Title: What do HTA agencies need for generating health-related quality of life evidence? Findings from a global survey

Authors: Annushiah Vasan Thakumar[1], Paula Lorgelly[2], Louise Longworth[3], Lucila Rey Ares[4], Fredrick Purba[5], Dominik Golicki[6], Federico Augustovski[7], Kim Rand[8], Rosalie Viney[9], Nick Bansback[10*], Nan Luo[1*]

[1]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore; [2]School of Population Health and Department of Economics, University of Auckland, Auckland, New Zealand; [3] Arrow Health Economics, London, England, UK; [4]Pfizer, Buenos Aires, Argentina; [5]Department of Psychology, Faculty of Psychology, Universitas Padjadjaran, Bandung, Indonesia; [6]Department of Experimental and Clinical Pharmacology, Medical University of Warsaw, Warsaw, Poland; [7] Health Technology Assessment and Health Economics Department, Institute for Clinical Effectiveness and Health Policy (IECS-CONICET) - University of Buenos Aires;[8]Maths in Health B.V., Netherlands; [9]Centre for Health Economics Research and Evaluation, University of Technology Sydney; [10]School of Population and Public Health, University of British Columbia, Vancouver, Canada

Objectives: Health technology assessment (HTA) agencies routinely utilize health-related quality of life (HRQoL) data to inform healthcare decision-making. While many have published methods guides, little is known about their preferences for HRQoL evidence, and their perspectives on the quality of the HRQoL data in HTA practice. The overall aim of this study is to understand the practices, views, and needs of HTA practitioners around the world regarding measurement, valuation, and use of HRQoL data.

Methods: We identified a total of 60 countries where national HTA agencies, bodies, or committees existed, and/or HTA was used to inform healthcare decision-making. Utilizing the EuroQol Group membership, we invited HTA practitioners in those countries to complete an online survey inquiring their experiences and views on: i) utility instruments; ii) utility elicitation methods, iii) health preference sources, iv) data quality and appropriateness, and v) perceived research priorities. For questions using a Likert-type response scale, we used the mode (or median if no or multiple modes) of all responses for that country, and reported the median based on all country responses. We performed descriptive analyses of the overall sample, and then we examined the response differences across six regions (Commonwealth - AU/CA/NZ/UK, Western Europe, Central/Eastern Europe, Asia, Latin America, and Middle East/Africa). We pooled responses to open-ended questions for content analysis.

Results: In total, 238 individuals from 45 countries completed the survey. The mean number of responses per country was 5.28 (SD: 4.45), with Vietnam, England, and South Korea being the top three. Overall, most responses came from government employees (72%) and 90% were involved in QALY-related work. The top three most frequently used utility instruments were EQ-5D (“very often”), SF-6D, (“occasionally”) and EQ-5D-Y (“occasionally”). The top three most frequently used utility elicitation methods were TTO, VAS, and SG, all of which were “often” used. Health-state preferences of the general public of another country (“often”) was more frequently used than the preferences of the local public (“occasionally”). The data quality issues that “often” arose across regions were the poor sample representativeness and small sample size of utility data (UD), poor matching of available health state UD with those of the CEA models, and the use of UD from multiple methods/instruments in a same model. In Asia and Western Europe, the top voted research priority was to develop utility instruments to capture the health care and social care impact; in the Middle East/Africa and Central/Eastern Europe, it was to make more recent UD available; in the Commonwealth countries, the priority was to develop utility instruments to capture the impact of treatment on carers and caregivers; in Latin America, the primary interest was to develop utility instruments address inequality in care. In four regions, utility instruments for children was the second highest research priority.

Conclusion: The survey filled important knowledge gaps with regard to the current practices of measurement and valuation of HRQoL in HTA and HTA practitioners’ views on instruments, methods, and data related issues and needs for generating HRQoL evidence.

Introduction

Health technology assessments (HTA) provide a comprehensive framework for integrating evidence of economic, social, and health consequences and effects into the decision-making process (Drummond et al., 2015). Globally, healthcare systems increasingly rely on HTA to guide resource allocation (Fontrier et al., 2022; Teerawattananon et al., 2019; World Health Organization, 2024). In HTA, cost-effectiveness analysis of new health technologies using quality-adjusted life years (QALYs) as the health effects measure, or cost-utility analysis (CUA), is the most recommended form of economic evaluation (Fontrier et al., 2022; Sharma et al., 2021) . However, estimating QALYs is technically challenging as it requires measuring and valuing patients' health-related quality of life (HRQoL). Consequently, analysts often rely on HRQoL and health-state utility (HSU, valuation of HRQoL) data from the literature, which can be limited in both quantity and quality.

HSU data can be estimated directly by describing a specific health condition using vignettes and capturing preferences through elicitation methods such as the standard gamble (SG), time trade-off (TTO), and discrete choice experiments (DCE). However, the indirect method of using standardised preference-weighted HRQoL instruments such as EQ-5D to obtain HSU data is more commonly used. In recent years, research on new methods and instruments for measuring and valuing HRQoL has been very active. For example, there is an increasing application of DCE in valuation of HRQoL (H. Wang et al., 2023). Since DCE can be used in self-administered online surveys, it significantly lowers costs and shortens data collection timelines compared to traditional preference elicitation methods, which entail interviewer administration. Consequently, it has enabled more preference-weighted instruments to be developed, including disease-specific instruments such as the QLU-C10D and FACT-8D (King et al., 2024; Shiroiwa, King, et al., 2024; Xu et al., 2024). Other examples include bolt-on research to enhance existing EQ-5D instruments (Kangwanrattanakul & Phimarn, 2019; P. Wang et al., 2023) and the EQ-HWB instrument to broaden the 'Q' in QALY to cover both health and well-being (Brazier et al., 2022). Despite their attractive features, these new methods and instruments also have disadvantages. For instance, DCE and widely used traditional preference elicitation method, such as TTO, have been shown to have poor agreement (Augustovski et al., 2020; Shiroiwa, Murata, et al., 2024). Given the wide range of methods and instruments available, and the proliferation of new ones, it remains unclear how existing instruments and methods are being used, and whether or when the new ones will be adopted by HTA practitioners for routine use.

HTA agencies or bodies (hereafter referred to as 'HTA agencies' for simplicity) are authoritative entities or divisions responsible for HTA evidence generation and/or evaluation. Although their size, capacity, and mandate vary, HTA agencies play a pivotal role in using HTA evidence to inform healthcare decisions. Moreover, because of their authority, HTA agencies' practices and views on evidence generation methods significantly influence practice and therefore are highly valuable to researchers. One source of information for understanding HTA agencies' views and preferences is the methods guide they publish (ISPOR, 2024). However, this approach is not optimal, particularly if the interest concerns HRQoL measurement and valuation methods. First, published methods guides may be outdated as practices and views constantly evolve, but the guides are not frequently updated. Second, the guides for certain methodological aspects may be ambiguous or missing. Third, some recommendations in methods guides may not reflect unanimous opinions within HTA agencies. Last but not least, many HTA agencies have not published methods guides.

An alternative way to understand HTA agencies' views and preferences regarding HRQoL measurement and valuation is to survey HTA agency personnel responsible for preparing or reviewing HTA dossiers. This approach offers the advantage of obtaining first-hand, contemporary, and detailed information, which could be very useful for HRQoL researchers to set their priorities. However, to the

best of our knowledge, such an approach has not been explored before. Therefore, the aim of this study was to conduct an international survey of HTA agency personnel. Our primary objective was to understand HTA practitioners' views on and needs for HRQoL-related methods and instruments. Our secondary objective was to understand their current practices and views on the availability and quality of HRQoL data and research priorities.

Methods

We conducted a cross-sectional online survey of HTA agency personnel from April 2023 to January 2024. In order to achieve a representative sample, we identified and invited survey respondents using personal contacts. The study received ethics approval from the Institutional Review Board of National University of Singapore (IRB number: NUS-IRB-2022-426).

Sampling and recruitment design

We employed a two-stage recruitment procedure. In the first stage, we identified target HTA agencies. HTA agencies were defined as independent organizations or governmental divisions authorized to generate and/or review HTA evidence for market access or reimbursement decisions at the health system level. We targeted to survey HTA agencies in 50 countries. We used the search strategy adopted by Kennedy-Martin et al. who identified 46 countries where HTA agencies existed in 2019 (Kennedy-Martin et al., 2020). We complemented this strategy with other sources including the Gear4Health database (Health Intervention and Technology Assessment Program, 2024), ISPOR's Pharmacoeconomic Guidelines Around the World database (ISPOR, 2024), the WHO Health Technology Assessment and Health Benefit Package Survey 2020/2021 webpage (World Health Organization, 2024), the INAHTA Members List (INAHTA, 2020), and consulted with EuroQol members and colleagues for additional input.

In the second stage, we identified members of the EuroQol Group or their acquaintance as recruiters for the countries we intended to survey. Recruiters sent invitations with country-specific survey links to potential respondents working in the target HTA agencies. Alternatively, they identified a contact person within each HTA agency to extend the survey invitations internally. A survey administrator monitored survey yields and prompted recruiters and/or contact persons to send reminders to potential respondents weekly for at least three consecutive weeks after the survey commenced in each country.

Participants

Rather than surveying official representatives of the target agencies, we sought to recruit all personnel involved in handling of CUA or HRQoL evidence, specifically, individuals responsible for reviewing, generating, and/or using QALY-based evidence. However, those involved in HTA-related work but not directly handling QALY-related tasks were also invited to participate if interested.

The inclusion criteria for the survey were: 1) being an employee of an HTA agency (e.g., governmental or public agency, division, body, or committee) whose responsibilities included evaluating or appraising health technologies for the purpose of listing/delisting, reimbursement, or pricing/repricing at the national level, or being a contracted professional, consultant, or advisor to such HTA organisation(s); 2) being able to understand the English survey form and complete open-ended questions in their language of choice; and 3) providing informed consent.

Survey form

After obtaining consent, participants were invited to complete an electronic survey form powered by Qualtrics anonymously at their convenience. The survey form was developed from scratch by the study team. The target of the development was a short survey that can be self-completed by most respondents in no more than 20 minutes. An iterative question drafting procedure was used, with multiple rounds of pilot-testing conducted with personnel from HTA agencies in Singapore, Indonesia, Canada, England, Norway, Australia, New Zealand, and Argentina until the development goal was achieved.

The final survey started off with screening questions followed by a consent-taking question. Eligible and consenting participants were invited to complete six sections of questions revolving around their experience with and opinions on Utility Instruments, Elicitation Methods, Data Source, Data Quality and Appropriateness, and Research Topics of Importance.

Section 1 inquired about the frequency of using or reviewing data collected from nine utility instruments (UI), including AQOL, EQ-5D (EQ-5D-3L/ EQ-5D-5L), EQ-5D-Y, EQ-HWB, Bolt-ons, HUI, PROPR, QWB, and SF-6D. Section 2 addressed the frequency of using or reviewing data collected using six elicitation methods (EM) including best-worst scaling (BWS), discrete choice experiment (DCE), person trade-off (PTO), standard gamble (SG), time trade-off (TTO), and visual analogue scale (VAS). In the Section 3, we surveyed the frequency of using preference data from different sources (DS), including data from the general population of one's own country or other countries, and data from patients of one's own country or other countries. Section 4 covered the frequency of encountering five concerns related to HRQoL/HSU data: i) The patient samples for data collection are not representative; ii) The health states for which HSU data was available do not match the health states in the CEA model; iii) The samples for data collection are too small; iv) The data is too old; The HSU values of different health states used in the same model are derived using different methods/instruments.

