



Real-world evidence: a practical toolbox for collecting health state utilities

Veronique Lambert-Obry^{*1} , Jean-Philippe Lafrance², Michelle Savoie¹ & Jean Lachaine¹

¹The Faculty of Pharmacy, Université de Montréal, 2940 Chemin de Polytechnique, Montréal, Québec, H3T 1J4, Canada

²The Faculty of Medicine, Université de Montréal, 2900 Boulevard Édouard-Montpetit, Montréal, Québec, H3T 1J4, Canada

*Author for correspondence: Tel.: +1 514 343 6111; veronique.lambert-obry@umontreal.ca

Journal of **Comparative Effectiveness Research**

Health state utilities (HSU) data collected in real-world evidence studies are at risk of bias. Although numerous guidance documents are available, practical advice to avoid bias in HSU studies is limited. Thus, the objective of this article was to develop a concise toolbox intended for investigators seeking to collect HSU in a real-world setting. The proposed toolbox builds on existing guidance and provides practical steps to help investigators perform good quality research. The toolbox aims at increasing the credibility of HSU data for future reimbursement decision making.

First draft submitted: 21 May 2021; Accepted for publication: 7 October 2021; Published online: 20 October 2021

Keywords: bias • health economics • health-related quality of life • observational research • real-world evidence

Health technology assessment (HTA) is a multidisciplinary process that encompasses diverse dimensions such as clinical, economic, organizational, social, cultural and ethical considerations [1]. Cost-utility analyses (CUA), which present results in terms of cost per quality-adjusted life-year, can be used to inform decision making [2]. Utility data are required to calculate quality-adjusted life-years. Health state utilities (HSU) describe the value of a health state on a scale where 1 represents full health, 0 represents dead and negative values represent states worse than death [2]. Well-designed randomized controlled trials (RCTs) can be a reliable source for generating utility data. However, RCTs often provide estimates by treatment arm rather than HSU, which refer to estimates for given economic model health states. Moreover, RCTs have a low external validity and may not capture all health states relevant to economic modeling. Therefore, HSU may have to be collected in real-world evidence (RWE) studies (herein defined as studies that are not conventional RCTs [3]), which are at higher risk of bias than RCTs. Regulatory and reimbursement agencies have shown growing interest in RWE to demonstrate the effectiveness of new therapies. Threats to internal validity of RWE studies have led to the publication of several guidance documents providing good procedural practices for comparative effectiveness research. In the same way as clinical outcomes, the estimation of HSU data can also be biased. Several publications showed that pharmacoeconomic data collected in RWE studies are at risk of bias, which can directly impact the incremental cost-effectiveness ratio (ICER) and lead to the wrong conclusions on cost-effectiveness [4–15]. Furthermore, sensitivity analyses of economic evaluations often show that the ICER is sensitive not only to clinical inputs but also to utility values [16,17]. Considering the risk of bias and the potential significant impact on the ICER, RWE studies collecting HSU should also entail good procedural practices.

According to stakeholder interviews, one of the challenges when using RWE for coverage decisions is the lack of investigator expertise (at interpreting data and adjusting for bias) [18]. Recommendations adapted for the collection of HSU in a real-life practice could help to estimate high-quality inputs for reimbursement decision making. To our knowledge, none of the existing guidelines focus on methodologies for generating HSU in real-life practice specifically. Thus, the objective of this article is to review current recommendations for HSU generation, and incorporate methodological standards into a single toolbox intended for investigators seeking to collect HSU in a real-world setting.

