International Journal of Technology Assessment in Health Care

cambridge.org/thc

Policy

Cite this article: Vreman RA, Strigkos G, Leufkens HGM, Schünemann HJ, Mantel-Teeuwisse AK, Goettsch WG (2022). Addressing uncertainty in relative effectiveness assessments by HTA organizations. *International Journal of Technology Assessment in Health Care* **38**, e17, 1–8. https://doi.org/10.1017/S026646232100177X

Received: 27 July 2021
Revised: 15 December 2021
Accepted: 27 December 2021

Key words:
Health technology assessment; Uncertainty; Relative effectiveness; Added benefit; Clinical benefit; Value; Risk of bias; Grade; Evidence

Author for correspondence:
Rick A. Vreman,
E-mail: r.a.vreman@uu.nl

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

CAMBRIDGE UNIVERSITY PRESS

Addressing uncertainty in relative effectiveness assessments by HTA organizations

Rick A. Vreman[1,2], George Strigkos[1], Hubert G. M. Leufkens[1], Holger J. Schünemann[3], Aukje K. Mantel-Teeuwisse[1] and Wim G. Goettsch[1,2]

[1]Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences (UIPS), Utrecht University, Universiteitsweg 99, 3584 CG Utrecht, The Netherlands; [2]The National Health Care Institute (ZIN), Diemen, The Netherlands and [3]Cochrane Canada; McMaster GRADE centre, WHO Collaborating Centre for Infectious Diseases, Research Methods and Recommendations, Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada

Abstract

This study outlines the ways in which different health technology assessment (HTA) organizations deal with uncertainty in relative effectiveness assessments (REAs), using the GRADE framework as a common reference. Guidelines regarding REA and uncertainty assessment methods and three most recent HTA reports (as of April 2020) of seven HTA organizations in Germany, England and Wales, France, the Netherlands, Europe (EUnetHTA), the USA, and Canada were included. First, it was analyzed how each organization addressed uncertainty on the following levels of evidence: (i) individual studies, (ii) body of evidence for one outcome, (iii) body of evidence across all outcomes, and (iv) added net benefit. Second, the extent to which HTA organizations considered the eight domains of certainty of evidence defined by GRADE was assessed. For individual studies, checklists were the most common approach to express uncertainty (4/7 organizations). Uncertainty in the body of evidence for all outcomes and in added benefit was combined in a single conclusion by five organizations. All organizations reported on at least 4/5 downgrading domains of GRADE, while the three upgrading domains were reported less. The operationalization of the assessment of multiple domains was unclear due to vague or absent guidelines. HTA organizations consider most domains of the GRADE framework, but approaches to assess uncertainty within REAs on different levels of evidence differ substantially between organizations. More alignment and guidance on the best methods to deal with uncertainty within HTA could lead to more clarity for stakeholders and to more aligned reimbursement recommendations.

Introduction

In order to benefit from approved drugs, patients usually require drugs to be reimbursed by publicly funded healthcare systems. Funding of drugs often follows a positive recommendation by a health technology assessment (HTA) organization. HTA bodies perform at least a relative effectiveness assessment (REA) compared with existing care standards (comparators), and sometimes also incorporate other aspects in their evaluations, for example, cost-effectiveness assessments (CEA). REAs reach an at least ordinal outcome that can inform pricing and reimbursement on a national level (1–3). Each HTA organization has its own set of preferred methods and processes and codifies its conclusions in different ways. Although HTA bodies follow evidence-based medicine (EBM) principles and endorse international frameworks assisting evidence assessment and decision-making [e.g., the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework], the criteria considered, and the weight attributed to each of them may vary widely (2;4).

Uncertainty in the effects of interventions is one such criterion, whose management varies not only between HTA bodies but within them as well and could be one possible explanation for discrepancies between HTA recommendations (5;6). These discrepancies have sparked the discussion for the implementation of more structured methodologies like Multi-Criteria Decision Analysis in various settings (7;8). However, even in these state-of-the-art methodologies, addressing the various types of uncertainty remains a major challenge (9).

The modern regulatory environment—with the availability of less clinical data at authorization (e.g., in the case of a conditional marketing authorization)—entails that decisions must be made with less confidence that the estimates regarding relative effectiveness are correct, that is, with less certainty (1;10–13). As a result, the way that (un)certainty is defined, comprehended, assessed, appraised, and expressed by HTA bodies is of great importance in the reimbursement decision-making process.

Within REAs, various terms are used to express (un)certainty from "strength or quality of evidence" to "confidence in the estimates" or the most recent "certainty of evidence" in the guidelines of the GRADE framework (14). GRADE has introduced and is building upon a holistic view toward the evaluation of the certainty of evidence. The framework is widely used and recognized worldwide (15). Other entities like the U.S. Preventive Services Task Force (USPSTF) or the Agency for Healthcare Research and Quality (AHRQ) refer to the same concept and use similar or identical domains with similar terminology, like "strength of evidence" (16;17).