In Section 5, respondents rated the importance of seven specific research topics related to HSUs, selecting up to three (of the seven research topics) they felt important. These topics were related to i) social care: to develop utility instruments to capture the impact of both health care and social care; ii) children: to develop utility instruments to capture the impact of treatment on children and adolescents; iii) caregivers: to develop utility instruments that capture the impact of a treatment on carers and caregivers; iv) specificity: to develop utility instruments that capture the impact of treatment on more specific aspects of health (e.g. vision, hearing, etc.); v) recent tariff: to make more recent utility data and value sets/tariffs available; vi) inequality: to develop utility instruments that can address inequality in care; and vii) minority: to develop utility instruments that can reflect the health preferences of minority groups (e.g. indigenous populations) or rural population. The final section assessed the participants' demographic and professional characteristics.

We used a 4-point Likert-type response scale to assess frequency in Sections 1 to 4 ("never"/ "not sure", "occasionally", "often", and "very often"). In Sections 1 to 5, text fields were provided for respondents to explain their responses and elaborate on other methods, instruments, concerns, or research topic not covered in the survey.

Statistical analysis

Descriptive analyses were performed for responses to closed-ended survey questions. For Likert-type questions in Sections 1 to 4, we first used the mode (or median if no or multiple modes were present) as the summary of the responses for each country, and then used the median of relevant country summaries as the summary of the responses for six regions (Commonwealth - Australia/Canada/New

Zealand/United Kingdom, Western Europe, Central/Eastern Europe, Asia, Latin America, and Middle East/Africa).

To analyse the nominated research priorities in Section 5, we excluded respondents (N=16) who did not endorse the importance of any research topics. For each of the remaining respondents who endorsed 1 to 3 research topics, each endorsed topic received a score of 1/N (N is the number of topics a respondent endorsed) and each not endorsed topic received a score of 0. Subsequently, we calculated a country-specific importance score for each research topic by averaging the scores from all respondents in the relevant country. Once the importance scores for all countries were calculated, a regional score was calculated by averaging scores from relevant countries in the region and a global score was calculated by averaging the regional scores.

All statistical analysis was performed using STATA v14 (StataCorp, College Station, TX, USA). We pooled qualitative responses to open-ended questions for content analysis.

Results

Sample characteristics

Of the 60 countries enlisted and approached, the survey was distributed in 49 countries. The remaining 11 countries were excluded for various reasons including non-responsiveness (N=2), infancy of HTA (N=2) or the non-use of CUA (N=3), declination/technical difficulty (N=2) or political turmoil (N=2). In 45 of these countries, we received at least one completed survey (median: 4; interquartile range: 2 to 6), while in the remaining 4 countries, there were zero responses despite multiple follow-ups. Appendix 1 outlines the distribution of responses and reasons for non-responses from the 11 countries approached. In total, 238 individuals in 45 countries and 65 HTA agencies completed and submitted the survey (Table 1). Overall, the majority of the responses came from government employees (71.9%), had at least 4 years of experience in HTA work (58.8%), were female (57.6%), aged 26-55 (88.2%), and holding a doctorate degree (51.7%). Additionally, 81.1% of the respondents reported presence of QALY estimation guidance in their work setting, and 89.5% were involved in QALY-related work responsibilities. More than half (61.3%) reviewed QALY-based cost-utility evidence submitted by industry or contractors, and 91.2% performed HTA work at the national level. Almost half (49.6%) identified themselves as health economists. Pharmaceuticals were the most common health technology appraised (83.6%), with oncology being the most frequently reported therapeutic area of HTA work (70.2%). Of the 238 respondents surveyed, 25 did not have work responsibilities related to reviewing, generating, and/or using QALY-based cost-utility evidence. These respondents mainly came from Vietnam (N=10), Slovenia (N=3), Austria (N=2), Colombia (N=2), and South Africa (N=2).

Use and importance of utility instruments

Overall, the top three most frequently used utility instruments by HTA practitioners involved with QALY-related work (n=213) were the EQ-5D (“very often”), SF-6D, (“occasionally”) and EQ-5D-Y (“occasionally”) (Table 2). Regionally, the use frequency trend was consistent with a few exceptions. In Western Europe, the use frequency for the EQ-5D-Y was “never”, while in Latin America, it was “often” for SF-6D (Appendix 3).

Collectively, respondents from the various HTA agencies listed 29 different measures/ techniques when asked to name other instruments they have encountered during their HTA work (Appendix 2). These included both preference-based measures (e.g., EORTC QLU-C10D), and profile-based measures (e.g., Patient Health Questionnaire-9) for specific populations or conditions.

Respondents across the regions (Table 3) generally agreed that the choice of UI matters (87.7%), ranging from 66.7% (Middle East/ Africa) to 91.2% (Western Europe). Content analysis revealed that the EQ-5D (EQ-5D-3L/ EQ-5D-5L) instrument was most often selected as the more fit-for-purpose instrument (Table 3) mainly due to its low respondent burden, good psychometric properties, availability of value sets, HTA guide recommendations, and its wide usage that promotes comparability and consistency in the HTA setting.

Use and importance of preference elicitation methods

The top three most frequently used utility elicitation methods were TTO, VAS, and SG, all of which were "often" used (Table 2). Across regions, the TTO was either "often" or "very often" used to inform decision-making. SG and VAS were only "occasionally" used in Western Europe and the Commonwealth. DCE was "occasionally" used in most regions, but its frequency ranged from "often" and "very often" in the Middle-East/Africa. The use frequency of BWS and PTO ranged between "never" and "occasionally" in all regions (Appendix 4).

Six respondents mentioned three other preference elicitation techniques they had come across in their work, including expert elicitation/qualitative methods, ranking methods, and online elicitation of Personal Utility Functions (OPUF).

Respondents across the regions generally agreed that the choice of EM matters (79.8%), ranging from 69.6% (East and Central Europe) to 86.7% (Western Europe). A total of 55 respondents mentioned the time trade-off (TTO) method as the more fit-for-purpose EM (Table 3). Common reasons (Table 3) included that the TTO had a strong theoretical foundation, involved trade-offs, produced cardinal utilities, was a validated method, and was easy to use. However, the high cognitive burden of the method was often recognized as a limitation of the technique. Many respondents (n=47) mentioned that the EM would need to depend on the disease area, the context of the study, and the availability of evidence.

Use and importance of health preference data sources

The general public of another country was more frequently used ("often") than the preferences of the local public ("occasionally") (Table 2). Region-wise, in Western Europe and the Commonwealth countries, the general population of one own's country were most frequently used ("often"). In Asia, the general public (own and other countries) and patients (own country) had parallel high usage ("often"). In Latin America, Middle-East/Africa, and Central/Eastern Europe, data of other countries (both general population and patient values) were "often" or "very often" used while the values of one's own country were only "occasionally" or even "never" utilized (Appendix 5).

Respondents across the regions had a strong consensus that health preference data source matters (91.2%), ranging from 86.7% (Middle East/Africa) to 96.8% (Western Europe). Of the 144 qualitative responses received, 81 respondents mentioned that the general public's preferences should determine the utility values. Commonly cited reasons include HTA/country guide's recommendations, consistency reasons, taxpayers being the most appropriate in a publicly funded healthcare system, and the tendency of patients to adapt to disease, thereby underestimating the disutilities. Conversely, 53 respondents felt that patients' preferences should determine the values, as they reflect the patient voice and capture the disease experience better. Ten people felt it should come from both, either combining both preferences or using them to address different research questions. A total of 79 respondents mentioned that these preferences should come from one's own country population as it reflects the culture and context of the preferences more accurately. None explicitly preferred utility values from other countries over their own.

Use of data with quality concerns

Data quality issues that were "often" encountered across regions included poor sample representativeness, small sample size, poor matching of available data with that needed for CEA models, and data used in the same CEA model generated from multiple EMs/IUs (Table 2). These concerns were generally shared across regions. The issue with using outdated data was less of a concern, with most regions reporting it only "occasionally" (Appendix 6).

Content analysis revealed that among the other issues related to data quality, the validity, generalizability, and availability of HSUs were the main concerns (Table 3). These concerns included whether the results remained valid when HSUs from another population's (other country, adult/children) were used for a different population. This was highlighted as a major concern given that country- or disease-specific utility data may not always be available. The generalizability of available HSUs was also a concern. When the targeted HSUs were not available, analysts resorted to other techniques including mapping and expert elicitation, but respondents questioned the appropriateness of these methods. Another concern was failing to address the adaptation of patients to the disease. This often resulted in poor responsiveness of utility values to change and sometimes led to illogical inconsistencies where the general public exhibited lower HSUs than patient groups.

Research priorities

The top three research priorities (Table 4) globally were i) to make more recent utility values available (recent tariff, IS=0.20), ii) to develop utility instruments for children (children, 0.19), and iii) to develop utility instruments to capture both healthcare and social care impact (social care, 0.17). Table 6 depicts the research priority by country and region. In Asia (importance score, 0.21) and Western Europe (0.33), the top-voted research priority was to related to social care; In the Middle East/Africa (0.33) and Central/Eastern Europe (0.31), the primary research priority was related to recent tariffs; In the Commonwealth (0.23), the priority was to develop utility instruments to capture the impact of treatment on carers; In Latin America (0.22), the top research topic was to develop utility instruments address inequality in care. In all regions except for Western Europe and Latin America, children were the second highest research priority (IS = 0.18 to 0.30). In Asia (0.17) and Western Europe (0.15), recent tariffs remain the third highest research priority.

Other research topics of priority (Table 5) mentioned by the respondents included developing population-specific HSUs (i.e. country-specific data, rare disorders, patient value sets), making HSUs transferable across countries and evidence more readily available, and addressing methodological issues including incremental cost-utility ratio threshold, capturing productivity losses and double-counting, capturing long-term effects, and impact of shifting to the EQ-5D-5L instrument.

Discussion

In this study, we obtained global insights into the practices, views, and needs of HTA agency personnel across six regions on a broad range of topics related to the measurement and valuation of health. Additionally, we explored data quality issues encountered by the HTA practitioners and research topics they perceived as important.

In general, the respondent's practices in terms of the choice of utility instruments and elicitation methods were consistent with HTA guide's recommendations of using EQ-5D instrument as the reference case and choice-based preference elicitation methods (ISPOR, 2024; Kennedy-Martin et al., 2020; Sharma et al., 2021) .

Interestingly, only in Western Europe and the Commonwealth were local public health-state preferences used more frequently than those of foreign sources, possibly highlighting the prevalent issue of data availability in the field of HTA (O'Rourke et al., 2020; Zisis et al., 2024). Additionally, patient preference data is only occasionally used in Western Europe and the Commonwealth, which is consistent with the recommendations of most of the HTA guide and findings from empirical studies (Hiligsmann et al., 2024; van Overbeeke et al., 2021). However, patient preference data is often used in regions outside Western Europe and the Commonwealth, perhaps motivated by interest in patients' view and/or unavailability of preference data from the general public. It is interesting that DCE was reported to be only occasionally used in all regions except for Middle East/Africa. It is surprising because DCE data is used to generate most of the EQ-5D-5L value sets which have become widely used. It is possible that HTA practitioners are not familiar with this new valuation method or the technical details of EQ-5D-5L valuation. While the views of the respondents regarding the choice of instrument, methods, and health preference data source generally reflect recommendations of the HTA guide, some respondents preferred patients' preferences and argued that the patient voice and disease experience are important. Additionally, some respondents expressed concerns about the shortcomings of the widely used EQ-5D and the TTO method. The main disadvantage of the EQ-5D, as cited by respondents, is poor responsiveness in certain health conditions, while TTO poses high cognitive burden to respondents. Similar concerns have been documented in the literature (Brazier et al., 2017; Feng et al., 2021; Qian et al., 2020) . Interestingly, those respondents did not consider using new instruments such as EQ-HWB or bolt-ons or new valuation methods such as DCE, perhaps because they were not aware of or familiar with those new alternatives.