Methods

To identify existing guidance on the collection of HSU, guidelines and methodology reports of countries with a long tradition of HTA [19] were screened: Australia [20], Canada [21], England and Wales [22], France [23], Germany [24], The Netherlands [25], Scotland [26], Spain (Catalonia) [27] and Sweden [28]. Guidelines published by international HTA organizations were also screened, including the European Network for Health Technology Assessment (EUnetHTA) [29], Health Technology Assessment International (HTAi) [30], International Health Economics Association (iHEA) [31], International Network of Agencies for Health Technology Assessment (INAHTA) [32] and International Society for Pharmacoeconomics and Outcomes Research (ISPOR) [33]. The search was conducted up to April 2021. Using the keywords ‘utility’ and ‘utilities,’ 58 potentially relevant publications were identified by titles and abstracts (when available). Full-text publications were then reviewed for eligibility, resolving any disagreements through discussion. Selection was limited to publications providing guidance on how to collect HSU. Following full-text screening, 49 publications did not specifically discuss HSU collection and were therefore excluded for the following reasons: focus on economic modeling/CUA methodologies ($n = 29$), focus on mapping ($n = 13$), specific to a certain population ($n = 4$) and specific to certain disease areas ($n = 3$). Among the nine remaining publications, two were duplicates (journal articles based on previously published Technical Support Document), leading to seven original publications: five from the NICE [34–38] and two from ISPOR [2,17]. Data extraction included key methodological elements affecting internal validity: definition of health state, choice of instrument, mode of administration, features of assessment (e.g., timing and frequency), choice of respondents, sampling, method of recruitment, response rate, missing data, study design and analytical techniques.

None of the seven publications provide a framework for investigators seeking to estimate HSU in a real-world setting (data extraction is summarized in the [Supplementary Table 1](#)). Three documents discuss the NICE reference case and alternative methods for measuring HSU, indicating preferred choices of instruments and respondents [34,35,38]. Two reports focus on secondary research, providing a systematic methodology to incorporate HSU obtained from the literature into economic models. These two reports provide a framework for the identification, review and synthesis of HSU, noting that original sources should be reviewed for data quality and relevance. Response rate and missing data are presented as part of the minimum quality checks [2,37]. A review by the NICE Decision Support Unit shows that reporting standards relating to basic information on regression models used in decision analytic models are poor, and provides recommendations for improvement [36]. The only publication providing an exhaustive framework for primary research on HSU is the ISPOR Task Force Report on the collection of HSU for economic models in clinical trials [17]. However, as the report mainly focuses on clinical trials, some aspects, like design and analytical considerations for RWE studies, are only briefly discussed. Overall, only a few methodological elements are extensively addressed: the health state definition for utility estimation should be representative of the economic model health state, the preferred method for HSU generation is generally a generic preference-based measure (PBM), and respondents should usually be patients themselves and be representative of the economic model population. Practical advice to avoid bias in HSU generation, particularly selection bias and confounding, is limited. Therefore, a second search of the HTA publications was conducted to identify good practices for observational research (regardless of type of outcomes) in order to complement recommendations for HSU collection. This was a necessary step to gather practical guidance on study conduct in a real-world setting. Seven publications providing recommendations to avoid bias were identified: one from NICE [39] and six from ISPOR [40–45].

As there is no single guideline for investigators seeking guidance on methodological steps to estimate HSU in a real-world setting, recommendations for HSU generation and for observational research were combined together to develop a concise toolbox. The toolbox recommendations aim to help investigators design and conduct a study as well as avoid methodological flaws. The toolbox is organized into four sections, as follows: the first section discusses confounding, the second and third sections address selection bias (study sample and missing data respectively) and the last section discusses information bias. Existing recommendations were summarized in each section of the toolbox, and were then synthesized into practical steps.

Results

The toolbox summarizing the practical steps for HSU generation in a real-world setting is presented in [Box 1](#).

Box 1. Toolbox to generate health state utilities in a real-world setting.**Confounding: study design and analytical techniques**

- Justify the choice of study design
- Explain how confounding variables are identified
- Describe how confounding variables are measured and the risk of residual confounding
- Control for confounding by design or through analytical methods
- Assess basic statistical assumptions and conduct sensitivity analyses
- Explore the risk of unmeasured confounding

Selection bias: study sample

- Determine sample to ensure representativeness and relevance
- Select participants by probability sampling or in a systematic manner
- Implement similar enrollment procedures at all study sites and for each exposure group
- Document participation rates to discuss risk of selection bias

Selection bias: missing data

- Take steps to minimize missing data
- Conduct a descriptive analysis of missing data
- Implement a method to handle missing data

Information bias: exposure and outcome measurement

- Justify the choice of instrument
- Implement standardized procedures for exposure and outcome measurement
- Select features of assessment to ensure efficiency and suitability

Confounding: study design & analytical techniques*Justify the choice of study design*

An important step is to choose between a cross-sectional design or a longitudinal design (e.g., before-after study and prospective study) [40]. Trade-offs of different study designs (e.g., speed, cost, quality, relevance) should be discussed to ensure the study provides useful evidence and is fit for purpose [40,45]. A well-known limitation of longitudinal studies is nonrandom missing data [17]. Conversely, cross-sectional data may provide biased estimates due to underlying heterogeneity across patients [17]. Indeed, there is evidence that a cross-sectional design may overestimate the impact of a health state on utility [17]. However, it is often impractical to wait for patients to progress through different health states [17]. The nature of the condition and health state (e.g., acute/episodic vs chronic disease, severity within a single health state and speed of progression) will influence the applicability of the different designs.

