates to individual model HSU estimates, using an early economic model to evaluate the importance of collecting high-quality data for each HSU estimate.

Evaluate the planned research and development program of the product for opportunities to collect the HSU data defined in the previous steps.

• Consider opportunities to use phase 2 trials, open-label extension studies, and/or other clinical or observational studies to collect long-term longitudinal data and/or collect additional data unlikely to be available from the phase 3 trial.

• Recognize the benefits of collecting all key HSU estimates within a single study (or at least using single methodology).* It can be helpful to prepare an HSU Research Plan.†

HRQOL, health-related quality of life; HSU, health-state utility; HTA, health technology assessment.

* If early planning has not been performed before the phase 3 trial, the items highlighted with an asterisk are recommended before developing the health-utility section of the trial protocol.

† The following content is suggested for the HSU research plan: a definition of the patient population; the HRQOL impact of the disease condition and potential HRQOL benefits of the new intervention; the definitions of the health states and events for which HSU estimates are required for the economic model; a summary of the appropriateness of alternative utility instruments for the condition of interest, availability of versions for the study countries, and acceptability to the economic model’s audience; identification of the selected measure and justification for its selection; a summary of HSU estimates already available and assessment of their quality, relevance to the model health states, and acceptability to the model’s audience; an assessment of the importance of collecting high-quality data for each HSU estimate (e.g., based on analyses using the early economic model); an evaluation of the product’s planned research and development program for opportunities to collect the HSU data; the identification of data gaps (HSU estimates that are unlikely to be able to be estimated in planned studies); a plan for collection of health-utility data within the planned research program; and an outline for any additional studies to bridge data gaps.

Considerations for the Design of Studies to Collect HSU Estimates for Economic Models

Recommendations for the Design of Health-Utility Data Collection during the Protocol Development for a Planned Clinical Trial

Trials often represent an important opportunity to measure health utility because they often provide a large sample of the
patient population of interest within a study that is performed and monitored to a high standard, thus maximizing data quality and completeness. In addition, the health-utility data may be linked to the end points that are used to estimate treatment effect. However, there are a number of pertinent issues that can determine the value of the health-utility data collected in this way.

**Number of possible assessments for a given health state**
There may be some HSU estimates required for which it is not feasible to collect health-utility data in the trial or for which the number of assessments possible is likely to be low—for example, estimates for rare health states or health states that tend to occur after the end of the trial’s follow-up period. Researchers should identify the economic model health states and events and determine for which of these it is feasible to estimate health utility within the trial. These should be defined clearly using (as far as possible) the same criteria that will be used to characterize them in the economic model. Any modeled health states or events for which data of sufficient quality may not be available from the trial should be identified, and alternative plans for their collection should be made.

In some cases, patients may enter the trial with different numbers of previous events experienced or different numbers of lines of treatment received. This may be problematic if the economic model requires HSU estimates for a first, second, or third event or treatment line. There may also be differences in combinations of previous events or treatments or in time since the event or treatment. These issues should be considered at the design stage. Collection of appropriate data on baseline characteristics, disease, and treatment history would allow separate estimates to be calculated as required.

If more than one trial is being designed for the target indication, it may be advisable to implement utility data collection across all trials, if feasible, to reduce uncertainty around the estimates and potentially capture additional health states, if such states are important for modeling and are supported by the variations in the trials being designed.

Considerations of whether the sample size for the utility estimates is likely to be sufficient should take into account the need for precision of HSU estimates rather than hypothesis testing because the main objective is to generate utility estimates, and the uncertainty around these estimates, for the model health states rather than to make statistical comparisons of utility between treatment arms.

---

**Figure 1 – Recommendations for HSU data collection strategy during the product development process.**

HRQOL, health-related quality of life; HSU, health-state utility; HTA, health technology assessment; PBM, preference-based measure; PRO, patient-reported measure.
Table 6 – Design of health-utility data collection in a clinical trial.

Design any health-utility data collection in a clinical trial during the development of the protocol, with reference to the anticipated needs of the economic model.

Health economists who understand the needs of the economic model should be involved in and able to influence the design of health-utility data collection in clinical trials, and the analyses of such data.

