

AdViSHE: A Validation-Assessment Tool of Health-Economic Models for Decision Makers and Model Users

P. Vemer^{1,2} · I. Corro Ramos³ · G. A. K. van Voorn⁴ · M. J. Al³ · T. L. Feenstra^{1,5}

Published online: 11 December 2015

© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract

Background A trade-off exists between building confidence in health-economic (HE) decision models and the use of scarce resources. We aimed to create a practical tool providing model users with a structured view into the validation status of HE decision models, to address this trade-off.

Methods A Delphi panel was organized, and was completed by a workshop during an international conference. The proposed tool was constructed iteratively based on comments from, and the discussion amongst, panellists. During the Delphi process, comments were solicited on the importance and feasibility of possible validation techniques for modellers, their relevance for decision makers, and the overall structure and formulation in the tool.

Results The panel consisted of 47 experts in HE modelling and HE decision making from various professional and international backgrounds. In addition, 50 discussants actively engaged in the discussion at the conference workshop and returned 19 questionnaires with additional comments. The final version consists of 13 items covering all relevant aspects of HE decision models: the conceptual model, the input data, the implemented software program, and the model outcomes.

Conclusions Assessment of the Validation Status of Health-Economic decision models (AdViSHE) is a validation-assessment tool in which model developers report in a systematic way both on validation efforts performed and on their outcomes. Subsequently, model users can establish whether confidence in the model is justified or whether additional validation efforts should be undertaken. In this way, AdViSHE enhances transparency of the validation status of HE models and supports efficient model validation.

Electronic supplementary material The online version of this article (doi:[10.1007/s40273-015-0327-2](https://doi.org/10.1007/s40273-015-0327-2)) contains supplementary material, which is available to authorized users.

✉ P. Vemer
p.vemer@rug.nl

¹ Department of Epidemiology, University Medical Center Groningen, University of Groningen, PO Box 30.001, 9700 RB Groningen, The Netherlands

² University of Groningen, Pharmacoepidemiology and Pharmacoeconomics (PE2), Groningen, The Netherlands

³ Institute for Medical Technology Assessment (iMTA), Erasmus University, Rotterdam, The Netherlands

⁴ Biometris, Wageningen University and Research, Wageningen, The Netherlands

⁵ Centre for Nutrition, Prevention and Health Services Research, National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands

Key Points

Model users can accept health-economic (HE) decision models as valid without further examination, thereby reducing model confidence, or they can validate models themselves, implying overlap with the validation efforts of the modelling team. Existing modelling and validation guidelines give little guidance in setting priorities for validation, nor do they address the issue of overlapping work by model developers and users.

Assessment of the Validation Status of Health-Economic decision models (AdViSHE) allows model developers to provide model users with structured information regarding the validation status of their HE decision model. Its main purpose is to avoid some of the current overlap in validation efforts and provide information on a list of priority validation items, selected by a Delphi consensus process. AdViSHE can be used to reproduce stated results and guide complementary validation efforts, which is expected to increase model users' understanding of, and confidence in, the model and its outcomes.

1 Introduction

The use of health-economic (HE) decision models can have extensive consequences for payers, patients, and practitioners alike. Since HE models have become a fixed part of the modern decision making process in healthcare policy [1], they should be validated before they are used. This is commonly done by the modelling team and sometimes extensively so. Since the cost of model validation can be significant, both in time and money [2], model users, that is, people using the outcomes of the model, such as reimbursement decision makers, could simply presume that the models are valid without further examination. However, this unquestioning acceptance may reduce the users' overall confidence in the model, especially when the modelling team has an economic interest in favourable outcomes [3]. Model users therefore often validate models themselves, leading to a possibly improved validation status of the model but also an overlap of work between the modelling team and model users.

We are thus presented with a trade-off between building confidence in the model and the use of scarce resources. Several guidelines and publications address model validity and quality assessment, both for simulation models in

general [4–6] and for HE decision models in particular [3, 7–11]. However, they do not address the trade-off referred to above or support modellers in setting validation priorities. A prioritized list of validation efforts with the general support of the research community may reduce possible waste of resources while improving the general validation status of HE models.

The aim of this study was therefore to create a practical tool for model developers to fill in during or shortly after model development. This tool provides model users with a structured view into the validation status of the model, according to a consensus on what good model validation entails. The tool may also provide guidance towards additional validation. This tool, called Assessment of the Validation Status of Health-Economic decision models (AdViSHE), may, for example, be part of dossiers sent to the (national) decision maker when applying for reimbursement, or it may be appended to manuscripts on modelling applications to support peer reviewers.

2 Methods

We defined validation as the act of evaluating whether a model is a proper and sufficient representation of the system it is intended to represent, in view of a specific application. Here, “proper” means that the model is in accordance with what is known about the system, and “sufficient” means that the results can serve as a solid basis for decision making [12].

2.1 Initial List

A literature search generated an initial gross list of validation techniques. Explicit attention was given to the inclusion of validation practices from outside the HE literature. Precise definitions were formulated to avoid confusion between terms that may be used interchangeably in daily practice.