Explain how confounding variables are identified & measured, & describe the risk of residual confounding

At the design stage, potential confounding domains should be identified based on existing literature and/or expert opinion [39,42,44]. Adjusting for intermediate and collider variables is inappropriate and causes bias [39,44]. The use of directed acyclic graphs (causal models that contain directed edges [arrows], linking nodes [variables] and their paths) can help to clarify the structural relationship among variables (confounder, mediator and collider) [43,44]. Naturally, to limit residual confounding, measures of the confounding variables should be valid and reliable.

Control for confounding by design or through analytical methods

Without the advantage of randomization, all observational study designs are vulnerable to time-invariant and/or time-varying confounding [39–45]. Study designs using patients as their own controls (i.e., individual exposure varies) protect against time-invariant confounding [40,41,44]. Any studies without a comparison group are vulnerable to confounding by secular trends or natural history (historical bias) [40,41,45]. Methods controlling for confounding most commonly include matching, restriction, stratification, multivariate regression analysis and propensity score analysis [39,40,42–45]. Methods for dealing with time-varying confounding (affected by prior exposure) include inverse probability weighting of marginal structural models and g-estimation [40,43,44].

Assess basic statistical assumptions & conduct sensitivity analyses

Although regression analysis is common practice in observational studies, it can lead to biased estimates if improperly performed [39,44,45]. Certainly, the functional form of the dependent variable drives the choice of the regression model (along with number of measures and/or clustering) [39,44,45]. However, it may not be an easy

choice considering the unique nature of HSU (left-skewed, censored and ceiling effect) [17]. One option is to transform HSU to the decrement scale (right-skewed), where the back transformation to the original scale does not impact standard errors [17]. Certainly, the original HSU distribution should be displayed using a histogram as a starting point for model selection [36]. The widely used ordinary least squares may not be the best option, and different model types might need to be tested [36,39,44,45]. For longitudinal studies, a common choice is generalized estimating equations [17,44]. As for independent variables, the choice of their inclusion into the model should be justified [36,39,44]. When possible, controlling for baseline utility is recommended [17]. It is also recommended to present the full regression model, not only the adjusted effect [42,44]. Coefficients should be presented along with their uncertainty parameters (e.g., standard errors and confidence intervals) [36,42]. The statistical assumptions that underlie the regression analysis should be tested (e.g., functional form, normality, heteroscedasticity, multicollinearity and overlap) [36,39,42,44,45]. Model performance and goodness of fit should be reported [36,42,44,45]. Key assumptions should be tested in sensitivity analyses, such as distribution of covariates (e.g., creation of categories, cutoffs), functional form (e.g., assumption of nonlinearity) and outliers (e.g., inclusion or exclusion) [36,39,42,44,45].

Explore the risk of unmeasured confounding

Unmeasured confounding may remain after adjustment, mainly due to unmeasured known confounders and unknown confounders. High-dimensional propensity score, additional adjustment (e.g., two-stage sampling and external adjustment) and instrumental variable analysis are proposed approaches to address unmeasured confounding [44]. The use of negative controls (outcomes not impacted by intervention) helps explore the presence of unmeasured confounding [40,42,43]. To substitute qualitative discussions of unmeasured confounding, sensitivity analyses should be performed [42,44]. Simple options include the array approach (explores how the observed association changes by varying the covariates distribution and the strength of association) and the rule-out approach (determines the strength of unmeasured confounding needed to fully explain away the observed association) [44].