Identify issues associated with collecting HSU estimates for the economic model in the planned trial and make plans to address them. The following issues may arise, for example:

• The timing of clinical assessments may not be optimal for utility measurement.
  - Assessments should be optimized to capture data for model health states and/or events and to maximize the amount and quality of data.
  - Consider the recall period of the utility instrument.
  - Consider acute events and the duration of their impact on HRQOL.
  - Consider changes in HRQOL over time within a single model health state (e.g., “postprogression” health states in cancer models).
• Consider the number of assessments that are likely to be feasible for each model health state and/or event and the likely precision of the HSU estimates.
• It may not be feasible to capture data for some model health states and/or events within the trial (e.g., rare events and events occurring after trial follow-up). Make alternative plans to collect these data.
• Consider the importance of acute events (i.e., those with a transient effect on HRQOL) for the economic model; if the collection of accurate estimates is important, consider whether measurement is feasible within the trial, or plan a separate study.
• The trial population may not be fully representative of model population, e.g., due to trial exclusion criteria or geographic footprint. Evaluate the potential for bias, consider adjusting trial selection criteria, collecting data for excluded patients, and/or adjusting estimates in the analyses.
• Respondents may be unable to complete utility assessments (e.g., young children, cognitively impaired or severely ill patients). Examine the relevant literature for any relevant research findings; if use of proxy respondents is expected to be the best solution, review relevant literature and examine the potential for bias.
• Identify potential causes of missing data (e.g., reasons for planned or unplanned loss of follow-up and types of patients who are less likely to complete assessments). Address these as far as possible by adjusting the study design and formulate plans to adjust for missing data in the analysis.

Development of a utility data collection protocol (or a section of the trial protocol) is recommended.

• Clearly define the objectives of the health-utility measurement in the trial:
  - Identify and define model health states and/or events for which it is planned to estimate health utility in the trial (considering the number of assessments feasible for each health state and whether HSU estimates for acute events are important for the economic model and feasible to estimate in the trial).
  - Specify whether statistical comparisons between treatment groups in overall utility over time also will (or will not) be performed.
• The following design features should be defined and justified in terms of the needs of the economic model:
  - Choice of instrument
  - Timing and frequency of assessments and period of follow-up
  - Variables that should be collected at baseline and at each health-utility assessment to determine health state at the time of assessment and to allow adjustment for covariates
  - Choice of respondents
  - Mode of administration
  - Methods to address any heterogeneity of the patient sample or issues with generalizability of results
• Analyses should be designed to provide HSU estimates for model health states or events, making any appropriate adjustments to generalize the results to the population of interest in the economic model, and to preserve valuable information available in the patient-level data (see Recommendations for Data Analysis and Reporting).
• Identify important HSU estimates that will not be measured in the trial or available from existing research and plan alternative studies.

HRQOL, health-related quality of life; HSU, health-state utility.

At each health-utility assessment, it is important to capture any other variables that will be needed to connect the patient’s assessment with the economic model health state the patient is experiencing at the time of the assessment. The nature of such additional data collection (e.g., categorical or continuous data, units) should be selected in view of the planned analyses of the data for the economic model.

Representativeness of the trial population to the population of interest in the economic model

Issues that may affect the representativeness of the trial population to the economic model population should be identified; for example, trial inclusion and exclusion criteria. If the economic model population includes individuals who are excluded from the trial, it is important to consider whether they are likely to differ in terms of health utility, both in absolute and in incremental terms. Researchers should evaluate the potential for bias arising from issues of generalizability and formulate plans to address any such issues. If a threat to generalizability exists, it is advisable to assess a possibility to adjust trial selection criteria or collect utility data for patients excluded from the trial who would be eligible for treatment in routine practice. The extent and direction of any bias may be examined (e.g., by comparing baseline characteristics with characteristics of real-world populations) and/or corrected for in the analysis (see the “Recommendations for Data Analysis and Reporting” section).

