Formatiert

eunethta

EUROPEAN NETWORK FOR HEALTH TECHNOLOGY ASSESSMENT

GUIDELINE

Endpoints used for Relative Effectiveness Assessment:

HEALTH-RELATED QUALITY OF LIFE and UTILITY MEASURES

Adapted version (2015)

based on

"Endpoints used for Relative Effectiveness Assessment:

HEALTH-RELATED QUALITY OF LIFE and UTILITY MEASURES" - February 2013

The primary objective of EUnetHTA JA1 WP5 methodology guidelines was to focus on methodological challenges that are encountered by HTA assessors while performing a rapid relative effectiveness assessment of pharmaceuticals.

This guideline has been elaborated during JA 1 by experts from KCE, reviewed and validated by HAS and all members of WP5 of the EUnetHTA network; the whole process was coordinated by HAS.

During Joint Action 2 the wording in this document has been revised by WP7 in order to extend the scope of the text and recommendations from pharmaceuticals only to the assessment of all health technologies. Content and recommendations remained unchanged.

This guideline represents a consolidated view of non-binding recommendations of EUnetHTA network members and in no case an official opinion of the participating institutions or individuals.

Gelöscht: is

Gelöscht: draft

Gelöscht: As such t

Gelöscht: e

This guideline gives general recommendations related to HRQoL that are applicable to all types of REA irrespective of their particular purpose. A specific addendum related to study design issues and interpretation of HRQoL in the context of assessment of an added therapeutic benefit might be elaborated in future if decided by EUnetHTA partners.

Table of contents

Acronyms – Abbreviations

AQoL	Australian Assessment of Quality of Life Instrument
CUA	Cost-utility analysis
EQ-5D	EuroQol 5 dimensions
HAS	Haute Authorité de Santé (FR)
HRQoL	Health-Related Quality of Life
HTA	Health Technology Assessment
HUI	Health Utilities Index
HYE	Healthy Years Equivalent
ICER	Incremental Cost-Effectiveness Ratio
IVR	Interactive Voice Response
KCE	Belgian Health Care Knowledge Centre / Federaal Kenniscentrum voor de Gezondheidszorg / Centre fédéral d'expertise des soins de santé
LYG	Life-Years Gained
MAUI	Multi-Attribute Utility Instrument
PRO	Patient-Reported Outcome
QALY	Quality Adjusted Life Year
QoL	Quality of Life
QWB	Quality of Well-Being
REA	Relative Effectiveness Assessment
SF-36	MOS Short Form 36
SG	Standard Gamble
TTO	Time Trade-off
VAS	Visual Analogue Scale
WHOQOL	World Health Organisation Quality of Life instrument
WP5	Work Package 5 of the EUnetHTA Joint Action on HTA 2010-2012

Summary and recommendations

SUMMARY

Health-related quality of life (HRQoL) is often considered to be an important endpoint of health care interventions. Because improvement in HRQoL is highly subjective, it does not necessarily correlate well with objectively measurable clinical benefits.

Different types of HRQoL measures exist. A first distinction can be made based on the content of the measures: generic measures cover dimensions that are considered important for HRQoL in general, disease- or population specific measures particularly focus on dimensions that are affected by a specific disease or population. Disease-specific measures are generally considered to be more sensitive to small changes in HRQoL than generic measures. Generic measures, on the other hand, are more comprehensive and therefore likely to pick up unexpected effects on HRQoL which are not measured by disease-specific instruments.

A second distinction can be made based on the result of the measurements: there are measures that result in a HRQoL profile, with separate scores per item or dimension of HRQoL, and measures that give an overall summary score as a result. The latter encompass the utility indices, which are typically used to calculate endpoints combining HRQoL outcomes with life expectancy outcomes, such as QALYs. Also for HRQoL profile measures a summary score is often calculated.

This guidance indicates which types of HRQoL measures are suitable for the demonstration of the relative effectiveness of health technologies and summarizes the caveats for interpreting HRQoL outcomes. It has a double purpose: (1) support assessors in identifying the strengths and weaknesses in the evidence provided and (2) inform researchers about the requirements regarding HRQoL assessment in order to allow them to anticipate the collection of the required data for REA when developing trial protocols.

Gelöscht: care interventions

Gelöscht: pharmaceuticals

1.1. FINDINGS

The main message of this guidance is that the appropriateness of the HRQoL measure used depends on the purpose of the relative effectiveness assessment (REA):

- Is the purpose of REA to inform patients and health care professionals about the HRQoL benefit of an intervention as compared to its comparator or is the purpose to inform health care policy makers about the relative value of a health technology? The level of decision making is clearly different, and therefore, different needs can be identified.

Gelöscht: product

In the latter case, the decision-making context also plays a crucial role:

- is cost-effectiveness taken into account in reimbursement decisions and

Gelöscht: drug

- are decisions taken within indications only or also comparing relative values across indications?

Given these variations in context, there is not always a consensus on the required HRQoL evidence. However, as these requirements are usually not mutually exclusive, a set of basic requirements applicable to all contexts could be identified. Where this is not the case, possible variations are mentioned and discussed.

1.2. REA TO INFORM REIMBURSEMENT DECISIONS

For the purpose of informing health care policy decisions with respect to resource allocation across indications, it is important to include a generic HRQoL measure in the REA allowing to make comparisons across indications and intervention types. Disease-specific measures are useful as complements in specific cases, for instance when no improvement on one of the generic HRQoL dimensions can be demonstrated but there possibly are improvements on disease-specific dimensions. In countries where cost-effectiveness of interventions is a consideration in the decision making process, it is moreover recommended to use a utility measure in the REA. It is recommended that utility values derived from the general public or patients and associated with an easily-administered generic descriptive HRQoL instrument are used in order to ensure consistency in the utility values used for REA and for cost-effectiveness analysis, and to ensure interpersonal comparability of HRQoL scores. However, some countries only take decisions with respect to resource allocation within the (licensed) indications, with or without the consideration of the relative cost-effectiveness of interventions. In this case, disease-specific HRQoL instruments can be considered sufficient because comparability across indications is not required. It should be noted, however, that even in this context, generic HRQoL instruments are useful to allow building up a reference framework for the determination of the societal value of HRQoL benefits. The value of the health benefits of the different interventions needs to be determined in the same way as for reimbursement decisions considering multiple indications at the same time. The systematic use of generic HRQoL instruments in all indications allows increasing the consistency in this value judgement (or appraisal) process. For the sake of legitimacy of the decision making process, consistency between decisions is important and, in the apparent absence of consistency with previous decisions, justification is required. The justification for an apparently inconsistent decision can be based on disease-specific outcomes or other relevant decision criteria. It is therefore recommended to use both a generic utility and disease-specific instrument in both policy contexts.

1.3. REA TO INFORM CLINICAL DECISION MAKING

For the purpose of informing patients and health care professionals about the HRQoL outcomes of an intervention, the use of disease-specific HRQoL instruments is recommended. Professionals and patients are first and foremost interested in the dimensions of life that are affected by a disease and will be affected by an intervention. However, it should be noted that the risk of focussing on disease-specific measures is that they exclude dimensions of HRQoL that are generally not affected by the disease or standard intervention, but might be affected by the new intervention, e.g. through side-effects that were not present with the standard intervention. Therefore, it is important to verify whether all affected domains of HRQoL are covered by the disease-specific HRQoL instrument.

1.4. METHODOLOGICAL CONSIDERATIONS

All methodological considerations related to the psychometric properties of patient-reported outcome measures apply to HRQoL measures. Due to the absence of a gold standard for HRQoL measurement, it is often difficult, however, to measure these properties for HRQoL instruments or to give general guidelines on what can be considered a valid and reliable HRQoL instrument. Nevertheless, a number of basic principles can be defined. For instance, to be appropriate and valid for the purpose of

informing resource allocation decisions across indications, generic instruments should encompass all dimensions considered important by the society. Disease-specific instruments used for reimbursement decisions within one indication should not only encompass dimensions expected to be positively affected by an intervention but also the dimensions in which deterioration or no change is expected. In other words, the instrument needs to be comprehensive in the HRQoL domains covered.

A number of caveats related to repeated measurements, the cultural adaptation and translation, missing data, modes of administration and evaluation by proxies are discussed and a position on each of these issues is taken.

RECOMMENDATIONS

1. HRQoL instruments used in the context of REA should first and foremost be **valid for the purpose the REA intends to serve**.(paragraph 1.2) REA assessors should thus first consider for what purpose the REA will be used: to inform reimbursement decisions or to inform clinical decision making. The recommendations apply to both full REA and rapid REA.

2. A general recommendation applicable to all types of REA irrespective of their particular purpose, is to require the inclusion of a **disease- or population specific and a generic HRQoL measure** for most adequately capturing the impact of a disease on daily life. In case there is a need for the calculation of QALYs, a utility measure (Time Trade-Off or Standard Gamble) or generic HRQoL instrument associated with a reference set of utility values (generic utility instrument) is recommended.

    a. For countries that require an economic evaluation to support a health technology reimbursement application, it is recommended to require data emerging from the administration of a generic utility instrument in the clinical trial(s). Utility values should be derived from the general public (indirect utility measurement) or from patients (direct utility measurement). There is no consensus across jurisdictions about the most appropriate source. The choice between the sources of utility values is a normative one and should be based on careful consideration of the expected consequences for the decisions for which the HRQoL measurements are used, especially in case of decisions across indications. Consistency in the application of the chosen source is required. In both decision contexts, the use of other estimates for the HRQoL benefit in the REA than in the economic evaluation should be avoided. To improve comparability and consistency, countries might also consider recommending the use of one particular instrument for national reimbursement requests that is widely used (e.g. the EQ-5D).(paragraph 2.1.3)

    b. For countries that do not require an economic evaluation to support a health technology reimbursement decision, a disease-specific or generic HRQoL measure may be sufficient. Utility measures remain useful for REA in this context, however, especially for the calculation of QALYs, which are particularly useful for comparing interventions affecting both mortality and morbidity.

3. **REA performed for informing resource allocation decisions across indications** should primarily be based on HRQoL data obtained with a **generic HRQoL** instrument**,** encompassing all HRQoL dimensions in which improvements are considered important by the general public. If no improvement on such generic HRQoL instrument is observed, the alleged benefit of an intervention is less likely to be considered meaningful *from a societal point of view, given the range of existing health problems between which public resources need to be allocated*. REA should consider the effect of an intervention on the HRQoL of a typical real life patient population, taking the impact of patient's co-morbidities on HRQoL into account.(paragraph 2.1)

4. **REA performed for informing resource allocation decisions within indications** can be based on **validated comprehensive disease-specific HRQoL** data, as comparability across indications is in this case less important. Nevertheless, the consideration of **generic HRQoL** data remains

Gelöscht: product

Gelöscht: product

useful for reasons of coherence in the valuation of health benefits, and in consequence, transparency of the decision-making process.(paragraph 2.1.2)

5. **REA performed for the purpose of informing health care professionals and patients** could be based on **disease-specific HRQoL instruments.** They can be considered as complementary to generic instruments in REA performed for policy purposes. Disease-specific HRQoL instruments may be useful for more in-depth assessment of the generic HRQoL dimensions affected by an intervention. It should be borne in mind that the burden imposed on respondents increases with the number of questionnaires used.(paragraph 2.1.2)
6. HRQoL benefits of interventions should be demonstrated by means of **repeated measurements** in both the intervention and the control group.(paragraph 2.1.5.1)
7. **Single item scores** for HRQoL alone are considered insufficient to demonstrate relative effectiveness because they are subject to bias and often too crude to detect changes in health. Single item scores are scores derived from one single question asking to value current overall health on a specific scale.(paragraph 2.1.5)

8. **Mapping** of disease-specific or generic instruments to preference-based instruments to obtain utility values is generally not recommended for REA. Authorities should encourage researchers to always include a preference-based instrument in their clinical trial protocol in order to avoid the need for mapping. (paragraph 2.1.3.3)

9. Documentation of the **validity, reliability, responsiveness** and **acceptability** of the HRQoL instruments used in REA should be provided, taking into account the applied mode of administration and possible cultural and/or language adaptations. (paragraphs 2.1.4, 2.1.5.2 and 2.1.5.3)
10. Evaluation of HRQoL by "**proxy judges**" is not recommended. Its acceptance is limited only to cases where the patient cannot contribute him/herself or where the use of proxies can be justified by the nature of the judgements to be made.(paragraph 2.1.5.4)
11. **Transparent reporting within due time** of the results of all HRQoL measurements is recommended. If not (yet) published, it is required to make these results accessible for HTA bodies to allow critical appraisal.
12. When changes in survival and HRQoL are combined in one outcome measure such as the **QALY**, separate reporting of changes in survival and HRQoL and a description of the methods to combine the measurements should be requested to allow for separate consideration of both endpoints.(paragraph 2.2)

1. INTRODUCTION

1.1. DEFINITIONS AND GENERAL INFORMATION

Quality of life has been defined by WHO as "individuals' perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations and concerns." This is a general definition, referring not only to quality of life related to individuals' health status but to life in general. Overall quality of life is affected by health, but also by income, environment and freedom.[1] Although income, freedom and environment may affect health, they are usually not the main focus of health policy measures.

The current guidelines relate to "health-related quality of life" (HRQoL), or quality of life related to factors that affect an individual's health.[2] It is considered that the primary aim of a health care system is to maintain or improve health and HRQoL of the population, rather than overall quality of life or well-being. There is little agreement in literature about what constitutes HRQoL,[3] even if the definition of HRQoL has as a common basis, being the definition of health given by the WHO. According to the definition of the WHO "health" is "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity".

HRQoL is a broad concept which can be defined as **a patient's general subjective perception of the effect of illness and intervention on physical, psychological and social aspects of daily life.**[4, 5] Multidimensionality is a key characteristic of HRQoL. Each domain (physical, psychological etc.) consists of several dimensions. *Physical functioning* refers to mobility, self-care, usual activities and other functional abilities. *Psychological health* includes elements like cognitive functioning, emotional distress and anxiety. Finally, *social health* refers to the quantity and quality of social contacts and interactions.[6] A single domain, e.g. physical functioning, is insufficient to cover HRQoL, even though it is an endpoint relevant to patients.

HRQoL assessment is important in the context of relative effectiveness assessment (REA) because objectively measurable clinical parameters such as mortality and some measures of morbidity (e.g. myocardial infarction) are felt to be insufficient to capture the full impact of an intervention from the patient's perspective. Objective clinical measures may correlate poorly to a patient's own feeling of wellness.[7] [8]In non fatal -but sometimes severe- diseases where an intervention does not increase survival, an improvement of HRQoL due to the intervention may be as important as the improvements in the efficacy endpoints (e.g. psoriasis, irritable bowel syndrome, asthma). Including HRQoL in clinical or epidemiological studies facilitates understanding patients' perspectives on what is gained or lost as a result of a disease or illness or a medical intervention. It can give insight into the balance between therapeutic benefits and adverse effects of an intervention from the perspective of patients.

HRQoL is a patient-reported outcome (PRO) and fits within the definition of patient-relevant outcomes (see guideline on "Endpoints used in REA: Clinical endpoints"). PRO is an umbrella term used to describe any outcome evaluated directly by the patient himself/herself, without interpretation by clinicians or others, and based on patients' perception of a disease and its intervention(s). HRQoL represents a specific type/subset of PROs, distinguished by its multi-dimensionality.

Gelöscht: of pharmaceuticals

Different types of instruments exist for the assessment of HRQoL (Table 1). Each type of instrument is used to collect information on patients' perceptions of their current health state.