2.2 Expert Input

In five e-mail rounds, HE experts commented on the initial list and drafts of AdViSHE. The setup of these rounds was based on the Delphi method, a structured communication technique in which experts answer questions in two or more rounds. The key element is that experts are encouraged to revise their earlier answers in light of the replies of other members of their panel in order to reach consensus [13–15]. The design of each round was not fixed beforehand, but was based on the outcomes of the previous round. A summary of the commentary from previous rounds was provided and every participant was actively encouraged to

Table 1 Source of contact information of health-economic experts

Source	Number of experts contacted
Personal network of the project team	129
Proposed replacements by invitees	10
Involved in the ISPOR-SMDM Good Modeling Practices Task Force	31
Authors of the CHEERS statement [11]	6
Involved in the organizing committee of at least one of the ISPOR conferences (International, European, Asia-Pacific, and Latin America) between 2008 and 2014	140
Involved in the ISPOR regional chapters	100
Identified by other experts	19
Identified by biomedexperts.com as experts in “Economic Models” and/or “Markov Chains”	35

CHEERS Consolidated Health Economic Evaluation Reporting Standards, *ISPOR* International Society For Pharmacoeconomics and Outcomes Research, *SMDM* Society for Medical Decision making

comment and provide suggestions for additions; all experts were allowed to refine or change their opinion. Steps were taken to include a wide variety of nationalities, work environments, and expertise (Table 1). In between rounds, new experts were approached to enhance international diversity and to counter attrition.

Comments on an early draft of AdViSHE were solicited from employees of Zorginstituut Nederland (the Dutch Healthcare Institute), the primary advisory council for the Dutch Ministry of Health regarding reimbursement. Zorginstituut Nederland is representative of the field of policy decision makers for whom AdViSHE might be useful.

A conference workshop was organized in Montreal, Canada, where attendees discussed the first full draft of the tool amongst themselves. Three of the authors (PV, GVV, ICR) actively approached groups of discussants. All participants were encouraged to comment using a questionnaire. All comments made during this workshop were collected and incorporated in the final draft, which was sent out to the Delphi panel in a final round of comments. It was then edited for language, after which the project group agreed on the final version of AdViSHE.

2.3 Case Studies

The applicability of AdViSHE was tested by applying it to two case studies. Both were HE decision models in which the study authors were involved. The first model was built specifically for a Diagnostic Assessment Report commissioned by the National Institute for Health Research Health Technology Assessment Programme [16]. It was programmed in Microsoft Excel and assessed devices used to

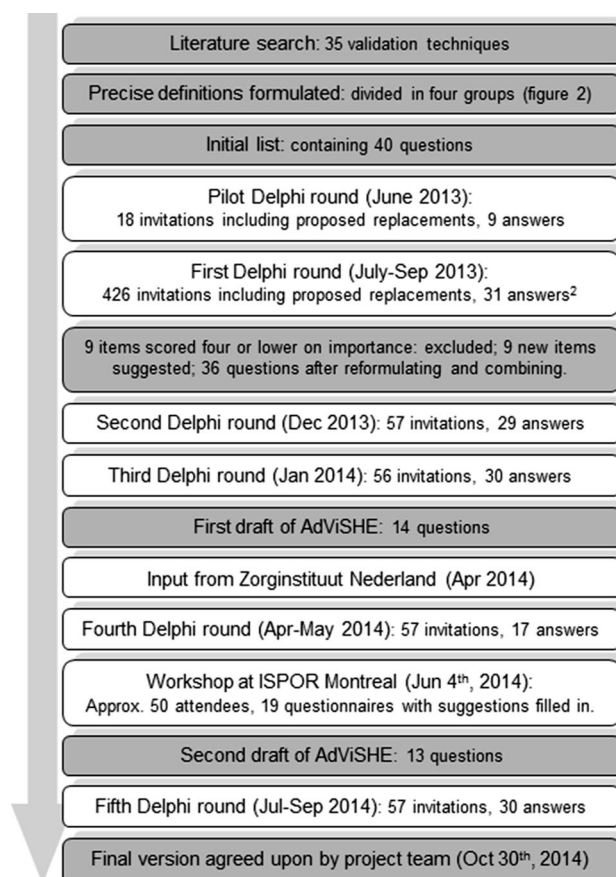


Fig. 1 Building the validation-assessment tool. Grey boxes display work by the project team; white boxes display input from outside sources. ¹High non-response since the invitations were sent out to a very wide range of people with the aim of selecting a suitable panel; see Table 1. *AdViSHE* Assessment of the Validation Status of Health-Economic decision models

assist with the diagnosis, management, and monitoring of haemostasis disorders. The second one, with a multi-use design and programmed in Wolfram Mathematica, is a dynamic Dutch population model of chronic obstructive pulmonary disease (COPD) progression [17]. AdViSHE was filled in using knowledge of the models and their development. The focus of this exercise was to identify any problems with AdViSHE that a model developer might encounter when applying it to a model.

3 Results

3.1 Building AdViSHE

The process of building AdViSHE is depicted in Fig. 1 and additional information is given in the online supplementary appendix (see electronic supplementary material, online resource 1). The literature search yielded 35 validation techniques [4–6, 9, 10, 18–22], which were then divided

Fig. 2 Typology of validation techniques, based on [4]

Conceptual validation:

Techniques used to determine whether the theories and assumptions underlying the **conceptual model** (the description of the underlying system using a mathematical, logical, verbal, or graphical representation) are correct and that the model's representation of the problem entity and the model's structure, logic and mathematical and causal relationships are "reasonable" for the intended purpose of the model.

Data validation:

Techniques used to determine whether the available **input data** is appropriate, accurate and sufficient and that data transformations were correctly performed.

Computerized model validation:

Techniques to ensure that the **implemented software program**, including code, mathematical calculations and implementation of the conceptual model, are performed correctly and are consistent with the model's specifications.