Selection bias: study sample

Determine sample to ensure representativeness & relevance

Except for special patient populations (e.g., young children and cognitively impaired), patients should rate their own health [2,17,34,35,37,38]. At design, the choice of eligibility criteria (e.g., disease severity, comorbidities and therapies) and study sites (e.g., primary, secondary and tertiary care) will directly impact generalizability and relevance for economic modeling [2,17,34–38]. To ensure the representativeness of an economic model, sampling should allow reflecting individual variability (i.e., patient heterogeneity) and variability over time (e.g., disease progression) within each health state [17]. The method of recruitment will also affect the representativeness, where an online sample (e.g., social media, recruitment panels and patient associations) may differ (e.g., sociodemographic and health profiles) from a sample of patients recruited through clinical sites [17]. The need to access detailed medical information certainly influences the choice of sample [17].

Select participants systematically, implement standardized enrollment procedures, & document participation rates to discuss risk of selection bias

For HSU generation, a significant threat to internal validity is self-selection bias (e.g., volunteer bias, nonresponse bias). Most often, healthier patients are more likely to participate, which may bias estimates for severe health states [17]. To limit the impact of self-selection bias, patients (exposed and unexposed) should be selected from the same population. Moreover, although information on nonparticipants is often limited, participation rates have to be recorded to document the risk of selection bias [2,17,37,42]. To select the study sample, there are several sampling techniques available. Although probability sampling is the gold standard to ensure sample representativeness, convenience samples are common practice for HSU generation in a real-world setting. To minimize selection bias, systematic participant recruitment such as consecutive sampling should be used [40].

Selection bias: missing data

Take steps to minimize missing data, & conduct a descriptive analysis of missing data

Minimizing missing data by taking preventive steps (e.g., reminders, incentives and multiple attempts) is the best way to limit bias [17,40]. The extent of missing data should be reported by outcome, exposure group and timepoint when applicable [2,17,36,37,42,44]. For longitudinal studies, missing data pattern should be described (i.e., monotonic

[i.e., no return after dropout] or non-monotonic [i.e., intermittent missing data]) [17]. Possible reasons for missing data should be discussed or baseline predictors should be explored [2,17,37,42].

Implement a method to handle missing data

One should explicitly report how missing data are handled [2,17,36,37,42,44]. There are several techniques to deal with missing data, including complete case analysis (CCA) (i.e., restricting analysis to patients without missing data). Drawbacks of CCA are reduced sample size and biased estimates if data are not missing completely at random (MCAR) [44]. Restricting to CCA when data are not MCAR will overestimate HSU, representing a healthier subsample of the target population [17]. The appropriate choice of the analytical approach to deal with missing data (e.g., inverse probability weighting, multiple imputation and Bayesian methods) depends on missing data pattern and mechanism assumption (i.e., MCAR, missing at random or missing not at random). Standard imputation such as last observation carried forward is not recommended for HSU not MCAR [17].

Information bias: exposure & outcome measurement

Justify the choice of instrument

The use of a validated instrument is recommended [2,17,34,35,37,38]. Generic PBM are preferred over condition-specific PBM (and over direct valuation), unless there is good evidence that generic measures are not able to measure changes in health [2,17,34,35,37,38]. When selecting a generic PBM, one should consider the psychometric assessment in the particular health condition and requirements from HTA agencies [2,17,34,35,37,38].

Implement standardized procedures for exposure & outcome measurement, & select features of assessment to ensure efficiency & suitability

The health state for utility estimation should be clearly defined and representative of the economic model health state [2,17,34,35,37,38]. It is not uncommon that the exposure is self-reported (rather than confirmed with medical records), which may impact categorization into health states depending on the complexity of the definitions (e.g., when progression through health states is defined by a laboratory or radiological marker) [17]. Although blinding is the optimal way to protect against information bias, patients are usually aware of their health condition at utility measurement. When applicable, one option is to schedule the timing of assessment before the patient is informed of disease evolution (e.g., at the beginning of a medical visit before receiving progression results), where standardization of the timing would be essential if other measures (e.g., test results) can affect HSU [17]. Frequency of assessments should also be carefully considered and adapted to disease progression, changes, events and duration [17]. One should anticipate the needs of a future economic model and optimize measurement accordingly (e.g., capture changes within and between health states and capture transient effects) [17]. The recall period of a questionnaire should also be considered when planning assessments, particularly when health states fluctuate over a short period of time (e.g., acute events) [2,17]. Moreover, one should choose the most suitable mode of administration (e.g., paper questionnaire, electronic questionnaire and interview), as they offer different advantages and disadvantages, which may best suit different populations and conditions [17]. To minimize the risk of bias, data collection, including mode of administration, should be standardized [17].