This guideline distinguishes four major **objectives of HRQoL measurement**. These are independent of the specific HRQoL instrument used. One instrument can be used for several objectives (e.g. generic utility instruments). HRQoL can be measured for the purpose of

1. describing the health status of a population (epidemiology of HRQoL),

2. assessing the relative effectiveness of a health technology (REA), and/or
3. assessing the cost-utility of a health technology (CUA).
4. Informing clinical decision making

Gelöscht: product

Gelöscht: product (

Although the requirements for HRQoL measures may depend on the objective(s) they intend to serve, they are not necessarily conflicting. For example, a HRQoL instrument used for the calculation of quality adjusted life years in the context of a CUA, might also be useful for REA.

For the **description of a population's health status**, as for instance in national health surveys, descriptive HRQoL instruments are generally used. Descriptive HRQoL instruments are generally more comprehensive, encompassing more items than **evaluative** instruments used for REA. Several instruments (e.g. SF-36, EQ-5D) are used for both descriptive and evaluative purposes. The use of HRQoL instruments for descriptive purposes will not be considered further in this guideline because the focus is on HRQoL measurement in the context of REA. Evaluative instruments used for **REA** encompass a moderate number of items (using several processes for reducing the number of items during the psychometric validation) to reduce the burden of completion by patients and enhance the response rate. To be useful for **cost-utility analysis**, evaluative HRQoL instruments should moreover be associated with utility values for the health states that can be described with the instruments. When an evaluative instrument is used for CUA, utilities are measured indirectly. Alternatively, utilities can also be measured directly using specific utility measurement techniques that do not require the use of a descriptive or evaluative HRQoL questionnaire, e.g. by means of the Time Trade-Off (TTO) or Standard Gamble (SG).

In terms of their content a distinction can be made between:

- generic HRQoL instruments (e.g. EQ-5D, SF-6D, SF-36, WHOQOL)
- disease-specific HRQoL instruments (e.g. Asthma Quality of Life Questionnaire, the St Georges Respiratory Questionnaire) and
- population-specific HRQoL instruments (e.g. the Child Health Questionnaire).

In terms of the results, a distinction can be made between:

- HRQoL profile measures, giving a separate score for each of the health state dimensions included in the questionnaire and in some cases a summary score, and
- summary scores, giving one single score for overall HRQoL. Scores can be expressed on any type of scale and do not necessarily have cardinal or interval properties.[a]
- utility measures, -which can be regarded as a specific case of summary scores but are subject to additional conditions for the resulting scores. In contrast to summary scores, utilities are values on a 0 to 1 scale, where 0 is the value of death and 1 the value of perfect health (negative values are possible for health states considered worse than death). The scores must have interval properties, i.e. a change of 0.2 is twice as good as a change of 0.1.

For utilities, a further distinction is made based on the measurement technique. There are two possibilities:

[a] Cardinal properties imply that a score of 80 on a scale from 0 to 100 is twice as good as a score of 40. Interval properties imply that changes in scores have cardinal properties; e.g. a change from 60 to 80 is twice as good as a change from 40 to 50.

- direct utility measurement, referring to the use of specific techniques to value the utility of health states (time trade-off (TTO) or standard gamble(SG)) rather than instruments or questionnaires,
- indirect utility measurement, referring to the use of multidimensional HRQoL instruments or questionnaires to which utility values that have been collected previously can be connected. The instrument can, in principle, be generic or disease-specific but is in practice most often generic (e.g. the EQ-5D).

Some HRQoL instruments combine a health status profile with a summary score or with a utility score. An example is the Health Utilities Index (HUI), which combines individual dimension scores with an overall score.

Single item HRQoL questions, asking to value current overall health on a specific scale (e.g. a Visual Analogue Scale) without a descriptive system accompanying the score, are not considered valid HRQoL measures for REA or cost-utility analysis and are therefore not considered further in this guideline.

Table 1: Examples of instruments and outcomes of HRQoL measures

	Descriptive instrument for HRQoL		Evaluation of HRQoL without using the intermediary of descriptive instrument
Result	**Disease-specific**	**Generic**	
Profile	Kidney Disease Quality of Life Short Form, Liver Disease Symptom Index	Sickness Impact Profile, Nottingham Health Profile, SF-36	
Summary score	End-stage Renal Disease Symptom Checklist*	Functional Limitations Profile (total summary score derived from dimension scores)	SF-36 question about change in HRQoL over 1 year (not scaled)**
Utility index (as a special case of a summary score)		EQ-5D, Health Utilities Index, SF-6D (indirect valuation)	Valuation of patients' health state with Time Trade-Off, Standard Gamble or Visual Analogue Scale calibrated on 0 to 1 scale (direct valuation).

* combines profile measures with summary score

** Is a single item question, answered on an ordinal scale

1.2. CONTEXT

1.2.1. Problem statement

The number of studies on HRQoL has been growing continuously over the past two decades. A bibliographic study of HRQoL measurement (2002) found that numerous HRQoL measures were developed, evaluated and used, with little standardisation even within specialties.[9]

HRQoL is a broad concept that has an important inter-individual variability and different meaning to each person.[10] HRQoL is by definition subjective, i.e. different respondents with the same health status might value their HRQoL differently, influenced or not by societal ideas of what constitutes minimal or optimal human happiness and well-being.[11] In addition, there might be measurement issues related to HRQoL. For example, different instruments give different results, reducing the comparability of results across studies. Another example is the problem of HRQoL assessment in specific population groups (e.g. children or people with cognitive impairments). These issues need to be dealt with when considering the inclusion of HRQoL measures in a REA. Therefore, suitable study designs are needed to measure the effects of interventions on HRQoL. The credibility and usefulness of any intervention-related improvement in HRQoL may be jeopardised by the lack of standardisation in HRQoL measurements.[12]

HRQoL can be measured for different purposes. The choice of the HRQoL instrument (generic versus disease-specific, utility versus profile measure) used will depend on the objective of the measurement. For cost-utility analyses, for instance, a utility measure is needed. For informing patients or clinicians, disease-specific HRQoL measures may be preferred over generic measures because they might capture better the specific impact of the disease and its intervention. For the REA of interventions that affect both HRQoL and survival, assessors might want to combine both outcomes into a single outcome measure,

such as quality-adjusted life years (QALYs) or Healthy Years Equivalents (HYEs). This allows comparisons between interventions with diverging results on HRQoL and survival respectively and may help decision makers in comparing the relative value of interventions.[13] QALYs are also frequently used as the outcome measure in cost-utility analyses, but may also be useful for REA.

1.2.2. Discussion (on the problem statement)

As explained in the 'background review paper of WP5'[14], the purpose of REA is "to inform health care professionals, patients and decision makers about the therapeutic added value of an intervention compared to already existing interventions". HRQoL assessment can be part of REA. Because the use of REA in reimbursement decisions differs between countries, the development of common guidelines for HRQoL assessment in REA is challenging. Three major reimbursement system/process characteristics determine the requirements for HRQoL measures in the context of REA:

- whether decision makers have to consider resource allocation across indications or only within indications,
- whether or not the relative cost-effectiveness of health technologies is considered during the reimbursement decision making process,[15]
- whether or not the REA must serve at the same time the decision makers and the professionals and patients.

When decision makers take resource allocation across indications into account (when deciding on the reimbursement of for example a particular pharmaceutical), comparability of the HRQoL measure across indications is important. It is less important when resource allocation decisions are made within the same indication.

When a cost-effectiveness analysis is required in the context of an application for a reimbursement decision, the HRQoL measure should allow the calculation of QALYs. The relevance of HRQoL measurement for economic evaluation does not preclude its relevance for REA. A synergy between both objectives may be found.

When a REA must serve at the same time the decision makers and the professionals and patients, it should include the relevant information for each of them. It can be argued that the same information should guide both clinical and reimbursement decisions. Although this is generally true, clinicians faced with an individual patient and different intervention strategies may still want more specific information on the HRQoL dimensions affected by a disease or its intervention.

In summary, the purpose of the REA and the policy context determine the best practice guidelines for HRQoL measurement in the context of REA.

1.3. SCOPE/OBJECTIVE(S) OF THE GUIDELINE

Guidelines on the way HRQoL should be assessed are needed to ensure that HRQoL measurements are relevant and useful for the REA of interventions in the context of health technology assessment. This guideline encompasses HRQoL measures used for assessing HRQoL as one of the **patient-reported outcomes** of an intervention targeting morbidity reduction as well as HRQoL measures used in combination with life expectancy. As such, it relates mainly to the use of HRQoL measures in clinical trials. The perspective taken is that of the assessor of the relative effectiveness of an intervention in the context of a reimbursement request.

HRQoL is also one of the **patient-relevant endpoints** (defined as morbidity, mortality, and HRQoL); detailed discussion on patient-relevant endpoints is out of scope of this guideline. For more information, please refer to the EUnetHTA guideline on "Endpoints used in REA: Clinical endpoints".

Gelöscht: drug

Gelöscht: drug

Gelöscht: care interventions

Gelöscht: pharmaceuticals

Gelöscht: of a pharmaceutical

Gelöscht: of pharmaceuticals

The guideline does not relate to the development of HRQoL instruments, nor to the use of HRQoL measures in clinical practice, for case-mix adjustments in the financing of health care services or for the assessment of the health status of the general population. Also measurement of specific dimensions of morbidity only (e.g. pain), without the measurement of their influence on physical, psychological or social functioning, falls outside the scope of this guideline. Finally, this guideline does not provide specific guidance on which disease-specific instruments are preferred for specific diseases.

The guideline for HRQoL measurement in the context of REA is formulated based on a literature review – done for the elaboration of this guideline during JA 1 – addressing the following questions:

- Which types of HRQoL measures are relevant for REA and what are their pros and cons?
- What are potential issues with HRQoL data that should be considered in a REA?
- What do existing guidelines say about HRQoL measurement in the context of a reimbursement request?

It has a double purpose: (1) support assessors in identifying the strengths and weaknesses in the evidence provided and (2) inform researchers about the requirements regarding HRQoL assessment in order to allow them to anticipate the collection of the required data for REA when developing trial protocols.

1.4. RELATED DOCUMENTS

This document should be read in conjunction with the following documents:

- EUnetHTA guideline: Endpoints used for REA: Clinical endpoints
- Methodological guidelines related to the assessment of patient-reported outcomes in the context of marketing authorisation applications:
    - U.S. Department of Health and Human Services, FDA Center for Drug Evaluation and Research, U.S. Department of Health and Human Services, FDA Center for Biologics Evaluation and Research, U.S. Department of Health and Human Services, FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. 2009. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf
    - European Medicines Agency – Committee for Medicinal Products for Human Use. Reflection Paper on the Regulatory Guidance for the Use of Health-related Quality of Life (HRQL) Measures in the Evaluation of Medicinal Products. London, 2005. EMEA/CHMP/EWP/139391/2004. www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003637.pdf

Gelöscht: of pharmaceuticals

2. SUMMARY OF THE ANALYSED LITERATURE[b]

2.1. HEALTH-RELATED QUALITY OF LIFE MEASUREMENT FOR REA

2.1.1. Profiles and summary measures

There is a diverse range of HRQoL instruments available, most often reflecting differences in objectives and focus. While some instruments focus on one particular dimension of HRQoL (e.g. physical functioning) others aim to measure HRQoL as a whole.[16] The common denominator of all instruments is that they address some aspect of the patient's subjective experience of health and the consequences of illness or its intervention.[3]

Descriptive measures for HRQoL usually consist of different items (questions), grouped into dimensions (e.g. physical functioning, cognitive functioning, anxiety and distress).[1] Each of the items is scored by the respondent. Scoring systems may use a binary scale (e.g. yes/no), an ordinal scale (e.g. 7-point Likert scale) or a continuous scale (e.g. a Visual Analogue Scale, VAS). A dimension usually consists of several items. Scores on separate items within a dimension are sometimes combined as a weighted or unweighted sum to create dimension-specific global measures. Descriptive measures for HRQoL find their theoretical basis in psychometric theory.

The separate items and dimensions can be considered HRQoL endpoints in themselves if they have been fully developed and validated. The study protocol should then ideally have specified that a given dimension will be the main focus of the HRQoL analyses and have taken this into account when determining the power of the trial. The statistical analysis will in that case mainly be descriptive. Interpretation problems may arise when an intervention performs better than its comparator on one item (or dimension) but worse on another. In REA the relative importance of the different HRQoL domains needs to be determined in order to draw conclusions with respect to the 'net' relative effectiveness of an intervention. This weighting –currently mostly implicit- might be controversial and subject to discussion.[2, 15] If the global score is calculated by simply summing the scores on all items in all dimensions, a dimension containing more items will get a relatively higher weight. If the global score is calculated as the mean of dimension scores, all the dimensions have the same weight. Relative weights may also be determined based on observational data, using scores from a factor analysis, summary item scores collected alongside separate item responses[17] or utility values obtained from a utility instrument administered alongside the descriptive measures.[2] A regression of summary or utility scores on individual item responses can reveal the relative weight of the individual items.

More frequently, researchers opt for one summary item, generating a single score for HRQoL, leaving the implicit weighing of the different dimensions to the individual patient. Although there is evidence for validity and reproducibility of summary items, such items do not allow the identification of opposing trends in different dimensions of health.[3] Especially if the response categories of simple summary items are limited to a few items, the response categories may be too crude to detect subtle but important changes in health.[18]

2.1.2. Generic and disease-specific instruments

HRQoL instruments can be generic or specific for a disease or population. Generic HRQoL instruments measure general aspects and are applicable to a broad range of indications, whereas specific instruments are only applicable to a specific indication, population or intervention. Table 2 briefly summarizes the

[b] Details on the original literature search conducted during JA1 and inclusion/exclusion criteria are presented in Appendix 1 and Appendix 2.

advantages and disadvantages of disease-specific and generic instruments, mainly based on a review by Fitzpatrick et al.[3].

Table 2: Advantages and disadvantages of disease-specific and generic HRQoL instruments

	Advantages	Disadvantages
Disease-specific instruments	Are expected to: - have relevant content - be more likely to detect important changes that occur over time in the disease studied - be more acceptable to patients and thus have a higher response rate	It is impossible to: - administer disease-specific instruments on samples who do not have the disease - make comparisons with HRQoL outcomes of interventions in other disease areas - may fail to capture unexpected change in HRQoL not addressed by the instrument[19]
Generic instruments	- Useful for broad range of health problems - Enables comparisons across interventions for patients with different diseases - May detect unexpected positive or negative effects of an intervention - Reduced patient burden if generic instrument replaces (battery of) disease-specific instruments - Potential to enhance the value and interpretability of HRQoL outcomes if trials generally converged on the use of (a) generic HRQoL instrument(s)	- Less detail in terms of relevance to specific illnesses - May sometimes be less sensitive to changes due to an intervention

Source: adapted from Fitzpatrick et al.[3]

Several authors have argued that the choice between using a generic or a disease-specific instrument depends on the purpose of the study and the future use of the data resulting from the study.[3, 20] In contrast to disease-specific instruments, generic HRQoL instruments allow comparisons of outcomes across a range of diseases. Therefore, they are generally considered to be of greater interest to policy makers having to allocate health care resources across different disease areas. However, even in situations where reimbursement decisions are made on a case-by-case basis and where policy makers assess resource allocations within one particular indication rather than across indications, generic HRQoL instruments are useful. In both contexts, the value of the health benefits of the different interventions needs to be determined. This value is by definition always relative and depending on several parameters. The systematic use of generic HRQoL instruments in all indications allows increasing the consistency in this value judgement (or appraisal) process. For the sake of legitimacy of the decision making process, coherence between decisions is important and, in the apparent absence of consistency with previous decisions, justification is required.[15] The justification for an apparently inconsistent decision can be based on disease-specific outcomes or other relevant decision criteria, such as the number of patients contributing to a specific increment in HRQoL.[15] The major argument in favour of the consideration of generic HRQoL measures in REA is

the potential for increasing the consistency between decisions across disease areas and hence transparency of the reimbursement system. Moreover, it should be borne in mind that disease-specific instruments might not capture unexpected changes in dimensions of HRQoL that are not included in the disease-specific instrument but are, nevertheless, important. Complementing disease-specific instruments with a generic instrument is therefore always useful.