Operational validation:

Techniques used to determine whether the behaviour of the **model outcomes** has sufficient accuracy for the model's intended purpose over the domain of the model's intended applicability.

into groups covering all aspects of model validation (Fig. 2). The Delphi panel ran from June 2013 until September 2014. Background information on the HE experts can be found in Table 2. The questions raised in each round are presented in Fig. 3.

In the pilot and first rounds, every respondent could comment on the full initial list, but was asked to at least comment on the techniques grouped within two of the four groups. Based on these rounds, nine techniques were dropped from the initial list, since they were deemed unimportant by the panel (Figs. 1, 3). Nine new techniques proposed by panellists were included. Several items were reformulated and combined.

To limit the burden, each panellist in the second and third rounds received a subset of five validation techniques to comment on, using a factorial design. The purpose was to improve the definitions of the techniques, make necessary clarifications, and hold an open discussion on the usefulness of each item. Contrary to the first round, no quantitative scoring was performed in these rounds. Based on the first draft of AdViSHE built after the third round, Zorginstituut Nederland suggested that the investigation of outliers, which was excluded in a previous round due to an

average "importance" of 3.8 (below 4) was explicitly mentioned in AdViSHE. The Delphi panel was asked to comment on the amended first draft.

The conference workshop also discussed the first draft. It was attended by approximately 50 participants; 19 filled-in questionnaires were returned. Three workshop participants indicated that they were also members of the Delphi panel.

The second draft was based on comments from the fourth round and the workshop. Based on the workshop, a final question was added, asking whether modellers have performed any validation techniques not covered in AdViSHE. The fifth Delphi round yielded no further substantial comments. The project team finalized the tool in October 2014.

3.2 Final Version

The final version of AdViSHE consists of 13 questions (Fig. 4). All questions are grouped to cover its various aspects: the conceptual model, the input data, the implemented software program, and the model outcomes (Fig. 2). The tool is designed to be filled in by modellers to

Table 2 Background information of participants who answered during at least one of the five Delphi rounds

Participant characteristics	Number (%)
Total number of individuals who answered in at least one round	47 (100)
Geographical region [26]	
Western Europe	28 (60)
Southern Europe	5 (11)
Northern Europe	4 (9)
Eastern Europe	3 (7)
North America	2 (4)
Central America	2 (4)
South America	1 (2)
Southern Asia	1 (2)
Australia and New Zealand	1 (2)
Field of work	
Academics	25 (53)
Consulting	8 (17)
Pharmaceutical industry	8 (17)
Government, decision making	6 (13)
Number of responses	
Provided comments three times or more	24 (50)
Provided comments twice	12 (26)
Provided comments once	12 (26)

report in a structured way on the efforts performed to improve the validation status of their HE decision model and the outcomes of these efforts. The information required to fill AdViSHE in is often available in other places, but is not collected systematically.

3.3 Application to Case Studies

Filling in AdViSHE took a little over 1 h for each model. No further issues with the formulation, structure or usability of AdViSHE were found. It was noted that filling in AdViSHE is best done during model development or soon after.

4 Discussion

4.1 Application

The validation-assessment tool AdViSHE allows the developers to provide the users with structured information regarding the validation status of their HE decision model. Its main purpose is to reduce the workload of model users and avoid some of the current overlap in validation efforts, thus saving resources. It does so by reporting which

validation efforts have been undertaken in a structured way, and giving the results of these efforts. AdViSHE is not intended to replace validation by model users, but rather to reproduce stated results and guide complementary validation efforts. By doing so, it is expected that the model users will gain a greater understanding of, and confidence in, the model and its outcomes.

AdViSHE can be particularly useful for decision makers who have to evaluate a reimbursement dossier. In that regard, the UK stands out internationally by providing independent experts with an 8-week window to validate HE decision models [23]; other jurisdictions have much shorter timelines. In the Netherlands, for instance, Zorginstituut Nederland has 3 weeks to comment on an HE model and its outcomes before the reimbursement submission is sent to the assessment committee (Wetenschappelijke adviesraad, WAR) [24]. Since manufacturer submissions rarely report on model validity, this often has to be assessed independently. Due to time and money constraints, model validation is sometimes insufficient to establish confidence among users. To that end, inclusion of AdViSHE in the reimbursement process could improve this process, in particular, because it reports on validation criteria based on consensus.

Participants in both the Delphi panel and the workshop specifically asked for a short, quick-scan version. AdViSHE could serve that purpose, using the answer options “Yes” or “No” for each of the questions. This checklist could be useful during the modelling process to ascertain whether all important validation efforts have been considered. It could also be useful during the process of research dissemination, by accompanying academic articles or conference presentations.

AdViSHE gives neither a validity score, nor a threshold for one. There are several reasons for this choice: a model may receive a passing score and yet have a defect that needs to be corrected; the subjective nature of this approach tends to be hidden so the assessment appears to be objective; scores may cause overconfidence in a model; and scores can be used to argue that one model is better than another [4]. Just as models must be tailored to an application, validation efforts must be tailored to a specific model. Therefore, no a priori “red flag” or “must do” labels have been defined for AdViSHE. A validation effort that a model user deems indispensable for one application may not be considered necessary for another.

4.2 Methodology

The methodology used in this study is not a Delphi panel in a strict sense. In a Delphi panel, a group of experts, usually small, is given one specific question to answer. Each participant is free to request additional data, which is then shared

Fig. 3 HE expert questions. *AdViSHE* Assessment of the Validation Status of Health-Economic decision models, *HE* health-economic. ¹ *ISPOR* International Society For Pharmacoeconomics and Outcomes Research

Pilot round

Qualitative assessment of familiarity with techniques and of their importance and feasibility.