Discussion

A review of guidelines published by HTA organizations was conducted to identify standards for HSU generation. Existing recommendations mainly discuss the choice of instruments and respondents. Practical advice to design and conduct a RWE study collecting HSU is limited. Currently, only guidance documents for observational research provide methodological steps to increase internal validity, which are adapted for comparative effectiveness research. Without specific recommendations, RWE studies may generate poor quality HSU data, which can affect the accuracy of assessment of cost-effectiveness. The recent ISPOR report on the identification, review and use of HSU in CUA highlights that the possible bias introduced into CUA may occur at different stages: searching, selecting, synthesizing and use in the model [2]. Considering the numerous steps that can lead to bias, a good starting point to ensure the validity of CUA is the quality of original sources (i.e., studies generating HSU estimates). Although it is recognized that the quality of HSU inputs can ultimately impact reimbursement decision making, none of the HTA organizations have published recommendations for investigators collecting HSU in a real-world setting. Existing guidelines focus either on HSU generation or on RWE, but a single framework is lacking. Therefore, the purpose was to develop a toolbox combining the different steps to help investigators avoid bias. The toolbox

recommendations build on existing guidance, and address important aspects such as confounding, selection bias and information bias. A limitation of the toolbox is that it may not cover all potential issues arising from HSU studies, as it is designed to be concise and practical for common investigators. Moreover, the toolbox is neither a protocol checklist nor a critical appraisal tool. The toolbox focuses on a narrow scope, and does not cover important topics, such as the conduct of economic evaluations, modeling techniques and decision-making processes. For in-depth discussion on a broad range of issues in HTA, further reading is available [33]. Although toolbox recommendations are supported by authoritative references, the next step would be that an international organization as ISPOR publishes an official framework.

Conclusion

The proposed toolbox provides practical steps to help investigators perform good quality research. It is specifically designed for HSU generation in a real-world setting and is complementary to existing good practices. The toolbox aims at increasing the credibility of HSU data for future reimbursement decision making.

Future perspective

In a resource-constrained healthcare system, CUA is essential to help decision making. Although clinical inputs for CUA are generally collected in RCTs, HTA agencies/regulatory bodies have shown growing interest in real-world comparative effectiveness research. Methodological principles for RWE studies have gained prominence in guidelines, and are highly topical in the era of ‘big data.’ Considering the opportunities created by RWE studies, a rising uptake is to be expected. Pitfalls of RWE studies are well-known, yet guidelines for pharmacoeconomics outcomes as HSU provide limited advice. Practical recommendations for investigators would help generate high quality estimates for CUA. Standardization of RWE studies collecting HSU has yet to come.

Executive summary

- Health state utilities (HSU) data collected in real-world evidence (RWE) studies are at risk of bias. This can directly impact the incremental cost–effectiveness ratio and lead to the wrong conclusions on cost–effectiveness.
- A review of the guidelines published by health technology assessment organizations was conducted. Data extraction included key methodological elements affecting internal validity: definition of health state, choice of instrument, mode of administration, features of assessment, choice of respondents, sampling, method of recruitment, response rate, missing data, study design and analytical techniques.
- None of the existing guidelines focus on methodologies for HSU generation in real-life practice. Therefore, current recommendations from the different publications were combined together into a single toolbox.
- The toolbox was developed to help investigators design and conduct a study as well as limit confounding, selection bias and information bias.
- The toolbox aims at increasing the credibility of HSU data for future reimbursement decision making.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/suppl/10.2217/cer-2021-0121

Author contributions

V Lambert-Obry contributed in the conceptualization, writing – original draft preparation. J-P Lafrance, M Savoie and J Lachaine contributed in the conceptualization, writing – reviewing and editing, and supervision.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