Gelöscht: drug

Generic HRQoL instruments are believed to be less responsive than disease-specific instruments, although empirical evidence confirming this belief is often missing. However, it should be kept in mind that the goal is not to measure the largest difference in *specific* HRQoL outcomes, but rather to measure the impact on *general* HRQoL. Nevertheless, as demonstrated by the development of the EQ-5D-5L (EQ-5D with 5 levels of severity in each dimension instead of 4), researchers acknowledge this critique and are looking for ways to increase the responsiveness of generic HRQoL instruments, while maintaining their advantage of being generic.

Disease-specific instruments may be specific to a disease (e.g. asthma), a site or region (e.g. the hip), a population (e.g. the elderly), a certain function (e.g. sleep) or a HRQoL dimension (e.g. pain). They are considered to be of greatest interest to patients and clinicians.[1, 3] However, disease-specific HRQoL instruments are also useful in the context of REA for policy making purposes. They can allow to justify reimbursement decisions that are, apparently, inconsistent with previous decisions. For example, a positive reimbursement decision for a pharmaceutical with the same effect on HRQoL than non-reimbursed pharmaceutical for another indication, ceteris paribus, might be justified by the disease-specific HRQoL effects. It might happen that the effect on a disease-specific HRQoL questionnaire –although important from the patients' point of view- is insufficient to bring about an observable effect on any of the domains of a generic HRQoL questionnaire. In that case, the information provided by the disease-specific HRQoL instrument is relevant for the REA.

In clinical trials that study an intervention's efficacy, lack of specificity of HRQoL measures may sometimes be considered to be a problem: if a person has multiple co-morbidities, changes in overall HRQoL or absence of changes in HRQoL may be related to aspects that have nothing to do with the intervention.[21] This argument is often used to justify the use of a disease-specific HRQoL instrument. However for REA, aiming to assess the intervention's efficacy in real life, co-morbidities and their impact on HRQoL are extremely relevant, in very much the same way as all-cause mortality is more important than disease-specific mortality. Whenever a significant improvement is observed on a disease-specific HRQoL measure and no effect on a generic HRQoL measure, the assessor should critically evaluate whether the changes in disease-specific HRQoL do not inflate the estimated effect of an intervention.

For a policy-oriented REA, it is recommended that the results of generic HRQoL measures are at least considered in the assessment. The generic instrument should include all HRQoL dimensions on which improvement is considered meaningful from a societal point of view. Any improvement on such a generic HRQoL instrument can then be considered meaningful according to society. In other words, a generic HRQoL instrument should be valid for its purpose of informing health care policy. This validity requirement is not different from the validity requirement imposed upon disease-specific instruments for their purpose (informing practitioners or health care policy makers assessing efficiency within indications). A specific caveat applies to disease-specific HRQoL instruments. Assessors should assess whether all potentially relevant dimensions are actually included in the disease-specific instrument used to demonstrate relative effectiveness, not only the dimensions on which an improvement is expected following the intervention. This is important so that the effects of unexpected side-effects are covered by HRQoL outcomes.

Disease-specific instruments can be considered relevant in the following cases:

- To get additional information on HRQoL for registration purposes, e. g. during the assessment of the risk-benefit of a pharmaceutical.[20]
- More in-depth assessment of how life is affected precisely if an **effect** is observed **on a generic HRQoL** instrument. In this case disease-specific HRQoL measures provide complementary information to generic HRQoL measures.
- Assessment of the HRQoL benefits of a health technology compared to a relevant comparator for the same indication if **no effect on a generic HRQoL instrument** is found but a HRQoL benefit is nevertheless assumed.
- Assessment of HRQoL benefits of a health technology if **no** adequate **generic instrument**, reflecting the HRQoL dimensions on which change is considered meaningful according to the society, is **available**. As most commonly used generic HRQoL instruments have been assessed on this point, this might be considered to be a rather theoretical possibility.

Gelöscht: for registration purposes

Gelöscht: product

Gelöscht: product

2.1.3. Utility measures

Utility measures for HRQoL are measures that lead to a single score for HRQoL with specific properties. In contrast to the summary item scores or aggregate HRQoL scores obtained by summing weighted or unweighted item or dimension scores, utilities are obtained using preference-based or choice-based methods. Utilities could reflect either patients' preferences for specific health states or the general public's preferences for these states.[c] The major significance of utility measures is that, first, a single index directly elicits the individual's overall preference for a health state and, second, they provide a simpler figure for the analysis of the net health benefit of interventions, compared with the many outcomes produced by multi-dimensional HRQoL measures.[3] The theoretical background for utility measures comes from the field of economics and decision theory.

2.1.3.1. Methods for utility measurement

Instruments used for utility measurement are the **time trade-off** (TTO), the **standard gamble** (SG) and the **visual analogue scale** (VAS). The SG and TTO are choice-based methods, requiring respondents to make a choice between two hypothetical situations and deriving utility values for health states based on the choices made by the respondent. The VAS is a preference-based method, not involving a choice but asking to reveal the relative value of health states on a thermometer-like scale. Utilities are measured on a continuous 0 to 1 scale, where 0 is the value for death and 1 the value for perfect health. Negative scores are possible. Utility values can be used for the calculation of Quality Adjusted Life Years (QALYs), a frequently used outcome measure in cost-utility analyses.

2.1.3.2. Direct and indirect methods

There are two different methods for assessing patient utilities: the direct method and the indirect method.[3] The **direct method** implies the elicitation of utilities directly from patients who are in the health states of interest using a TTO, SG or VAS. Because utility scores obtained in this way do not provide information on which domains of HRQoL are affected by an intervention, this method is often used in combination with a descriptive generic or disease-specific[22] HRQoL instrument in which the patient describes his current health state. As such, descriptive information on the health state as well as the value of that health state for the patient is obtained. For example, the EQ-5D, a descriptive measure consisting of five HRQoL dimensions with three levels of severity in each dimension, can be used to describe a patient's health state, while the TTO can be used to derive the utility for this health state from the patient. The same health

c The terms 'preference' and 'utility' are frequently used as synonyms, although technically, 'utilities' are preferences obtained by methods that involve uncertainty (i.e. the standard gamble approach).

state on this descriptive system might be valued differently by different individuals, depending on the relative importance of each of the dimensions for these individuals. These interpersonal differences in preferences need to be taken into account when analysing and presenting HRQoL data. For example, comparing utility scores of individual patients over time make sense, while comparing utility scores, obtained through direct elicitation, of different people is less meaningful.

The **indirect method** involves the use of a descriptive generic HRQoL instrument, on which patients report their health states. The utility values subsequently attached to these health states come from prior survey data, in which utilities have been measured from appropriate samples of respondents. A well-known and frequently used instrument used for this approach is the EuroQol EQ-5D with the EuroQol "tariff". The "tariff" is a list of the utility values of every health state that can be described with the EQ-5D. The utility values are derived from the general public. The public values are derived based on a sample of the general public valuing hypothetical health states described by means of the EQ-5D. In clinical trials, patients simply complete the 5-dimensional EQ-5D to describe their health state. Subsequently, the corresponding utility value from the "tariff" is assigned to the patient's health state. Other examples of generic descriptive instruments for which utility values have been collected from the general public in some countries are the Quality of Well-Being Scale (QWB), the Health Utilities Index (HUI), the SF-6D (6 dimensions of the SF-36), the 15D and the Australian Assessment of Quality of Life Instrument AQoL. In all cases a multi-attribute utility function gives the utility value corresponding to each of the health states that can be described with the instrument. For some instruments, various tariffs exist. This is due to the fact that the selection of the multi-attribute utility function that will be used to generate the tariff values is not straightforward and is often a matter of choice. Results of the assessment may vary depending on the utility function applied. Moreover, different generic utility instruments may yield different results. It is therefore recommended to select one instrument with one tariff and apply this to all assessments to ensure consistency in the REA. The disadvantages of the indirect methods for assigning utilities to health states are the same as the disadvantages of the generic descriptive HRQoL instruments (see Table 2).

Gelöscht: O

2.1.3.3. Mapping

The indirect approach mostly uses a generic descriptive instrument, although the use of a disease-specific instrument is theoretically possible. However, utility values for disease-specific descriptive instruments are rarely available. Therefore, **mapping** of disease-specific data to a generic HRQoL measure is sometimes applied in order to assign utility values generated with the generic instrument to the disease-specific health state descriptions.[22] This approach is also used when generic HRQoL data have not been collected in a trial or a generic instrument was used in a trial for which no utility values exist.[22, 23] Utility data are required to assess the cost-utility of an intervention,[24] but because cost-utility assessment is usually not the primary purpose of a trial, generic HRQoL instruments allowing translation of health state descriptions to utility scores are often missing in the trial protocol. Mapping of disease-specific to generic or generic non-utility measures to generic utility measures always introduces uncertainty.[23, 25, 26]

Considering the arguments for mapping, it can be argued that mapping is actually a compensation for an imperfect trial protocol. It demonstrates the need for careful consideration of the purpose and future use of trial data when designing the protocol of a study. If HRQoL is expected to be an important outcome and trial data are meant to be informative for policy makers assessing the relative effectiveness of interventions, the appropriate HRQoL measures should be included in the study protocol.[23] As for clinical data, post-hoc solutions to solve the problem of uncollected HRQoL data such as mapping, should be scrutinised in REA, because such solutions are by definition inferior to adequate data collection.

2.1.3.4. *From whom to derive utility values*

The choice of the people from whom to derive utility is a crucial one. It is determined by the purpose of the data collection. The two main options are patients and the general public. In general, it seems most appropriate to use utility values from patients if the objective is to address a particular clinical question, while it seems more appropriate to use values from the general public if the objective is to inform resource allocation decisions (decisions across indications).[3]

In the context of *economic evaluations*, intended to inform the process of assessing the societal value of an intervention, health state utilities used have therefore usually been derived from the general public. This has also been proven to be a practical approach. However, there is no overall consensus about whether patients' utility values or the general public's values should be used. It is recognized that both approaches have their advantages and disadvantages.[16] Different countries may have different opinions. For decisions within indications, utility values from patients could be used for economic evaluations without any clear objection.

For *REA*, however, the choice is not straightforward. On the one hand, REA is concerned with patient-relevant outcomes and thus patient preferences are most relevant. On the other hand, REA is performed to serve a certain goal, being to make a decision about the appropriateness of reimbursement. In this case, utilities from the general public become more relevant. When using individual patient utility values, issues related to the analysis of patient preferences should not be neglected (e.g. problem of inter-personal comparability, making it less straightforward to assess baseline differences in health status utility between experimental and control group). This problem is avoided when using public utility values, because the same health state description cannot have different utility values. It goes without saying that the use of public utility values is not without problems either (e.g. lack of distributions in scores, giving a false image of preciseness of the utility values). The only fixed health states are death (value of 0) and perfect health (value of 1).

It is important to be clear about *how to handle the choice between patient utilities and utilities of the general public* in REA, in order to avoid confusion with the public. How to deal with a situation where the gain in utility is much higher if measured directly in patients than if measured indirectly, using preference values from the general public? The opposite might also happen: the gain in utility might be much higher if measured indirectly than if measured directly in patients because patients (especially those with chronic diseases) might have coped with their condition. It makes little sense to use a different health benefit estimate for the assessment of the relative effectiveness than for the assessment of the value for money. The distinction between decisions across and within indications becomes relevant again. The effect of coping with a disease might have important consequences for the outcomes of cost-utility analyses. If patients are coping well with their disease, their directly measured utility will be relatively high compared to the utility assigned to this health state by the general public. As a consequence, the potential utility gain will be lower when directly measured utilities are used than when indirectly measured utilities are used in cost-utility analyses. For decisions within an indication, this is not a problem, as utilities are measured in patients suffering from the same condition. For decisions across indications, the conclusion might be that the interventions targeted to these conditions with which the patients cope well are not cost-effective, because of the limited potential gain in utility. As coping is more frequently occurring in patients with chronic conditions than in patients with acute conditions, equity issues may arise. It is therefore a social choice whether a country prefers to use utility values directly measured in patients or utility values derived from the general public. Awareness of the pros and cons of each approach and the possible consequences when using them for decision making is crucial. Moreover, consistency in the applied approach is essential. Once it has been decided to use utilities derived directly from patients, this approach should be

applied to all evaluations. The feasibility of this choice, as compared to using a generic utility instrument, should be carefully considered.

2.1.4. Psychometric properties of HRQoL measures

The psychometric requirements for HRQoL measures are similar to those for other patient-relevant endpoints such as patient satisfaction. To be useful for relative effectiveness assessment, HRQoL measures must be valid, reliable, responsive and acceptable.[1, 27] An overview of these concepts, based on the framework used by the FDA for its guidelines on PROs,[5] is given in Appendix 3.

The problem is that the criteria of validity, reliability, responsiveness and acceptability are not consistently defined in the literature. It is therefore difficult, if not impossible, to make explicit statements about the extent to which a HRQoL measure used in a particular trial satisfies these criteria.[3] Nevertheless, it is reasonable to expect at least an indication for the performance of the HRQoL measure on each of these criteria in the REA. Several guidelines for assessing the psychometric properties of HRQoL instruments have been developed by international societies such as the ISOQOL[d] and ISPOR[e]. In addition, the FDA has published recommendations for assessing the psychometric properties of PROs.[5] We refer to these references for further information on this aspect.

2.1.5. Measurement issues

2.1.5.1. Repeated measurements

In the context of REA, one is interested in the change in HRQoL due to an intervention. This implies repeated measurement of HRQoL in groups of patients, at least before and after intervention and at crucial events (e.g. occurrence of serious side-effects or complications). The results should provide information concerning statistical differences within groups and among groups and rates of response for the HRQoL dimensions. For utility measures, used in a reimbursement decision context, this is less relevant, as the decision will in that case be based on the *societal value* of the demonstrated improvement in utility, reflected by how much society is willing to pay for the increase in utility.

Single item scores asking about changes in HRQoL compared to the past are prone to many biases and should not be used to draw conclusions about the relative effectiveness of interventions. It has been shown that individuals tend to recall poorer health states than actually experienced and that the degree of improvement tends to be exaggerated, and that respondents' answer may be influenced by their current health state when asked to compare current with past health states.[3, 28]

When dealing with longitudinal multidimensional HRQoL data, multilevel analysis can be applied. The first level relates to the analysis of the various HRQoL dimensions. The second level involves the analysis of the observations over time. The third level is the level of the individual patients.[29] The advantage of multilevel modelling is that it provides estimates of the intervention effect for each dimension separately as well as – if appropriate – an overall summary estimate and the corresponding test statistics.

Multilevel modelling should only be applied if it can be assumed that the missing data mechanism is 'ignorable', i.e. missing data are missing completely at random. This is often not the case in longitudinal HRQoL studies.[2] Often, missing data are informative, for instance, if data are missing due to drop-out as a consequence of illness or death. A more complete discussion on missing data is provided in paragraph 2.1.6.1.