It is currently unclear how different HTA organizations evaluate uncertainty within their REAs. This study therefore aims to outline the ways in which different HTA agencies evaluate and express uncertainty in drugs' relative effectiveness and to analyze commonalities and differences, using the GRADE framework as a common reference.

Methods

Inclusion of Jurisdictions

Pharmaceutical markets from Northwest Europe and North America with HTA guidelines publicly available through Web sites in English were selected. This led to the selection of the following seven HTA jurisdictions: Germany (Institute for Quality and Efficiency in Health Care—IQWiG), England and Wales (National Institute for Health and Clinical Excellence—NICE), France (High Authority for Health—HAS), the USA (Institute for Clinical and Economic Review—ICER), the Netherlands (National Health Care Institute—ZIN), Canada (Canadian Agency for Drugs and Technologies in Health—CADTH), and Europe as a whole (European Network for Health Technology Assessment—EUnetHTA). It was already known to the authors that ZIN used the GRADE framework and that EUnetHTA endorsed it. However, it was decided not to actively exclude these jurisdictions from the sample but include them in order to determine whether their guidelines indeed clearly reported used methods and the role of GRADE.

Included Data

To assess the methods employed by the different HTA organizations, we aimed to analyze and compare the role that uncertainty played within their REAs, based on publicly available guidelines and a set of three assessment reports per organization. Guidelines in this paper are defined as documents that describe the processes and methodologies used by the different organizations.

Regarding the guidelines, we included those published by the agencies regarding clinical evaluation or REA methods. The guidelines were searched for on the Web sites of the agencies. In the case of multiple complementary guidelines, each was included. Guidelines not (fully) available in English were translated through Google Translate.

Regarding the assessment reports, the practical application of the guidelines in each organization's REA processes was evaluated by assessing three recently published REA reports. The three most recent reports from the date of retrieval (April 2020) that were available in English were selected. Conclusions on the certainty of evidence were searched for in the relevant relative effectiveness chapters of the assessments, the approaches were compared with those stated in the guidelines, and deviations were recorded. Supplementary Appendix 1 reports the included assessment reports per jurisdiction.

In order to verify that all relevant guidelines had been included and that the interpretation of the guidelines and assessment reports was correct, we contacted the HTA bodies under research. We received comments from all assessed HTA organizations except from HAS.

Analysis

Different HTA organizations have different names for the "certainty of evidence." Similarly, certainty and its antonym, uncertainty, may be used interchangeably. These names are reported within the results section that describes approaches of each organization, but in this study, we further use only the term certainty of evidence.

From the guidelines and assessment reports retrieved for each HTA organization, profiles were established that clarified each organization's approach to expressing the certainty of evidence within their REA procedure. These profiles aimed to provide a narrative description of the employed methods, thereby providing inputs for three analyses:

First, a qualitative discussion of the role that each HTA organization assigns to uncertainty within their REAs. This discussion included whether the evaluation of uncertainty is explicitly mentioned as a goal of the REA.

Second, an assessment of the evaluation of uncertainty on different levels of evidence. The following four levels of evidence were defined, where the certainty of evidence can be assessed:

(1) individual studies, which refers to certainty of individual studies;
(2) body of evidence for one outcome, which refers to multiple studies that constitute the body of evidence for one outcome;
(3) body of evidence across all outcomes, which refers to multiple studies (body of evidence) across all outcomes; and
(4) added benefit, which refers to the final verdict on added net health benefit versus a comparator, balancing all benefits and harms.

Third, we assessed the extent to which HTA organizations aimed to consider each of the domains of uncertainty of the GRADE framework. The GRADE framework was used as a common reference in order to make comparisons across HTA organizations (14). The certainty of evidence term used by GRADE includes eight domains that affect its grading of the uncertainty. In the GRADE framework, these domains affect the certainty grading either downward (five domains) or upward (three domains) on a four-level scale (very low, low, moderate, and high) *via* fixed but flexible relationships that allow for structured assessments. More information can be drawn from Table 1. More information on the GRADE approach can be retrieved from Supplementary Appendix 2.

For this third analysis, it was reported for each of the HTA organizations which domains of GRADE they considered in their assessments. Five categories were used to report how each domain was addressed within REAs of HTA organizations. The first stated whether the domain was considered as part of the certainty of evidence grading (called "Considered as part of certainty grading"), the second whether it was addressed separately (called "Considered separately"), the third whether it was reported to be

https://doi.org/10.1017/S026646232100177X Published online by Cambridge University Press

Table 1. The GRADE domains for rating the certainty of evidence (with higher levels expressing more certainty), the direction of effect of each domain, and the magnitudes of effect.