References

Papers of special note have been highlighted as: ●● of considerable interest

1. O'Rourke B, Oortwijn W, Schuller T; International Joint Task Group. The new definition of health technology assessment: a milestone in international collaboration. *Int. J. Technol. Assess. Health Care* 36(3), 187–190 (2020).
2. Brazier J, Ara R, Azzabi I *et al.* 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 (2019).
- **Contains a checklist of minimum reporting standards of systematic review of utilities for cost–effectiveness models.**
3. Garrison LP Jr., Neumann PJ, Erickson P, Marshall D, Mullins CD. Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. *Value Health* 10(5), 326–335 (2007).
4. Kreif N, Grieve R, Sadique MZ. Statistical methods for cost–effectiveness analyses that use observational data: a critical appraisal tool and review of current practice. *Health Econ.* 22(4), 486–500 (2013).
5. Guertin JR, Conombo B, Langevin R *et al.* A systematic review of methods used for confounding adjustment in observational economic evaluations in cardiology conducted between 2013 and 2017. *Med. Decis. Making* 40(5), 582–595 (2020).
6. Handorf EA, Heitjan DF, Bekelman JE, Mitra N. Estimating cost–effectiveness from claims and registry data with measured and unmeasured confounders. *Stat. Methods Med. Res.* 28(7), 2227–2242 (2019).
7. Lambert-Obry V, Lafrance JP, Savoie M, Henri S, Lachaine J. Review of real-world evidence studies in Type 2 diabetes mellitus: lack of good practices. *Int. J. Technol. Assess. Health Care* doi:10.1017/S0266462320000392 1–8 (2020) (Epub ahead of print).
8. Faries D, Peng X, Pawaskar M, Price K, Stamey JD, Seaman JW Jr. Evaluating the impact of unmeasured confounding with internal validation data: an example cost evaluation in Type 2 diabetes. *Value Health* 16(2), 259–266 (2013).
9. Frappier J, Tremblay G, Charny M, Cloutier LM. Costing bias in economic evaluations. *J. Med. Econ.* 18(8), 596–599 (2015).
10. Fukuda H, Imanaka Y. Assessment of transparency of cost estimates in economic evaluations of patient safety programmes. *J. Eval. Clin. Pract.* 15(3), 451–459 (2009).
11. Hakanen T, Maljanen T, Lindfors O, Virtala E, Knekt P. Confounding and missing data in cost–effectiveness analysis: comparing different methods. *Health Econ. Rev.* 3(1), 8 (2013).
12. Guertin JR, Bowen JM, De Rose G, O'reilly DJ, Tarride JE. Illustration of the impact of unmeasured confounding within an economic evaluation based on nonrandomized data. *MDM Policy Pract.* 2(1), 2381468317697711 (2017).
13. Clement Nee Shrive FM, Ghali WA, Donaldson C, Manns BJ. The impact of using different costing methods on the results of an economic evaluation of cardiac care: microcosting vs gross-costing approaches. *Health Econ.* 18(4), 377–388 (2009).
14. Evans CJ, Crawford B. Data collection methods in prospective economic evaluations: how accurate are the results? *Value Health* 3(4), 277–286 (2000).
15. Franklin M, Lomas J, Walker S, Young T. An educational review about using cost data for the purpose of cost–effectiveness analysis. *Pharmacoeconomics* 37(5), 631–643 (2019).
16. Ara R, Hill H, Lloyd A, Woods HB, Brazier J. are current reporting standards used to describe health state utilities in cost–effectiveness models satisfactory? *Value Health* 23(3), 397–405 (2020).
17. Wolowacz SE, Briggs A, Belozeroff V *et al.* Estimating health-state utility for economic models in clinical studies: an ISPOR Good Research Practices Task Force Report. *Value Health* 19(6), 704–719 (2016).
- **Provides guidance on how to collect health state utilities in clinical trials.**
18. The Institute for Clinical and Economic Review (ICER) and The Office of Health Economics (OHE). Real world evidence for coverage decisions: opportunities and challenges (2018)
<https://icer-review.org/wp-content/uploads/2018/03/ICER-Real-World-Evidence-White-Paper-03282018.pdf>
19. Rowen D, Azzabi Zouraq I, Chevrou-Severac H, Van Hout B. International regulations and recommendations for utility data for health technology assessment. *Pharmacoeconomics* 35(Suppl. 1), 11–19 (2017).
20. Department of Health and Ageing, Australian Government, Pharmaceutical Benefits Advisory Committee (PBAC).
<https://pbac.pbs.gov.au/>
21. Canadian Agency for Drugs and Technologies in Health (CADTH). <http://www.cadth.ca/>
22. National Institute for Health and Care Excellence (NICE). <http://www.nice.org.uk/>
23. Haute Autorité de Santé (HAS). <http://has-sante.fr>
24. German national institute for quality and efficiency in health care (IQWiG). <http://www.iqwig.de/index.736.en.html>
25. National Health Care Institute (ZIN). <http://www.zorginstituutnederland.nl>
26. Scottish Medicines Consortium (SMC). <http://www.scottishmedicines.org.uk>
27. CatSalut Catalan Health Service. <https://web.gencat.cat/en/temes/salut/>
28. The Dental and Pharmaceutical Benefits Agency (TLV). www.tlv.se