This study found pervasive suboptimal use of HRQoL and HSU data in current HTA practice across regions. These data related issues included sample representativeness, sample size, and use of matched data and data generated using different instruments and methods. These issues are mostly likely due to scarcity of quality and appropriate data, underscoring the need for research to make such data available (Claire et al., 2023) . Respondents' comments suggested that HTA practitioners are aware of the data quality issues and the validity of the methods used to address these issues. This finding echoes the increasing concerns about the methodological rigor in using HSU data for cost-utility analysis (Ara et al., 2018; Ara et al., 2020; Brazier et al., 2019) . However, the magnitude of such issues is largely unknown. A recently conducted systematic review of published cost-utility analyses from Asia found that the overall reporting quality for HRQoL or HSU data was very poor (Yang et al., 2023).

Regarding research priorities, notably, developing instruments to capture the impact of treatments on children and adolescents is an important topic in most regions. This may reflect a real unmet need for fit-for-purpose instruments all around the world. Instruments assessing effects of social care, caregiver needs, and specific health problems are on the top of the wish list of HTA practitioners from many regions. The need for recent tariffs globally as the top research priority further strengthens the importance of valuation work. This need was highlighted especially in the Middle East/Africa and Central/Eastern Europe, where value set generation is only starting to gain momentum (Al-Jedai et al., 2024; Al Shabasy et al., 2022; Prevolnik Rupel & Ogorevc, 2023) . Countries in these regions generally lack preference-based values. In line with the growth of HTA in these countries, the presence of value sets become essential in expanding the use of CUAs and in implementing HTA for wider coverage of healthcare decision-making (Callenbach et al., 2023; Falkowski et al., 2023; Kaló et al., 2016) . An interesting research topic proposed by a respondent is to develop public depositories of HSU data. Such a depository would act as a library, storing data from different population groups, facilitating crosswalks to other country value sets, and being referenced by HTA practitioners as needed. A properly regulated HSU depository would alleviate the

issue of scarce data faced by HTA practitioners globally. These research topics, along with insights from the content analysis, highlight the global need for greater generation of HSUs in different areas to better capture the health preferences of populations.

The above findings about the current practice and views of HTA personnel on instruments, methods, and data for generating QALY-based evidence may provide useful guidance for future research. First, research on instruments targeting children and adolescents such as EQ-5D-Y may be prioritized since there is a global need for such instrument. Compared to HRQoL instruments for adults, instruments for children and adolescents are fewer and less developed. The recently completed methods review by NICE found insufficient evidence for recommending any existing HRQoL instruments for use in paediatric HTA and therefore called for research in this area. Second, researchers on new instruments and elicitation methods may consider to shift from a pure academic approach to a user-oriented approach by engaging stakeholders particularly HTA agencies in the whole development process. Such a collaborative approach may increase the chance of developing a product that will be accepted or adopted sooner for use in HTA practice. This approach to involving stakeholders such as patient advocacy groups and decision makers into instrument development has been used to developing the EQ-HWB instrument (Carlton et al., 2022). However, as instrument development is a long, multi-stage process and HTA agencies are prudent in endorsing new instruments and methods, perhaps only continuous engagement and long-time collaboration may lead to tangible impact. Moreover, given that the established HTA agencies are more concerned about maintaining consistency and standardization, new instruments and methods may be more likely to be endorsed and accepted by burgeoning HTA agencies. Last but not least, research on methods for making more HRQoL and HSU data available or making better use of existing data seems equally or even more important than making new instruments available. This is because data scarcity for endorsed HRQoL instruments such as EQ-5D may be a greater issue than the lack of more fit-for-purpose instruments such as EQ-HWB because those instruments are routinely used. Such research work may involve systematically collecting and publishing HRQoL data from health systems, collating and compiling HRQoL data published in the literature, and developing tools for modifying or transforming HRQoL data for use across health systems. Databases providing HSU data such as and guidelines promoting appropriate use of HSU data (Brazier et al., 2019; Wolowacz et al., 2016) have been available. However, those may not be sufficient and more work is needed to fill in this data gap and need.

This study had several limitations. In 12 of the 45 countries, there were fewer than three responses despite repeated reminders, limiting the representativeness of these countries. Another limitation concerns the snowball-type recruitment method we employed. Potential respondents were identified through the network of EuroQol Group members. As a result, HTA personnel who are familiar with or in favour of EuroQol instruments may be overrepresented in our study sample. However, EuroQol members come from diverse backgrounds and regions and many of them actively participate in HTA development in their respective countries, which makes them ideal recruiters for this by-invitation-only global survey. Lastly, we were not able to verify the eligibility of the respondents. Country-specific survey links were distributed by recruiters to potential respondents in the target HTA agencies. Although screening questions were included at the start of the survey, personal identifiers such as respondent names or the agencies they worked for were not collected to encourage more candid responses.

Conclusion

This study filled important knowledge gaps regarding the current practices of measuring and valuing HRQoL in HTA and the views on the challenges and needs of HTA agency personnel around the world. Findings from this study may be used to guide research aimed at developing tools and methods for providing high-quality QALY evidence for economic evaluations.

Acknowledgement

We extend our deepest gratitude to the following individuals who acted as recruiters for the study and contributed to the success of our study: Aureliano Finch, Claire Gudex, Fanni Rencz, Fatima Al Sayah, Gan Yan Nee, Haarathi Chandriah, Jan Busschbach, Jazmin Joanna Pinzon, Jennifer Jelsma, Juan M. Ramos-Goñi, Marisa Santos, Mimi Astrom, Min-Woo Jo, Patricia Antunes, Petra Došenovič Bonča, Princess Allyza, Sarah Dewilde, Sivaraj Raman, Takeru Shiroiwa, Tommy Chen, Victor Zarate, Victor Zarate, Victor Zarate, Vu Quynh Mai, Wang Pei, and Wanrudee Isaranuwatchai. We are also thankful to the following recruiters whose attempts were not successful: Gerard de Pouvourville, Henry Bailey, John Yfantopoulos, Oliver Rivero Arias, Sarah Dewilde, and Wolfgang Greiner.

References

Al-Jedai, A., Almudaiheem, H., Al-Salamah, T., Aldosari, M., Almutairi, A. R., Almogbel, Y., AlRuthia, Y., Althemery, A. U., Alluhidan, M., Roudijk, B., Purba, F. D., Awad, N., & O'Jeil, R. (2024). Valuation of EQ-5D-5L in the Kingdom of Saudi Arabia: A National Representative Study. *Value Health*. https://doi.org/10.1016/j.jval.2024.01.017

Al Shabasy, S., Abbassi, M., Finch, A., Roudijk, B., Baines, D., & Farid, S. (2022). The EQ-5D-5L Valuation Study in Egypt. *Pharmacoeconomics*, *40*(4), 433-447. https://doi.org/10.1007/s40273-021-01100-y

Ara, R., Brazier, J., Lloyd, A., & Chevrou-Severac, H. (2018). How health state utilities used in cost-effectiveness models are currently identified, reviewed and reported. *Value Outcomes Spotlight*, *4*(5), 31e33.

Ara, R., Hill, H., Lloyd, A., Woods, H. B., & Brazier, J. (2020). Are Current Reporting Standards Used to Describe Health State Utilities in Cost-Effectiveness Models Satisfactory? *Value Health*, *23*(3), 397-405. https://doi.org/10.1016/j.jval.2019.12.004

Augustovski, F., Belizán, M., Gibbons, L., Reyes, N., Stolk, E., Craig, B. M., & Tejada, R. A. (2020). Peruvian Valuation of the EQ-5D-5L: A Direct Comparison of Time Trade-Off and Discrete Choice Experiments. *Value Health*, *23*(7), 880-888. https://doi.org/10.1016/j.jval.2020.05.004

Brazier, J., Ara, R., Azzabi, I., Busschbach, J., Chevrou-Séverac, H., Crawford, B., Cruz, L., Karnon, J., Lloyd, A., Paisley, S., & Pickard, A. S. (2019). Identification, Review, and Use of Health State Utilities in Cost-Effectiveness Models: An ISPOR Good Practices for Outcomes Research Task Force Report. *Value Health*, *22*(3), 267-275. https://doi.org/10.1016/j.jval.2019.01.004

Brazier, J., Ara, R., Rowen, D., & Chevrou-Severac, H. (2017). A review of generic preference-based measures for use in cost-effectiveness models. *Pharmacoeconomics*, *35*(Suppl 1), 21-31. https://doi.org/10.1007/s40273-017-0545-x

Brazier, J., Peasgood, T., Mukuria, C., Marten, O., Kreimeier, S., Luo, N., Mulhern, B., Pickard, A. S., Augustovski, F., Greiner, W., Engel, L., Belizan, M., Yang, Z., Monteiro, A., Kuharic, M., Gibbons, L., Ludwig, K., Carlton, J., Connell, J., . . . Rejon-Parrilla, J. C. (2022). The EQ-HWB: Overview of the Development of a Measure of Health and Wellbeing and Key Results. *Value Health*, *25*(4), 482-491. https://doi.org/10.1016/j.jval.2022.01.009

Callenbach, M. H. E., Ádám, L., Vreman, R. A., Németh, B., Kaló, Z., & Goettsch, W. G. (2023). Reimbursement and payment models in Central and Eastern European as well as Middle Eastern countries: A survey of their current use and future outlook. *Drug Discov Today*, *28*(1), 103433. https://doi.org/10.1016/j.drudis.2022.103433

Carlton, J., Peasgood, T., Mukuria, C., Johnson, J., Ogden, M., & Tovey, W. (2022). The role of patient and public involvement and engagement (PPIE) within the development of the EQ Health and Wellbeing (EQ-HWB). *J Patient Rep Outcomes*, *6*(1), 35. https://doi.org/10.1186/s41687-022-00437-y

Claire, R., Elvidge, J., Hanif, S., Goovaerts, H., Rijnbeek, P. R., Jónsson, P., Facey, K., & Dawoud, D. (2023). Advancing the use of real world evidence in health technology assessment: insights from a multi-stakeholder workshop. *Front Pharmacol*, *14*, 1289365. https://doi.org/10.3389/fphar.2023.1289365

Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the economic evaluation of health care programmes* (Fourth ed.). Oxford University Press.