First round

Familiarity Assessed qualitatively
Importance Allocating a number of points to all techniques within a group; maximum number of points was 5 times the number of techniques within the group.
Feasibility Considering a typical case of developing an HE decision model, was each technique “unfeasible”, “feasible only under conditions” or “always feasible”? Comments were requested when a technique was deemed “feasible only under conditions”.

Second and third round

Techniques that scored 4 or less on “importance” in the first round were excluded in these rounds; techniques proposed by panellists were included. General comments were made on each item, while showing comments from earlier rounds and a summary of feasibility results from the first round. Each panel member was allocated five validation techniques in a factorial design. Comments from panellists were grouped according to clarity of the question, suggestions for reformulation, feasibility of the technique, relevance of the technique for decision makers and suggestions for operationalization.

Fourth round

First draft of *AdViSHE*. Comments were requested on the following issues:

- Informative for decision makers
- Level of detail
- Length of the tool
- As a modeller, would you feel comfortable filling in this tool?
- Missing items
- Redundant items
- Items that are unclear or need reformulation
- Further comments

Workshop

Group discussion was held in two rounds during a workshop at the *ISPOR*¹ 19th Annual International Meeting in Montreal, Canada. Participants were asked to form groups of 5-6 people each and discuss the following issues:

- Missing or redundant items regarding validation of HE models
- Any items that need reformulation
- Can the structure of the questionnaire be improved?
- For decision makers: Will a filled-in version of *AdViSHE* be informative?
- For modellers: Will you be able to complete *AdViSHE* for your models?

Fifth round

Participants were asked for comments on second draft of *AdViSHE*.

with the rest, along with their opinions [13–15]. In our study, no single question could be posed in a meaningful way. We therefore recruited a relatively large group of respondents, each being asked to answer a subset of questions. Thereby we mimicked the Delphi method as closely as possible. In some rounds, we applied a factorial design to reduce the number of questions presented to each respondent while keeping several respondents for each question. The added value of interaction between respondents that the Delphi method provides was explicitly incorporated.

AdViSHE assumes that the modelling process is performed with generally accepted modelling and reporting techniques. This could mean that the model builders are adhering to the International Society For

Pharmacoeconomics and Outcomes Research (*ISPOR*)—Society for Medical Decision Making (*SMDM*) Modeling Good Research Practices and to the Consolidated Health Economic Evaluation Reporting Standards (*CHEERS*) statement [1, 11]. *AdViSHE* does not evaluate the model itself, nor the building process.

The intention of our study was to obtain a useful tool that promotes the validation status of HE models by putting resources to good use and providing a consensus on the reporting on this effort. Close to 100 people have actively participated and discussed (parts of) *AdViSHE*. This large, diverse group of respondents was one of the biggest strengths of our study. In our Delphi panel, we have representatives of many different geographical regions and

Fig. 4 AdViSHE: Assessment of the Validation Status of Health-Economic decision models. ¹ *ISPOR* International Society For Pharmacoeconomics and Outcomes Research, *SMDM* Society for Medical Decision making, *CHEERS* Consolidated Health Economic Evaluation Reporting Standards

AdViSHE

Assessment of the Validation Status of Health-Economic decision models

AdViSHE contains 13 items that modellers can complete to report on the efforts performed to improve the validation status of their health-economic (HE) decision model. The tool is not intended to replace validation by model users but rather to inform the direction of validation efforts and to provide a baseline for replication of the results. In addition to using it after a model is finished, AdViSHE can be used to guide validation efforts during the modelling process.

The modellers are asked to comment on the validation efforts performed while building the underlying HE decision model and afterwards. Many of the items can be answered simply by referring to the model documentation. AdViSHE is divided into five parts, each covering an aspect of validation:

- Part A: Validation of the conceptual model (2 questions)
- Part B: Input data validation (2 questions)
- Part C: Validation of the computerized model (4 questions)
- Part D: Operational validation (4 questions)
- Part E: Other validation techniques (1 question)

No final validation score is calculated, as the assessment of the answers and the overall validation effort is left to the model users. It is assumed that the model has been built according to prevailing modelling and reporting guidelines. For instance, the model builders would presumably adhere to the ISPOR-SMDM¹ Modeling Good Research Practices (Caro et al., 2010) and/or CHEERS¹ Statement (Husereau et al., 2013). Some questions may not be applicable to a particular model. If this is the case, the model builder should take the opt-out option and provide a justification of why this item is not deemed applicable.

Part A: Validation of the conceptual model (2 questions)

Part A discusses techniques for validating the conceptual model. A conceptual model describes the underlying system (e.g., progression of disease) using a mathematical, logical, verbal, or graphical representation. Please indicate where the conceptual model and its underlying assumptions are described and justified.

<p>A1/ Face validity testing (conceptual model): Have experts been asked to judge the appropriateness of the conceptual model?</p> <p>If yes, please provide information on the following aspects:</p> <ul style="list-style-type: none"> - Who are these experts? - What is your justification for considering them experts? - To what extent do they agree that the conceptual model is appropriate? <p>If no, please indicate why not.</p>
<p>Aspects to judge include: appropriateness to represent the underlying clinical process/disease (disease stages, physiological processes, etc.); and appropriateness for economic evaluation (comparators, perspective, costs covered, etc.).</p>
<p>A2/ Cross validity testing (conceptual model): Has this model been compared to other conceptual models found in the literature or clinical textbooks?</p> <p>If yes, please indicate where this comparison is reported.</p> <p>If no, please indicate why not.</p>

working fields relevant to HE modelling. Their input gave useful information both of a qualitative and a quantitative nature, and our design allowed for the suggestion of other methods not yet included in the list. The workload of the Delphi panel was relatively low due to the factorial design in the second and third rounds. Although this meant that participants did not comment on all validation techniques, it did keep participants interested.