29. European Network for Health Technology Assessment (EUnetHTA). Methodology guidelines. <https://eunethta.eu/methodology-guidelines/>
30. Health Technology Assessment International (HTAi). HTAi vortal. <https://vortal.htai.org>
31. International Health Economics Association (iHEA). Health preference research. <https://www.healtheconomics.org/page/HealthPreference>
32. International Network of Agencies for Health Technology Assessment (INAHTA). HTA tools & resources. <https://www.inahta.org/hta-tools-resources/>
33. International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Good practices for outcomes research. <https://www.ispor.org/heor-resources/good-practices-for-outcomes-research>
34. Brazier J. Valuing health States for use in cost–effectiveness analysis. *Pharmacoeconomics* 26(9), 769–779 (2008).
35. Brazier JE, Rowen D. NICE DSU Technical Support Document 11: alternatives to EQ-5D for generating health state utility values. The National Institute for Health and Care Excellence (2011) <http://www.nicedsu.org.uk>
36. Kearns B, Ara R, Wailoo AG. A review of the use of statistical regression models to inform cost effectiveness analyses within the NICE technology appraisals programme: report by the decision support unit. The National Institute for Health and Care Excellence (2012) www.nicedsu.org.uk
37. Papaioannou D, Brazier JE, Paisley S. NICE DSU Technical Support Document 9: the identification, review and synthesis of health state utility values from the literature. The National Institute for Health and Care Excellence (2011) <http://www.nicedsu.org.uk>
38. Rowen D, Brazier JE, Wong R, Wailoo AG. Measuring and valuing health-related quality of life when sufficient EQ-5D data is not available. The National Institute for Health and Care Excellence (2020) <http://nicedsu.org.uk>
39. Faria R, Alava MH, Manca A, Wailoo AG. NICE DSU Technical Support Document 17: the use of observational data to inform estimates of treatment effectiveness in technology appraisal: methods for comparative individual patient data. The National Institute for Health and Care Excellence (2015). <http://www.nicedsu.org.uk>
- **Provides guidance on how to conduct real-world evidence studies for comparative effectiveness research.**
40. Berger ML, Dreyer N, Anderson F, Towse A, Sedrakyan A, Normand SL. Prospective observational studies to assess comparative effectiveness: the ISPOR good research practices task force report. *Value Health* 15(2), 217–230 (2012).
- **Provides guidance on how to conduct prospective comparative effectiveness studies.**
41. Berger ML, Mamdani M, Atkins D, Johnson ML. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report – Part I. *Value Health* 12(8), 1044–1052 (2009).
42. Berger ML, Martin BC, Huserau D *et al.* A questionnaire to assess the relevance and credibility of observational studies to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. *Value Health* 17(2), 143–156 (2014).
- **Contains a 33-item questionnaire to assess the relevance and credibility of comparative effectiveness studies.**
43. Cox E, Martin BC, Van Staa T, Garbe E, Siebert U, Johnson ML. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report–Part II. *Value Health* 12(8), 1053–1061 (2009).
44. Johnson ML, Crown W, Martin BC, Dormuth CR, Siebert U. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report–Part III. *Value Health* 12(8), 1062–1073 (2009).
- **Discusses analytical methods to avoid bias in retrospective comparative effectiveness studies.**
45. Motheral B, Brooks J, Clark MA *et al.* A checklist for retrospective database studies–report of the ISPOR Task Force on Retrospective Databases. *Value Health* 6(2), 90–97 (2003).
- **Contains a 27-item questionnaire to assess the relevance and validity of retrospective database studies.**