

Model Validation: Has it's Time Come?

Jonathan Karnon¹

Published online: 25 May 2016
© Springer International Publishing Switzerland 2016

There is broad agreement that effort should be made to validate cost-effectiveness models. The International Society for Pharmacoeconomics and Outcomes Research–Society for Medical Decision Making (ISPOR–SMDM) Modeling Good Research Practices Task Force considered model validation to be ‘vital’, while recognising that it is “not possible to specify criteria that a model must meet to be declared “valid”” [1] (page 736).

Guidelines for submissions to reimbursement agencies make common references to model validation but provide limited guidance on expectations regarding the application of alternative validation approaches. National Institute for Health and Care Excellence (NICE) guidelines request that sponsors provide the rationale for the chosen validation methods but provide no further guidance other than to note that sponsors should consider if and why presented results differ from the published literature [2]. Canadian guidelines describe alternative validation approaches, that the validation process should be documented, and ideally undertaken by ‘someone impartial’ [3]. In Australia, current Pharmaceutical Benefits Advisory Committee (PBAC) guidelines note that sponsors should “Consider developing and presenting any approaches to validate the results of a modelled economic evaluation” [4]. More specifically, the PBAC guidelines request that sponsors compare “model traces that correspond with observed or empirical data (e.g. overall survival or partitioned survival) as a means of validating the model”.

Personal experience reviewing PBAC submissions is that model validation is rarely reported. Sponsors present model traces describing the proportions of the intervention and the comparator cohorts in alternative health states over time, but few compare the traces to observed data. In this issue of *Pharmacoeconomics*, De Boer et al. have reviewed the reporting of efforts to validate cost-effectiveness models in seasonal influenza and early breast cancer [5], while an earlier paper by Afzali and colleagues reviewed approaches to evaluate the performance of decision analytic models in cardiovascular disease [6].

The two reviews report similar findings with respect to cross-model validation of model outputs, which was by far the most commonly reported form of validation. De Boer et al. report that 57 and 51 % of the 53 seasonal influenza models and 41 early breast cancer models referred to cross-model validation, respectively [5]. Afzali et al. found that 55 % of the 81 reviewed cardiovascular models reported on cross-model validation [6]. De Boer et al. do not comment further on the cross-model validation efforts, while Afzali et al. found significant variation across the cardiovascular studies. Of the 45 references to cross-model validation, 16 studies noted that no relevant published studies were identified, 13 did not report quantitative comparisons of model outputs, and only 16 (20 % of all 81) studies provided a quantitative assessment of cross-model outputs.

There was variation across the disease areas with respect to dependent or internal validation, in which model outputs are compared with outputs derived from data used to estimate one or more input parameter. Thirty-seven percent of early breast cancer models, 12 % of cardiovascular models, and no seasonal influenza models reported on this form of validation [5, 6]. This discord may be partly explained by variation in the primary data sources for

✉ Jonathan Karnon
jonathan.karnon@adelaide.edu.au

¹ School of Public Health, University of Adelaide, 178 North Terrace, Adelaide, SA, Australia

economic evaluations in these three disease areas. Some form of dependent validation should generally be possible for cost-effectiveness models that extrapolate beyond the follow-up of a clinical trial. In the case of early breast cancer, clinical trials often inform input parameters describing the probability of cancer recurrence, but trials also report on overall survival, which is a model output parameter.

Some form of validation to external output data (i.e. from a data source not used to inform input parameter values) was more common in the cardiovascular models (16 %) than in the seasonal influenza (6 %) or early breast cancer (5 %) models [5, 6]. The availability of relevant external data against which a model can be validated will vary across disease areas, but in common areas such as cardiovascular disease and early breast cancer, external data describing comparator model outputs are generally available. The Early Breast Cancer Trialists' Collaborative Group updates meta-analyses of therapies for early breast cancer on a regular basis that could inform the external validity of cost-effectiveness models (<https://www.ctsu.ox.ac.uk/research/meta-trials/ebctcg>). Reasons for the irrelevance of such data to the validation of a cost-effectiveness model for early breast cancer should at least be stated.

Low levels of reporting of approaches to validate the conceptual model, the input data, and to verify or test the implemented model were identified, although there may be underreporting of validation in these areas. De Boer et al. found that six of seven authors who responded to a request for further information confirmed they had undertaken but not reported these three forms of validation [5].