One of the limitations of our study was that the time path to consensus was more than a year and it was labour intensive to filter all information after each round. Since this filtering of information was to some extent subjective, there is no complete certainty that consensus was unanimous, although the reactions to the full drafts were positive. A final limitation of the study is that the original search for validation techniques was not based on a

Fig. 4 continued

Part B: Input data validation (2 questions)

Part B discusses techniques to validate the data serving as input in the model. These techniques are applicable to all types of models commonly used in HE modelling.

Please indicate where the description and justification of the following aspects are given:

- search strategy;
- data sources, including descriptive statistics;
- reasons for inclusion of these data sources;
- reasons for exclusion of other available data sources;
- assumptions that have been made to assign values to parameters for which no data was available;
- distributions and parameters to represent uncertainty;
- data adjustments: mathematical transformations (e.g., logarithms, squares); treatment of outliers; treatment of missing data; data synthesis (indirect treatment comparison, network meta-analysis); calibration; etc.

B1/ Face validity testing (input data): Have experts been asked to judge the appropriateness of the input data?

If yes, please provide information on the following aspects:

- Who are these experts?
- What is your justification for considering them experts?
- To what extent do they agree that appropriate data have been used?

If no, please indicate why not.

Aspects to judge may include but are not limited to: potential for bias; generalizability to the target population; availability of alternative data sources; any adjustments made to the data.

B2/ Model fit testing: When input parameters are based on regression models, have statistical tests been performed?

If yes, please indicate where the description, the justification and the outcomes of these tests are reported.

If no, please indicate why not.

Examples of regression models include but are not limited to: disease progression based on survival curves; risk profiles using regression analysis on a cohort; local cost estimates based on multi-level models; meta-regression; quality-of-life weights estimated using discrete choice analysis; mapping of disease-specific quality-of-life weights to utility values.

Examples of tests include but are not limited to: comparing model fit parameters (R^2 , Akaike information criterion (AIC), Bayesian information criterion (BIC)); comparing alternative model specifications (covariates, distributional assumptions); comparing alternative distributions for survival curves (Weibull, lognormal, logit); testing the numerical stability of the outcomes (sufficient number of iterations); testing the convergence of the regression model; visually testing model fit and/or regression residuals.

Part C: Validation of the computerized model (4 questions)

Part C discusses various techniques for validating the model as it is implemented in a software program. If there are any differences between the conceptual model (Part A) and the final computerized model, please indicate where these differences are reported and justified.

systematic review of the literature. We started with a list of ten guidelines from inside and outside the HE field and listed all techniques mentioned in these guidelines. By allowing the Delphi panel and the workshop participants to add techniques they considered useful, we have used an alternative approach to sufficiently guarantee inclusion of all relevant techniques on the list of items considered.

4.3 Comparison to Other Tools

Several tools that have been published in the past few years deal with the quality assessment of HE decision models [3,

8]. Others deal with the quality of reporting of HE decision models [7, 9, 11]. Only one of these recent tools refers explicitly to validation, namely that of Caro et al., which briefly discusses validation as a part of the tool's overall "credibility" [3]. The Drummond and Jefferson [7] and Consensus on Health Economic Criteria (CHEC) [8] checklists, and the Philips framework [9] were intended to be filled in by modellers to help them in the model development process, although it is implied that model users can fill them into evaluate models [7, 9]. The CHEERS checklist was built to be used by both model developers and model users, in particular, editors and peer reviewers evaluating publication

Fig. 4 continued

<p>C1/ External review: Has the computerized model been examined by modelling experts?</p> <p>If yes, please provide information on the following aspects:</p> <ul style="list-style-type: none"> - Who are these experts? - What is your justification for considering them experts? - Can these experts be qualified as independent? - Please indicate where the results of this review are reported, including a discussion of any unresolved issues. <p>If no, please indicate why not.</p>
<p>Aspects to judge may include but are not limited to: absence of apparent bugs; logical code structure optimized for speed and accuracy; appropriate translation of the conceptual model.</p>
<p>C2/ Extreme value testing: Has the model been run for specific, extreme sets of parameter values in order to detect any coding errors?</p> <p>If yes, please indicate where these tests and their outcomes are reported.</p> <p>If no, please indicate why not.</p>
<p>Examples include but are not limited to: zero and extremely high (background) mortality; extremely beneficial, extremely detrimental, or no treatment effect; zero or extremely high treatment or healthcare costs.</p>
<p>C3/ Testing of traces: Have patients been tracked through the model to determine whether its logic is correct?</p> <p>If yes, please indicate where these tests and their outcomes are reported.</p> <p>If no, please indicate why not.</p>
<p>In cohort models, this would involve listing the number of patients in each disease stage at one, several, or all time points (e.g., Markov traces). In individual patient simulation models, this would involve following several patients throughout their natural disease progression.</p>
<p>C4/ Unit testing: Have individual sub-modules of the computerized model been tested?</p> <p>If yes, please provide information on the following aspects:</p> <ul style="list-style-type: none"> - Was a protocol that describes the tests, criteria, and acceptance norms defined beforehand? - Please indicate where these tests and their outcomes are reported. <p>If no, please indicate why not.</p>
<p>Examples include but are not limited to: turning sub-modules of the program on and off; altering global parameters; testing messages (e.g., warning against illegal or illogical inputs), drop-down menus, named areas, switches, labelling, formulas and macros; removing redundant elements.</p>

Part D: Operational validation (4 questions)

Part D discusses techniques used to validate the model outcomes.