Falkowski, A., Ciminata, G., Manca, F., Bouttell, J., Jaiswal, N., Farhana Binti Kamaruzaman, H., Hollingworth, S., Al-Adwan, M., Heggie, R., Putri, S., Rana, D., Mukelabai Simangolwa, W., & Grieve, E. (2023). How Least Developed to Lower-Middle Income Countries Use Health Technology Assessment: A Scoping Review. *Pathog Glob Health*, *117*(2), 104-119. https://doi.org/10.1080/20477724.2022.2106108

Feng, Y. S., Kohlmann, T., Janssen, M. F., & Buchholz, I. (2021). Psychometric properties of the EQ-5D-5L: a systematic review of the literature. *Qual Life Res*, *30*(3), 647-673. https://doi.org/10.1007/s11136-020-02688-y

Fontrier, A. M., Visintin, E., & Kanavos, P. (2022). Similarities and Differences in Health Technology Assessment Systems and Implications for Coverage Decisions: Evidence from 32 Countries. *Pharmacoecon Open*, *6*(3), 315-328. https://doi.org/10.1007/s41669-021-00311-5

Health Intervention and Technology Assessment Program. (2024). *Guideline Comparison: What can I learn from the existing health economic evaluation guidelines?* Guide to Economic Analysis and Research (GEAR) Online Resource. http://www.gear4health.com/gear/health-economic-evaluation-guidelines

Hiligsmann, M., Liden, B., Beaudart, C., Germeni, E., Hanna, A., Joshi, M., Koola, C. P., Stein, B., Tonkinson, M., Marshall, D., & Fifer, S. (2024). HTA community perspectives on the use of patient preference information: lessons learned from a survey with members of HTA bodies. *Int J Technol Assess Health Care*, *40*(1), e17. https://doi.org/10.1017/s0266462324000138

INAHTA. (2020). *INAHTA Members List*. https://www.inahta.org/members/members_list/

ISPOR. (2024). *Pharmacoeconomic Guidelines Around the World* ISPOR–The Professional Society for Health Economics and Outcomes Research: HEOR Resources. https://www.ispor.org/heor-resources/more-heor-resources/pharmacoeconomic-guidelines/pe-guideline-detail/

Kaló, Z., Gheorghe, A., Huic, M., Csanádi, M., & Kristensen, F. B. (2016). HTA Implementation Roadmap in Central and Eastern European Countries. *Health Econ*, *25 Suppl 1*(Suppl Suppl 1), 179-192. https://doi.org/10.1002/hec.3298

Kangwanrattanakul, K., & Phimarn, W. (2019). A systematic review of the development and testing of additional dimensions for the EQ-5D descriptive system [Review]. *Expert Review of Pharmacoeconomics and Outcomes Research*, *19*(4), 431-443. https://doi.org/10.1080/14737167.2019.1637736

Kennedy-Martin, M., Slaap, B., Herdman, M., van Reenen, M., Kennedy-Martin, T., Greiner, W., Busschbach, J., & Boye, K. S. (2020). Which multi-attribute utility instruments are recommended for use in cost-utility analysis? A review of national health technology assessment (HTA) guidelines. *Eur J Health Econ*, *21*(8), 1245-1257. https://doi.org/10.1007/s10198-020-01195-8

King, M. T., Revicki, D. A., Norman, R., Müller, F., Viney, R. C., Pickard, A. S., Cella, D., & Shaw, J. W. (2024). United States Value Set for the Functional Assessment of Cancer Therapy-General Eight Dimensions (FACT-8D), a Cancer-Specific Preference-Based Quality of Life Instrument. *Pharmacoecon Open*, *8*(1), 49-63. https://doi.org/10.1007/s41669-023-00448-5

O'Rourke, B., Werkö, S. S., Merlin, T., Huang, L. Y., & Schuller, T. (2020). The 'Top 10' Challenges for Health Technology Assessment: INAHTA Viewpoint. *Int J Technol Assess Health Care*, *36*(1), 1-4. https://doi.org/10.1017/s0266462319000825

Prevolnik Rupel, V., & Ogorevc, M. (2023). EQ-5D-5L Value Set for Slovenia. *Pharmacoeconomics*, *41*(11), 1515-1524. https://doi.org/10.1007/s40273-023-01280-9

Qian, X., Tan, R. L. Y., Chuang, L. H., & Luo, N. (2020). Measurement Properties of Commonly Used Generic Preference-Based Measures in East and South-East Asia: A Systematic Review [Review]. *Pharmacoeconomics*, *38*(2), 159-170. https://doi.org/10.1007/s40273-019-00854-w

Sharma, D., Aggarwal, A. K., Downey, L. E., & Prinja, S. (2021). National Healthcare Economic Evaluation Guidelines: A Cross-Country Comparison. *Pharmacoecon Open*, *5*(3), 349-364. https://doi.org/10.1007/s41669-020-00250-7

Shiroiwa, T., King, M. T., Norman, R., Müller, F., Campbell, R., Kemmler, G., Murata, T., Shimozuma, K., & Fukuda, T. (2024). Japanese value set for the EORTC QLU-C10D: A multi-attribute utility instrument based on the EORTC QLQ-C30 cancer-specific quality-of-life questionnaire. *Qual Life Res*, *33*(7), 1865-1879. https://doi.org/10.1007/s11136-024-03655-7

Shiroiwa, T., Murata, T., Morii, Y., Hoshino, E., & Fukuda, T. (2024). Comparison of four value sets derived using different TTO and DCE approaches: application to the new region-specific PBM, AP-7D. *Health Qual Life Outcomes*, *22*(1), 16. https://doi.org/10.1186/s12955-024-02233-2

Teerawattananon, Y., Rattanavipapong, W., Lin, L. W., Dabak, S. V., Gibbons, B., Isaranuwatchai, W., Toh, K. Y., Cher, B. P., Pearce, F., Bayani, D. B. S., Nakamura, R., Pwu, R. F., Shafie, A. A., Adhikari, D., Prinja, S., & Babidge, W. (2019). Landscape analysis of health

technology assessment (HTA): systems and practices in Asia. *Int J Technol Assess Health Care*, *35*(6), 416-421. https://doi.org/10.1017/s0266462319000667

van Overbeeke, E., Forrester, V., Simoens, S., & Huys, I. (2021). Use of Patient Preferences in Health Technology Assessment: Perspectives of Canadian, Belgian and German HTA Representatives. *Patient*, *14*(1), 119-128. https://doi.org/10.1007/s40271-020-00449-0

Wang, H., Rowen, D. L., Brazier, J. E., & Jiang, L. (2023). Discrete Choice Experiments in Health State Valuation: A Systematic Review of Progress and New Trends. *Appl Health Econ Health Policy*, 1-14. https://doi.org/10.1007/s40258-023-00794-9

Wang, P., Chong, S. L., Tan, R. L., & Luo, N. (2023). A hearing bolt-on item increased the measurement properties of the EQ-5D-5L in a community-based hearing loss screening program. *Eur J Health Econ*, *24*(3), 393-398. https://doi.org/10.1007/s10198-022-01479-1

Wolowacz, S. E., Briggs, A., Belozeroff, V., Clarke, P., Doward, L., Goeree, R., Lloyd, A., & Norman, R. (2016). Estimating Health-State Utility for Economic Models in Clinical Studies: An ISPOR Good Research Practices Task Force Report. *Value Health*, *19*(6), 704-719. https://doi.org/10.1016/j.jval.2016.06.001

World Health Organization. (2024). *Health Technology Assessment and Health Benefit Package Survey 2020/2021* World Health Organization Health Systems Governance and Financing. https://www.who.int/teams/health-systems-governance-and-financing/economic-analysis/health-technology-assessment-and-benefit-package-design/survey-homepage

Xu, R. H., Wong, E. L., Luo, N., Norman, R., Lehmann, J., Holzner, B., King, M. T., & Kemmler, G. (2024). The EORTC QLU-C10D: the Hong Kong valuation study. *Eur J Health Econ*, *25*(5), 889-901. https://doi.org/10.1007/s10198-023-01632-4

Yang, Z., Zeng, X., Huang, W., Chai, Q., Zhao, A., Chuang, L. H., Wu, B., & Luo, N. (2023). Characteristics of health-state utilities used in cost-effectiveness analyses: a systematic review of published studies in Asia [Review]. *Health and Quality of Life Outcomes*, *21*(1), Article 59. https://doi.org/10.1186/s12955-023-02131-z

Zisis, K., Pavi, E., Geitona, M., & Athanasakis, K. (2024). Real-world data: a comprehensive literature review on the barriers, challenges, and opportunities associated with their inclusion in the health technology assessment process. *J Pharm Pharm Sci*, *27*, 12302. https://doi.org/10.3389/jpps.2024.12302

Table 1 Characteristics of respondents (N=238)

	Region, n (%)						Total, n(%)
	Common wealth	**Western Europe**	**Central/ Eastern Europe**	**Asia**	**Latin America**	**Middle East/ Africa**	
Government Employee							
Yes	21 (61.8)	24 (77.4)	19 (76.0)	77 (81.1)	20 (52.6)	10 (66.7)	**177 (71.9)**
No	13 (35.2)	7 (22.6)	6 (24.0)	18 (19.0)	18 (47.4)	5 (33.3)	**67 (28.2)**
Contracted Professional							
Yes	17 (50.0)	14 (45.2)	8 (32.0)	33 (34.7)	26 (68.4)	5 (33.3)	**103 (43.3)**
No	17 (50.0)	17 (54.8)	17 (68.0)	62 (65.3)	12 (31.6)	10 (66.7)	**135 (56.7)**
Experience with HTA (years)							
Less than a year	3 (8.8)	3 (9.7)	6 (24.0)	31 (32.6)	1 (2.6)	5 (33.3)	**49 (20.6)**
1 - 3 years	6 (17.7)	3 (9.7)	7 (28.0)	19 (20.0)	12 (31.6)	2 (13.3)	**49 (20.6)**
4 - 6 years	7 (20.6)	3 (9.7)	2 (8.0)	9 (9.5)	9 (23.7)	0 (0)	**30 (12.6)**
7 - 9 years	0 (0)	0 (0)	1 (4.0)	7 (7.4)	1 (2.6)	1 (6.7)	**10 (4.2)**
10 years or more	18 (52.9)	22 (71.0)	9 (36.0)	29 (30.5)	15 (39.5)	7 (46.7)	**100 (42.0)**
Gender							
Female	17 (51.5)	23 (74.2)	17 (70.8)	55 (57.9)	17 (44.7)	7 (46.7)	**136 (57.6)**
Male	16 (48.5)	8 (25.8)	7 (29.2)	40 (42.1)	21 (55.3)	8 (53.3)	**100 (42.4)**
Age group (years)							
<= 25	1 (2.94)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	**1 (0.4)**
26- 35	10 (29.4)	3 (9.7)	9 (36.0)	34 (35.8)	8 (21.1)	2 (13.3)	**66 (27.7)**
36 - 45	9 (26.5)	8 (25.8)	6 (24.0)	39 (41.1)	11 (29.0)	4 (26.7)	**77 (32.4)**
46 - 55	9 (26.5)	14 (45.2)	6 (24.0)	17 (17.9)	13 (34.2)	8 (53.3)	**67 (28.2)**
56 - 65	5 (14.7)	6 (19.4)	2 (8.0)	3 (3.2)	5 (13.2)	1 (6.7)	**22 (9.2)**
>=66	0 (0)	0 (0)	2 (8.0)	2 (2.1)	1 (2.6)	0 (0)	**5 (2.1)**
Education Attainment							
Bachelors	2 (5.9)	0 (0)	0 (0)	11 (11.6)	0 (0)	1 (6.7)	**14 (5.9)**
Masters	15 (44.1)	7 (22.6)	12 (48.0)	38 (40.0)	24 (63.2)	4 (26.7)	**100 (42.0)**
Doctorate	17 (50.0)	24 (77.4)	13 (52.0)	46 (48.4)	14 (36.8)	9 (60.0)	**123 (51.7)**
Decline to disclose	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (6.7)	**1 (0.4)**
Professional Identity							
Health economist	31 (91.2)	21 (67.7)	9 (36.0)	29 (30.5)	19 (50.0)	9 (60.0)	**118 (50.0)**
Pharmacist	0 (0)	1 (3.2)	9 (36.0)	26 (27.4)	3 (7.9)	1 (6.7)	**40 (16.8)**
Public health professional	0 (0)	4 (12.9)	3 (12.0)	21 (22.1)	6 (15.8)	0 (0)	**34 (14.3)**
Clinician/Medical doctor	0 (0)	1 (3.2)	3 (12.0)	2 (2.1)	1 (2.6)	2 (13.3)	**9 (3.8)**
Epidemiologist	0 (0)	0 (0)	0 (0)	3 (3.16)	6 (15.8)	0 (0)	**9 (3.8)**
Statistician	0 (0)	1 (3.2)	0 (0)	4 (4.21)	0 (0)	0 (0)	**5 (2.1)**
Other	3 (8.8)	3 (9.7)	1 (4.0)	10 (10.5)	3 (7.8)	3 (20.0)	**23 (9.7)**
Presence of QALY Estimation Guidance							
Yes	31 (91.2)	28 (90.3)	17 (68.0)	85 (89.5)	21 (55.3)	11 (73.3)	**193 (81.1)**
No	3 (8.8)	3 (9.7)	8 (32.0)	10 (10.5)	17 (44.7)	4 (26.7)	**45 (18.9)**
QALY-based Responsibilities							
Yes	33 (97.1)	27 (87.1)	22 (88.0)	84 (88.4)	34 (89.5)	13 (86.7)	**213 (89.5)**
No	1 (2.9)	4 (12.9)	3 (12.0)	11 (11.6)	4 (10.5)	2 (13.3)	**25 (10.5)**

