

Appendix: Details of methodology

In this appendix, we provide additional information on the results of building AdViSHE. See also figure 1 of the main text.

The initial list of validation techniques was based on a range of guidelines on validation from the HE economics literature.[1-3] This was amended with several guidelines from outside the HE literature.[4-10] The project group organized a number of meetings to categorize the techniques (take out doubles, classify techniques in several groups, etc.). This resulted in an initial gross list consisting of 40 different techniques, divided over the four types of validation as defined by Sargent (see also figure 2) [4], plus one category “techniques designed to ease overall validation”. Precise definitions for each technique were formulated by the project team. The complete initial list is shown in abbreviated format in table S1.

Table S1 Initial list of validation techniques presented in the pilot and first rounds.

Techniques designed ...to ease overall validation	Precise problem formulation Continuous feedback from experts Continuous feedback from users Technical documentation Instruction manual Starting with an existing model Using explicit programming languages Start with a “moderately detailed” model Model maintenance system
...for conceptual model validation	Flow diagram Conducting structured “walk-throughs” Content Validity Testing Face validity testing
...for computerized model validation	Examining the structure of the program Conducting structured “walk-throughs” External review Compare to hand calculations Identifying unnecessary detail Internal consistency testing Double programming Operational Graphics Traces
...for data validation	Debug in modules and subprograms Data description Descriptive statistics Face validity testing Distributional testing Heterogeneity tests Investigation of outliers
...for operational validation	Graphical plots Dynamic animations Naïve benchmarking Cross validation testing Event validity testing Comparison of intermediate outcomes and data Internal consistency testing Sensitivity analyses Face validity testing Historical validation testing Predictive validation testing

In the pilot round, participants were presented with full descriptions and asked to qualitatively assess their familiarity with the techniques and their importance and feasibility. In the first round, familiarity with each technique was assessed qualitatively. The perceived importance of each technique was assessed by allocating a number of points to all techniques within a group. The total number of points in each group was 5 times the number of techniques within the group. In for example the group of techniques designed for validation of the conceptual (four techniques), a participant could give each technique 5 points, or one technique 20 points and none to the other techniques. Feasibility was assessed by asking whether, if a participant considers a typical case of developing an HE decision model, each technique was “unfeasible”, “feasible only under conditions” or “always feasible”. If a technique was deemed “feasible only under conditions” extra comments were requested. Participants were also asked for suggestions of other techniques. The original surveys used are available from the authors upon request.

In the first round, 9 items scored on average over our responders four or lower on importance: starting with an existing model, conducting structured “walk-throughs” (for conceptual model validation, not for computerized model validation), identifying unnecessary detail, double programming, heterogeneity tests, investigation of outliers, dynamic animations, naive benchmarking and historical validation testing. These were excluded. 9 New items were also suggested which were all included in the new list: a full list of assumptions, comparing the model to other conceptual models, testing of all model functions, such as drop down menus, restore defaults, and macros for opening a model, commenting your code, statistical tests for model fit, validation against an alternative dataset, calibration, convergence criteria and numerical stability testing. After reformulating and combining very similar items, 36 techniques remained.

During the second and third Delphi rounds, the participants discussed the items in the list and provided suggestions for improvement. The number of questions was reduced, and the group of techniques “designed to ease model validation” was dropped. This was because the techniques mentioned in this section were not validation techniques, but rather “modelling good practices”. It was discussed with the Delphi panel, and agreed in the project group that this was not part of the scope of AdViSHE. It is already extensively covered in other guidelines. Instead, it is assumed that the modeler takes their own responsibility with adherence to these prevailing modelling and reporting guidelines and this was stated in the tool (see introductory text for AdViSHE in figure 4).

The first draft was made after the third round. Representatives from the Zorginstituut Nederland (Dutch Healthcare Institute) commented on the draft and requested the inclusion of the treatment of outliers. After discussion with Zorginstituut Nederland, it was agreed that this was included in the introductory text of the section on data validation.

This first draft was also presented at a workshop at ISPOR Montreal. During this workshop participants could discuss the draft amongst themselves, and were asked to fill out a questionnaire reviewing the draft. Based on workshop results a final question was added, asking whether modellers have performed any validation techniques not covered in AdViSHE (Part E). The discussion was with a total of 50 workshop participants and 19 filled-in questionnaires were returned, which helped to inform the fifth Delphi round.

In the fifth round the final version was presented for comments to the full panel of Delphi participants. No further changes were proposed. The tool was then sent to an English speaking editor, who proposed several linguistic changes. The whole project group approved the final version of AdViSHE during a meeting in October 2014.

LITERATURE

- [1] 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):355-371.
- [2] 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 Sep;15(6):843-850.
- [3] Halpern MT, Luce BR, Brown RE, Geneste B. Health and economic outcomes modelling practices: a suggested framework. *Value Health* 1998;1(2):131-47.
- [4] Sargent RG. Validation and verification of simulation models. 2004 Winter Simulation Conference Piscataway, NJ. Ingalls R, Rossetti M, Smith J, Peters B, editors. IEEE Press; 2004.
- [5] Law AM. How to build valid and credible simulation models. 2001 Winter Simulation Conference Arlington, VA. Peters B, Smith J, Medeiros D, Rohrer M, editors. IEEE Press; 2001.
- [6] Law AM, Kelton WD. *Simulation Modeling and Analysis*. 4th ed. New York: McGraw-Hill; 2007.
- [7] Kleijnen JPC. Ethical Issues in Engineering Models: An Operations Researcher's Reflections *Sci Eng Ethics*, 2011, 17, 3, 539-552
- [8] Zeigler BP, Praehofer H, Kim TG. *Theory of modeling and simulation 2000*, Academic Press, San Diego, CA, 2nd
- [9] Fairley, R.E. *Dynamic testing of simulation software 1976*, IEEE Press, Piscataway, NJ, Gaithersburg, MD. Available from <http://www.computer.org/csdl/mags/co/1978/04/01646907.pdf>
- [10] Schruben, L.W. Establishing the credibility of simulations *Simulation*, 1980, 34, 3, 101-105