<p>D1/ Face validity testing (model outcomes): Have experts been asked to judge the appropriateness of the model outcomes?</p> <p>If yes, please provide information on the following aspects:</p> <ul style="list-style-type: none"> - Who are these experts? - What is your justification for considering them experts? - To what extent did they conclude that the model outcomes are reasonable? <p>If no, please indicate why not.</p>
<p>Outcomes may include but are not limited to: (quality-adjusted) life years; deaths; hospitalizations; total costs.</p>

potential of economic evaluations [11]. The checklist by Caro et al. was specially built to be filled in by model users [3]. Using these tools for their intended purpose will hence often add to the workload of model users and may overlap with work already done by the developers. In addition to the mentioned checklists, AdViSHE was specifically intended to be filled in by modellers, while its outcome is immediately useful to model users.

There are also several tools that deal with model validation for simulation models in general [4–6]. However, these present ideals rather than a priority list of feasible acceptability criteria. In addition, most recommendations are necessarily general and not geared towards validating HE models [4–6]. The limited number of validation techniques in AdViSHE is a consensus between what is feasible and what is necessary in HE modelling. For specific

D2/ Cross validation testing (model outcomes): Have the model outcomes been compared to the outcomes of other models that address similar problems?

If yes, please provide information on the following aspects:

- Are these comparisons based on published outcomes only, or did you have access to the alternative model?
- Can the differences in outcomes between your model and other models be explained?
- Please indicate where this comparison is reported, including a discussion of the comparability with your model.

If no, please indicate why not.

Other models may include models that describe the same disease, the same intervention, and/or the same population.

D3/ Validation against outcomes using alternative input data: Have the model outcomes been compared to the outcomes obtained when using alternative input data?

If yes, please indicate where these tests and their outcomes are reported.

If no, please indicate why not.

Alternative input data can be obtained by using different literature sources or datasets, but can also be constructed by splitting the original data set in two parts, and using one part to calculate the model outcomes and the other part to validate against.

D4/ Validation against empirical data: Have the model outcomes been compared to empirical data?

If yes, please provide information on the following aspects:

- Are these comparisons based on summary statistics, or patient-level datasets?
- Have you been able to explain any difference between the model outcomes and empirical data?
- Please indicate where this comparison is reported.

If no, please indicate why not.

D4.A/ Comparison against the data sources on which the model is based (dependent validation).

D4.B/ Comparison against a data source that was not used to build the model (independent validation).

Part E: Other validation techniques (1 question)

E1/ Other validation techniques: Have any other validation techniques been performed?

If yes, indicate where the application and outcomes are reported, or else provide a short summary here.

Examples of other validation techniques: structured "walk-throughs" (guiding others through the conceptual model or computerized program step-by-step); naïve benchmarking ("back-of-the-envelope" calculations); heterogeneity tests; double programming (two model developers program components independently and/or the model is programmed in two different software packages to determine if the same results are obtained).

Fig. 4 continued

applications, additional items may of course be very important, which can be reported in the last part of AdViSHE. As a priority list, AdViSHE thus supplements existing tools and guidelines with different purposes.

Despite efforts to make the evidence as objective as possible, the judgment of model validity (and confidence) will ultimately be subjective. It is therefore of paramount importance that model users can make their own assessment. AdViSHE makes this possible in an efficient way: it asks not only which validation aspects were tested but also how they were tested and where the outcomes are reported. Other tools just provide general suggestions for which aspects should be discussed. The exception is the CHEERS checklist, which also asks specifically where certain items are reported [3, 7–9, 11].

4.4 Terminology

There is little if any consensus on terminology in the validation literature [25], even in the field of HE. The problem of ambiguity is exacerbated by the different meanings of the same terms in computer science and psychometrics. For example, conceptual model validation is sometimes called content validity [9, 18], but in psychometrics, this term indicates whether a measure represents all facets of a given social construct. Computerized model validity is sometimes called verification, internal validity, internal consistency, technical validity, and debugging; moreover, all of these terms have additional and divergent meanings. Notably, internal validity was interpreted differently by several members of the Delphi panel.

In AdViSHE, we have attempted to steer clear of terminology that may be considered confusing. We present a lucid overview of possible techniques, with clear definitions, to be used in the validation of HE decision models. This explains the discrepancy between our terms and the classification of validation types by the recent ISPOR-SMDM Modeling Good Research Practices Task Force [10].

5 Conclusion

A validation-assessment tool for HE models called Assessment of the Validation Status of Health-Economic decision models (AdViSHE) has been developed to address the trade-off that model users potentially experience, between a loss of confidence resulting from lacking or unreported validation efforts, and an inefficient use of resources resulting from overlapping validation efforts by the modelling team and model users. In addition, it presents a certain consensus among model users and model developers on what is good validation. The tool is tailored for

the validation of HE models through the involvement of a large group of HE experts, coming from many backgrounds and countries. In AdViSHE, model developers comment on the validation efforts performed while building the underlying HE decision model. This information can subsequently be applied by model users, such as people involved in decision making or peer reviewers, to establish whether confidence in the model is warranted or additional validation efforts should be undertaken. The tool thus reduces the overlap between the validation efforts of model developers and those of model users without leading to a loss of confidence in the model or its outcomes.