	Region, n (%)						Total, n(%)
	Common wealth	**Western Europe**	**Central/ Eastern Europe**	**Asia**	**Latin America**	**Middle East/ Africa**	
Role							
Review Industry	31 (91.2)	17 (54.8)	18 (72.0)	52 (54.7)	17 (44.7)	11 (73.3)	**146 (61.3)**
Review Public	13 (38.2)	17 (54.8)	9 (36.0)	47 (49.5)	22 (57.9)	8 (53.3)	**116 (48.7)**
Primary study	17 (50.0)	10 (32.3)	2 (8.0)	34 (35.8)	8 (21.1)	6 (40.0)	**77 (32.4)**
Recommend method	9 (26.5)	7 (22.6)	2 (8.0)	12 (12.6)	5 (13.2)	5 (33.3)	**40 (16.8)**
None of the above	1 (2.9)	4 (12.9)	2 (8.0)	7 (7.4)	8 (21.1)	1 (6.7)	**23 (9.7)**
Level of HTA Work							
National	33 (97.1)	30 (96.8)	25 (100)	89 (93.7)	30 (79.0)	10 (66.7)	**217 (91.2)**
Regional provincial or state-level	3 (8.8)	9 (29.0)	1 (4.0)	15 (15.9)	3 (7.9)	6 (40.0)	**37 (15.6)**
Hospital	1 (2.9)	3 (9.7)	3 (12.0)	7 (7.4)	7 (18.4)	0 (0)	**21 (8.8)**
Health plan	1 (2.9)	5 (16.1)	2 (8.0)	13 (13.7)	6 (15.8)	5 (33.3)	**32 (13.5)**
Health Tech Appraised							
Pharmaceuticals	34 (100)	23 (74.2)	23 (92.0)	72 (75.8)	35 (92.1)	12 (80.0)	**199 (83.6)**
Medical devices	17 (50.0)	18 (58.1)	8 (32.0)	39 (41.1)	23 (60.5)	9 (60.0)	**114 (47.9)**
Vaccines	10 (29.4)	15 (48.4)	6 (24.0)	29 (30.5)	20 (52.6)	6 (40.0)	**86 (36.1)**
Diagnostics	14 (41.2)	17 (54.8)	4 (16.0)	25 (26.3)	22 (57.9)	6 (40.0)	**88 (37.0)**
Surgical procedures	11 (32.4)	19 (61.3)	2 (8.0)	18 (19.0)	10 (26.3)	4 (26.7)	**64 (26.9)**
Public health professionals	3 (8.8)	11 (35.5)	3 (12.0)	23 (24.2)	5 (13.2)	2 (13.3)	**47 (19.8)**
Other	2 (5.9)	6 (19.4)	3 (12.0)	9 (9.5)	2 (5.3)	0 (0)	**22 (9.2)**
Therapeutic Area							
Oncology	26 (76.5)	14 (45.2)	19 (76.0)	67 (70.5)	29 (76.3)	12 (80.0)	**167 (70.2)**
Cardiovascular Disease	16 (47.1)	15 (48.4)	19 (76.0)	40 (42.1)	16 (42.1)	5 (33.3)	**111 (46.6)**
Diabetes/ Hypertension/ Dyslipidaemia	9 (26.5)	12 (38.7)	11 (44.0)	43 (45.3)	13 (34.2)	7 (46.7)	**95 (39.9)**
Respiratory Disease	9 (26.5)	0 (0)	3 (12.0)	21 (22.1)	6 (15.8)	3 (20.0)	**42 (17.7)**
Musculoskeletal/ Rheumatology	8 (23.5)	8 (25.8)	6 (24.0)	7 (7.4)	11 (29.0)	0 (0)	**40 (16.8)**
Gynaecology/Obstetrics	1 (2.9)	3 (9.7)	1 (4.0)	6 (6.3)	3 (7.9)	3 (20.0)	**17 (7.1)**
Infections Disease/ HIV/ AIDS	1 (2.9)	6 (19.4)	0 (0)	17 (17.9)	10 (26.3)	6 (40.0)	**40 (16.8)**
Neurology	6 (17.7)	3 (9.7)	2 (8.0)	7 (7.4)	7 (18.4)	5 (33.3)	**30 (12.6)**
Psychiatric Disorders/ Substance Abuse	2 (5.9)	6 (19.4)	3 (12.0)	7 (7.4)	1 (2.6)	1 (6.7)	**20 (8.4)**
Gastrointestinal Disease	7 (20.6)	4 (12.9)	0 (0)	7 (7.4)	4 (10.5)	0 (0)	**22 (9.2)**
Endocrine	4 (11.8)	2 (6.5)	3 (12.0)	7 (7.4)	3 (7.9)	0 (0)	**19 (8/0)**
Surgery/Transplantation	2 (5.9)	4 (12.9)	1 (4.0)	6 (6.3)	1 (2.6)	0 (0)	**14 (5.9)**
Urology/Nephrology	0 (0)	4 (12.9)	2 (8.0)	7 (7.4)	2 (5.3)	0 (0)	**15 (6.3)**
Dermatology	5 (14.7)	3 (9.7)	0 (0)	2 (2.1)	1 (2.6)	3 (20.0)	**14 (5.9)**

Notes: Role: Review Industry: I review QALY-based cost-effectiveness evidence submitted by industry or contractors; Review Public: I review publicly available QALY-based cost-effectiveness evidence; Primary study: I conduct primary studies to generate QALY-based cost-effectiveness evidence; Recommend method: I develop or recommend methods for generating QALY-based cost-effectiveness evidence; I do none of the above

Total number of responses from each region and country: Asia=95 (China:4, India:5, Indonesia:6, Japan:3, Malaysia:9, Philippines:3, Singapore:15, South Korea:16, Taiwan:11, Thailand:5, Vietnam:18), Central/Eastern Europe=25 (Bulgaria:6, Croatia:1, Czech Republic:1, Estonia:1, Hungary:5, Latvia:1, Poland:3, Romania:1, Slovenia:6), Western Europe=31 (Austria:3, Denmark:4, Italy:2, Netherlands:6, Portugal:6, Spain:7, Sweden:3), Latin America=38 (Argentina:2, Brazil:10, Chile:3, Colombia:12, Ecuador:6, Mexico:3, Peru:2), Middle East/ Africa=15 (Egypt:2, Saudi Arabia:1, South Africa:5, Tunisia:3, UAE:4), Commonwealth=34 (Australia:7, Canada:4, England:17, New Zealand:4, Scotland:1, Wales:1)

Table 2 Median (IQR) responses by region

	Response Frequency, Median (IQR)						Total
	Common wealth (n=6)	Western Europe (n=7)	Central/ Eastern Europe (n=9)	Asia (n=11)	Latin America (n=7)	Middle-East/ Africa (n=5)	
UI use frequency							
Total responses (N)	33	27	22	83	33	13	211
AQOL	1 (0)	1 (0)	1 (1.0)	1 (1.5)	1 (1.5)	1 (0.5)	6.0
EQ-5D	4 (0)	4 (0)	4 (0)	4 (0)	3.5 (1.0)	4 (0.5)	23.5
EQ-5D-Y	1.75 (0.5)	1 (1)	2 (1.5)	2 (1.0)	2 (1.0)	2 (0.5)	10.75
EQ-HWB	1 (0)	1 (0)	1 (0.75)	1 (0)	1 (1.0)	1 (0.5)	6.0
Bolt-ons	1 (0)	1 (0)	1 (0.25)	1 (0)	1 (0)	1 (0)	6.0
HUI	2 (0.5)	1 (1.0)	1 (1)	1.5 (1.0)	1 (0)	2 (0.5)	8.5
PROPR	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0.5)	6.0
QWB	1 (0)	1 (0)	1 (0.5)	1 (0)	1 (0)	1 (0)	6.0
SF-6D	2 (0)	2 (1.0)	2 (1.0)	2 (1.0)	3 (1.0)	2.5 (1.0)	13.5
UI used matters, n(%)							
Yes	31 (91.2)	28 (90.3)	22 (88.0)	83 (88.3)	33 (89.2)	10 (66.7)	207 (87.7)
No/ not sure	3 (8.8)	3 (9.7)	3 (12.0)	11 (11.7)	4 (10.8)	5 (33.3)	29 (12.3)
EM use frequency							
Total responses (N)	33	27	21	78	32	11	202
BWS	1 (0)	1 (0)	1.5 (2.0)	1 (1.0)	1.5 (1.0)	2 (0.5)	8.0
DCE	2 (0.5)	1.5 (1.0)	2 (1.0)	2 (1.5)	2 (1.0)	3.5 (1.75)	13.0
PTO	1.25 (1.0)	1 (1.0)	2 (0.5)	2 (2.0)	2 (0.5)	2 (1.0)	10.25
SG	2.25 (1.0)	2 (1.0)	2.75 (1.5)	3 (1.0)	3 (1.0)	3.5 (1.5)	16.5
TTO	4 (1.0)	3.5 (1.0)	3 (1.0)	3 (1.0)	3 (1.0)	3.5 (1.0)	20.0
VAS	2 (1.0)	2 (1.0)	3 (1.5)	3 (1.5)	3 (1.5)	3.5 (1.0)	16.5
EM used matters, n(%)							
Yes	27 (79.4)	26 (86.7)	16 (69.6)	71 (81.6)	27 (75.0)	11 (84.6)	178 (79.8)
No/ not sure	7 (20.6)	4 (13.3)	7 (30.4)	16 (18.4)	9 (25.0)	2 (15.4)	45 (20.2)
HPS use frequency							
Total responses (N)	33	27	22	84	34	13	213
General population own	3.25 (2.0)	3 (2.0)	2 (1.0)	3 (1.5)	2 (0.5)	1 (1.0)	14.25
General population other	2.5 (1.0)	2 (2.0)	4 (1.0)	3 (1.0)	3 (1.0)	3 (1.5)	17.5
Patient own	2 (0)	2 (1.0)	1 (1.0)	3 (1.0)	2 (1.0)	1 (1.0)	11.0
Patient other	2 (2.0)	2 (0.5)	3 (1.5)	2 (1.0)	3 (1.0)	3 (1.5)	15.0
HPS used matter, n(%)							
Yes	30 (88.2)	30 (96.8)	23 (92.0)	86 (90.5)	35 (92.1)	13 (86.7)	217 (91.2)
No/ not sure	4 (11.8)	1 (3.2)	2 (8.0)	9 (9.5)	3 (7.9)	2 (13.3)	21(8.9)
Data quality issue frequency							
Total responses (N)	34	31	25	95	38	15	238
Patient samples	3 (2.0)	3 (1.5)	2 (0)	3 (1.0)	3 (2.0)	2.5 (1.0)	16.5
Health states	3 (1.0)	2 (1.5)	2 (1.0)	3 (0)	4 (1.0)	3 (1.0)	17.0
Sample size	3 (0.5)	2 (1.0)	3 (1.0)	3 (0)	3 (1.0)	2 (1.0)	16.0
Old data	2 (0)	2 (0.5)	2 (1.25)	2 (1.0)	2 (2.0)	2.5 (0.5)	12.5
Different methods	2.75 (1.0)	2 (1.0)	3 (1.0)	3 (1.0)	3 (2.0)	3 (1.5)	16.75