Domain	Explanation by GRADE	Effect on level of certainty ranking
Risk of bias (limitations in study design or execution)	Also referred to as problems with (internal) validity by others. Regards study design or conduct.	−1, −2, or −3 levels
Imprecision (random error)	Regards how narrow the CIs are (sample size & number of events) and whether "they cross a clinical decision threshold that dictates recommending versus not recommending the intervention".	−1 or −2 levels
Inconsistency	Regards differences in the magnitude of effect (heterogeneity) and direction of the effect.	−1 or −2 levels
Indirectness	Regards the applicability (different populations or interventions), surrogate outcomes, and indirect comparisons.	−1 or −2 levels
Publication bias	Regards cases when studies are left out leading to a downward (positive studies left out) or an upward (negative studies left out) bias in the estimate of effect.	−1 or −2 levels
Large effect	Refers to cases of "extreme confidence in effectiveness".	+1 or +2 levels
Dose-response gradient	A long-recognized important factor for establishing a cause-effect relationship.	+1 level
Plausible confounding	Also "residual confounding or biases". Regards cases when the final effect might be larger than the data suggest.	+1 level

More details on the GRADE approach are provided in Supplementary Appendix 2.

considered but the operationalization was unclear (called "Considered but operationalization unclear"), the fourth when the domain was explicitly not considered ("Not considered"), and the fifth when nothing was reported regarding the domain ("Not reported").

Results

The Role of Uncertainty within Each Organization's REA Process

In the guidelines of *IQWiG*, certainty is explicitly stated as a goal of the REA process, with the relevant guidelines giving a clear picture on the approach that IQWiG pursues. EBM guidelines are followed, and special reference is made to the GRADE framework as an example of these. Quantitative and qualitative certainty is being assessed on a three-level scale for the individual study level. The assessment of uncertainty is among others referred to as the "certainty of conclusions." Uncertainty both for one and across all outcomes is reported based on a different three-level scale in combination with the magnitude of added benefit (hint, indication, proof + magnitude of added benefit), or none of these three certainty ratings apply (in the case of a lack of data or when none of the other conclusions can be drawn). The conclusion for certainty across all outcomes already includes a balance between benefits and harms and as a result serves as the same to that of added benefit (18).

In the *HAS* guidelines regarding the assessment and appraisal of drugs for reimbursement purposes, assessing certainty is not explicitly stated as a goal of the REA process. HAS introduced its own, custom evidence assessment framework in a document nonspecific to the pricing and reimbursement processes of medicinal products (19). Terminologies that are used are "quality of research evidence" and "level of proof." In the HAS framework, uncertainty is explicitly assessed on scales both for the individual studies (low, intermediate, strong) and body of evidence (4, 3, 2, 1) level, without however making any distinction between one or across all outcomes. It should be noted that special reference to uncertainty is made in lower (added) benefit ratings (20), both in the guidelines and the assessments included in this study. That is to say that a low (added) benefit grade may imply a low magnitude of added benefit as well as uncertainties identified in the evidence. Reference to the scales is not made in recent guidelines of the HAS "Transparency Committee" or in recent assessments (20).

In the *NICE* HTA process, the available evidence is synthesized in a systematic review, following the extensive and publicly available guidance provided by the University of York (21). Terminologies used by NICE are "quality of the relevant clinical effectiveness evidence" and "critical appraisal." The quality

https://doi.org/10.1017/S026646232100177X Published online by Cambridge University Press

assessment of evidence in technology appraisals (nonspecific to pharmaceuticals) should be a part of the systematic review (22). In their user guide for the Single Technology Appraisal (STA) process (23), a checklist is cited containing the minimum criteria, which should be considered per study for all appraised technologies. Based on the assessments included, it is indicated that uncertainty is referred to in a narrative way for the body of evidence level. Finally, uncertainty may be implied in the final recommendation in cases of, for example, "only in research" conclusions, where at the time-point of appraisal there is not enough clinical evidence. The above should be viewed alongside the general perspective of NICE of considering clinical effectiveness simultaneously with cost-effectiveness (22).

For *ZIN*, the estimation of the certainty is stated as a major goal of the REA process, alongside the estimation of the magnitude of benefit (24;25). Terminologies used are "the probability of the effect" and the "quality of evidence." The probability of the effect is assessed based on the GRADE approach whenever possible. On the individual study level, checklists are used, based on EBM literature. GRADE evidence profiles (tables expressing the certainty of evidence on a per outcome basis) are produced whenever possible. This was also verified by recent REA reports, where two out of the three reports included followed GRADE minutely. The one assessment report not reporting uncertainty based on the GRADE framework was performed in the context of a transnational initiative (BeNeLuxA). Narrative approaches express uncertainty on an added benefit level (24).