Special patient populations

Collecting health-utility data in certain patient populations poses significant issues (e.g., patients with dementia [50], younger
Biases should be explored using sensitivity analyses in the direction and extent of bias that might result and potential patient and proxy respondents for the instrument and condition. Evidence examining inter-rater reliability or agreement between proxy respondents, researchers should review the available research. In which the patients themselves are unable to complete assessments. Using proxy respondents in cases in which the patients themselves are unable to fill out questionnaires is an acceptable practice; it is common practice to use a nominated informant for each patient to ensure consistency across multiple assessments for individual patients. For example, in cognitively impaired patients who have experienced stroke and cannot complete questionnaires, research suggests that family caregivers can reliably complete the utility assessment [53]. However, health preference studies of children (including those with brain tumors) have reported higher utilities for parent proxy-report compared with child self-report, whereas other studies have reported lower utilities [54]. Therefore, when using proxy respondents, researchers should review the available evidence examining inter-rater reliability or agreement between patient and proxy respondents for the instrument and condition in question. Researchers should also examine the potential for direction, and extent of bias that might result and potential biases should be explored using sensitivity analyses in the economic model.

Timing of assessments
Clinical trial assessments are often made at regular scheduled intervals. The optimum timing of health-utility assessments may not coincide with the clinical assessments. For example, health-utility assessments in cancer therapy trials are often made alongside clinical assessments at the chemotherapy administration visit (e.g., approximately every 3 weeks) and a short time after disease progression. This often results in numerous utility assessments before disease progression, but these are unlikely to capture the impact of chemotherapy toxicity (because this occurs primarily between treatment administrations) and few assessments are available after progression and very few or none during terminal illness.

As a general rule, the number and timing of assessments should be planned to optimize the amount and quality of data collected and its relevance to the economic model, recognizing the health states and events for which HSU estimates are required in the model. It is also important to recognize the instrument’s recall period, which can range widely, and plan the frequency of utility assessments in the context of the recall period. For example, the EQ-5D describes health today, whereas the SF-6D Health Survey describes health over the last 4 weeks. HRQL is expected to vary over time within an individual economic model health state and the HSU estimate is conceptually an average for all patients and all times in that state (e.g., as in the case of postprogression utility in many cancer models), care should be taken when scheduling assessments to ensure that the variability between patients and over time is adequately sampled. Data quality and relevance should take precedence over convenience, as far as is possible, and it may be necessary to schedule assessments specifically for health-utility measurement. For example, in cancer trials, it would be valuable to schedule assessments throughout the period after disease progression until death; this may be possible particularly in trials that include overall survival as an end point. Use of modern technology that avoids the need for an actual visit to complete an assessment, such as electronic diaries, may be helpful in minimizing patient and investigator burden and use of proxy respondents may be considered if patients are too ill to complete assessments. Economic models are often concerned with changes in HRQL as patients move between health states. Therefore, it is important to schedule assessments that allow these changes to be estimated (e.g., at baseline and when certain clinical outcomes are reached, which define the economic model health states), Trials that involve a change from baseline (e.g., trials of surgical interventions) may benefit from multiple assessments before baseline, to reduce the uncertainty around the treatment effect.

When HSU data are collected during a visit involving other interventions and/or assessments, it may be important to standardize the timing of the utility data collection (e.g., at the beginning of the visit), bearing in mind whether and how any of the other interventions and/or assessments may affect HRQL.

Acute events
An additional challenge for collecting health-utility estimates is in circumstances in which much of the quality-of-life gain from an intervention is in the reduction of acute episodes, that is, those with a transient effect on HRQL (e.g., heart attacks, asthma exacerbations, and hemophilia bleeds) or of short-term (but often severe) treatment-related adverse events. These circumstances represent a particular challenge in health-utility assessment because of the practical and potential ethical considerations in requiring patients to complete a questionnaire during an acute episode (e.g., during a hospitalization or disease