Abbreviations: Responses: 1: Never/not sure; 2: Occasionally; 3: Often; 4:Very often; Patient samples: The patient samples from which HRQoL/utility data was collected were inappropriate (e.g. poor representativeness); Health states: The health states (e.g. the vignettes) for which utility data was available do not match the health states in the CEA model; Sample size: The population samples from which HRQoL/utility data was collected were too small; Old data: The HRQoL/utility data was too old; Different methods: The utility values of different health states used in the same model were derived using different methods/instruments

Countries in each region: Asia (China, India, Indonesia, Japan, Malaysia, Philippines, Singapore, South Korea, Taiwan, Thailand, Vietnam), Central/Eastern Europe (Bulgaria, Croatia, Czech Republic, Estonia, Hungary, Latvia, Poland, Romania, Slovenia:6), Western Europe (Austria, Denmark, Italy, Netherlands, Portugal, Spain, Sweden), Latin America (Argentina, Brazil, Chile Colombia, Ecuador, Mexico, Peru), Middle East/ Africa (Egypt, Saudi Arabia, South Africa, Tunisia, UAE), Commonwealth (Australia, Canada, England, New Zealand, Scotland, Wales)

Table 3 More fit-for-purpose tool and their pros and cons and data source issues encountered

Instruments	Counts	Pros	Cons
EQ-5D (EQ-5D-3L/ EQ-5D-5L)	148	guidelines, value sets, comparability/ consistency in use, widely used, low respondent burden, familiarity, validated/ good psychometric properties, established in HTA	not sensitive in certain health conditions
Depends	37	population/condition, standardisation/ consistency/ comparability	
SF-6D	18	value sets, comparability, established in HTA	
HUI/ HUI3	14	more discriminative in certain conditions, value sets	expensive, seldom used
AQoL	13	more discriminative in certain conditions	
EQ-5D-Y	13	child population	
Bolt-ons	7	improve sensitivity of EQ-5D in certain health conditions	
EQ-HWB	3	broader QoL outcomes	limited experience, insufficient evidence currently (psychometric properties, value sets)
QWB	3	focus on well-being	limited experience
CHU-9D	2	use in child population, value sets.	
PROMIS 10	1	government use, value sets	
QLQ-C30	1	value sets, disease specific	
Methods	**Counts**	**Pros**	**Cons**
TTO	55	valid method, ease of use, involves trade-offs, strong theoretical foundation, produces cardinal utilities	high cognitive burden
Depends	47	depends on disease area, context, available evidence	
SG	29	involves trade-offs, strong theoretical foundation, produces cardinal utilities	high cognitive burden
DCE	28	ease of understanding, involves trade-offs, online feasibility	latent utilities, and converting them to QALY-based values
VAS	24	low respondent burden	not-choice based
Best-worst scaling	4	ease of understanding	
Person trade-off	4		low familiarity
Issue	**Counts**	**Description**	
Data validity	15	Whose data is more valid (poor quality own country or high-quality other country), logical inconsistencies (the sick have higher values than the general public/ less sick), clinical significance of minute differences observed in disutilities, selection of healthier patients, use of adult values for children	
Generalisability	13	Cross-culture, representativeness of different sub-groups within the country	
Data availability	11	Country-level data, norms, rare disease	
Data collection	9	Timing of interview, short follow-ups, mong-term change not considered	
Technique appropriateness	9	Mapping, expert elicitation	
Data analysis	6	Suboptimal methods to obtain utilities, combining utilities of comorbidities, age adjustments, extrapolation to longer time horizon	
Poor reporting	4	Source of utilities, modelling of data, methods to derive utilities	
Data appropriateness	3	Use of other available data	
Utility responsiveness	3	Sensitivity to change and adaptation to disease	
Feasibility	3	Missing values	
EE technique concerns	2	CUA not used to inform decision making, familiarity of decision makers to available techniques	
QALY validity	2	Equity issues	
Study design	2	Real world data, single arm trials	
Content validity	1	Dimensions included in a measure	
Technical	1	Decimal point of utility for CEA models	

Table 4 Research priority by mean sum score

Country	Social Care	Children	Care-givers	Heath Specificity	Recent Tariff	Care Inequality	Minority/ rural
Asia (n=11)	**0.21**	**0.18**	**0.09**	**0.15**	**0.17**	**0.13**	**0.08**
China	0.25	0.17	0.00	0.17	0.17	0.17	0.08
India	0.27	0.07	0.00	0.13	0.33	0.20	0.00
Indonesia	0.20	0.13	0.07	0.13	0.20	0.13	0.13
Japan	0.22	0.22	0.11	0.11	0.22	0.00	0.11
Malaysia	0.17	0.17	0.08	0.13	0.13	0.29	0.04
Philippines	0.22	0.22	0.11	0.11	0.00	0.00	0.33
Singapore	0.16	0.20	0.13	0.13	0.31	0.04	0.02
South Korea	0.19	0.27	0.08	0.22	0.22	0.02	0.00
Taiwan	0.28	0.12	0.07	0.17	0.20	0.07	0.10
Thailand	0.13	0.20	0.20	0.20	0.00	0.27	0.00
Vietnam	0.17	0.19	0.12	0.10	0.13	0.20	0.10
Central/Eastern Europe (n=8)	**0.10**	**0.30**	**0.09**	**0.12**	**0.31**	**0.05**	**0.03**
Bulgaria	0.17	0.22	0.06	0.11	0.11	0.06	0.28
Croatia	0.33	0.33	0.00	0.00	0.33	0.00	0.00
Czech Republic	0.00	0.33	0.33	0.00	0.33	0.00	0.00
Hungary	0.00	0.22	0.00	0.28	0.39	0.11	0.00
Latvia	0.00	0.50	0.00	0.00	0.50	0.00	0.00
Poland	0.00	0.42	0.17	0.17	0.25	0.00	0.00
Romania	0.00	0.33	0.00	0.33	0.33	0.00	0.00
Slovenia	0.33	0.00	0.17	0.06	0.22	0.22	0.00
Commonwealth (n=6)	**0.07**	**0.21**	**0.23**	**0.20**	**0.15**	**0.05**	**0.10**
Australia	0.14	0.14	0.10	0.38	0.05	0.05	0.14
Canada	0.08	0.08	0.17	0.17	0.17	0.17	0.17
England	0.17	0.15	0.31	0.13	0.17	0.06	0.02
New Zealand	0.00	0.25	0.17	0.17	0.17	0.00	0.25
Scotland	0.00	0.33	0.33	0.00	0.33	0.00	0.00
Wales	0.00	0.33	0.33	0.33	0.00	0.00	0.00
Western Europe (n=7)	**0.33**	**0.13**	**0.12**	**0.10**	**0.15**	**0.17**	**0.01**
Austria	0.39	0.22	0.00	0.00	0.17	0.22	0.00
Denmark	0.17	0.13	0.13	0.00	0.17	0.42	0.00
Italy	0.33	0.00	0.17	0.17	0.33	0.00	0.00
Netherlands	0.42	0.06	0.17	0.11	0.00	0.25	0.00
Portugal	0.19	0.25	0.11	0.25	0.08	0.06	0.06
Spain	0.28	0.11	0.17	0.06	0.17	0.22	0.00
Sweden	0.50	0.11	0.11	0.11	0.17	0.00	0.00
Latin America (n=7)	**0.16**	**0.08**	**0.15**	**0.16**	**0.11**	**0.22**	**0.12**
Argentina	0.17	0.00	0.17	0.17	0.00	0.33	0.17
Brazil	0.07	0.10	0.15	0.18	0.18	0.22	0.10
Chile	0.22	0.11	0.28	0.11	0.00	0.28	0.00
Colombia	0.17	0.14	0.11	0.25	0.08	0.17	0.08
Ecuador	0.22	0.11	0.22	0.22	0.06	0.11	0.06
Mexico	0.28	0.11	0.11	0.00	0.28	0.11	0.11
Peru	0.00	0.00	0.00	0.17	0.17	0.33	0.33
Middle East/ Africa (n=5)	**0.12**	**0.23**	**0.14**	**0.03**	**0.33**	**0.12**	**0.03**
Egypt	0.17	0.00	0.17	0.17	0.17	0.33	0.00
Saudi Arabia	0.00	0.33	0.33	0.00	0.33	0.00	0.00
South Africa	0.13	0.13	0.07	0.00	0.37	0.17	0.13
Tunisia	0.17	0.42	0.00	0.00	0.42	0.00	0.00
UAE	0.11	0.28	0.11	0.00	0.39	0.11	0.00
Total Average Score	**0.17**	**0.19**	**0.14**	**0.13**	**0.20**	**0.12**	**0.06**

Note: Social care: To develop utility instruments to capture the impact of both health care and social care; Children: To develop utility instruments to capture the impact of treatment on children and adolescents; Caregivers: To develop utility instruments that capture the impact of a treatment on carers and caregivers; Health Impact: To develop utility instruments that capture the impact of treatment on more specific aspects of health (e.g. vision hearing etc.); Recent Tariff: To make more recent utility data and value sets/tariffs available; Care Inequality: To develop utility instruments that can address inequality in care; Minority/Rural: To develop utility instruments that can reflect the health preferences of minority groups (e.g. indigenous populations) or rural population

For example, respondent A from Thailand endorsed two research priorities; social care and children. Each of these two topics received a score of 0.5 and remain topics received a score of 0. To calculate the score for the social care research priority for Thailand, these scores belonging to the individuals from Thailand were averaged. If 10 respondents from Thailand endorsed at least one topic, and the score for social care was 0.5 for five respondents and 0.2 for the remaining five respondents, the importance score for social care in Thailand is 0.35 (i.e., ([0.5 x 5] + [0.2 x 5]) / 10).