Acknowledgments The project team would like to thank the participating health-economic experts from around the world who have contributed to this study as part of the Delphi panel. Their names are listed in the Sect. 7. Participants of workshop W2 during the ISPOR 19th International Meeting in Montreal, Canada, are thanked for their active contribution. Three anonymous reviewers are thanked for their helpful comments. The first author would also like to thank <http://acronymcreator.net/for> suggesting the name of the tool.

AdViSHE study group includes H. Boshuizen, professor¹, P. F. M. Krabbe, associate professor², C. van der Meijden, advisor pharmacoeconomics³, D. Hamerlijnck, advisor patient perspective⁴

¹Biometris, Wageningen University and Research, Wageningen, The Netherlands

²Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

³Dutch Healthcare Institute, Diemen, The Netherlands, at the time of being involved with this project.

⁴Lung Foundation Netherlands, Amersfoort, The Netherlands

Compliance with Ethical Standards

Competing interests PV does not report competing interests. ICR does not report competing interests. GVV does not report competing interests. MA does not report competing interests. TF does not report competing interests. HB does not report competing interests. PK does not report competing interests. CM does not report competing interests. DH does not report competing interests.

Financial statement This study was financially supported by the Netherlands Organization for Health Research and Development (ZonMW), under number 80-82500-98-12211.

Author contributions GVV, MA, TF, HB, and PK conceived the design of the study. PV designed and performed the Delphi panel, communicated with panel members, and summarized and analysed the data. PV, ICR, GVV, MA, and TF interpreted the data and discussed the design of the various Delphi rounds. The workshop was performed by PV, ICR, and GVV. PV and TF drafted the article, which was revised critically by ICR, GVV, and MA. All authors have given final approval of the version to be published.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

HE Experts

The authors would like to acknowledge the following HE experts (in alphabetical order) for their participation in the Delphi Panel, thereby enabling the preparation of AdViSHE:

- Dr. F. Antoñanzas, Dep. de Economía y Empresa, University of La Rioja, Logroño, Spain
- Dr. C. Asseburg, PhD, MSc, ESOR Oy, Kuopio, Finland
- Dr. A. Beresniak, MD, MPH, PhD, Data Mining International SA, Geneva, Switzerland
- Mr. S. Borg, MSc, The Swedish Institute for Health Economics, Lund, Sweden, and Lund University, Lund, Sweden
- Mr. Cao Q., MSc, Department of Epidemiology, UMC Groningen, Groningen, The Netherlands
- Mr. G. L. Colombo, S.A.V.E. Studi Analisi Valutazioni Economiche, Milano, Italy
- Dr. M. Costa-Scharplatz, Health Economics, Novartis Sweden, Stockholm, Sweden
- Dr. G. W. J. Frederix, Department of Pharmaceutical Sciences, Div. of Pharmcoepi. and Clinical Pharmacol., University of Utrecht, Utrecht, The Netherlands
- Dr. M. Hoogendoorn, iMTA/iBMG, Erasmus University Rotterdam, Rotterdam, The Netherlands
- Dr. H. Koffijberg, Department of Epidemiology, Division of Julius Center, UMC Utrecht, Utrecht, The Netherlands
- Dr. M. Kretzschmar, Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands, and Department of Epidemiology, Division of Julius Center, UMC Utrecht, Utrecht, The Netherlands
- Mr. M. Lamotte, MD, Modelling Center of Excellence, IMS Health, Vilvoorde, Belgium
- Ms A. Leunis, MSc, iMTA/iBMG, Erasmus University Rotterdam, Rotterdam, The Netherlands
- Dr. S. K. Lhachimi, Dipl.-Pol., MPP, MSc, PGD, GStat, Institute for Public Health and Nursing, Health Sciences Bremen, University Bremen, Germany, and Research Group Evidence-Based Public Health, Leibniz Institute for Prevention Research and Epidemiology-BIPS, Bremen, Germany
- Dr. J. A. Mauskopf, RTI Health Solutions, Research Triangle Park, NC, USA
- Dr. J. F. Mould-Quevedo, PhD, MSc, MBA, Bayer Healthcare, Berlin, Germany
- Dr. B. Nagy, Syreon Research Institute, Economic Modelling Division, Budapest, Hungary

- Dr. E. A. Over, Center for Nutrition, Prevention and Health Services, National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands
- Prof A. J. Palmer, Menzies Research Institute Tasmania, University of Tasmania, Australia
- Ms. B. Parkinson, Centre for Health Economics, Research and Evaluation, University of Technology, Sydney, Sydney, Australia
- Mr. L. Pradelli, MD, AdRes HEOR, Turin, Italy
- Dr. B. L. T. Ramaekers, Dept of Clinical Epidemiology and Medical Technology Assessment, Maastricht University Medical Centre, Maastricht, The Netherlands
- Mr. D. Rosselli, MD, MEd, MHP, Associate Professor, Health Economics, Javeriana Papal University, Bogotá, Colombia
- Mr. V. Sonathi, Economic Modelling Team, GBS-GMO, Novartis
- Dr. L. M. A. Steuten, Dept of Health Technology and Services Research, University of Twente and CEO Panaxea bv, The Netherlands
- Dr. B. Stollenwerk, Helmholtz Zentrum München (GmbH), Germany
- Mr. J. Tosh, MSc, HEDS, School of Health and Related Research (ScHARR), University of Sheffield, UK
- Dr. M. Willis, The Swedish Institute for Health Economics (IHE), Lund, Sweden
- Nineteen respondents who preferred to remain anonymous.