<table>
<thead>
<tr>
<th>Table 7 – Analysis and reporting.</th>
</tr>
</thead>
<tbody>
<tr>
<td>The data analysis approach should reflect needs of the economic analysis and should not be constrained by the traditional approach to analyzing data for regulatory purposes (e.g., comparison between treatment arms).</td>
</tr>
<tr>
<td>Data are constrained to values &lt; 1 and are left-skewed. Consider a simple linear transformation (X = 1 - U) and then use familiar methods for right-skewed data, e.g., generalized linear modeling [62] or generalized estimating equations for longitudinal data [59,63,64], with the possibility of two-part model to handle excess zeros [65]. Consider explicit modeling of prognostic factors. This has the potential to increase the generalizability of results by adjusting clinical trial data to the characteristics of populations in routine clinical practice.</td>
</tr>
</tbody>
</table>
| Regarding the transferability of data in multinational studies:
- There are predictable differences in the way patients map to the index score of the health-utility instrument.
- Country or region effects can be handled as a covariate in statistical modeling.
- Use the country value set or tariff that is appropriate for the economic model’s audience (separate analyses may be needed for each country-specific economic model). |
| Careful reporting can maximize the value of research for future economic evaluations. Covariance can be used to retain the integrity of the logical ordering of health-utility estimates under conditions of uncertainty. Alternatively, consider specifying a functional form that maintains the logical ordering of HSUs. This is important for other analysts seeking to apply the results in future models, including appropriate assessment of uncertainty. |
| HSU, health-state utility. |

VALUE IN HEALTH 19 (2016) 704–719
exacerbation). The recall period of the utility measure should be carefully considered in relation to the timing of assessments and the occurrence of acute events. There is the potential, at least in theory, for double counting if the utility observed in a trial is a blend of exacerbating and nonexacerbating states and this is used in a model that layers on additional utility deficits of acute events. In practice, careful analysis, of the sort described in the next section, should minimize this potential issue. Another challenge facing researchers is in measuring the period of time over which the event impacts HRQOL. From a practical standpoint, clinical trial personnel may not be aware of a patient’s acute event until the patient’s scheduled trial visit, by which time some recovery may have occurred.

Researchers therefore should think about whether their economic model needs health-utility estimates for acute events or states. If the answer is yes, further consideration should be made as to whether these estimates can be realistically measured in the context of a clinical trial. As a practical consideration, the more the acute events are related to the intervention of interest and the more they are expected to affect the cost-effectiveness estimate, the more accurately they should be described. Important acute events (e.g., heart attacks, asthma exacerbations, and hemophilia bleeds) are commonly included among the trial end points that trigger collection of clinical data, and economic models often focus on grade 3 or greater adverse events because these require treatment. Both types of event, therefore, require clinical contact and data collection; consideration could be given to assessment of health utility during this clinical contact.

The approach to collecting health-utility data for acute events and adverse events should be considered carefully and informed by an understanding of the expected nature and duration of the HRQOL impacts of the event (e.g., by consultation with clinical experts). When considering acute states, the following approaches may be helpful: 1) asking proxy respondents to complete the assessment, 2) asking patients to recall their experience with the acute events of interest from the recent past, or 3) conducting additional health-utility elicitation outside of the trial. It may be possible to measure the impact of adverse events on HRQOL in aggregate, that is, by sampling the time during which patients are on treatment in such a way that the average effect of adverse events over time is captured. For example, in cancer trials, patients could be asked to complete the health-utility measure on randomly assigned days after chemotherapy administration (the random time allocation could be performed as part of the process of random allocation to treatment arm).

**Missing data** Where there is loss to follow-up from the trial or other mechanism resulting in data missing not at random, the patients remaining in the trial may represent a subpopulation that is distinct from that which was lost to follow-up (e.g., patients with less severe disease, response to treatment, or fewer adverse events). Therefore, measurement of the health utilities in the patients who remain in trial may not accurately capture the health utilities of the target population. In these situations, standard imputation procedures such as the last observation carried forward or treating the utility data as if they are censored may not be appropriate. Where deliberate plans are in place not to follow up certain patients (e.g., if it is no longer necessary for the purposes of the efficacy and safety assessments), the relevance of these patients from a health-utility perspective should be assessed and continuing follow-up should be performed as appropriate and when possible. If assessing patients lost to follow-up is not practical, it may be valuable to contrast the baseline general characteristics and HSU estimates of those who subsequently drop out and those who do not, to explore whether missing data for those patients might be expected to bias the results.

Recognizing that missing data often is more prevalent in patient-reported outcomes than clinical end points, plans should be put in place to minimize missing data resulting from loss to follow-up and other causes. The extent and pattern of missing data should be described when reporting health-utility data.