EUnetHTA states that assessing the certainty of evidence is one of the two goals of the clinical effectiveness domain of the HTA Core Model, alongside an assessment of the magnitude of added benefit (26). A general preference for the GRADE approach is expressed, without this being binding. Terminologies used are the "certainty of the evidence," the "quality of evidence," and the "level of evidence." The methods regarding the assessment of the individual domains contributing to uncertainty are further described in extensive guidelines separately (27–30). These methods are borrowed from the approaches of various international frameworks depending on the domain, including AHRQ, GRADE or Cochrane, and mainly concern a per study approach. Risk of bias, external validity, and directness of evidence, which correspond to the separate guidelines mentioned above, are appraised in separate chapters in the assessments included, mainly on a per study basis. A narrative reflection on the available evidence is presented in the "Conclusions" section of the REA, based on the assessments included. EUnetHTA assessments principally do not draw conclusions regarding the added benefit (31).

ICER explicitly states certainty as being a primary goal alongside establishing the magnitude of added benefit. The terminologies used by ICER include the "level of certainty in the evidence." The assessment of the certainty of evidence is based on multiple guidance documents (AHRQ, USPSTF, and GRADE). Based on the assessments included in this study, the USPSTF scale is used for assessing the Risk of Bias on a per study basis. Explicit provisions regarding assessing and reporting uncertainty on the individual or the body of evidence level were not identified. The "ICER Matrix" allows for systematic reporting of certainty on the added benefit level. However, the certainty rankings within the ICER Matrix have a purposely unconstrained relationship with the uncertainty domains, and their influence is completely left to the assessors' judgment, on an *ad hoc* basis, followed by transparent reporting (32). The ICER matrix is comprised of two axes: an *x*-axis of ascending "comparative net health benefit" and a *y*-axis of ascending "level of certainty in the evidence." When making conclusions on overall net benefit, ICER assigns a "joint rating" (consisting of a letter and possibly a +/− symbol) that corresponds to a predefined space on the ICER matrix and gives information both for the magnitude of added benefit as well as the certainty surrounding it (32).

Finally, for *CADTH*, uncertainty assessment is part of the drug REA process, in the critical appraisal of evidence chapters (33). The terminologies used by CADTH include the "strength of the body of evidence" and the "critical appraisal." CADTH does not explicitly use checklists nor GRADE. Instead, an annotated template is used for evaluating uncertainty in individual studies and on an individual outcome and across all outcomes levels. A narrative reflection on the available evidence is presented in the discussion and conclusion sections of the REA. Guidelines referring specifically to uncertainty are not publicly available. Last but not least, it is worth noting that in their final recommendations, and more specifically in the "reimburse with conditions" recommendation, uncertainty can be implied in a way similar to the "only in research" recommendation of NICE or low (added) benefit ratings of HAS. Detailed profiles of all HTA jurisdictions can be found in Supplementary Appendix 1.

The Levels of Evidence at Which Certainty Is Assessed

A summary of the approaches for evaluating uncertainty on the different levels of evidence can be found in Table 2. For individual studies, most organizations use a checklist or some sort of template to evaluate and express uncertainty. On the three evidence levels that include a body of evidence, uncertainty is either expressed on a scale or narratively. Of note, the organizations that provide an uncertainty rating on a scale usually complement that rating with a narrative discussion of uncertainty.

Domains of GRADE Addressed

Table 3 shows for each of the domains of GRADE whether it is evaluated in the REAs of HTA organizations. Most of the domains defined by GRADE are also addressed by HTA organizations. The relationship of the domains to the certainty of evidence grading is however rarely fixed in a way such as it is for GRADE, either purposefully or at least it is not reported. For instance, ICER explicitly states that different domains may play a larger or smaller role on the final certainty rating in the ICER Matrix. On the other hand, IQWiG implements a more restrained approach, having *a priori* defined, for example, the role that the number of studies or direction of results might have in the certainty of conclusions (see Supplementary Appendix 1). No domains were found that were explicitly not considered by HTA organizations, but for several organizations, nothing was reported about some domains.

Discussion

Summary of Results

Most of the studied HTA bodies state assessment of uncertainty to be one of the main goals of REAs, next to establishing the magnitude of added net benefit. All of the included organizations assess uncertainty based on EBM standards on multiple levels of evidence. The specific levels of evidence at which uncertainty

https://doi.org/10.1017/S026646232100177X Published online by Cambridge University Press

Table 2. Levels of evidence and methods of expressing uncertainty in relative effectiveness assessment

Certainty of evidence on the level of:	HAS	IQWiG	NICE	ZIN	EUnetHTA	ICER	CADTH
Added benefit	N	S	N	N		S	N
Body of evidence for all outcomes	S				N		
Body of evidence for one outcome		S		S	S		N
Individual studies	S	S	C	C	C	S	C

: Explicitly reported, : Not reported

S: Scale, **N**: Narrative, **C**: Checklist/template

(and added benefit) is assessed differ between organizations, and the measures to express uncertainty are equally different (checklists, scales narrative). Most of the uncertainty domains as defined by the GRADE approach were covered. The less reported domains were the three upgrading domains and publication bias out of the downgrading domains.