Model validation can increase confidence in the relevance and accuracy of a cost-effectiveness model, but model validation is not a cure-all. In particular, it is difficult to validate extrapolated treatment effects. However, appropriate validation approaches can establish the applicability of the model structure to capture all important differences in costs and outcomes, assess the robustness of the expected value and sampling distribution of the model's input parameters, and confirm the absence of implementation errors. Where applicable, cross-model, dependent and independent validation approaches provide additional confidence in the accuracy and relevance of the model outputs for the comparator and, in some cases, for the intervention too.

There is clearly scope for substantial improvement in the conduct and reporting of validation approaches for cost-effectiveness models. General reporting checklists such as the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) refer to the description of 'approaches to validate' a model [7], but consistent critical appraisal of model validation efforts requires more detailed

guidance around the use of alternative validation approaches.

De Boer et al. used the Assessment of the Validation Status of Health-Economic decision models (AdViSHE) tool to inform their review of the reporting of validation efforts [5, 8]. AdViSHE is a validation-assessment tool comprising 13 items covering four broad areas: the conceptual model, input data, model implementation, and model outcomes. The tool was developed via an adapted Delphi process involving 47 experts in cost-effectiveness modelling and healthcare decision making. AdViSHE is intended for completion by analysts to report on the methods and results of applied validation approaches or to justify why validation approaches that were not applied were not undertaken.

The latest draft of the updated PBAC guidelines (<http://www.pbs.gov.au/info/reviews/pbac-guidelines-review>) refers to the AdViSHE tool [8], noting that sponsors should justify the non-completion of any of the 13 items. The completion of the tool by sponsors should ease the burden on the independent review teams who have approximately 8 weeks to review full submissions to the PBAC. Review teams can replicate applied validation approaches and focus more on model components that were not subject to validation, or for which validation results were least convincing.

The AdViSHE tool was designed to balance what is feasible and what is necessary for the validation of cost-effectiveness models [8], but the resources required to undertake the composite validation approaches remain an important consideration. In particular, face validation and external review require the involvement of relevant experts, while independent validation involves the identification and analysis of external data and potential adaptations to the model and its inputs.

It seems reasonable to expect multinational companies to allocate resources to complete the full set of applicable forms of validation to support multi-million dollar funding decisions. Lack of time and resources may be a reasonable justification for the exclusion of the more time-consuming validation approaches in other cases, such as government-sponsored contract research working towards a tight decision-making deadline. In such cases, completion of the AdViSHE tool can highlight areas in which validation might be possible but was not conducted, and therefore inform decision makers of the potential value of delaying or revisiting funding decisions to allow full model validation.

How should journal editors and peer reviewers view or test claims of a lack of resources as a justification for the non-application of model validation approaches? Do we need to borrow terminology from our information

specialist colleagues and start referring to ‘full’ and ‘rapid’ cost-effectiveness models?

Model validation is an issue that has been discussed but not acted on for too long. The recently published AdViSHE tool is a credible validation-assessment tool that supports the consistent reporting and review of key applied validation approaches [8]. Will other institutions follow the PBAC in using the AdViSHE tool to guide the validation of cost-effectiveness models?

Compliance with Ethical Standards

Conflict of interest Jonathan Karnon did not receive any funding to support the drafting of this manuscript and has no conflicts of interest to declare.

References

1. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. *Med Decis Making*. 2012;32(5):733–43.
2. National Institute for Health and Care Excellence (NICE). Single technology appraisal: user guide for company evidence submission template. January 2015 Available at: <https://www.nice.org.uk/article/pmg24/chapter/5-Cost-effectiveness>. Accessed 23 May 2016
3. Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies: Canada. 3rd ed. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2006.
4. Pharmaceutical Benefits Advisory Committee. Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (version 4.5). Australian Government Department of Health; 2015.
5. de Boer PT, Frederix GW, Feenstra TL, Vemer P. Unremarked or unperformed? Systematic review on reporting of validation efforts of health economic decision models in seasonal influenza and early breast cancer. *Pharmacoeconomics*. 2016. (**Epub ahead of print**)
6. Afzali HHA, Gray J, Karnon J. Model performance evaluation (validation and calibration) in model-based studies of therapeutic interventions for cardiovascular diseases. a review and suggested reporting framework. *Appl Health Econ Health Policy*. 2013;11:85–93.
7. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)—explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. *Value Health*. 2013;16:231–50.
8. Vemer P, Ramos IC, van Voorn GAK, Al MJ, Feenstra TL. AdViSHE: a validation-assessment tool of health-economic models for decision makers and model users. *Pharmacoeconomics*. 2016;34:349–61.