Table 5 Other research topic of importance-related to utility values

Research Topic
To assess the validity of QALYs in capturing outcomes
To capture productivity losses and double counting with QoL measures
To develop a public depository of HSUs
To develop guidelines on instrument use to increase comparability
To develop patient-specific utilities
To ensure instruments used have content validity- relevant domains/health states are captured
To generate living HSUVs over long time horizons
To generate utility values of rare disease
To make country-specific utility data available
To make cross-country preference and utility evidence available
To produce more research on the impact of shifting to EQ-5D-5L instrument

Appendix 1 Responses by country

	Country	Responses
1	Argentina	2
2	Australia	7
3	Austria	3
4	Belgium	0
5	Brazil	10
6	Bulgaria	6
7	Canada	4
8	Chile	3
9	China	4
10	Colombia	12
11	Croatia	1
12	Czech Republic	1
13	Denmark	4
14	Ecuador	6
15	Egypt	2
16	England	17
17	Estonia	1
18	Hungary	5
19	India	5
20	Indonesia	6
21	Ireland	0
22	Italy	2
23	Japan	3
24	Latvia	1
25	Lithuania	0
26	Malaysia	9
27	Mexico	3
28	Netherlands	6
29	New Zealand	4
30	Peru	2
31	Philippines	3
32	Poland	3
33	Portugal	6
34	Romania	1
35	Saudi Arabia	1
36	Scotland	1
37	Singapore	15
38	Slovakia	0
39	Slovenia	6
40	South Africa	5
41	South Korea	16
42	Spain	7
43	Sweden	3
44	Taiwan	11
45	Thailand	5
46	Tunisia	3
47	UAE	4
48	Vietnam	18
49	Wales	1

	Country	Reason for non-response
1	Cuba	Survey was not carried out because of current political turmoil in the country
2	Finland	Survey was not carried out because potential contact person was non-responsive
3	France	Survey was not carried out because the HTA agency declined to participate
4	Germany	Survey was not carried out because Germany does not use CEA to inform country-level decision making
5	Greece	Survey was not carried out because potential contact person was non-responsive
6	Hong Kong	Survey was not carried out because Hong Kong does not use CEA for HTA decision making
7	Israel	Survey was not carried out because of the current political turmoil in the country
8	Jamaica	Survey was not carried out because HTA is in its infancy in the Caribbean
9	Norway	Survey was not carried out because of the red tape involved with distributing the survey
10	Trinidad and Tobago	Survey was not carried out because HTA is in its infancy in the Caribbean
11	USA	Survey was not carried out because CEA is not frequently used to inform HTA decision making

Appendix 2 Other instruments respondents have come across during their HTA related work

Instrument	Count Overall	Country
Disease-specific instruments	11	Singapore (1), Vietnam (1), Hungary (1), Denmark (1), Netherlands (2), Brazil (2), Colombia (1), Australia (1), England (1)
Direct utility elicitation (SG, TTO, VAS, DCE)	7	Japan (1), Thailand (1), Thailand (1), Poland (1),Netherlands (1),Australia (1)
EORTC QLQ-C30	6	Taiwan (1), Denmark (1), Portugal (1), Colombia (1), Australia (1), England (1)
Mapping	6	Japan (1), Netherlands (1), Portugal (1), Brazil (1), Australia (1), England (1)
CHU9D	5	Australia (3); England (1); Wales (1)
Health-Related Quality of Life Instrument with 8 Items (HINT-8)	4	South Korea
Asia PBM 7	2	Thailand
Vignettes	2	Portugal (1), England (1)
PedsQL	1	Australia
Patient Health Questionnaire-9 (PHQ-9)	1	Vietnam
Beck Depression Inventory (BDI)	1	Australia
EORTC QLU-C10D	1	Australia
Visual Function Questionnaire - Utility Index (VFQ-UI)	1	Brazil
Asthma Quality of Life Questionnaire (AQLQ)	1	Colombia
Cambridge Pulmonary Hypertension Outcome Review (CAMPHOR)	1	Singapore
The Functional Assessment of Cancer Therapy (FACT)	1	Colombia
Functional Assessment of Cancer Therapy - Lung (FACT-L)	1	Australia
MOS Social Support Survey	1	Vietnam
St. George's Respiratory Questionnaire (SGRQ)	1	Australia
Western Ontario and McMaster Universities Arthritis Index (WOMAC)	1	Australia
Functional Assessment of Chronic Illness Therapy (FACIT)	1	Poland
Mini-Mental Adjustment to Cancer (MAC)	1	Vietnam
Depression Anxiety and Stress Scale 21- Short Form (DASS-21)	1	Vietnam
Six-minute walking test (6MWT)	1	Australia
15D	1	Denmark
Veterans RAND 12 Item Health Survey (VR-12)	1	Canada
WHOQOL-BREF	1	India
ICECAP A and O	1	Netherlands
Acceptance and Action Questionnaire version 2 (AAQ-2)	1	Vietnam

Appendix 3 Utility instrument use frequency

Country	Total responses (N)	Value	Statistic	Value	Statistic	Value	Statistic	Value	Statistic	Value	Statistic	Value	Statistic	Value	Statistic	Value	Statistic	Value	Statistic
		AQOL		EQ-5D		EQ-5D-Y		EQ-HWB		Bolt-ons		HUI		PROPR		QWB		SF6D	
Argentina	2	3	mode	3	mode	3	mode	3	mode	1	mode	2	p50	1	mode	1	mode	3	p50
Australia	6	2	mode	4	mode	1.5	p50	1	mode	1	mode	3	p50	1	mode	1	mode	2.5	p50
Austria	1	1	mode	4	mode	1	mode	1	mode	1	mode	1	mode		nil	1	mode	2	mode
Brazil	10	1	mode	4	mode	2	mode	1	mode	1	mode	1	mode	1	mode	1	mode	3	p50
Bulgaria	6	1	mode	4	mode	2	mode	2	mode	1	mode	1	mode	1	mode	1.5	p50	4	mode
Canada	4	1	mode	4	mode	1	mode	1	mode	1	mode	2.5	p50	1	mode	1	mode	2	mode
Chile	3	2	mode	4	mode	2	mode	1	mode	1	mode	1	mode	1	mode	1	mode	3	mode
China	4	2.5	p50	4	mode	1	mode	1	mode	2	mode	1	mode	1	mode	1	mode	3	mode
Colombia	10	1	mode	4	mode	1	mode	1	mode	1	mode	1	mode	1	mode	1	mode	3	mode
Croatia	1	1	mode	4	mode		nil	1	mode	1	mode	1	mode	1	mode	1	mode	2	mode
Czech Republic	1		nil	4	mode		nil	1	mode	1	mode	2	mode		nil		nil	2	mode
Denmark	4	1	mode	4	mode	1.5	p50	1	mode	1	mode	1	mode	1	mode	1.5	p50	2	mode
Ecuador	4	1	mode	3	p50	2	mode	2	mode	1	mode	1	mode	1	mode	2	p50	2	mode
Egypt	2	1	mode	4	mode	3	mode	2	mode	2	mode	2	mode	2	mode	1	mode	2.5	p50
England	17	1	mode	4	mode	2	mode	1	mode	2	mode	2	mode	1	mode	1	mode	2	mode
Estonia	1		nil	4	mode		nil		nil		nil	2	mode		nil		nil	3	mode
Hungary	5	2	mode	4	mode	2.5	p50	1	mode	1	mode	2	mode	1	mode	1	mode	2	mode
India	5	3	mode	4	mode	4	mode	3	mode	2	p50	3	mode	2	mode	3	mode	3	p50
Indonesia	6	1	mode	4	mode	1	mode	1	mode	1	mode	1	mode	1	mode	1	mode	2	p50
Italy	1	1	mode	4	mode	2	mode	1	mode	1	mode	1	mode	1	mode	1	mode	2	mode
Japan	3	1	mode	4	mode	2	mode	1	mode	1	mode	2	mode	1	mode	1	mode	2	p50
Latvia	1	1	mode	4	mode	1	mode	1	mode	1	mode	2	mode	1	mode	1	mode	2	mode
Malaysia	8	1.5	p50	4	mode	2	mode	1	mode	1	mode	2	mode	1	mode	1	mode	3	mode
Mexico	2	2.5	p50	3.5	p50	2	p50	1	mode	1	mode	1	mode	1	mode	1	mode	2.5	p50
Netherlands	6	1	mode	4	mode	1	mode	1	mode	1	mode	2	p50	1.5	p50	1	mode	3	mode
New Zealand	4	1	mode	4	mode	1.5	p50	1	mode	1	mode	2	mode	1	mode	1	mode	2	mode
Peru	2	1	mode	3.5	p50	3	mode	1	mode	1	mode	1	mode	1	mode	1	mode	2	p50
Philippines	3	1	mode	4	mode	2	p50	1	mode	1	mode	1.5	p50	1	mode	1	mode	2	mode

Country	Total responses (N)	Value	Statistic	Value	Statistic	Value	Statistic	Value	Statistic	Value	Statistic	Value	Statistic	Value	Statistic	Value	Statistic	Value	Statistic
		AQOL		EQ-5D		EQ-5D-Y		EQ-HWB		Bolt-ons		HUI		PROPR		QWB		SF6D	
Poland	3	1.5	p50	4	mode	3	mode	2	p50	1.5	p50	1	p50	2	mode	2	mode	4	mode
Portugal	6	1	mode	4	mode	1	mode	1	mode	1	mode	1	mode	1	mode	1	mode	2	mode
Romania	1	1	mode	1	mode	1	mode	1	mode	1	mode	1	mode	1	mode	1	mode	1	mode
Saudi Arabia	1	1	mode	4	mode	2	mode	1	mode	1	mode	2	mode	1	mode	2	mode	3	mode
Scotland	1	1	mode	4	mode	2	mode	1	mode	1	mode	2	mode	1	mode	1	mode	2	mode
Singapore	14	1	mode	4	mode	1	mode	1	mode	1	mode	1	mode	1	mode	1	mode	2	p50
Slovenia	3	2.5	p50	3	mode	2	mode	1.5	p50	2	mode	1	mode	1	mode	1	mode	2.5	p50
South Africa	3	1	mode	3.5	p50	1	mode	1	mode	1	mode	1.5	p50	1	mode	1	mode	2	mode
South Korea	16	1	mode	4	mode	2	p50	1	mode	1	mode	1	mode	1	mode	1	mode	2	mode
Spain	6	1	mode	4	mode	2.5	p50	1	mode	1	mode	1	mode	1	mode	1	mode	3	mode
Sweden	3	1	mode	4	mode	1	mode	1	mode	1	mode	2	mode	1	mode	1	mode	2	mode
Taiwan	11	1	mode	4	mode	2	p50	1	mode	1	mode	1	mode	1	mode	1	mode	3	p50
Thailand	5	1	mode	4	mode	2	mode	1	mode	1	mode	2	mode	1	mode	1	mode	2	mode
Tunisia	3		nil	4	mode	2	p50	1.5	p50	1	mode	2.5	p50	1	mode	1	mode	3	p50
UAE	4	2	p50	2	p50	1.5	p50	1	mode	1	mode	1	mode	1.5	p50	1	mode	1	mode
Vietnam	8	3	mode	3.5	p50	3	p50	1	mode	1	mode	1.5	p50	1	mode	1	mode	3	mode
Wales	1	1	mode	4	mode	4	mode	1	mode	1	mode	2	mode	1	mode	1	mode	2	mode
MEDIAN (IQR)		1	(1-1.5)	4	(4-4)	2	(1-2)	1	(1-1)	1	(1-1)	1.5	(1-2)	1	(1-1)	1	(1-1)	2	(2-3)

Notes: p50: median; The number of responses column reflects IQR for median responses. Responses: 1: Never; 2: Occasionally; 3: Often; 4:Very often