[d] http://www.isoqol.org/
[e] http://www.ispor.org/

2.1.5.2. *Cultural and language barriers*

Questionnaires developed and tested in a specific language cannot simply be translated and supposed to have the same psychometric properties as in the language and country of origin.[30] Translated versions might be interpreted differently and cultural differences might adversely affect an instrument's measurement properties.[1] Therefore, translated HRQoL instruments should be properly validated and tested before use in clinical studies that aim to demonstrate improved relative effectiveness of an intervention.[31]

A literature review of methods to translate HRQoL questionnaires concluded that there is no empirical evidence in favour of one specific method for translating HRQoL instruments.[32] The translation and cultural adaptation process should cover an assessment of equivalence. Equivalence covers several concepts, e.g. conceptual, item, semantic, operational, measurement and functional equivalence.[33]

For REA it is important that the assessments of equivalence are documented when HRQoL data are derived with translated and adapted instruments. The assessor of the relative effectiveness should evaluate the methodological rigour of the translation and cultural adaptation process, as well as the psychometric properties of the translated and adapted version, if this has not been done before, e.g. at the level of EMA if new HRQoL data is presented for a pharmaceutical compared to the registration document. Several questionnaires have been translated, adapted and tested for cross-cultural applicability (e.g. SF-36[30, 34], EQ-5D[35]).

Gelöscht: done

Gelöscht: (e.g.

Gelöscht:)

2.1.5.3. *Modes of administration*

HRQoL data can be obtained by administering HRQoL instruments through different modes: interview, mailing, telephone or self-administration. It has been demonstrated that the mode of administration can have an impact of HRQoL scores (see, for example, a study by Lyons et al.[36]) The advantages and disadvantages of the different modes of administration from the perspective of the researcher have been described by Guyatt et al.[1], Jackowski et al.[7], Coons et al.[37] and Hacker[38]. Possible sources of bias related to the mode of administration, to be considered for the REA, are described in Table 3.

Table 3: Possible sources of bias related to modes of administration of HRQoL instruments

	Possible type of bias
Self-administration	Selection bias (non-response, exclusion of illiterate, less educated, other language) Respondent may misunderstand the questions Researcher may misunderstand the answers
Interview	Interviewer bias Reporting bias Characteristics of the interviewer (voice inflections, age, race, gender) may introduce bias
Telephone with live interviewer	Selection bias (only respondents with a telephone can be surveyed) Voice inflections of the interviewer may introduce bias
Mailing	Selection bias (non-response, exclusion of illiterate, less educated, other language) Respondent may misunderstand the question Researcher may misunderstand the answers
Telephone with interactive voice response	Selection bias
Computer-based technology	Selection bias possible, although it might also reduce selection bias, by applying easily accessible formats with touch screen and audio components.
Web/internet-based	Selection bias (only patients with access to the internet)

technology	

Today, interactive voice response (IVR), computer-based and web-based technology are increasingly used for collecting HRQoL data. Several HRQoL instruments initially developed for paper-and-pencil administration are now available in IVR version, e.g. the EQ-5D.[f38] The use of multimedia tools for the collection of HRQoL data has several advantages: it may reduce the number of missing data by notifying respondents of unanswered questions, allows for the application of features that help people with low literacy, allows the language to be chosen, can reach populations in a variety of geographic locations and enables 24 hours data collection.[38] Computer adaptive testing programmes for HRQoL assessment, where the type and order of the questions depends on the answers of the respondent to previous questions, are a recent development within HRQoL research.[39] The experience with this approach is still limited and therefore the relevance of data collected in this way for REA is still unclear.

It is recommended to document the psychometric properties of a HRQoL instrument, given the mode of administration. Because the mode of administration may have an impact on the psychometric properties of an instrument, these need to be re-evaluated whenever a different mode of administration is applied.[38, 40] For example, an interview-based standard gamble may yield different results from a paper-based standard gamble.[41] For electronic versions of paper HRQoL questionnaires, the measurement equivalence should be addressed using the appropriate techniques. The appropriateness of techniques for measurement equivalence testing depends on the magnitude of modifications to the content and format of the original paper version of the questionnaire required during the migration process.[37] Ideally, such measurement equivalence testing should have been performed before application of the electronic version of an originally paper questionnaire in a clinical trial that is performed to inform REA processes.

Some researchers suggest that, given the differences in responses depending on the mode of administration, mode-specific population norms should be established and used when HRQoL data from patients are compared to those of the general public as the standard population.[42]

2.1.5.4. Evaluation by patients versus proxies

It is recommended that HRQoL, as a patient reported outcome, be assessed by patients themselves (self-report). The use of proxies, such as caregivers or family, should be avoided where possible. However, the use of proxies for the measurement of HRQoL is unavoidable in some cases, e.g. cognitively impaired patients, small children. Sometimes patients are too ill to complete HRQoL questionnaires. Recording this as missing data would potentially bias the results. Using proxy judges may be an option to this. However, evaluators should be aware that the correspondence between patient and proxy response to HRQoL measures varies depending on the domain assessed and the choice of the proxy. A review of empirical studies concluded that proxy responses on more observable domains, such as physical functioning and cognition, are generally more highly correlated with responses from patients, whereas proxies tend to overestimate patients' functional limitations (proxies tend to overestimate patient dysfunction relative to the patients themselves).[1] Medical professionals may be inclined to focus on the limitations a particular functional impairment presents, whereas patients may emphasise the possibilities still left to them.[43]

Because of the demonstrated lack of agreement between patients-reported and proxy-reported HRQoL, proxy valuation is generally discouraged and accepted only if the patient cannot contribute him/herself.[3] There might be scope for proxy judgements of HRQoL if the reason for patients not contributing is ill health *and*

[f] http://www.euroqol.org/eq-5d/eq-5d-products/eq-5d-3l-translations/alternative-modes.html

the judgements are relatively simple. For instance, it might be acceptable to let proxies fill out a simple generic utility instrument and subsequently assign public utility values to these health states if patients are not able to fill out the descriptive questionnaire themselves. It is less evident to assume that a multidimensional HRQoL questionnaire, where items are to be valued on a VAS reflecting patient's feelings, filled out by a proxy is valid and hence useful for REA. By using public utility values for the health states, the influence of differences in subjective perceptions about HRQoL between the proxy and the patient is limited to potential differences in the health state descriptions. Because the proxy judges do not *value* the health state of the patients but only *describe* the patients' health state, differences in preferences do not influence the judgements. In principle, it is recommended not to use proxy data if important differences between patient and proxy assessment are possible.

When designing a study protocol, researchers should carefully consider the possibility of non-response due to ill health and possible solutions to this risk. For example, alternative modes of administration might be considered for HRQoL instrument(s) if there is a risk of a higher drop-out rate with self-completed paper questionnaires due to ill health.

In this context, it should be noted that low literacy or visual impairment does not justify the use of proxy judges or the exclusion of these patients from the study. It would lead either to bias in the results or lack of representativeness of the sample, which are both important considerations in REA. In these cases, the administration mode for the HRQoL instruments could be changed to allow patients to participate in the study (e.g. interview or self-administrated multimedia survey[40] instead of self-completed paper survey).

Reasons for using proxy judges should always be justified.

2.1.6. Data analysis issues specific for HRQoL

2.1.6.1. Missing data

Missing data in a longitudinal HRQoL study induce similar problems as in other types of longitudinal studies. However, there are some specific issues with missing data in multi-dimensional and multi-item HRQoL instruments. Types of missing data are item non-response (responses on some items are missing) and unit non-response (the whole questionnaire is missing). Unit non-response can be due to patient drop-out from the study, intermittent missing questionnaires and late entry into the study.[44] The major undesirable effects of both types of missing data are the introduction of bias due to inadequate modes of analysis and the loss of efficiency due to reduced sample sizes (loss of power) and, as a consequence, the diminished ability to draw useful conclusions from the study.[45] The particular problem with item non-response in HRQoL studies is that they can drastically reduce the number of patients for analyses which assume availability of complete patient data. This may be a major issue where repeated HRQoL measurements are required to demonstrate relative effectiveness.

The assessor should be able to evaluate whether the researchers have done everything to minimize avoidable missing data: e.g. by maintaining confidentiality at all times, clearly describing the purpose of the assessment to patients, planning for sufficient time in good material conditions, explaining that there are no incorrect answers. It is recommended to follow the HRQoL-instrument-specific guidelines for handling missing data if these are available. The assessors should be able to access the reference to these instrument-specific guidelines.

At the level of the statistical analysis, it is essential to distinguish between data missing completely at random, data missing at random and data missing not at random.[44, 46, 47] There is a difference between data missing completely at random and data missing at random. A HRQoL questionnaire is missing completely at random if the probability of having a missing questionnaire is independent of scores on previous observed questionnaires and independent of the current and

future scores had they been observed. It means that the reason for missingness must be independent of the patient's HRQoL. For missingness at random, the probability of having a missing questionnaire may depend on previous scores but must be independent of the current and future scores, i.e. current HRQol should not be the reason for the missingness, although previous poor HRQoL may have an impact on the likelihood of missingness at the current assessment.[44] Theoretically, it can be tested whether data are missing completely at random or at random.[46] However, this is not trivial and relies on fundamentally un-testable assumptions.[44] Informative drop-out – i.e. patient drop-out due to ill health or death – should be recorded as such, as it is not random. Several approaches can be used for adjusting for informative dropout (e.g. generalized linear mixed models, conditional linear models).[48]

In general, it can be recommended that missing data should be avoided and if not, they should be replaced in the analysis with a value derived from hypotheses about the HRQoL of patients with missing data.

Missing data on single items in a HRQoL study mainly cause problems in the calculation of global scores. Values for items missing in the dataset need to be imputed or the calculation of global scores needs to accommodate them.[2, 45] Values can be imputed if the number of items on which data are missing is limited, e.g. from values of the other items in the patient's HRQoL questionnaire, from the patient's values of the other items within the same dimension or from the patient's values of the item on t-1 and t+1.[2, 45] Alternatively, the global score could be expressed as a percentage of the maximum achievable score over all completed items. A final option is to record the global score as missing if one of the item scores is missing.[45] The latter approach should always be applied if the item non-reponse is non-random.

2.1.6.2. Multiple testing

Due to the multi-dimensional nature of many descriptive HRQoL instruments, the problem of multiple testing may arise if a hypothesis is formulated for the outcome of each of the dimensions included in the instrument.[2] The multiple testing problem refers to the increasing probability of finding a false-positive result as the number of tests increases. Suggested ways to deal with this problem are:

1. to limit the number of hypotheses tested (i.e. specify a priori the dimensions of particular interest, which serve as the basis for the principal analysis on HRQoL) and analyse the remaining variables descriptively,
2. to combine dimension scores to create a summary score (if this is provided for in the HRQoL questionnaire),
3. to combine results of univariate tests on multiple outcomes (e.g. Bonferroni correction or other methods), and
4. to apply a hierarchical approach.[2, 5]

The hierarchical approach has been recommended by the FDA. It implies a clear definition of the hierarchy of endpoints and relationships between them in the study protocol and sequential testing, i.e. testing the secondary endpoints only after success on the primary endpoint.[5] Multilevel modelling is also an example of the hierarchical approach for longitudinal multi-dimensional HRQoL data.[2]

2.1.7. Presentation of the results of HRQoL studies

Results of HRQoL studies can be presented in various ways, depending on the type of instrument (descriptive multi-dimensional or generating a summary measure for HRQoL) and the design of the study (longitudinal or cross-sectional). Specific guidelines for reporting results of quality of life assessments in clinical trials have been published.[19, 49, 50]

Important to retain is that means and medians are not meaningful in the case of ordinal HRQoL scales. For ordinal or binary data, proportions of patients with a specific score should be used.

For longitudinal studies using global measures for HRQoL, individual data on each time point could be plotted. The individual summary measures could then also be stratified by survival time to reveal a consistent pattern across patients.[18] This approach may, however, not be practical for large studies.[2]

For groups of patients, the mean and the median of a summary measure for HRQoL can be presented or the proportion of patients with a certain level of HRQoL (in both cases with their appropriate confidence interval), depending on the properties of the scores obtained with the HRQoL instrument. The same can be done for separate domains of HRQoL, to see whether the overall pattern of response to the various dimensions differs between treatments.[51]

2.2. MEASURES COMBINING HEALTH-RELATED QUALITY OF LIFE AND LIFE EXPECTANCY[19, 52]

In health care, it is important to consider not only HRQoL (morbidity) but also life years (mortality). An intervention may reduce mortality at the expense of a worse HRQoL (e.g. intensive end-of-life treatment) or vice versa (e.g. an invasive surgical procedure with a high operative mortality but good HRQoL for survivors). Combining these two dimensions results in a combined outcome measure, that is the number of life years adjusted for HRQoL. The most frequently used outcome measure combining HRQoL and survival is the QALY. This measure is also often used in economic evaluations to support rational decision making. QALYs allow outcomes to be compared across different disease areas and are therefore useful for health care policy systems aiming to allocate resources efficiently across disease areas, where efficiency is defined as maximising health given the available resources. For systems focussing on efficiency within disease areas, QALYs might be used as a way to combine HRQoL and mortality outcomes in one measure.

A published review of reimbursement agency requirements for HRQoL data in Australia, Canada, England & Wales, Germany, Scotland, and Sweden revealed many differences between agencies' requirements regarding methods for deriving utilities.[53] The authors conclude that standardisation of approaches to the collection of utility data would reduce variation in REA and in economic evaluations. They further observe that for utilities, there seems to be a general agreement that choice-based methods to collect preferences are to be prefered, that a societal perspective should be taken that includes national preferences rather than the preferences from other countries. Generic measures such as the EQ-5D, HUI and SF-6D seem to be the favorites.[53]

All guidelines reviewed for the first version of this guideline in JA1 recommend that HRQoL be considered if it is a clinical relevant or principal intended outcome and mention its use to develop a cost-utility analysis (CUA) when meaningful differences in HRQoL between intervention and comparators have been demonstrated. There is also a consensus that health effects should be expressed in terms of QALYs in economic evaluations. QALYs are preferred for CUA because of their clarity, simplicity, ease of application, and face validity.[54, 55] The strengths and weaknesses of alternative measures such as the healthy-year equivalent are considered not to be fully established.[56]

Gelöscht: included in the current review

While most HTA institutes use QALYs as an outcome measurement, there is less consensus on which instrument to use to measure HRQoL weights when calculating QALYs. QALYs require the use of preference weights for HRQoL.[57] Appendix 5 provides an overview of opinions and/or recommendations regarding which instrument can/should be used.

Most guidelines on utility measures recommend explicitly the use of a preference-based measure (i.e. a generic instrument with an index measure), especially if the data are to be used for the calculation of QALYs. These instruments measure health on a cardinal scale with death being scored 0 and perfect health 1 and allow scores less than 0 for health states worse than death. The following instruments are most often recommended / mentioned in the retrieved guidelines:

- EQ-5D[13, 54-68]
- Health Utilities Index (HUI)[13, 54-58, 61, 63, 65-67, 69, 70]
- Quality of Well-being (QWB)[55, 63, 65, 69]
- SF-6D[54, 55, 57, 58, 62, 64]
- 15D[13, 54]
- AQoL[13, 55, 57, 58]

The overview shows that there is no gold standard for HRQoL measurement. Nevertheless, to maximise comparability across submissions, it is frequently recommended that a generic HRQoL instrument associated with 'off-the-shelf' utility values be consistently administered in randomised trials.[57, 60] Currently, only NICE explicitly identifies a specific instrument (the EQ-5D) to be used.[56] Also ISPOR recommends that analysts collect preference weights as part of clinical trials.[65] To support availability of HRQoL data for economic evaluations and to improve comparability across disease areas, the use of a preference-based utility measure in clinical trials should be encouraged. This does not preclude the use of a complementary disease-specific measure in trials.