Implications of Findings

All guidelines of the studied HTA bodies praise transparency in reporting their rationale during the decision-making process. However, the operationalization of uncertainty assessment into the processes of organizations is frequently unclear, as has been previously reported (6), while guidelines vary widely in structure and clarity when it comes to specifications regarding uncertainty assessment.

In most organizations, there was no predefined relationship of uncertainty rankings with the domains or components that contribute to uncertainty, contrary to the GRADE approach where the domains are operationalized in a way that they can operate independently by judgments on one domain leading to rating down or combined judgments (e.g., imprecision and inconsistency) that lead to lowering the certainty by one or more levels. This could be a possible explanation why the weight attributed to various uncertainty components may be unclear, even within one specific HTA organization (5;34). Taking into account the varied picture across jurisdictions, one could hypothesize that this diversity would multiply when the picture is viewed across various organizations.

This study further reveals some interesting findings regarding the terminology used to communicate concepts of uncertainty: similar words may often be used to describe either identical or closely related concepts, which may however essentially differ. The quality of research evidence (used by HAS, including four domains, see the black circles in Table 3) and the quality of evidence (used by ZIN, including all eight domains) are only some examples of similar terms expressing closely related concepts that vary in their level of complexity. Similarly, there are cases where different terms are used to refer to an identical concept. These discrepancies re-emphasize the lack of a common philosophy previously reported in the HTA environment (5;6).

More systematic research on the possible sources of the mixed picture observed in this study could provide useful insights, as some may be informed by true national preferences while other differences may simply be a result of the gradual development of uncertainty assessment in each jurisdiction without a universal theoretical basis grounded in generic risk science (35). Nevertheless, even though differences are evident, all HTA organizations emphasize the relevance of uncertainty assessments and the incorporation of qualitative reflection next to numerical estimations and probabilities.

Future Outlook

The inconsistencies revealed in this study could at least partly be resolved by providing more structured and broader guidelines that offer a holistic approach to uncertainty. A useful addition could be that of defining more clearly the effects that various uncertainty components have on the conclusions of uncertainty in the different levels of evidence, without that meaning that an algorithmic approach is encouraged. The results further indicate that guidelines on uncertainty assessment and expression could benefit from being better defined and structured. Pre-specifying structure and wording preferences, as is the case for NICE in their "guidelines" process (36), when the expression of uncertainty is based on a narrative approach could prevent misinterpretations caused by the ambiguity of language (37).

https://doi.org/10.1017/S026646232100177X Published online by Cambridge University Press

Table 3. Domains included in the certainty of evidence evaluations by HTA organizations

	HAS	IQWiG	NICE	ZIN	EUnetHTA	ICER	CADTH
Risk of bias	●	●	●	●	● / ◐[b]	●	●
Imprecision	●	●	●	●	● / ◐[b]	●	●
Inconsistency	●	●	●	●	● / ◐[b]	●	●
Indirectness	●	●	●	●	● / ◐[b]	●	●
Publication bias	-	●	●	● /○[a]	● / ◐[b]	●	●
Large effect	◐	●	-	●	●	●	○
Dose-response gradient	-	○	-	●	●	●	○
Plausible confounding	-	●	-	●	●	-	○

● Considered as part of certainty grading; ◐ Considered separately;
○ Considered but operationalization unclear; - Not reported

[a]ZIN uses the GRADE framework and refers to all its domains in their guidelines. However, publication bias was not found to be reported in the recent reports included in the study.
[b]EUnetHTA urges assessors to use the GRADE approach. However, they have also developed separate guidelines for various of the domains.

The role of GRADE should be highlighted. It offers a platform to systematically reflect, discuss, and conclude on aspects relevant to the certainty of evidence that arise during decision-making in HTA, in particular when its tools such as the GRADEpro app are used to perform the assessment. GRADE is not the only tool to assess uncertainty, and as the results have shown, various HTA organizations have used other tools next to GRADE to inform the construction of their own uncertainty assessment process. Other well-established tools include the one of the Evidence-based Practice Center (EPC) program—but this tool is based on GRADE principles and contains similar domains—and that of the USPSTF, which includes the domains research design, internal validity, applicability, precision, consistency, and additional factors. These domains essentially match the GRADE domains, although GRADE combines research design and execution in a single domain (risk of bias) and uses slightly different terminologies (applicability is called indirectness, precision is imprecision, consistency is inconsistency). The large congruency between different tools (even those that were developed independently of each other) is an indication that each of these tools really does cover the most important domains. This is reassuring because a limitation of our approach to use GRADE as a common reference is that we did not capture any domains that would fall outside the scope of GRADE. However, larger discrepancies are noted within the upgrading domains of GRADE. These are not included within the USPSTF method and also our results indicate that now all HTA organizations systematically address these domains. This begs the question whether HTA organizations might need to advance their methods or that these domains are simply not relevant to all HTA organizations. There is no clear-cut answer to this question, but the endorsement of EUnetHTA of GRADE establishes at least in principle that these domains are found to be relevant to some extent by HTA organizations. The benefit of having a common method is that it provides the shared philosophy and language and that it promotes consistency between and within jurisdictions (whether it would be GRADE or another method). The restrained preference toward the GRADE approach provided by EUnetHTA is an important step for reaching a unified solution for uncertainty assessment. EUnetHTA is working on a new statement regarding the use of GRADE, which may be helpful in international alignment on the assessment of uncertainty (38).