**Recommendations for minimizing missing data** are available in the published literature [55].

**Patient and investigator burden** A common objection to the inclusion of health-utility data collection in clinical trials is the additional burden imposed upon patients and investigators completing another assessment. This is also a consideration in determining the number and timing of assessments. In this context, the importance of high-quality HSU estimates to the HTA and reimbursement processes that govern the availability of the product after launch should be recognized and balanced with the requirements for other data collected in the trial. It should also be recognized that health-utility measures often use very short questionnaires (e.g., the EQ-5D is scored from only five questions). Use of an HRQOL instrument that has a subset of items for which utility has been described (e.g., SF-36/SF-6D, EORTC QLQ C-30/EORTC-8D) may be helpful in providing HRQOL and utility estimates from the administration of a single questionnaire. Consideration of the burden imposed on patients and investigators underlines the importance of good planning and design to align the utility data collection with the needs of the economic model, ensuring that the burden is minimized and that the resulting data are useful.

**Optimization of health-utility data collection** Using good design and appropriate analysis should allow researchers to account for many of the issues raised above in Recommendations for the design of health-utility data collection during the protocol development for a planned clinical trial. Health economists who understand the needs of the economic model should be involved in, and able to influence the design of health-utility data collection in clinical trials, and the analysis of such data. The aim of their participation should be to optimize the design of the health-utility data collection in the trial to meet the needs of the economic model and the model’s audience and to identify any HSU estimates that may not be estimated within the trial so that other plans can be made for their estimation. It is important to plan any utility data collection during protocol development, that is, before the trial design has been finalized.

The objectives of health-utility measurement in trials should be clearly defined, to aid in the selection of the instrument and measurement schedule. The economic model health states for which it is feasible to estimate health utility in the trial should be identified and clearly defined. Careful consideration should be given as to whether it is appropriate to reflect differences between treatment groups within health states—for example, because of different patterns of adverse events or different intensities of therapeutic response.

**Recommendations for the Design of Prospective or Cross-Sectional Observational Studies for Health-Utility Estimation** For various reasons, it may not be possible to collect health-utility data within a clinical trial program (Table 6). In this section, we discuss the merits of conducting separate observational studies to collect these data. We also review factors to consider in the design and conduct of such research. Any
planned studies should be designed to be representative of the population considered in the economic model, free from known sources of bias and valid in their design and execution.

Prospective and cross-sectional health-utility studies can range in complexity, from complex studies performed at clinical sites to very simple online surveys. Many of the issues and recommendations discussed in Recommendations for the design of health-utility data collection during the protocol development for a planned clinical trial also apply to these studies and should be evaluated during the study’s design phase. This section considers additional issues that are specific to prospective observational studies in clinical centers and cross-sectional surveys. Other study designs are also possible for which similar considerations apply. For example, cross-sectional studies performed at clinical centers may be appropriate if longitudinal data are not required or if there is insufficient time for a longitudinal study, but detailed and verifiable clinical data are needed. Longitudinal patient surveys with repeated administrations over time may be appropriate if longitudinal data are required and access to patients via clinical centers is difficult.

Cross-sectional surveys
Cross-sectional surveys can be set up and run quite rapidly. Most HRQOL measures have validated Web-based or tablet-based versions available for use and guidance exists regarding this process [37]. Cross-sectional surveys have a number of important advantages. They are a relatively quick and cost-effective method for capturing HRQOL data. Surveys can be conducted nationally or internationally, which can improve the representativeness of data. Online surveys may be considered more discreet and, thus, a better format for capturing sensitive data [56]. In the study by Kerr et al. [56], the impact of sexual dysfunction on HRQOL was assessed solely via an online survey that consisted of a combination of self-complete clinical assessment scales and HRQOL measures. Online surveys can recruit through different sources, such as patient advocacy groups, social media, and recruitment panels, as well as from health care providers.