In the base-case, i.e. the minimally required analysis, it is often preferred that patients describe their health state[57, 64] (or proxy judges –their carers rather than healthcare professionals[56] – if appropriate). The preference weights connected with these health states are preferably generated by a representative sample of the general public.[54, 56, 60, 71] A US guideline[70] mentions that the choice depends on the perspective of the study. Values can be provided by the population at large or by a sample of patients with the condition for which the intervention is being evaluated. If the issue is allocating resources between competing programmes the former might be used; if it is deciding the best way to treat a given condition the latter might be used. In contrast to other guidelines, Sweden recommends these weights to be derived from persons in the health state described.[72] The survival data (i.e. length of life) and assessment of the health state (i.e. the quality weight) should be reported separately.[54, 64, 67, 71] The procedure to combine these two elements should be reported transparently.[54, 64, 71]

While it is generally preferred that utility scores are derived from the country's own population,[55] it is recognised that these scores are often not available. Moreover, primary data on patients' individual health state descriptions are often also not publicly available, as a consequence of which it becomes impossible to assign national utility values to patients' reported health states. For example, a REA might be partly based on results of studies published in scientific literature, in which case the results might not be country-specific. Therefore, it is often allowed to use utility scores from the general public of other countries with similar cultural or political backgrounds and economic circumstances.[57, 63, 64]

Several HTA institutes mention several methodological disadvantages linked to QALYs as well as disadvantages related to their practical use in the context of reimbursement decisions. Concerns relate to the distributive justice and equity of the resource allocation resulting from the use of the cost-per-QALY as an absolute decision criterion. Therefore, cost-per-QALY is more often considered as one of the decision criteria, amongst others.[15, 73] Full elaboration of these issues is outside the scope of this guideline.

In conclusion, although there is no consensus on the most appropriate instrument, the use of a standard instrument to measure HRQoL would improve

comparability and reliability of economic evaluations. If HRQoL aspects seem to be important, systematically adding a generic utility HRQoL instrument associated with utility values from the general public (e.g. EQ-5D), could be a big step forward in countries where economic evaluations are used in reimbursement decision making. Investigators using instruments that do not use a single index need to think carefully about the future use of the results of their study.[74] If one of the (future) aims might be to calculate an intervention's cost effectiveness to support a reimbursement request, measuring the impact on HRQoL with a generic utility instrument may improve the comparability of the outcomes of these analyses.[75]

3. DISCUSSION AND CONCLUSION

The measurement of HRQoL is especially important when assessing the impact of long-term illness or chronic disease where the goal of intervention is to improve how people are able to function.[7] It can have different aims: to inform clinicians about the intervention most likely to improve a patients' HRQoL, to inform patients about the expected impact of an intervention on his or her HRQoL as compared to alternative interventions, to inform regulatory authorities about the relative benefits of a health technology, from the perspective of the patient and to inform policy makers about the relative effectiveness and/or cost-effectiveness of a health technology, compared to an alternative intervention for the same disease or compared to alternative courses of action elsewhere in the health care sector.

Gelöscht: product

Gelöscht: product

HRQoL assessment in the context of a relative efficacy assessment for registration purposes is distinct from REA for reimbursement purposes. Guidelines for HRQoL measurement for e.g. pharmaceutical product registration purposes can therefore not be simply extended to REA. While safety and benefit-risks are the prior concerns for registration e.g. in the case of pharmaceuticals, REA is primarily concerned with relative effectiveness compared to existing alternatives. The role of HRQoL is different in case of risk-benefit assessments than in REA. For example, HRQoL is only considered in a registration dossier at EMA if efficacy and safety have been demonstrated on the primary endpoint (hierarchical testing), while in REA HRQoL is one of clinical endpoints assessed together with other clinical endpoints to substantiate an added benefit of a pharmaceutical (see EunetHTA guideline on clinical endpoints). Nevertheless, there are also similarities between the requirements for HRQoL measures for product registration and the requirements for REA.

For example, guidelines related to HRQoL assessment for clinical trial protocols have been published.[19, 52] Proof of validity of the HRQoL instrument for the condition studied is required as a prerequisite of its use as well as a definition of clinical meaningful changes in HRQoL scores. Statistical analyses of HRQoL outcomes should be of the same rigor as for other clinical efficacy endpoints. Ways of handling missing data in the statistical analyses should be described in the study protocol. The same requirements could be imposed on the evidence for claims of HRQoL improvements in REA: results from clinical trials, based on a protocol specifying with which validated HRQoL instrument HRQoL would be measured and which hypothesis would be tested, including definitions of meaningful improvements in HRQoL scores, should be presented. It needs to be defined whether decrements in any domain are acceptable or not. There is no scientific guidance for this matter, as it is basically a matter of weighting. The outcome of the weighting process depends for instance on the relative importance of the domain that deteriorates compared to the domain that improves.

In REA, the definition of the hierarchy of endpoints seems to be crucial. HRQoL data will often not even be considered if an intervention shows a deterioriation on the primary endpoints such as overall survival. When HRQoL does become a consideration in the REA, there are still different options for HRQoL measurement, depending on whether or not countries wish to make comparisons across indications and whether or not countries consider cost-utility in their decision making process.

Disease-specific HRQoL instruments are generally considered to be more sensitive to small changes in HRQoL. If no benefit in HRQoL is observed on a generic HRQoL instrument, a disease-specific instrument might still show an added benefit of the intervention. Such instruments might be useful for comparisons within one indication, but still need to be treated with caution as they may ignore changes in domains of HRQoL that are not included in the disease-specific questionnaire but are nevertheless important to patients.

REA is often used in a policy context to assess if the product should be paid for from public resources. Some countries use REA to inform decisions about the allocation of resources *across* indications, while others use REA to assess the relative value of interventions *within* indications.

For informing resource allocation decisions across indications, REA, being one element in the decision making process, is mainly concerned with the value of a therapeutic benefit from a societal point of view. Value is a relative concept, i.e. the value of a particular therapeutic benefit depends on the benefits that can be obtained elsewhere in the health care sector. This implies the need for comparisons across disease areas and types of interventions. Only generic HRQoL instrument, covering a broad range of HRQoL dimensions, allow such comparisons. Disease-specific instruments are useful complements to provide more detailed information on the HRQoL dimensions that have improved/deteriorated. This is especially relevant if besides HRQoL, interventions have an equal weight on all other relevant decision making criteria.

For informing **resource allocation decisions within indications**, disease-specific HRQoL instruments are often preferred, because comparability with HRQoL outcomes in other indications is considered less important. However, it should be noted that generic HRQoL instruments remain relevant and useful for decision makers in this case. When judging the efficiency of different interventions for the same indication, policy makers still have to define the *value* of the health benefits. By using the same generic HRQoL instruments across different indications, decision makers can build up reference cases in order to determine this value. Even though the value of an equal benefit on a generic HRQoL can differ between indications (e.g. because of the weight given to the disease-specific outcomes), it potentially increases the transparency of the appraisal process.[15]

BIBLIOGRAPHY

1. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Intern Med. 1993;118(8):622-9.
2. Billingham LJ, Abrams KR, Jones DR. Methods for the analysis of quality-of-life and survival data in health technology assessment. Health Technol Assess. 1999 3(10):1-152.
3. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess. 1998;2(14):i-iv, 1-74.
4. European Medicines Agency. Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products. London: EMA; 2006 January 2006. EMEA/CHMP/EWP/139391/2004 Available from: http://www.ispor.org/workpaper/emea-hrql-guidance.pdf
5. Food and Drug Administration. Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009.
6. Goodman CS. Healthcare technology assessment: methods, framework, and role in policy making. Am J Manag Care. 1998 4 Spec No:SP200-6.
7. Jackowski D, Guyatt G. A Guide to Health Measurement. Clin. Orthop. Relat. Res. 2003(413):80-9.
8. Jonsson B. Changing health environment: the challenge to demonstrate cost-effectiveness of new compounds. PharmacoEconomics. 2004 22 Suppl 4:5-10.
9. Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. British Medical Journal. 2002;324(7351):1417.
10. Lubeck DP. Health-related quality of life measurements and studies in rheumatoid arthritis. Am J Manag Care. 2002 8(9):811-20.
11. Hoedemaekers R. Introduction: Towards Better Integration of Normative Judgements in Health Care Package Decisions. Health Care Anal. 2003;11(4):275-8.
12. Contopoulos-Ioannidis DG, Karvouni A, Kouri I, Ioannidis JPA. Reporting and interpretation of SF-36 outcomes in randomised trials: systematic review. Bmj. 2009;338(7687):152-4.
13. Kristensen F, Sigmund H. Health Technology Assessment Handbook. Copenhagen: Danish Centre for Health Technology Assessment, National Board of Health; 2007.
14. Kleijnen S, Goettsch W, d'Andon A, Vitre P, George E, Goulden S, et al. EUnetHTA JA WP5: Relative Effectiveness Assessment (REA) of Pharmaceuticals. Background review. July 2011 (version 5B). 2011. Available from: http://www.eunethta.eu/sites/5026.fedimbo.belgium.be/files/Final%20version%20of%20Background%20Review%20on%20Relative%20Effectiveness%20Assessment+appendix.pdf
15. le Polain M, Franken M, Koopmanschap M, Cleemput I. Drug reimbursement systems : international comparison and policy recommendations. 147C ed. Bruxelles: KCE (Federaal Kenniscentrum voor de gezondheiszorg / Centre fédéral d'expertise des soins de santé / Belgian Health Care Knowledge Centre); 2011.
16. Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. Measuring and Valuing Health Benefits for Economic Evaluation. Oxford: Oxford University Press; 2007.
17. Lumley T, Simes RJ, Gebski V, Hudson HM. Combining components of quality of life to increase precision and evaluate trade-offs. Statistics in Medicine. 2001;20:3231-49.
18. Machin D, Weeden S. Suggestions for the presentation of quality of life data from clinical trials. Statistics in Medicine. 1998;17(5-7):711-24.
19. Calvert MJ, Freemantle N. Use of health-related quality of life in prescribing research. Part 2: methodological considerations for the assessment of health-related quality of life in clinical trials. Journal of Clinical Pharmacy & Therapeutics. 2004;29(1):85-94.
20. Papanicolaou S, Sykes D, Mossialos E. EMEA and the evaluation of health-related quality of life data in the drug regulatory process. INT. J. TECHNOL. ASSESS. HEALTH CARE. 2004;20(3):311-24.
21. Greenfield S. Health profiles and quality of life. Acad Radiol. 1999 6 Suppl 1:S89-99.
22. Petrillo J, Cairns J. Converting condition-specific measures into preference-based outcomes for use in economic evaluation. Expert Rev. Pharmacoecon. Outcomes Res. 2008;8(5):453-61.

23. Chuang LH, Kind P. Converting the SF-12 into the EQ-5D: an empirical comparison of methodologies. PharmacoEconomics. 2009;27(6):491-505.
24. Brazier J, Yang Y, Tsuchiya A. Review of methods for mapping between condition specific measures onto generic measures of health. Sheffield: University of Sheffield; 2007. Report prepared for the Office of Health Economics Available from: http://www.ohe.org/lib/liDownload/588/5%20Mapping%20disease%20specifc%20to%20generic%20measures%20research%20report.pdf?CFID=4642125&CFTOKEN=20179228
25. Franks P, Lubetkin EI, Gold MR, Tancredi DJ, Jia H. Mapping the SF-12 to the EuroQol EQ-5D Index in a national US sample. Medical Decision Making. 2004;24(3):247-54.
26. Lawrence WF, Fleishman JA. Predicting EuroQoL EQ-5D preference scores from the SF-12 Health Survey in a nationally representative sample. Medical Decision Making. 2004;24(2):160-9.
27. Frei A, Svarin A, Steurer-Stey C, Puhan MA. Self-efficacy instruments for patients with chronic diseases suffer from methodological limitations--a systematic review. Health Qual Life Outcomes. 2009;7:86.
28. Stull DE, Leidy NK, Parasuraman B, Chassany O. Optimal recall periods for patient-reported outcomes: challenges and potential solutions. Curr Med Res Opin. 2009;25(4):929-42.
29. Beacon HJ, Thompson SG. Multi-level models for repeated measurement data: application to quality of life data in clinical trials. Statistics in Medicine. 1996;15(24):2717-32.
30. Cheung YB, Machin D, Fong KY, Thio ST, Thumboo J. Discriminative ability of the short-form 36 health survey: a tale of two versions. Quality of Life Research. 2005;14(2):555-9.
31. Wild D, Eremenco S, Mear I, Martin M, Houchin C, Gawlicki M, et al. Multinational trials-recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: the ISPOR Patient-Reported Outcomes Translation and Linguistic Validation Good Research Practices Task Force report. Value Health. 2009;12(4):430-40.
32. Acquadro C, Conway K, Hareendran A, Aaronson N. Literature review of methods to translate health-related quality of life questionnaires for use in multinational clinical trials. Value in Health. 2008;11(3):509-21.
33. Herdman M, Fox-Rushby J, Badia X. A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. Qual Life Res. 1998;7(4):323-35.
34. Ware JE, Jr., Keller SD, Gandek B, Brazier JE, Sullivan M. Evaluating translations of health status questionnaires. Methods from the IQOLA project. International Quality of Life Assessment. International Journal of Technology Assessment in Health Care. 1995;11(3):525-51.
35. Gao F, Ng GY, Cheung YB, Thumboo J, Pang G, Koo WH, et al. The Singaporean English and Chinese versions of the EQ-5D achieved measurement equivalence in cancer patients. Journal of Clinical Epidemiology. 2009;62(2):206-13.
36. Lyons RA, Wareham K, Lucas M, Price D, Williams J, Hutchings HA. SF-36 scores vary by method of administration: implications for study design. J Public Health Med. 1999;21(1):41-5.
37. Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value Health. 2009;12(4):419-29.
38. Hacker ED. Technology and quality of life outcomes. Semin Oncol Nurs. 2010 26(1):47-58.
39. Cook KF, O'Malley KJ, Roddey TS. Dynamic assessment of health outcomes: time to let the CAT out of the bag? Health Serv Res. 2005 40(5 Pt 2):1694-711.
40. Hahn EA, Bode RK, Du H, Cella D. Evaluating linguistic equivalence of patient-reported outcomes in a cancer clinical trial. Clinical Trials. 2006;3(3):280-90.
41. Buskirk TD, Stein KD. Telephone vs. mail survey gives different SF-36 quality-of-life scores among cancer survivors. Journal of Clinical Epidemiology. 2008;61(10):1049-55.