References

1. Caro JJ, Briggs AH, Siebert U, Kuntz KM. ISPOR-SMDM Modeling Good Research Practices Task force. Modeling good research practices—overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. *Value Health*. 2012;15(6):796–803.
2. Balci O. Golden rules of verification, validation, testing, and certification of modeling and simulation applications. *SCS M&S Mag* 2010;4.
3. Caro JJ, Eddy DM, Kan H, Kaltz C, Patel B, Eldessouki R, Briggs AH; ISPOR-AMCP-NPC Modeling CER Task Forces. Questionnaire to assess relevance and credibility of modeling studies for informing health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force Report. *Value Health*. 2014;17(2):174–82.
4. Sargent RG. Validation and verification of simulation models. In: Ingalls R, Rossetti M, Smith J, Peters B, editors. 2004 Winter Simulation Conference Piscataway, NJ. IEEE Press; 2004.
5. Law AM. How to build valid and credible simulation models. Peters B, Smith J, Medeiros D, Rohrer M, editors. 2001 Winter Simulation Conference Arlington, VA. IEEE Press; 2001.
6. Law AM, Kelton WD. Simulation modeling and analysis. 4th ed. New York: McGraw-Hill; 2007.

7. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. *BMJ*. 1996;313:275–83.
8. Evers S, Goossens M, de Vet H, van Tulder M, Ament A. Criteria list for assessment of methodological quality of economic evaluations: consensus on health economic criteria. *Int J Technol Assess Health Care*. 2005;21(2):240–5.
9. Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. *Pharmacoeconomics*. 2006;24(4):355–71.
10. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB, et al. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. *Value Health*. 2012;15(6):843–50.
11. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated health economic evaluation reporting standards (CHEERS) statement. *Pharmacoeconomics*. 2013;31(5):361–7. doi:10.1007/s40273-013-0032-y.
12. Vemer P, van Voorn GA, Ramos IC, Krabbe PF, Al MJ, Feenstra TL. Improving model validation in health technology assessment: comments on guidelines of the ISPOR-SMDM Modeling Good Research Practices Task Force. *Value Health*. 2013;16(6):1106–7.
13. Helmer-Hirschberg O. Analysis of the future: the Delphi method. RAND Corporation: Santa Monica, CA. 1967;3558. Available from <http://www.rand.org/pubs/papers/P3558.html>
14. Dalkey N, Helmer-Hirschberg O. An experimental application of the Delphi method to the use of experts. *Manage Sci*. 1963;9(3):458–67.
15. Linstone HA, Turoff M, editors. *The Delphi method: techniques and applications*. New Jersey Institute of Technology: Newark; 2002.
16. Whiting P, Al MJ, Westwood ME, Corro Ramos I, Ryder S, Armstrong N, Misso K, Ross J, Severens J, Kleijnen J. Viscoelastic point-of-care testing to assist with the diagnosis, management and monitoring of haemostasis: a systematic review and cost-effectiveness analysis. 2014.
17. Hoogendoorn M, Rutten-van Mölken MPMH, Hoogenveen RT, van Genugten MLL, Buist AS, Wouters EFM, et al. A dynamic population model of disease progression in COPD. *Eur Respir J*. 2005;26(2):223–33.
18. Halpern MT, Luce BR, Brown RE, Geneste B. Health and economic outcomes modelling practices: a suggested framework. *Value Health*. 1998;1(2):131–47.
19. Kleijnen JPC. Ethical issues in engineering models: an operations researcher's reflections. *Sci Eng Ethics*. 2011;17(3):539–52.
20. Zeigler BP, Praehofer H, Kim TG. *Theory of modeling and simulation*. 2nd ed. San Diego: Academic Press; 2000.
21. Fairley RE. *Dynamic testing of simulation software*. 1976. Piscataway, NJ, Gaithersburg, MD: IEEE Press. <http://www.computer.org/csdl/mags/co/1978/04/01646907.pdf>.
22. Schruben LW. Establishing the credibility of simulations. *Simulation*. 1980;34(3):101–5.
23. NICE. Guide to the single technology appraisal process. NICE: London/Manchester. 2009. ISBN 978-1-84936-104-0. Available from: <http://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/Guide-to-the-single-technology-appraisal-process.pdf>
24. van der Meijden C. Procedure beoordeling extramurale geneesmiddelen [Procedure for the judgement of extramural drugs]. CVZ: Diemen. 2012;2010117719. Available from: <http://www.zorginstituutnederland.nl/binaries/content/documents/zinl-www/documenten/rubrieken/pakket/pakketbeheer/1206-procedure-beoordeling-extramurale-geneesmiddelen/1206-procedure-beoordeling-extramurale-geneesmiddelen/Procedure+beoordeling+extramurale+geneesmiddelen.pdf>
25. Chilcott J, Tappenden P, Rawdin A, Johnson M, Kaltenthaler E, Paisley S, et al. Avoiding and identifying errors in health technology assessment models: qualitative study and methodological review. *Health Technol Assess*. 2010;14(25):iii–iv ix–xii, 1–107.
26. United Nations Statistics Division (UNSD). United Nations Geoscheme. 2013. <http://millenniumindicators.un.org/unsd/methods/m49/m49regin.htm>. 28 Accessed Aug 2014.