Further research into performed assessments and the evaluation of uncertainty within them could provide an interesting basis for comparisons across HTA jurisdictions and could give valuable insights on aspects such as jurisdictions' relative risk willingness.

Limitations

This study has several limitations. A limited set of countries has been selected in order to eliminate language and reporting issues. Practices within Western-European and North-American countries do not necessarily reflect the practices employed by other countries. The conclusions of this study therefore do not necessarily apply to other jurisdictions, although it may be expected that discrepancies in approaches to handling uncertainty will be similar or greater due to the absence of systematic approaches

https://doi.org/10.1017/S026646232100177X Published online by Cambridge University Press

within some of the excluded countries. Especially the upgrading domains of GRADE were not considered by all HTA organizations, even more variability in these domains might be expected for organizations that fell outside our inclusion criteria. The review of guidelines as a method is not without its limitations. Guidance documents are essentially provisions that practical work could deviate from. The level of detail, structure, and consistency of guidelines significantly vary across jurisdictions. Three recent REA reports were therefore included for every jurisdiction to offer complementary information and validate the implementation of guidelines in practice, and the organizations themselves were consulted. Nevertheless, a larger assessment report sample followed by systematic analysis could shed more light on the approaches of HTA organizations on uncertainty in practice. We used GRADE as a method to compare all other methods. This implies that aspects not covered by GRADE fall outside of the scope of our assessment and therefore may not have been included in the comparison. We are somewhat reassured by the fact that other evidence grading tools cover essentially the same downgrading domains, which would be an indication that it is not highly likely that other aspects related to uncertainty would be assessed, but it can nevertheless not be excluded that some HTA organization assesses an aspect that we did not capture. Our study focused on pharmaceuticals but most of the HTA organizations also assess other technologies. Guidelines are sometimes not specific for pharmaceuticals. GRADE may play a different role within assessments of other technologies than the one it plays for pharmaceuticals; therefore, our results apply to drugs only.

Conclusions

Approaches to assess uncertainty within REAs on different levels of evidence differ substantially between HTA organizations. The expression of uncertainty varies from structured (explicit, checklists, or scales) to freer approaches (implicit, narrative). Discrepancies in the uncertainty domains evaluated demonstrate a lack of a universal definition on the certainty of evidence. More alignment and guidance on the best methods to deal with uncertainty within HTA could lead to more clarity for stakeholders that wish to generate relevant evidence, and to more aligned recommendations regarding relative effectiveness.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S026646232100177X.

Author Contribution Statement. R.A.V., G.S., A.K.M.-T., H.J.S., and W.G.G. conceived the study. G.S. and R.A.V. analyzed the data. R.A.V., G.S., A.K.M.-T., H.G.M.L., H.J.S., and W.G.G. interpreted the data. All authors drafted and revised the manuscript. All authors approved the final manuscript. R.A.V. is the guarantor.

Funding. No funding was received for this study.

Disclaimer. The views expressed in this article are the personal views of the authors and may not be understood or quoted as being made on behalf of or reflecting the position of the agencies or organizations with which the authors are affiliated.

Conflicts of Interest. The authors report no conflicts of interest relevant to this study. H.G.M.L. reports he is a member of the Lygature Leadership Team. H.J.S. reports that he is a co-chair of the GRADE working group.