42. Bowling A, Bond M, Jenkinson C, Lamping DL. Short Form 36 (SF-36) Health Survey questionnaire: which normative data should be used? Comparisons between the norms provided by the Omnibus Survey in Britain, the Health Survey for England and the Oxford Healthy Life Survey. Journal of Public Health Medicine. 1999;21(3):255-70.
43. Hoedemaekers R, Dekkers W. Key concepts in health care priority setting. Health Care Anal. 2003 11(4):309-23.
44. Curran D, Bacchi M, Schmitz SF, Molenberghs G, Sylvester RJ. Identifying the types of missingness in quality of life data from clinical trials. Statistics in Medicine. 1998;17(5-7):739-56.
45. Fayers PM, Curran D, Machin D. Incomplete quality of life data in randomized trials: missing items. Statistics in Medicine. 1998;17(5-7):679-96.
46. Troxel AB, Fairclough DL, Curran D, Hahn EA. Statistical analysis of quality of life with missing data in cancer clinical trials. Statistics in Medicine. 1998;17(5-7):653-66.
47. Fielding S, Fayers PM, Ramsay CR. Investigating the missing data mechanism in quality of life outcomes: a comparison of approaches. Health and Quality of Life Outcomes. 2009;7:57.
48. Wu MC, Borkowf CB, Albert PS. Analysis of longitudinal quality of life data with informative dropout. In: Mesbah M, Cole BF, Lee MT, editors. Statistical Methods for Quality of life Studies, Design, Measurements and Analysis. Netherlands: Kluwer Academic Publishers; 2002.
49. Staquet M, Berzon R, Osoba D, Machin D. Guidelines for reporting results of quality of life assessments in clinical trials. Quality of Life Research. 1996;5(5):496-502.
50. EuroQol Group How to report. [2013 [cited 21/02/2013]. Available from: http://www.euroqol.org/about-eq-5d/how-to-use-eq-5d/how-to-report.html
51. Fayers PM, Machin D. Quality of life: assessment, analysis, and interpretation. Chichester, New York: John Wiley; 2000.
52. Fayers PM, Hopwood P, Harvey A, Girling DJ, Machin D, Stephens R. Quality of life assessment in clinical trials - guidelines and a checklist for protocol writers: the UK Medical Research Council experience. European Journal of Cancer. 1997;33(1):20-8.
53. Lloyd A, Wild D, Gallop K, Cowell W. Reimbursement agency requirements for health-related quality-of-life data: A case study. Expert Rev. Pharmacoecon. Outcomes Res. 2009;9(6):527-37.
54. Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies: Canada. Ottawa: Canadian Agency for Drugs and Technologies in Health (CADTH); 2006. 3rd ed
55. Pharmaceutical management Agency (PHARMAC). Prescription for Pharmacoeconomic Analysis - Methods for cost-utility analysis. May 2007. Version 2
56. National Institute for Health and Clinical Excellence (NICE). Guide to the methods of technology appraisal. June 2008.
57. Pharmaceutical Benefits Advisory Committee. Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (Version 4.3). Australian Government, Department of Health and Ageing; December 2008.
58. Medical Services Advisory Committee. Economics Section of the MSAC Guidelines. Australian Government, Department of Health and Ageing; August 2008.
59. Behmane D, Lambot K, Irs A, Steikunas N, Hill S, Freemantle N. Baltic guidelines for economic evaluation of pharmaceuticals (pharmacoeconomic analysis). August 2002.
60. Cleemput I, Van Wilder P, Vrijens F, Huybrechts M, Ramaekers D. Guidelines for Pharmacoeconomic Evaluations in Belgium. Health Technology Assessment (HTA). Brussels: Health Care Knowledge Centre (KCE); 2008. KCE Reports 78C.
61. Haute Autorité de Santé. L'évaluation économique à la Haute Autorité de Santé Principes et méthodes. Décembre/janvier 2010.
62. Graf von der Schulenburg JM, Greiner W, Jost F, Klusen N, Kubin M, Leidl R, et al. German recommendations on health economic evaluation: third and updated version of the Hanover Consensus. Value Health. 2008;11(4):539-44.
63. Szende Á, Mogyorósy Z, Muszbek N, Nagy J, Pallos G, Dózsa C. Methodological guidelines for conducting economic evaluation of healthcare interventions in Hungary: a Hungarian proposal for methodology standards. Eur J Health Econom. 2002;3:196–206.
64. Health Information and Quality Authority. Guidelines for the Economic Evaluation of Health Technologies in Ireland. 2010.

65. Ramsey S, Willke R, Briggs A, Brown R, Buxton M, Chawla A, et al. Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report. Value Health. 2005;8(5):521-33.
66. Capri S, Ceci A, Terranova L, Merlo F, Mantovani L, Attanasio E, et al. Guidelines for economic evaluations in Italy: Recommendations from the Italian group of pharmacoeconomic studies. Drug Inf. J. 2001;35(1):189-201.
67. College voor zorgverzekeringen. Guidelines for pharmacoeconomic research, updated version. 2006.
68. Norwegian Medicines Agency. Norwegian guidelines for pharmacoeconomic analysis in connection with applications for reimbursement. 2005.
69. Collège des Économistes de la Santé. French guidelines for the economic evaluation of health care technologies. September 2004.
70. Academy of managed care pharmacy (AMCP). The AMCP Format for Formulary Submissions, version 2.1. A Format for Submission of Clinical and Economic Data in Support of Formulary Consideration by Health Care Systems in the United States. April 2005.
71. López-Bastida J, Oliva J, Antoñanzas F, García-Altés A, Gisbert R, Mar J, et al. Spanish recommendations on economic evaluation of health technologies. European Journal of Health Economics. 2010;11(5):512-20.
72. Pharmaceutical Benefits Board. General guidelines for economic evaluations from the Pharmaceutical Benefits Board. 2003.
73. Cleemput I, Neyt M, Thiry N, De Laet C, Leys M. Threshold values for cost-effectiveness in health care. Brussel: Federaal Kenniscentrum voor de Gezondheiszorg/Centre fédéral d'expertise des soins de santé/Belgian Health Care Knowledge Centre (KCE); 2008. 100C
74. Deverill M, Brazier J, Green C, Booth A. The use of QALY and non-QALY measures of health-related quality of life. Assessing the state of the art. PharmacoEconomics. 1998;13(4):411-20.
75. Neyt M. Towards more consistent use of generic quality-of-life instruments. PharmacoEconomics. 2010;28(4):345-6.
76. Walter E, Zehetmayr S. Guidelines on Health Economic Evaluation. Consensus paper. IPF Institut für Pharmaökonomische Forschung; April 2006.
77. IQWIG, Institute for Quality and Efficiency in Health Care. General Methods (Version 3.0 of 27.05.2008). 2008.
78. Agency for Health Technology Assessment. Guidelines for conducting Health Technology Assessment (HTA). Warsaw: April 2009.
79. da Silva E, Pinto V, Sampaio C, Pereira J, Drummond M, Trindade R. Guidelines for Economic Drug Evaluation Studies. Infarmed; 1998.
80. Hiebinger C, Pertl D, Stürzlinger H, Zsifkovits J. Vorarbeiten zum Methodenhandbuch für HTA-Berichte: Übersicht zu nationalen und internationalen HTA Methodenhandbüchern (Austria). Mai 2009.
81. Stürzlinger H, Fröschl B, Hiebinger C, Pertl D, Zsifkovits J. Prozesshandbuch für Health Technology Assessment (Austria). Mai 2009.
82. No author. Metodická pomôcka: pre vykonávanie farmako-ekonomického rozboru lieku a medicínsko-ekonomického rozboru zdravotníckej pomôcky ku vyhláške Ministerstva zdravotníctva Slovenskej republiky (Slovakia). 2008.
83. Manso de Mello Vianna C, Caetano R. Diretrizes Metodológicas para Estudos de Avaliação Econômica de Tecnologias para o Ministério da Saúde (Brazil). 2007.
84. Gálvez González A. Guía metodológica para la evaluación económica en salud. Cuba. 2003.
85. Ministério da Saúde. Secretaria de Ciência, Tecnologia e Insumos Estratégicos. Departamento de Ciência e Tecnologia. Diretrizes Metodológicas : estudos de avaliação econômica de tecnologias em saúde (Brasília). 2009.
86. Health Insurance Review Agency. 의약품 경제성평가지침 (Korea). 2006.
87. No author. Guidelines of Methodological Standards for Pharmacoeconomic Evaluations (Taiwan). 2006.
88. Drummond MF, Schwartz JS, Jonsson B, Luce BR, Neumann PJ, Siebert U, et al. Key principles for the improved conduct of health technology assessments for resource allocation decisions. Int J Technol Assess Health Care. 2008;24(3):244-58; discussion 362-8.

89. IQWIG, Institute for Quality and Efficiency in Health Care. Technical Document Modelling (Version 1.0 of 09.10.2008). 2008.
90. No author. The Standardization System in the Russian Federation Health Care System Clinico-Economic Studies. 2002.

APPENDIX I: METHODS OF THE LITERATURE SEARCH (CONDUCTED DURING ORIGINAL GUIDELINE ELABORATION IN JA 1)

KEYWORDS

The keywords used for HRQoL were defined as broadly as possible, in order to be as sensitive as possible. Besides HRQoL, keywords such as well-being, utility and preferences were used to retrieve relevant literature. These were then combined with a set of keywords related to REA. The different sets of keywords used in the search strategy are presented in Table 4.

Table 4: Topics and keywords used for the search strategy on HRQoL and relative effectiveness assessment

Topic	Keywords
Quality of life	quality of life QoL life quality quality life well being wellbeing
Utilities	Utilities, utility preference(s)
Measures combining quality of life and life expectancy	quality adjusted quality of life adjusted QALY, QALD, QALE, HYE, HYEs, HYE's quality survival time healthy life expectancy healthy years equivalent(s)
Relative effectiveness	technology assessment relative effectiveness comparative effectiveness drug reimbursement
Guidelines for quality of life research	Guidelines

All keywords for quality of life, utilities and combined measures were combined using the Boolean expression "or" in order to capture all references related to these topics. Subsequently, the yield was reduced by requiring that at least one of the keywords related to relative effectiveness or guidelines were relevant for the reference. This was done by using the Boolean expression "and".

SEARCH ENGINES AND SOURCES OF INFORMATION

Guidelines, reports, recommendations from regulatory agencies

European Medicines Agency (EMA): "Reflection Paper On The Regulatory Guidance For The Use Of Health-related Quality Of Life (Hrql) Measures In The Evaluation Of Medicinal Products", 2005.[4]

FDA: "Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims", 2009.[5]

(Pharmacoeconomic) guidelines, reports and recommendations

Guidelines and templates for reimbursement request files for pharmaceuticals were screened for specific guidance on HRQoL measures in the context of the demonstration of relative effectiveness. They were retrieved through screening of web-sites of reimbursement agencies. The search was not limited to Europe - guidelines from Canada, Australia and New Zealand were also included. In addition, guidelines for economic evaluations of health interventions were

screened in order to make meaningful recommendations for HRQoL measurement for the calculation of measures combining HRQoL and life expectancy.

Besides reimbursement agencies also other governmental, semi-governmental or private organisations develop guidelines for HRQoL. These have been retrieved through screening the websites of HTA agencies and research groups.

A list of institutes' and organisations' websites that were searched as well as a complete list of all guidelines included in the review are presented in Appendix 4 and 5 respectively. If several guidelines were retrieved from the same institute, the most recent one was included. The guidelines retained were often broader than just relating to HRQoL, e.g. offering guidance for full economic evaluations. The actual guidelines for HRQoL measurement, as part of these broader guidelines, are cited in Appendix 6.

Bibliographic databases

The following bibliographic databases were searched:

- Medline (OVID)
- Embase
- Cochrane methodology register

Others

NICE's guide to the Methods of Technology Appraisal:

http://www.nice.org.uk/niceMedia/pdf/TAP_Methods.pdf

NICE's Briefing paper for methods review, related to key issues in utility measurement:
http://www.nice.org.uk/aboutnice/howwework/devnicetech/technologyappraisalprocessguides/selectedfurtherreadingguidetothemethodsoftechnologyappraisal.jsp?domedia=1&mid=4A655B27-19B9-E0B5-D45D0B46FC59F61C

Handsearching, based on reference lists of retained articles.

SEARCH STRATEGIES

The Medline search was performed in Medline (OVID) on January 12th, 2011. References published in 1995 or later were retrieved. The full search strategy is presented in Table 5.

Table 5: Medline Search Strategy

Database: Ovid MEDLINE(R) 1948 to Present with Daily Update Search Strategy: performed on 12/01/2011 --- 1 exp Comparative Effectiveness Research/ (261) 2 exp Technology Assessment, Biomedical/ (8213) 3 relative effectiveness.mp. (1949) 5 exp Insurance, Health, Reimbursement/ or exp Fees, Pharmaceutical/ (34492) 6 exp "Quality of Life"/ (85774) 7 exp Health Status/ (77780) 8 "well being".mp. (27220) 9 "wellbeing".mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (2967) 10 "well-being".mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (27220) 11 exp Quality-Adjusted Life Years/ (4638) 12 QAL*.mp. (3154) 13 HYE*.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (523)

14 quality survival time.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (3)
15 Healthy life expectancy.mp. (90)
16 HLE.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (1763)
17 HYE.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (33)
18 Healthy years equival*.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (26)
19 utilit*.mp. (80435)
20 exp Patient Preference/ (496)

23 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 (258111)
35 2 or 3 or 4 or 5 (44404)

38 23 and 35 (1305)

44 limit 38 to (english language and yr="1995 -Current") (960)

The Embase search was performed on January 19th, 2011. References published in 1995 or later were retrieved. The full search strategy is presented in Table 6.

Table 6: Embase search strategy

No.	Query	Results
#1	comparative effectiveness research'/exp OR 'comparative effectiveness research'	686
#2	technology assessment biomedical'/exp OR 'technology assessment biomedical'	10,541
#3	'relative effectiveness'	2,375
#4	comparative effectiveness research'/exp OR 'comparative effectiveness research'	686
#5	'reimbursement'/exp OR reimbursement	33,281
#6	'quality of life'/exp OR 'quality of life'	197,922
#7	'health state'/exp OR 'health state'	92,934
#8	'wellbeing'/exp OR wellbeing	24,703
#9	quality adjusted life years'/exp OR 'quality adjusted life years'	7,628
#10	qal*	4,681
#11	hye*	18,642
#12	'quality survival time'	5
#13	'healthy life expectancy'	118
#14	hle	766
#15	hye	8,751
#16	'healthy years equivalent'	15
#17	'healthy years equivalents'	18
#18	utilit*	220,006
#19	'patient preference'	2,262
#20	# 2 OR # 3 OR # 4 OR # 5	46,512
#21	# 6 OR # 7 OR # 8	292,85
#22	# 9 OR # 10 OR # 11 OR # 12 OR # 13 OR # 14 OR # 15 OR # 16 OR # 17	28,568
#23	# 18 OR # 19	222,161
#24	# 21 OR # 22 OR # 23	526,757
#25	# 20 AND # 24	2,655

#27	practice guideline'/exp OR 'practice guideline'	233,69
#28	Guideline	243,447
#29	# 28 NOT # 27	9,757
#30	# 24 AND # 29	872
#31	# 25 OR # 30	3,52
#32	# 25 OR # 30 AND [english]/lim AND ([article]/lim OR [article in press]/lim OR [conferencepaper]/lim OR [conference review]/lim OR [editorial]/lim OR [review]/lim OR [short survey]/lim) AND [1995-2011]/py	2,617

The Cochrane Methodology register was searched using the keyword "quality of life". The search was limited to references published after 1995 in English.

In contrast to the database searches, the search of the (pharmacoeconomic) guidelines was not restricted to guidelines published in English but also included guidelines published in Dutch or French.