Appendix 4 Elicitation method use frequency

Country	N	BWS		DCE		PTO		SG		TTO		VAS	
		Value	Statistic	Value	Statistic	Value	Statistic	Value	Statistic	Value	Statistic	Value	Statistic
Argentina	2	1.5	p50	1.5	p50	2.5	p50	2.5	p50	2.5	p50	2.5	p50
Australia	6	1.5	p50	4	mode	1.5	p50	2	mode	3	mode	2	mode
Austria	1		nil		nil	2	mode	2	mode	4	mode	2	mode
Brazil	10	2	mode	3	mode	1	mode	3	mode	4	mode	4	mode
Bulgaria	6	1	mode	1	mode	2	p50	2	mode	2.5	p50	4	mode
Canada	4	1	mode	1	mode	1	mode	2.5	p50	4	mode	1	mode
Chile	3	2	mode	3	mode	2.5	p50	3	p50	3	p50	4	mode
China	4	2	mode	3.5	p50	2	mode	2	mode	4	mode	2.5	p50
Colombia	10	2	mode	2	mode	2	mode	3	mode	3	p50	4	mode
Croatia	1		nil	3	mode		nil		nil		nil		nil
Czech Republic	1		nil	2	mode		nil	2	mode	2	mode	2	mode
Denmark	4	1	mode	1	mode	1	mode	1	mode	3	p50	2.5	p50
Ecuador	4	1	mode	2	mode	2	p50	3	mode	3	p50	2	mode
Egypt	2	3.5	p50	3	p50		nil	3	p50	3	p50	4	mode
England	17	1	mode	2	mode	1	mode	3	mode	4	mode	1	mode
Estonia	1	3	mode	2	mode	3	mode	4	mode	4	mode	4	mode
Hungary	5	2	mode	2	mode	1.5	p50	3	mode	3	p50	3	p50
India	5	2	mode	4	mode	4	mode	4	mode	4	mode	4	mode
Indonesia	6	1	mode	1	mode	2	p50	2	mode	2	p50	4	mode
Italy	1	1	mode	1	mode	1	mode	2	mode	2	mode	3	mode
Japan	3	1	mode	4	mode	1	mode	2	mode	3	p50	1	mode
Latvia	1	1	mode		nil		nil	3	mode	3	mode	3	mode
Malaysia	8	3	mode	2	mode	2	p50	4	mode	3	mode	4	mode
Mexico	2	1	mode	2	p50	2	mode	2	mode	2	mode	3	mode
Netherlands	6	1	mode	2	mode	1	mode	2	mode	3.5	p50	2	mode
New Zealand	4	1	mode	2.5	p50	2	mode	2	mode	2.5	p50	4	mode
Peru	2	1	mode	2.5	p50	2	mode	1.5	p50	2	p50	2.5	p50
Philippines	3	1.5	p50	3	mode	3	mode	3	mode	2	mode	3	mode
Poland	3	3	p50	2	mode	2	p50	4	mode	3	p50	4	mode
Portugal	6	1	mode	1.5	p50	1.5	p50	3	p50	4	mode	3	p50
Romania	1	1	mode	1	mode	1	mode	1	mode	1	mode	2	mode
Saudi Arabia	1	2	mode	4	mode	2	mode	4	mode	4	mode	4	mode
Scotland	1	1	mode	2	mode	1	mode	2	mode	4	mode	2	mode
Singapore	14	1	mode	2	mode	1	mode	3	mode	3	mode	3	p50
Slovenia	3	1.5	p50	1	mode	2	p50	2.5	p50	3.5	p50	3	mode
South Africa	3	1.5	p50	1.5	p50	1	mode	2	mode	3	p50	3	p50
South Korea	16	1	mode	3	mode	1	mode	3	mode	3	mode	3	mode
Spain	6	1	mode	4	mode	2	p50	4	mode	4	mode	2	p50
Sweden	3	1	mode	1.5	p50	1	mode	3	mode	3	p50	2	mode
Taiwan	11	1	mode	1	mode	3	mode	2	mode	2	mode	4	mode
Thailand	5	1	mode	2	mode	1	mode	2	mode	2	mode	2	mode
Tunisia	3	1	mode	4	mode	2	mode	4	mode	4	mode	3.5	p50
UAE	4	2	mode		nil		nil		nil		nil	2	mode
Vietnam	8	1	mode	2	mode	2	p50	3	p50	2.5	p50	4	mode
Wales	1	1	mode	2	mode	2	mode	3	mode	4	mode	2	mode
MEDIAN (IQR)		1	(1-2)	2	(1.5-3)	2	(1-2)	3	(2-3)	3	(2.5-4)	3	(2-4)

Notes: p50: median; The number of responses column reflects IQR for median responses. Responses: 1: Never; 2: Occasionally; 3: Often; 4:Very often

Appendix 5 Frequency of using data from different sources

Country	N	Value	Statistic	Value	Statistic	Value	Statistic	Value	Statistic
		General Pop Own Country		General Pop Other Country		Patient Own Country		Patient Other Country	
Argentina	2	2	mode	3	mode	2	mode	3	mode
Australia	6	2.5	p50	2	mode	3	mode	3	mode
Austria	1	1	mode	4	mode	2	mode	2	mode
Brazil	10	2	p50	4	mode	2	mode	4	mode
Bulgaria	6	2	mode	4	mode	2	mode	4	mode
Canada	4	4	mode	3	mode	2	mode	2	mode
Chile	3	3	mode	4	mode	2	mode	3	mode
China	4	3	mode	3	p50	3	mode	2	mode
Colombia	10	2	p50	3	p50	2	mode	3	mode
Croatia	1	1	mode	3	mode	1	mode	3	mode
Czech Republic	1	1	mode	4	mode	1	mode	4	mode
Denmark	4	4	mode	2	mode	2	mode	1.5	p50
Ecuador	5	2	p50	4	mode	1	mode	4	mode
Egypt	2	1	mode	2.5	p50	1	mode	4	mode
England	17	4	mode	2	mode	2	mode	1	mode
Estonia	1	1	mode	4	mode	1	mode	2	mode
Hungary	5	2	mode	4	mode	1	mode	3.5	p50
India	5	3	p50	2	mode	4	mode	2	mode
Indonesia	6	3.5	p50	3	mode	3	p50	3	mode
Italy	1	2	mode	2	mode	2	mode	2	mode
Japan	3	4	mode	3	p50	2	p50	2	p50
Latvia	1	1	mode	4	mode	1	mode	3	mode
Malaysia	8	2	mode	3	p50	2	mode	3	p50
Mexico	2	2.5	p50	2	mode	1.5	p50	2	p50
Netherlands	6	3	mode	2	mode	1	mode	2	mode
New Zealand	4	2	mode	3	mode	2	mode	3	mode
Peru	2	1	mode	3	p50	1	mode	3.5	p50
Philippines	3	2	mode	3	mode	2	mode	3	mode
Poland	3	3	mode	3	mode	2.5	p50	3.5	p50
Portugal	6	4	mode	2	mode	1	mode	2	p50
Romania	1	2	mode	2	mode	2	mode	2	mode
Saudi Arabia	1	2	mode	4	mode	2	mode	3	mode
Scotland	1	4	mode	2	mode	2	mode	1	mode
Singapore	15	2	p50	4	mode	1	mode	2	mode
Slovenia	3	3	mode	2	mode	2	mode	2	p50
South Africa	3	1	mode	3	p50	1	mode	2.5	p50
South Korea	16	4	mode	2	mode	3	mode	2	mode
Spain	6	4	mode	4	mode	4	mode	3	p50
Sweden	3	2	p50	3	p50	2	mode	2.5	p50
Taiwan	11	2	mode	3	mode	2	mode	3	mode
Thailand	5	3	p50	2	p50	4	mode	2	p50
Tunisia	3	1	mode	4	mode	1	mode	4	mode
UAE	4	2	mode	2	mode	2	p50	1.5	p50
Vietnam	8	3	mode	2	mode	3	mode	2	mode
Wales	1	2	mode	4	mode	2	mode	2	mode
MEDIAN (IQR)		2	(2-3)	3	(2-4)	2	(2-3)	1	(1-1)

Notes: p50: median; The number of responses column reflects IQR for median responses. Responses: 1: Never; 2: Occasionally; 3: Often; 4:Very often

Appendix 6 Data quality issue frequency: Patient samples, health states, sample size, old data, different methods

Country	N	Patient samples		Health states		Sample size		Old data		Different methods	
		Value	Statistic	Value	Statistic	Value	Statistic	Value	Statistic	Value	Statistic
Argentina	2	2	mode	2	mode	3	mode	1	mode	2	mode
Australia	7	2	mode	3	mode	2	mode	2	mode	3	p50
Austria	3	2	mode	2	mode	2	mode	1.5	p50	2	mode
Brazil	10	4	mode	4	mode	3	mode	2.5	p50	4	mode
Bulgaria	6	2	mode	3	p50	2	mode	3	p50	2	mode
Canada	4	4	mode	3	mode	3	mode	2.5	p50	2	mode
Chile	3	4	mode	4	mode	3	p50	3	mode	3	mode
China	4	3	mode	3	mode	3	mode	3	mode	3	mode
Colombia	12	4	mode	3	mode	3	mode	3	mode	3	mode
Croatia	1	3	mode	3	mode	3	mode	3	mode	3	mode
Czech Republic	1	2	mode	2	mode	3	mode		nil	2	mode
Denmark	4	4	mode	2	mode	2	mode	2	mode	1	mode
Ecuador	6	3	p50	4	mode	2	mode	2	p50	2	p50
Egypt	2	2.5	p50	3	p50	1.5	p50	2.5	p50	3	mode
England	17	3	mode	2	mode	3	mode	2	mode	2	mode
Estonia	1	2	mode	3	mode	2	mode	2	mode	3	mode
Hungary	5	2	mode	2	mode	3	mode	2	mode	3	p50
India	5	3	mode	3	mode	3	mode	3	mode	3	mode
Indonesia	6	3	mode	2.5	p50	3	mode	2	mode	2	p50
Italy	2	3	p50	1.5	p50	1.5	p50	2	p50	3	mode
Japan	3	3	mode	4	mode	3	mode	2	mode	3	mode
Latvia	1	2	mode	2	mode	1	mode	1	mode	3	mode
Malaysia	9	2	p50	3	mode	3	p50	3	mode	2	mode
Mexico	3	3	mode	3	mode	1	mode	1	mode	3	mode
Netherlands	6	3	mode	3	mode	3	mode	2	p50	4	mode
New Zealand	4	2	mode	2	mode	2.5	p50	2	mode	2.5	p50
Peru	2	2	mode	4	mode	2	mode	1.5	p50	4	mode
Philippines	3	4	mode	3	mode	3	mode	2	mode	3	mode
Poland	3	3	mode	2	mode	3	p50	2	p50	3	mode
Portugal	6	2.5	p50	2.5	p50	2.5	p50	1.5	p50	2	mode
Romania	1	2	mode	2	mode	2	mode	1	mode		nil
Saudi Arabia	1	4	mode	4	mode	4	mode	4	mode	4	mode
Scotland	1	3	mode	3	mode	3	mode	2	mode	3	mode
Singapore	15	2	mode	3	mode	3	mode	3	mode	3	mode
Slovenia	6	2	mode	2	mode	3	mode	2.5	p50	2	mode
South Africa	5	3	p50	3	mode	2	mode	2	p50	2	mode
South Korea	16	2	mode	3	mode	2	mode	2	mode	3	mode
Spain	7	4	mode	3.5	p50	2	mode	3	mode	3	mode
Sweden	3	3	mode	1	mode	3	mode	1	mode	2	mode
Taiwan	11	3	p50	3	p50	3	p50	2	mode	2	mode
Thailand	5	2	mode	2	mode	3	mode	2	mode	3	mode
Tunisia	3	2	mode	2	p50	3	mode	2.5	p50	3.5	p50
UAE	4	2	mode	2	mode	2	mode	2	mode	2	mode
Vietnam	18	2	mode	3	mode	3	p50	2	mode	3	p50
Wales	1	4	mode	4	mode	4	mode	2	mode	4	mode
MEDIAN (IQR)		3	(2-3)	3	(2-3)	3	(2-3)	2	(2-2.5)	3	(2-3)

Abbreviations: p50: median; The number of responses column reflects IQR for median responses. Responses: 1: Never; 2: Occasionally; 3: Often; 4:Very often; Patient samples: The patient samples from which HRQoL/utility data was collected were inappropriate (e.g. poor representativeness); Health states: The health states (e.g. the vignettes) for which utility data was available do not match the health states in the CEA model; Sample size: The population samples from which HRQoL/utility data was collected were too small; Old data: The HRQoL/utility data was too old; Different methods: The utility values of different health states used in the same model were derived using different methods/instruments