There are certain limitations to cross-sectional surveys that should be considered. As well as capturing HRQOL data, any survey needs to capture the correct variables for categorizing participant assessments into health states. Studies in which the health state is defined by a laboratory or radiological marker would most likely need to be conducted at a clinical site. In studies in which participants are recruited through advocacy groups or recruitment panels, it is likely that only basic clinical information (diagnosis, time since diagnosis, treatments, etc.) can be captured. Depending on the recruitment method, it may not even be possible to accurately verify diagnosis of participants. Some self-complete tools are available that can be used to measure severity, but this is not always possible. Last, studies that are conducted exclusively online may be criticized for excluding participants who do not have access to or are less familiar with using the Internet (e.g., elderly people), although, with ever increasing access to such technology, this has become less of a concern.

Prospective observational studies in clinical sites
More formal observational studies can be conducted at clinical sites to capture HRQOL data. This is recommended when the potential participants can be considered a hard-to-reach group because of prevalence or other reasons. For example, an observational study of patients receiving treatment for a life-threatening condition such as cancer is probably best conducted through clinical sites because it is usually important to have detailed diagnostic, staging, and treatment history information, which requires access to medical records. Lloyd et al. [57] reported a prospective HRQOL study, conducted at four clinical sites, of people with moderate to severe allergic asthma that necessitated access to clinical data. However, recent experience from Sheffield, United Kingdom, shows the challenge of this type of research [58]. In the Keetharuth study, the authors attempted to recruit a sample of women with breast cancer to measure the disease’s impact on HRQOL. Despite the fact that the study was cross-sectional only, the authors were not able to recruit sufficient numbers of patients in the more advanced stages of the disease.

Prospective studies are often complex to undertake. Clinical sites may not see the value of this research, compared with an investigational trial, and thus it can be difficult to recruit participants. Usually patients cannot be in two studies simultaneously (even if one is an observational study), and patients may be more likely to be invited into a clinical trial. Thus, observational studies can be slow to recruit patients and, more importantly, may not be representative of patients generally [59]. Careful study site selection and engagement with medical specialists and advocacy groups may help to avoid these problems.

Last, it is also possible to undertake longitudinal research without recruiting participants through sites but instead by recruiting them through other sources (e.g., patient advocacy groups). This is particularly relevant if detailed clinical data are not required. Monthly or even annual data collection could be undertaken to understand changes over time. Endorsement from patient charities or medical specialists may help to minimize patient dropout in such studies.

Alternative Study Types for Estimation of HSU
In situations in which HSU may not be estimated using methods described in previous sections, alternative methods may be considered.

Early-access or compassionate-use–type programs, phase 4 studies, registries, and other postlicensing research activities can provide opportunities for including HRQOL assessment. The design issues and recommendations discussed in the previous sections also apply to these studies. These types of studies can be an efficient way to capture HRQOL data, but these may be performed too late in the product development program to provide HSU estimates for HTA submissions.

Valuation of health-state descriptions (vignettes) by the general population may be an option when other methods are not possible, for example, if administration of an existing instrument to a cohort of patients is not feasible or impractical. The quality of the health-state descriptions is critical in vignette studies. Published guidelines recommend extensive qualitative work with patients, independent psychometric validation of the vignette descriptions, and use of quantitative HRQOL data to inform the content [20].

Basing HSU estimates on clinical opinion should be avoided. However, the opinion of clinical experts can be helpful in evaluating the plausibility of alternative available estimates and/or the expected rank order of HSU estimates, from best to worst health state, on the health-utility scale.

Recommendations for Data Analysis and Reporting
This section presents recommendations for statistical analyses and reporting of the results of health-utility analyses conducted alongside clinical studies to maximize the potential of the HSU estimates for current and future economic models.

Estimation of HSU for Economic Models
When clinical trials are used to collect HSU data, it is often the case that the approach to the analysis is strongly influenced by
the standard, between-arm comparisons that are used for regulatory submissions. However, if reimbursement submissions are to be made using an economic model, then it is more appropriate that the health-utility data in a clinical trial be analyzed to inform that model, rather than be analyzed by treatment arm. Consider the following example of health-utility data collected as part of the UK Prospective Diabetes Study (UKPDS) [44]. When the study started in 1977, economic evaluation was barely used and health-utility measures had not been invented. However, toward the end of this trial, which reported in 1996, there was much interest in performing economic evaluations alongside clinical trials. Therefore, in 1996, as the study was drawing to a close, the EQ-5D was administered cross-sectionally to all 3667 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 intensive blood glucose control had been demonstrated to significantly reduce the long-term sequelae of diabetes in this landmark trial. A subsequent analysis therefore regressed EQ-5D scores against the long-term clinical outcomes that formed the primary end point of the UKPDS. As expected, the results showed very clearly that these long-term outcomes had a significant detrimental effect on HRQOL as measured by the EQ-5D.