References

1. **Vreman RA, Bouvy JC, Bloem LT, Hövels AM, Mantel-Teeuwisse AK, Leufkens HGM, et al.** Weighing of evidence by health technology assessment bodies: Retrospective study of reimbursement recommendations for conditionally approved drugs. *Clin Pharmacol Ther*. 2019;**105**:684–91.
2. **Vreman RA, Mantel-Teeuwisse AK, Hövels AM, Leufkens HGM, Goettsch WG**. Differences in health technology assessment recommendations among European jurisdictions: The role of practice variations. *Value Health*. 2020;**23**:10–6.
3. **Kelly MP, Moore TA**. The judgement process in evidence-based medicine and health technology assessment. *Soc Theory Health*. 2012;**10**:1–19.
4. **Angelis A, Lange A, Kanavos P**. Using health technology assessment to assess the value of new medicines: Results of a systematic review and expert consultation across eight European countries. *Eur J Health Econ*. 2018;**19**:123–52.
5. **Allen N, Walker SR, Liberti L, Salek S**. Health technology assessment (HTA) case studies: Factors influencing divergent HTA reimbursement recommendations in Australia, Canada, England, and Scotland. *Value Health*. 2017;**20**:320–8.
6. **Akehurst RL, Abadie E, Renaudin N, Sarkozy F**. Variation in health technology assessment and reimbursement processes in Europe. *Value Health*. 2017;**20**:67–76.
7. **Towse A, Barnsley P**. Approaches to identifying, measuring, and aggregating elements of value. *Int J Technol Assess Health Care*. 2013;**29**:360–4.
8. **Marsh K, IJzerman M, Thokala P, Baltussen R, Boysen M, Kaló Z, et al.** Multiple criteria decision analysis for health care decision making—Emerging good practices: Report 2 of the ISPOR MCDA emerging good practices task force. *Value Health*. 2016;**19**:125–37.
9. **Oliveira MD, Mataloto I, Kanavos P**. Multi-criteria decision analysis for health technology assessment: Addressing methodological challenges to improve the state of the art. *Eur J Health Econ*. 2019;**20**:891–918.
10. **Eichler H-G, Pignatti F, Flamion B, Leufkens H, Breckenridge A**. Balancing early market access to new drugs with the need for benefit/risk data: A mounting dilemma. *Nat Rev Drug Discov*. 2008;**7**:818–26.
11. **Sculpher M, Claxton K**. Establishing the cost-effectiveness of new pharmaceuticals under conditions of uncertainty—When is there sufficient evidence? *Value Health*. 2005;**8**:433–46.
12. **Vreman RA, Naci H, Goettsch WG, Mantel-Teeuwisse AK, Schneeweiss SG, Leufkens HGM, et al.** Decision making under uncertainty: Comparing regulatory and health technology assessment reviews of medicines in the United States and Europe. *Clin Pharmacol Ther*. 2020;**108**:350–7.
13. **Vreman RA, Bloem LT, van Oirschot S, Hoekman J, van der Elst ME, Leufkens HG, et al.** The role of regulator-imposed post-approval studies in health technology assessments for conditionally approved drugs. *Int J Health Policy Manag*. 2020.
14. **Hultcrantz M, Rind D, Akl EA, Treweek S, Mustafa RA, Iorio A, et al.** The GRADE working group clarifies the construct of certainty of evidence. *J Clin Epidemiol*. 2017;**87**:4–13.
15. **Institute of Health Economics** [Internet] Evidence grading systems used in health technology assessment practice. Institute of Health Economics. Canada. 2018. [cited 2020 Dec 8]. Available from: https://www.ihe.ca/advanced-search/evidence-grading-systems-used-in-health-technology-assessment-practice.
16. **Berkman ND, Lohr KN, Ansari M, McDonagh M, Balk E, Whitlock E, et al.** Grading the strength of a body of evidence when assessing health care interventions for the effective health care program of the agency for healthcare research and quality: An update. In: *Methods guide for effectiveness and comparative effectiveness reviews [Internet]*. Rockville (MD): Agency for Healthcare Research and Quality (US). 2008 [cited 2020 Dec 8]. (AHRQ Methods for Effective Health Care). Available from: http://www.ncbi.nlm.nih.gov/books/NBK174881/.
17. **Sawaya GF, Guirguis-Blake J, LeFevre M, Harris R, Petitti D, U.S. Preventive Services Task Force**. Update on the methods of the U.S. Preventive Services Task Force: Estimating certainty and magnitude of net benefit. *Ann Intern Med*. 2007;**147**:871–5.
18. **Institute for Quality and Efficiency in Healthcare (IQWiG) [Internet]**. General methods - Version 5.0. Köln (Germany); 2017. Available from: https://www.iqwig.de/download/General-Methods_Version-5-0.pdf.