INCLUSION AND NON-INCLUSION CRITERIA

Inclusion criteria:

- Critical analyses of HRQoL measurement
- English language
- General reflections, theoretical considerations

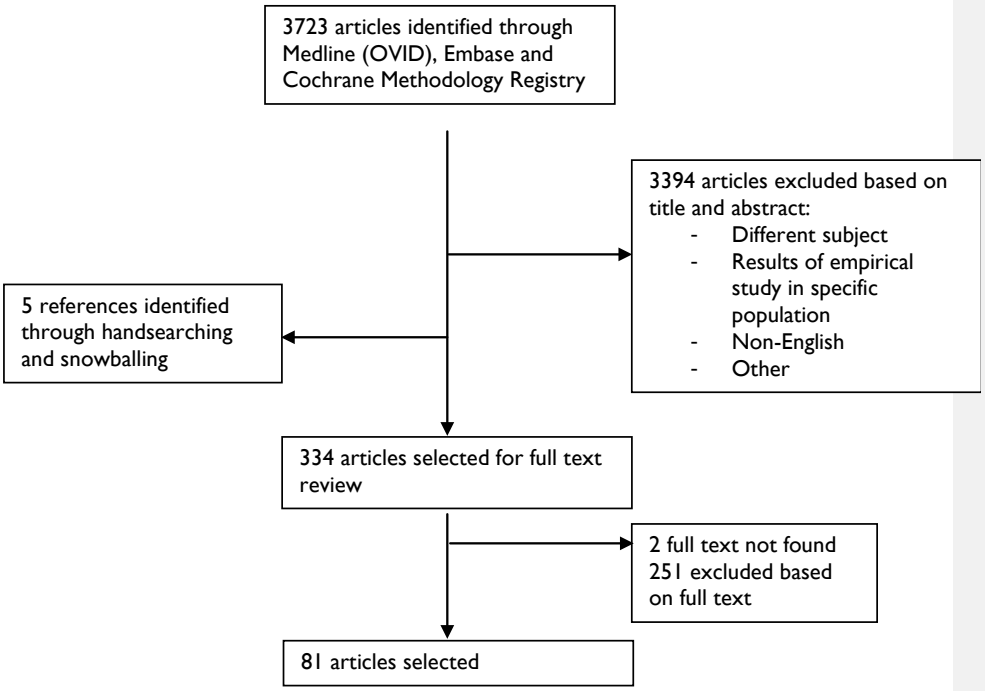
Exclusion criteria:

- Studies on specific interventions, one specific instrument or specific populations
- Letters, conference abstracts
- Studies about quality of care
- Studies on the use of HRQoL measures for case-mix adjustments for financing or for purposes not related to the reimbursement of specific interventions.

APPENDIX 2: ANALYSIS AND SYNTHESIS OF THE LITERATURE (CONDUCTED DURING GUIDELINE ELABORATION IN JA 1)

LITERATURE SEARCH RESULTS

The search in Medline, Embase and the Cochrane Methodology Registry resulted in 3723 references. On a first selection, based on title and abstract, we retained 332 references of which the full text was obtained. Selection based on full texts reduced the number of relevant papers to 81.



This flow chart does not include the guidelines for reimbursement submissions or guidelines developed by other agencies or research groups. Twenty-four existing guidelines were included for review (see appendix 4).

The relevant references were classified according to their main topic and the subsequent analysis was performed per topic. The following topics were defined:

- Descriptive HRQoL measures: profiles and summary measures, disease-specific and generic instruments,
- Utility/preference-based HRQoL measures
- Psychometric properties
- Measurement issues: evaluation by patients versus proxies, practical measurement issues
- Data analysis issues: missing data, multiple testing
- Presentation of HRQoL study results
- Guidelines for HRQoL measurement in specific diseases
- Guidelines for HRQoL measurement for product registration purposes
- HRQoL measurement for the calculation of measures combining HRQoL and life expectancy as described in existing guidelines.

Several general papers on HRQoL were found. These were mainly used for the introduction. Two extensive reviews on HRQoL measurement were used as the basis of this review.[2, 3] They encompassed all relevant topics for the current guideline. Other references were used if they included additional information fitting within the scope of this guideline.

APPENDIX 3: PSYCHOMETRIC PROPERTIES OF HRQOL INSTRUMENTS: CONCEPTS AND DEFINITIONS*

Property	Type	What is assessed?
Reliability	Test-retest reliability and intra-interviewer reliability (for interviewer-administered questionnaires)	Stability of scores over time when no change is expected in the concept of interest. [7]
	Internal consistency	- Extent to which items comprising a scale measure the same concept - Intercorrelation of items that contribute to a score - Internal consistency
	Inter-interviewer reliability (for interviewer-administered questionnaires)	Agreement among responses when the PRO is administered by two or more different interviewers.
Validity	Content validity	Evidence that the instrument measures the concept of interest including evidence from qualitative studies that the items and domains of an instrument are appropriate and comprehensive relative to its intended measurement concept, population, and use. Testing other measurement properties will not replace or rectify problems with content validity.
	Construct validity	Evidence that relationships among items, domains and concepts conform to *a priori* hypotheses concerning logical relationships that should exist with measures of related concepts or scores produced in similar or diverse patient groups. It involves the establishment of a model or theoretical framework defining the logical relations that should exist between changes observed on a HRQoL measure and changes observed on other (e.g. clinical) measures.[1]
Responsiveness/ ability to detect change		Evidence that a PRO instrument can identify differences in scores over time in individuals or groups (similar to those in the clinical trials), who have changed with respect to the measurement concept. Responsiveness of HRQoL instruments might be influenced by ceiling effects and floor effects. In case of a ceiling effect, a relatively large deterioration can be observed in patients with a good initial health (highest score), while the floor effect might imply that further deteriorations cannot be observed anymore in patients with an initially bad health state (lowest score).[1]
Acceptability		Evidence on the extent to which an instrument is considered acceptable for respondents to complete.[3] In this context it is important to consider the burden associated with the administration and processing of an instrument or a batch of HRQoL instruments.[3] Accumulation of HRQoL

		questionnaires with the aim of increasing the amount of information obtained, might be counterproductive. Moreover, each addition of a HRQoL measure increases the number of statistical analyses and therefore the probability of significant effects arising by chance. The latter problem may be solved by requiring the research protocol defining the hypothesis that will be tested.

1 * Based on FDA Guidance for Industry for Patient-Reported Outcome measures.[5]

2

APPENDIX 4: LIST OF ORGANISATIONS SEARCHED TO RETRIEVE RELEVANT GUIDELINES (CONDUCTED DURING GUIDELINE ELABORATION IN JA 1)

Organisation	Full name	Country
AETMIS	Agence d´Évaluation des Technologies et des Modes d´Intervention en Santé	Canada
AETS	Agencia de Evaluación de Tecnologias Sanitarias	Spain
AETSA	Andalusian Agency for Health Technology Assessment	Spain
AGENAS	L'Agenzia nazionale per i servizi sanitari regionali - The Agency for Regional Healthcare	Italy
AHRQ	Agency for Healthcare Research and Quality	USA
AHTA	Adelaide Health Technology Assessment	Australia
AHTAPol	Agency for Health Technology Assessment in Poland (Agencja Oceny Technologii Medycznych)	Poland
ASERNIP-S	Australian Safety and Efficacy Register of New Interventional Procedures -Surgical	Australia
AVALIA-T	Galician Agency for Health Technology Assessment (Axencia de Avaliación de Tecnoloxías Sanitarias de Galicia)	Spain
CADTH	Canadian Agency for Drugs and Technologies in Health	Canada
CAHTA	Catalan Agency for Health Technology Assessment and Research	Spain
CDE	Center for Drug Evaluation	Taiwan, Republic of China
CEDIT	Comité dÉvaluation et de Diffusion des Innovations Technologiques	France
CENETEC	Centro Nacional de Excelencia Tecnológica en Salud Reforma	Mexico
CNHTA	Committee for New Health Technology Assessment	Korea
CRD	Centre for Reviews and Dissemination	United Kingdom
CVZ	College voor Zorgverzekeringen	The Netherlands
DACEHTA	Danish Centre for Health Technology Assessment	Denmark
DAHTA @DIMDI	German Agency for HTA at the German Institute for Medical Documentation and Information	Germany
DECIT-CGATS	Secretaria de Ciëncia, Tecnologia e Insumos Estratégicos, Departamento de Ciência e Tecnologia	Brazil
DSI	Danish Institute for Health Services Research (Dansk Sundhedsinstitut)	Denmark
ETESA	Department of Quality and Patient Safety of the Ministry Health of Chile (Evaluación de tecnologías de la Salud)	Chile
EUnetHTA	European Network for HealthTechnology Assessment	Europe
FinOHTA	Finnish Office for Health Care Technology Assessment	Finland
GÖG	Gesundheit Österreich GmbH	Austria
GR	Gezondheidsraad	The Netherlands
HAS	Haute Autorité de Santé	France
HIQA	Health Information and Quality Authority	Ireland
HSAC	Health Services Assessment Collaboration	New Zealand
HTAi	Health Technology Assessment International	International
ICTAHC	Israel Center for Technology Assessment in Health Care	Israel
IECS	Institute for Clinical Effectiveness and Health Policy (Instituto de Effectividad Clinical y Sanytaria	Argentina
IHE	Institute of Health Economics	Canada

iHEA	International Health Economics Association (iHEA)	International
INAHTA	International Network of Agencies for Health Technology Assessment	International
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen	Germany
ISPOR	International Society for Pharmacoeconomics and Outcomes Research	International
KCE	Belgian Federal Health Care Knowledge Centre (Federaal Kenniscentrum voor de Gezondheidszorg)	Belgium
LBI of HTA	Ludwig Boltzmann Institut für Health Technology Assessment	Austria
MaHTAS	Health Technology Assessment Section, Ministry of Health Malaysia (	Malaysia
MAS	Medical Advisory Secretariat	Canada
MSAC	Medicare Services Advisory Committee	Australia
MTU-SFOPH	Medical Technology Unit - Swiss Federal Office of Public Health	Switzerland
NCCHTA	National Coordinating Centre for Health Technology Assessment	United Kingdom
NHS QIS	Quality Improvement Scotland	United Kingdom
NHSC	National Horizon Scanning Center	United Kingdom
NICE	National Institute for Health and Care Excellence	United Kingdom
NOKC	Norwegian Knowledge Centre for Health Services	Norway
NZHTA	New Zealand Health Technology Assessment	New Zealand
OSTEBA	Basque Office for Health Technology Assessment (Osasun Teknologien Ebaluazioa)	Spain
PHARMAC	Pharmaceutical Management Agency	New Zealand
SBU	Swedish Council on Technology Assessment in Health Care (Statens beredning för medicinsk utvärdering)	Sweden
UETS	Unidad de Evaluación de Tecnologías Sanitarias	Spain
UVT	HTA Unit in A.Gemelli University Hospital (Unità di Valutazione delle Tecnologie)	Italy
VASPVT	State Health Care Accreditation Agency under the Ministry of Health of the Republic of Lithuania (Valstybinė akreditavimo sveikatos priežiūros veiklai tarnyba prie Sveikatos apsaugos ministerijos)	Lithuania
VATAP	VA Technology Assessment Program	USA
ZonMw	The Medical and Health Research Council of The Netherlands	The Netherlands

Gelöscht: linical

APPENDIX 5: GUIDELINES FOR HRQOL MEASUREMENT FROM REIMBURSEMENT AGENCIES, HTA AGENCIES AND RESEARCH GROUPS INCLUDED IN THE REVIEW (CONDUCTED DURING GUIDELINE ELABORATION IN JA 1)

Country	Reference
Australia (PBAC)	Pharmaceutical Benefits Advisory Committee. Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (Version 4.3). Australian Government, Department of Health and Ageing; December 2008.[57]
Australia (MSAC)	Medical Services Advisory Committee. Economics Section of the MSAC Guidelines. Australian Government, Department of Health and Ageing; August 2008.[58]
Austria	Walter E, Zehetmayr S. Guidelines on Health Economic Evaluation. Consensus paper. IPF Institut für Pharmaökonomische Forschung; April 2006.[76]
Baltic countries	Behmane D, Lambot K, Irs A, Steikunas N, Hill S, Freemantle N. Baltic guidelines for economic evaluation of pharmaceuticals (pharmacoeconomic analysis). August 2002.[59]
Belgium	Cleemput I, Van Wilder P, Vrijens F, Huybrechts M, Ramaekers D. Guidelines for Pharmacoeconomic Evaluations in Belgium. Health Technology Assessment (HTA). Brussels: Health Care Knowledge Centre (KCE); 2008. KCE Reports 78C.[60]
Canada	Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies: Canada. Ottawa: Canadian Agency for Drugs and Technologies in Health (CADTH); 2006. 3rd ed.[54]
Denmark	Kristensen F, Sigmund H: Health Technology Assessment Handbook. Copenhagen: Danish Centre for Health Technology Assessment, National Board of Health; 2007.[13]
France (CES)	Collège des Économistes de la Santé: French guidelines for the economic evaluation of health care technologies. September 2004.[69]
France (HAS)	Haute Autorité de Santé. L'évaluation économique à la Haute Autorité de Santé Principes et méthodes. Décembre/janvier 2010.[61]
Germany (Hanover Consensus Group)	Graf von der Schulenburg JM, Greiner W, Jost F, Klusen N, Kubin M, Leidl R, Mittendorf T, Rebscher H, Schoeffski O, Vauth C, Volmer T, Wahler S, Wasem J, Weber C, Hanover Consensus Group: German recommendations on health economic evaluation: third and updated version of the Hanover Consensus. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research 2008, 11:539-544.[62]
Germany (IQWIG)	IQWIG, Institute for Quality and Efficiency in Health Care. General Methods (Version 3.0 of 27.05.2008). 2008.[77]
Hungary	Szende A, Mogyorósy Z, Muszbek N, Nagy J, Pallos G, Dózsa C: Methodological guidelines for conducting economic evaluation of healthcare interventions in Hungary: a Hungarian proposal for methodology standards. Eur J Health Econom 2002, 3:196–206.[63]
Italy	Capri S, Ceci A, Terranova L, Merlo F, Mantovani L, Attanasio E, Benzi G, Berto P, Bruzzi P, Bruzzone M, Colombo G, Fattore G, Massotti M, Negrini C, Palazzo F, Paoletti R, Pasotti V, Reggio S, Santi L, Serra G: Guidelines for economic evaluations in Italy: Recommendations from the Italian group of pharmacoeconomic studies. Drug Inf J 2001, 35:189-201.[66]
Ireland	Health Information and Quality Authority. Guidelines for the Economic Evaluation of Health Technologies in Ireland. 2010.[64]
ISPOR	Ramsey S, Willke R, Briggs A, Brown R, Buxton M, Chawla A, Cook J, Glick H, Liljas B, Petitti D, Reed S: Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task

	Force report. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research 2005, 8:521-533.[65]
The Netherlands	College voor zorgverzekeringen: Guidelines for pharmacoeconomic research, updated version. 2006.[67]
New Zealand	Pharmaceutical management Agency (PHARMAC). Prescription for Pharmacoeconomic Analysis - Methods for cost-utility analysis. May 2007. Version 2.[55]
Norway	Norwegian Medicines Agency. Norwegian guidelines for pharmacoeconomic analysis in connection with applications for reimbursement. 2005.[68]
Poland	Agency for Health Technology Assessment. Guidelines for conducting Health Technology Assessment (HTA). Warsaw: April 2009.[78]
Portugal	da Silva E, Pinto V, Sampaio C, Pereira J, Drummond M, Trindade R. Guidelines for Economic Drug Evaluation Studies. Infarmed; 1998.[79]
Spain	López-Bastida J, Oliva J, Antoñanzas F, García-Altés A, Gisbert R, Mar J, Puig-Junoy J: Spanish recommendations on economic evaluation of health technologies. European Journal of Health Economics 2010, 11:512-520.[71]
Sweden	Pharmaceutical Benefits Board. General guidelines for economic evaluations from the Pharmaceutical Benefits Board. 2003.[72]
UK	National Institute for Health and Clinical Excellence (NICE). Guide to the methods of technology appraisal. June 2008.[56]
US	Academy of managed care pharmacy (AMCP). The AMCP Format for Formulary Submissions, version 2.1. A Format for Submission of Clinical and Economic Data in Support of Formulary Consideration by Health Care Systems in the United States. April 2005. [70]

Guidelines in other languages[80-87] or that do not include guidance on HRQoL measurement[88-90] have not been included in this overview.