The UKPDS example is compelling. The “story” told by the model, that treatment impacts the long-term risk of events and that those events impact HRQOL, is confirmed by the analysis of the data. Yet with all the “noise” in the observed data, 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 those experiencing an event could have had a lower starting health utility than those who did not experience an event, and this was shown in an analysis of the extended follow-up data from the UKPDS [44]. However, in contrast to the UKPDS, most studies collect health utility at randomization and at intervals throughout the study. With longitudinal data, more sophisticated analyses are possible that control for baseline health utility and that are also able to examine the effect on utility over time. An example is the HSU analysis conducted on the Evaluation Of Cinacalcet HCl Therapy to Lower CardioVascular Events study [60]. Figure 2 shows the results in terms of the immediate (short-term) and prolonged (long-term) effect of clinical events on utility, based on longitudinal data analysis after controlling for baseline utility. It is clear from the figure, with associated confidence intervals, that the impact of clinical events on utility is highly significant. However, a traditional by-arm comparison failed to show a significant difference. In contrast, in the statistical model, after controlling for the impact of clinical events, a small but significant treatment effect on utility was found [60]. The lesson is that the approach to analyzing data for reimbursement outcomes should reflect the needs of the economic analysis and should not be constrained by the traditional approach to analyzing data for regulatory purposes.

### Statistical Considerations

When examining health-utility data, it is clear that they are often subject to left-skewness and can often have a point probability mass at 1. This is because 1 represents full health and generates a ceiling effect in that values above 1 are not possible. In contrast, states valued as worse than death are allowable, and therefore there is no theoretical lower limit for the utility scale (although there is a practical lower limit according to the valuation protocols of time trade-off and standard gamble of ~1.0). From a statistical analysis perspective, left-skewed data are more problematic than right-skewed data [61]. However, a very simple solution exists. A simple linear transformation of $X = 1 - U$ facilitates a move to the “health-utility decrement” scale. Health-utility decrements are right-skewed with a possible point probability mass at 0. The key is that, being right-skewed, health-utility decrements are similarly constrained as cost data, and all the usual methods analysts will be familiar with in analyzing right-skewed cost data can be used, for example, generalized linear modeling [62] or generalized estimating equations for longitudinal data [60,63,64], with the possibility of a two-part model to handle excess zeros [65]. Although utility data can exhibit multimodality, this is likely as a result of the nature of the disease state and or clinical events under study, such that using events as explanatory variables in a regression context (as in the example above) will help explain the shape of the distribution. Having fitted a model to the health-utility decrement scale, returning to the original health-utility scale after estimation is trivial: because the original transformation was linear, the back transformation of $U = 1 - X$ does not impact any of the estimated quantities, such as standard errors. Alternatively, transformation of the data on the health-utility decrement scale is possible but would require a smearing correction [66] when returning back to the original scale.

---

**Figure 2** - Utility decrements estimated by regression analysis of longitudinal data EQ-5D data collected alongside the EVOLVE trial. EVOLVE, Evaluation Of Cinacalcet HCl Therapy to Lower CardioVascular Events; HF, heart failure; HUA, hospitalization for unstable angina; MI, myocardial infarction; PVD, peripheral vascular disease.
Modeling Heterogeneity and Increasing Generalizability

It has long been recognized by statisticians that statistical modeling of the effect of prognostic variables on the outcome of interest in clinical trials can improve the precision of estimated treatment effects, even where randomization is used to control for unobserved heterogeneity [67,68]. In nonrandomized clinical studies, the ability to control for observed prognostic variables is critical for the estimation of treatment effects. Furthermore, modeling health utility as a function of clinical end points offers the potential to describe a causal pathway of how treatment affects intermediate end points, which, in turn, affect health utility. It is well known that clinical trials trade external validity for internal validity. However, the explicit modeling of prognostic factors offers the potential at least to increase the generalizability of the results by offering a method of adjusting clinical trial data to the characteristics of real-world populations.