19. **Haute Autorité de Santé (HAS) [Internet].** Niveau de preuve et gradation des recommandations de bonne pratique. Saint-Denis (France). 2013. Available from: https://www.has-sante.fr/upload/docs/application/pdf/2013-06/etat_des_lieux_niveau_preuve_gradation.pdf.
20. **Haute Autorité de Santé (HAS) [Internet].** Medicinal products assessment. Transparency Committee Doctrine. Principles of medicinal products assessment and appraisal for reimbursement purposes.. Saint-Denis (France). 2018. Available from: https://www.has-sante.fr/upload/docs/application/pdf/2019-07/doctrine_de_la_commission_de_la_transparence_-_version_anglaise.pdf.
21. **Centre for Reviews and Dissemination (CRD) [Internet]**. CRD's guidance for undertaking reviews in healthcare. CRD, University of York. 2009. Available from: https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf.
22. **National Institute for Health and Care (NICE) [Internet]**. Guide to the methods of technology appraisal 2013 (PMG9). 2013. Available from: https://www.nice.org.uk/process/pmg9/.
23. **National Institute for Health and Care Excellence (NICE) [Internet]**. Single technology appraisal: User guide for company evidence submission template (PMG24). 2018. Available from: https://www.nice.org.uk/process/pmg24/.
24. **Zorginstituut Nederland (ZIN) [Internet]**. Assessment of "established medical science and medical practice" (English summary). Diemen (The Netherlands). 2015. Available from: https://english.zorginstituutnederland.nl/binaries/zinl-eng/documents/reports/2015/01/19/assessment-of-established-medical-science-and-medical-practice/Assessment+of+established+medical+science+and+medical+practice.pdf.
25. **Zorginstituut Nederland (ZIN) [Internet]**. Package advice in practice. Diemen (The Netherlands). 2017. Available from: https://english.zorginstituutnederland.nl/binaries/zinl-eng/documents/reports/2018/09/05/package-advice-in-practice---deliberations-for-arriving-at-a-fair-package/Package+advice+in+practice+-+Deliberations+for+arriving+at+a+fair+package.pdf.
26. **European Network for Health Technology Assessment (EUnetHTA) [Internet]**. EUnetHTA Joint Action 2, Work Package 8. HTA Core Model ® version 3.0 (Pdf). 2016. Available from: www.htacoremodel.info/BrowseModel.aspx.
27. **European Network for Health Technology Assessment (EUnetHTA) [Internet].** Levels of evidence: Internal validity of randomised controlled trials. 2015. Available from: https://www.eunethta.eu/wp-content/uploads/2018/01/16_WP7-SG3-GL-int_val_RCTs_amend2015.pdf.
28. **European Network for Health Technology Assessment (EUnetHTA) [Internet]**. Levels of evidence: Applicability of evidence for the context of a relative effectiveness assessment. 2015. Available from: https://www.eunethta.eu/wp-content/uploads/2018/01/Levels-of-Evidence-Applicability-of-evidence-for-the-context-of-a-relative-effectiveness-assessment_Amended-JA1-Guideline_Final-Nov-2015.pdf.
29. **European Network for Health Technology Assessment (EUnetHTA) [Internet]**. Internal validity of non-randomised studies (NRS) on interventions. 2015. Available from: https://www.eunethta.eu/wp-content/uploads/2018/01/Internal-validity-of-non-randomised-studies-NRS-on-interventions_Guideline_Final-Jul-2015_0.pdf.
30. **European Network for Health Technology Assessment (EUnetHTA) [Internet]**. Comparators & Comparisons: Direct and indirect comparisons. 2015. Available from: https://eunethta.eu/wp-content/uploads/2018/01/Comparators-Comparisons-Direct-and-indirect-comparisons_Amended-JA1-Guideline_Final-Nov-2015.pdf.
31. **European Network for Health Technology Assessment (EUnetHTA) [Internet]**. Pharmaceutical submission: What is a health technology assessment (HTA)? [cited 2020 Jun 11]. Available from: https://eunethta.eu/services/submission-guidelines/pharmaceutical-submission/.
32. **Ollendorf DA, Pearson SD. ICER evidence rating matrix: A user's guide [Internet].** 2017. Available from: https://icer-review.org/wp-content/uploads/2016/02/Rating-Matrix-User-Guide-UPDATED-06.30.17.pdf.
33. **Canadian Agency for Drugs and Technologies in Health (CADTH) [Internet]**. Therapeutic review framework and process [Internet]. 2019. Available from: https://cadth.ca/.
34. **Allen N, Liberti L, Walker SR, Salek S**. A comparison of reimbursement recommendations by European HTA agencies: Is there opportunity for further alignment? *Front Pharmacol*. 2017;**8**. [cited 2018 Feb 14]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5491965/.
35. **Aven T.** How risk science can support decision-making under uncertainty. In: *Global c*hallenges *s*cience *w*eek [Internet]. Grenoble: Université Grenoble Alpes; 2019. Available from: https://www.youtube.com/watch?v=YH8x-xcakYM.
36. **National Institute for Health and Care Excellence (NICE) [Internet]**. Developing NICE guidelines: The manual (PMG20). 2014. Available from: http://www.nice.org.uk/article/pmg20.
37. **EFSA Scientific Committee**, Benford D, Halldorsson T, Jeger MJ, Knutsen HK, More S, et al. The principles and methods behind EFSA's guidance on uncertainty analysis in scientific assessment. EFSA J. 2018;16:e05122.
38. **European Network for Health Technology Assessment (EUnetHTA) [Internet]**. Assembly & Forum 2020. 2020. Available from: https://eunethta.eu/wp-content/uploads/2020/03/2020-EUnetHTA-Assembly-Forum- Welcome-Guide.pdf.

https://doi.org/10.1017/S026646232100177X Published online by Cambridge University Press