APPENDIX 6: RETRIEVED RECOMMENDATIONS ON HOW TO MEASURE QOL[g] (CONDUCTED DURING GUIDELINE ELABORATION IN JA 1)

Guidelines with explicit or implicit preference for (a) specific (types of) HRQoL instrument(s)	
Australia (PBAC)[57] and (MSAC)[58]	The generally preferred method of measuring QALYs is by the *repeated application of a valid, reliable and responsive multi-attribute utility instrument (MAUI) questionnaire* to participants in a randomized double-blind trial, together with the application of an appropriate scoring algorithm. *Acceptable MAUIs are the Health Utilities Index (HUI2 or HUI3), the EQ5D ('EuroQol'), the SF-6D* (a subset of the Short Form 36, or SF-36) *or the Assessment of Quality of Life (AQoL) instrument*. Currently, there is insufficient basis for a preference to be expressed between these MAUIs.
Baltic countries[59]	Origin of the utilities used in the analysis should be explained and the instrument, whether generic or disorder-specific, used for measurement of quality of life has to be validated. It is *recommended to use the EuroQoL and the Health Utility Index methods*.
Belgium[60]	In the reference case, *a generic health-related quality of life measure should be used for the description of health states*. The health state description should be done by patients on a generic descriptive system, such as the EQ-5D or SF-36. Other instruments exist, e.g. the HUI or QWB scale, but these have not been validated in Dutch or French for Belgium. Health state descriptions with the EQ-5D or SF-36 in similar patient populations in other countries may be used, provided that the criteria for valuation are fulfilled. If it is thought that a generic instrument is insufficiently sensitive to relevant changes in health in a specific disease, *additional (disease-specific) quality of life results can be described in separate analyses*.
Canada[54]	*Preference-based measures provide a summary score that numerically reflects the HRQL, and are the only approaches that are suitable for use in a cost-utility analysis (CUA).* If HRQL is being measured in a prospective study, it is *advisable to include a preference-based measure where the intention is to undertake an economic evaluation*. Preference-based scores (i.e., the quality-weight for a CUA) can be measured directly or indirectly. - Three methods are used for the direct measurement of preferences: standard gamble, time trade-off, and visual analogue scale. Analysts prefer the standard gamble approach because of its strong normative foundation in von Neumann-Morgenstern utility theory. There are arguments against the superiority of the standard gamble approach. Visual analogue scales are inappropriate to use alone because of well known biases. - *"Off the shelf" instruments* are available for obtaining utilities without undertaking direct measurement. Some widely used instruments in this category are the *Health Utilities Index* (HUI), the *EQ-5D*, the *SF-6D*, and the *15D*. These instruments use preferences from the "informed" general public, which is the appropriate source to use for collective resource allocation purposes. *Analysts are encouraged to use indirect measurement instruments*, because they are easy to obtain, compare, and interpret. Some studies use expert judgment with an extensive sensitivity analysis as the source of quality-weights. This approach is not favoured.
Denmark[13]	Depending on the objective of the study, it is *often recommended to use both a generic and a disease-specific instrument* in the same study. *Certain generic instruments are also index measures and can be used for calculating QALYs* in cost-utility analyses: Besides functioning as a profile measure for describing patients' self-assessed health, certain health status instruments can also be used as utility measures in economic

[g] Literal quotes from the original guidelines. Specific (types of) instruments that are mentioned explicitly are underlined and *italic*, even though they are not underlined or italic in the original guidelines.

	evaluations. In this context, the instrument must be able to generate a simple preference-based index score (on a scale of 0-1) for health status, e.g. for various patient groups or treatment alternatives. The original document presents five generic instruments, which have either been developed primarily as utility measures (*EQ-5D, 15D, HUI, AQoL*) or aim at this on the basis of broad use as a profile measure (SF-36). Whereas 15D and AQoL are relatively new, the other instruments are widely used and perform well with respect to validity and reliability.
France (CES)[69]	Utilities are derived from surveys of the general population using preference revelation methods. The utility so calculated is thus an evaluation of the average preference rating that would be attributed to this health state by a random sample of the general population. *The main methods used are QWB (Quality of Well-being), HUI (Health Utility Index) and Euroqol EQ-5D*… These methods cumulate the assumptions of the techniques used to determine utility scores for the general population (standard gamble or time trade-off) and the assumptions of multi-attribute utility, but they have the advantage of simplicity in that they avoid the need for population surveys.
France (HAS)[61]	Dans l'analyse de référence, les scores de préférence devraient être obtenus grâce à une méthode reposant, d'une part, sur *un instrument descriptif générique et validé en France* et, d'autre part, sur un système de scorage garantissant une mesure d'utilité ou de valeur également validé en France. De nombreux instruments sont disponibles pour décrire les états de santé, sous le vocable de questionnaires de qualité de vie. *Très peu d'instruments sont en fait directement utilisables dans une évaluation coût-utilité* : la plupart d'entre eux peuvent être valides pour décrire les états de santé, mais *ils ne disposent pas d'un système de scorage adapté à l'évaluation coût-utilité (l'exemple le plus connu étant le SF 36*). En aucun cas, des scores issus d'un tel instrument descriptif de qualité de vie associée à la santé, non fondés sur les préférences, ne peuvent être utilisés pour le calcul d'un nombre de QALYs. La description des états de santé repose en priorité sur une étude prospective auprès d'un échantillon français, à partir d'*un instrument générique disposant d'une fonction de scorage adaptée à l'évaluation coût-utilité, tels que l'EQ-5D ou le HUI 3* . D'autres instruments existent (QWB, SF6D), mais n'ont pas été validés pour la France. Les *instruments* de mesure de la qualité de vie *spécifiques à une pathologie ne sont pas recommandés dans l'analyse de référence*.
Germany (Hanover Consensus Group) [62]	When applying cost-effectiveness and cost-utility analyses, the selection of outcome parameters is of key importance. If utility values (e.g., QALYs) are included in a study, these should be *preferably determined through direct generation of individual values via standard gamble, the time-trade-off method, or with validated, preference-based, generic instruments (e.g., EQ-5D or SF-6D)*. The validation and preferences of these questionnaires should be based on a representative population sample from Germany. In specific study situations the application of a visual analog scale (VAS) can also be appropriate, if the validity of the information can be justified. In well-substantiated exceptions, it is acceptable to deviate from population-based preference values.
Hungary[63]	Disease-specific and non-utility-based generic quality-of-life (QoL) measurement that expresses health improvements in scores or in clinically important minimal changes is increasingly used in cost-effectiveness studies. Validated versions exist of several disease-specific and generic (such as the Medical Outcomes Study 36-Item Short-Form, SF-36) QoL questionnaires in Hungary. *Analyses using non-utility-based generic questionnaires are unsuitable for cost-utility studies*. [However, there are new mapping studies that calculate a formula between the non-utility-based questionnaire results and utilities, and these are increasingly used for cost-utility studies. The most commonly applied mapping formula was developed by John Brazier (1998 and 2001) and enables researchers to elicit utility values (SF-6D) from the SF-36 questionnaire. Because a Hungarian version is available for the SF-36, this method might become important for Hungarian cost-utility studies in the future.] It is *preferable* for health status weights for QALY calculations to be derived from the *use of utility-based health-related QoL questionnaires*, for which preference values were elicited by general population surveys. Internationally recommended questionnaires include the EQ-5D, Health Utilities Index, Quality of Well-Being Scale, and Years of Healthy Life.
Ireland[64]	*Use of an indirect preference-based measure, such as the EQ-5D or SF-6D, is recommended for the reference case* as these measures have wide-spread availability,

	are easy to use and interpret and because they are based on preferences of the general public. The population from which these preferences are derived should be clearly described along with their relevance to the Irish population. *Alternatively, direct HRQoL methods such as time trade-off or standard gamble may be used* provided these have been gathered in a relevant population.
ISPOR[65]	Because cost–utility analyses are widely accepted, we recommend that analysts collect preference weights as part of clinical trials. The *most common method of assessing preferences is the use of a preference-weighted health state classification system such as the EuroQol-5D, one of the three versions of the Health Utilities Index, or the Quality of Well-Being Scale*. Analysts may also consider the inclusion of a rating scale to measure patient-based preferences. Frequency and timing of these assessments should capture changes in patients' quality of life that may be affected by the treatment but will be influenced by the disease severity of the study population, the study duration, the timing of trial visits, and patient burden. Other options for collecting preference data include direct-elicitation methods such as standard gamble or time-tradeoff exercises. These methods have certain theoretical advantages; however, their use in clinical trials is often difficult.
Italy[66]	It is *suggested to simultaneously apply*, if possible, *general instruments*, such as, for example, the Short Form 36 (a widely used quality of life questionnaire which in36 questions gives the health profile according to six attributes: physical, role-emotional, social, mental health, health perceptions, and pain), *specific instruments* for the group of patients being analyzed or for the pathology, *and instruments for surveying preferences/utility* such as the Health Utility Index, similar in principle to the EuroQol but more complex with seven attributes and up to five levels for each of them, the EuroQol […], quality of well being, and so forth. For the economic evaluation which uses the cost-utility analysis, a utility value is attributed to health conditions using specific techniques, preferably "standard gamble" and "time trade-off".
The Netherlands[67]	*Descriptive quality-of-life questionnaires* (generic, illness-related and domain-related) *cannot be used as a measurement of effect in pharmacoeconomic evaluations*. It is often useful to add such questionnaires to the study, particularly in order to determine the health domains where alterations occur. In the case of empirical studies, *health assessment systems such as EQ-5D, HUI 2/3* which are completed by patients or by proxy, *can be used*. The replies to the questions are subsequently used to calculate assessments with the aid of algorithms.
New Zealand[55]	There has been much debate in the literature regarding the most appropriate tool for measuring preferences in health gains. Given the multidimensional nature of HRQOL, *no single measure has been (or is likely to be) accepted as the gold standard*. Instruments available include (but are not limited to) the EuroQol 5D (EQ-5D); Health Utility Index (HUI); Short-Form 36 (SF-36); Short-Form 6D (SF-6D); Quality of Well Being index (QWB); Quality of Life and Health Questionnaire (QLHQ); Rosser-Kind Index; Assessment of Quality of Life instrument (AQOL); Sickness Impact Profile (SIP); and Index of Health Related Quality of Life (IHRQOL). *The New Zealand EQ-5D Tariff 2 should be referred to first when measuring health-related quality of life*, and should be used to describe the health states. The Global Burden of Disease disability weights should be used to check for consistency with the estimated EQ-5D values. The New Zealand EQ-5D Tariff 1 should be included in the sensitivity analysis. Utility values may be obtained through questioning the general public, patients, physicians, and/or related health professionals and caregivers. This can be done using the Standard Gamble (SG), Time Trade-Off (TTO) or VAS techniques. However time constraints mean this is often not a feasible option at PHARMAC.
Norway[68]	A number of *indices* have been developed to assist in performing QALY-calculations, for scoring complex health profiles on a life quality scale from zero to one (e.g. 'EuroQol').
Portugal[79]	Descriptive tools should be presented, as they are an asset to any assessment study. Whenever possible, it is advisable to *present results based on generic measurements* (such as the SF-36, Sickness Impact Profile or Nottingham Health Profile) *and specific instruments* (i.e. those designed to measure concrete health problems) at the same time. Descriptive instruments cannot replace value-based ones, however, and do not constitute an adequate base for a cost-utility study. The literature on the comparative advantages of any of the value-based methods does

	not enable us to say that any one of them is better than the others. We cannot, therefore, exclude the possibility of using any of them, provided that it has been validated for Portugal and we can justify that the choice is appropriate for the study.
Spain[71]	*Indirect methods to measure utilities are preferable*, as they are easier to obtain, compare, and interpret. However, these considerations do not rule out direct measurements when their use and scientific validity is justified for the study in question.
Sweden[72]	QALY-weightings should be based on methods such as the Standard Gamble (SG) or Time-Trade-Off (TTO) methods. In a second instance, QALY-weightings should be based on the rating scale method. QALY-weightings can be based either on direct measurements with the above-mentioned methods or indirect measurements (where a health classification system such as EQ-5D is linked to QALY-weightings). *QALY weightings based on appraisals of persons in the health condition in question are preferred before weightings calculated from an average of a population estimating a condition depicted for it* (e.g. the "social tariff" from EQ-5D). Using weightings for current health conditions collected from previous studies may be a solution.
UK[56]	For the reference case, the measurement of changes in HRQL should be reported directly from patients and the value of changes in patients' HRQL (that is, utilities) should be based on public preferences using a choice-based method. … *To quantify the effects of technologies on HRQL, the EQ-5D (a standardised and validated generic instrument) is preferred*. Different classification systems produce different utility values; therefore, results from the use of different systems cannot always be compared. Given the comparative nature of the Institute's work and the need for consistency across appraisals, a single classification system, the EQ-5D, is preferred for the measurement and valuation of HRQL. … The methods to elicit EQ-5D utility values should be fully described. When EQ-5D data are not available or are inappropriate for the condition or effects of treatment, the valuation methods should be fully described and comparable to those used for the EQ-5D. Data collected using condition-specific, preference-based measures may be presented in separate analyses. It is recognised that the current version of the EQ-5D has not been designed for use *in children*. When necessary, *consideration should be given to alternative standardised and validated preference-based measures of HRQL, such as the Health Utility Index 2 (HUI 2)*, that have been designed specifically for use in children.
US[70]	Estimates obtained by time trade off methods reflect respondents' attitudes to time as well as their attitudes to the health state being valued. Likewise, estimates obtained by standard gamble methods reflect respondents' attitudes to risk as well as their attitudes to the health state being valued. *Economists are still debating which approach is most desirable*. Another *cheaper approach* is to include in the clinical trial a generic health state preference instrument, such as the *EuroQoL (EQ5D) or McMaster health utilities index*. The responses from patients to a simple questionnaire can then be expressed as a health state preference value by reference to pre-scaled responses (obtained by standard gamble or time trade oft) from a relevant reference group. *Values can be provided by the population at large* or by a sample of patients with the condition for which the treatment is being evaluated. The choice depends on the perspective of the study. *If the issue is allocating resources between competing programmes* the former might be used; if it is deciding the best way to treat a given condition the latter might be used. In reporting their results authors should explain why a particular source of values has been used.
Guidelines without explicit or implicit preference for (a) specific (types of) HRQoL instrument(s)	
Austria[76]	If the quality of life is to serve as an outcome variable, it must be ensured that the variable measured is also an appropriate measure for comparing the chosen treatment alternatives. Outcomes of this kind, in other words utilities, can be determined in the following way: - specific scales (rank scales), - game theory procedures (e.g. standard gamble, time-trade off, etc), - psychometric scale procedures which include generic and disease-specific procedures as well as one-dimensional and multidimensional instruments. These individual measures are suitable for combining with quantitative objective measurements such as survival time in the form of quality adjusted life years (QALYs).
Germany (IQWIG)[77]	In the assessment of QoL and patient satisfaction, only instruments should be used that are suited for application in clinical trials and have been evaluated accordingly.

Poland[78]	The state of health utility values can be sought based on data from published research. It is admissible to perform the quality of life measurement in the patient population or the preference measurement in the general population. It is a requirement to maintain the standards accepted in the literature and to present a detailed description of the methods used.