Transferability in Multinational Studies

In multinational studies, analyses of data collected via PBMs should use the country tariff that is most appropriate for the economic model’s audience or jurisdiction. A particular form of heterogeneity relates to the multinational aspect of many clinical studies. For example, a recent analysis of the Action in Diabetes and Vascular Disease: Preterax and Diamicron MR Controlled Evaluation trial, which randomized more than 10,000 patients across 20 countries, showed substantial variation across regions in reporting on functional health problems. This variation could not be explained by differences in demographic variables, clinical risk factors, or rates of complications [69]. Although regional variation appears to affect the reported level of quality of life, a recent analysis of the same study [70] suggested that changes in EQ-SD tariffs associated with disease progression in diabetes are not significantly different across regions or a wide variety of other patient characteristics. Hence, tests for heterogeneity in reporting of both levels and changes should be undertaken, and when country- or region-level effects manifest, these can be handled using a covariate in a standard statistical modeling approach [71].

Reporting to Maximize Value for Future Economic Models

When reporting the analysis of health-utility data, care should be taken to always report variance and covariance estimates from the statistical model. Alongside point estimates, these data will be important for other analysts seeking to apply the results of the reported health-utility analysis to their own economic models and model settings, including the appropriate assessment of uncertainty. In particular, when health utility is regressed against ordered levels of health states representing disease severity, it would be inconsistent to have higher utility estimates placed on more severe health states. Inclusion of covariance information can be crucial to ensure the integrity of the ordering of health-utility estimates under conditions of uncertainty. Alternatively, consideration should be given to specifying a functional form for the modeling that maintains the logical ordering of HSUs, perhaps by parameterizing the more severe health states in terms of how much worse they are than a less severe health state and choosing appropriate distributional assumptions.

It should be recognized that this covariance information relates to the covariance between model parameters. For some applications of individual simulation models, it may be important to also have estimates of the level of variation due to the individual variation. In a standard regression model, this is given by the error term and in a generalized linear modeling or generalized estimating equations framework, this is given by the residual deviance. Consideration also should be given to reporting these measures, especially when it is known that the economic model will be designed as an individual simulation model (Table 7).

Conclusions

The quality of HSU estimates is critical to the decisions being made by HTA, pricing, and reimbursement authorities that affect patients’ access to new treatments, physicians’ ability to use them, and the return that manufacturers are able to realize on their investment in developing new products. Health economists with an understanding of the economic model should be influential in the design of healthy utility data collection in clinical studies, to ensure that methods are optimized to meet the needs of economic models.

Careful planning is needed, beginning early in the product development process, to define the HSU estimates that will be needed for the economic model, determine which of these may be measured in the planned trial or trials, establish the optimal design for the data collection and analyses, and plan any appropriate additional health-utility research.

The design of health-utility studies should start with a clear statement of the objectives, framed in terms of the needs of the economic model. The following aspects should be carefully considered: the choice of instrument and respondents, the mode of administration, the timing and frequency of health-utility assessments, the additional data that should be collected at each assessment, the follow-up period, and any issues related to heterogeneity of the sample and generalizability of results. Analyses should be designed to provide HSU estimates for model health states, making any appropriate adjustments to generalize the results to the population of interest in the economic model and to preserve valuable information available in the patient-level data.

Acknowledgments

The individual contribution by Jennifer Petrillo, Pierre Lévy, Sheri L. Pohar, Allan Wailoo, Charlie Nicholls, and Florence Joulin are gratefully acknowledged.

We thank the reviewers who commented during our forums at the ISPOR Philadelphia International Meeting and the ISPOR Amsterdam European Congress. We especially thank the following individuals who reviewed drafts of the report and submitted written comments. Their feedback has both improved the manuscript and made it an expert consensus ISPOR task force report.


Finally, many thanks to Theresa Tesoro for her assistance in developing this task force report.

References


