Validating Health Economic Models With the Probabilistic Analysis Check dashBOARD

Pouwels, Xavier G.L.V.; Kroeze, Karel; van der Linden, Naomi; Kip, Michelle M.A.; Koffijberg, Hendrik

DOI
10.1016/j.jval.2024.04.008

Publication date
2024

Document Version
Final published version

Published in
Value in Health

Citation (APA)
Pouwels, X. G. L. V., Kroeze, K., van der Linden, N., Kip, M. M. A., & Koffijberg, H. (in press). Validating Health Economic Models With the Probabilistic Analysis Check dashBOARD. *Value in Health*, *27*(8), 1073-1084. https://doi.org/10.1016/j.jval.2024.04.008

ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Methodology

Validating Health Economic Models With the Probabilistic Analysis Check dashBOARD

Xavier G.L.V. Pouwels, PhD, Karel Kroeze, MSc, Naomi van der Linden, PhD, Michelle M.A. Kip, PhD, Hendrik Koffijberg, PhD

ABSTRACT

Objectives: Health economic (HE) models are often considered as "black boxes" because they are not publicly available and lack transparency, which prevents independent scrutiny of HE models. Additionally, validation efforts and validation status of HE models are not systematically reported. Methods to validate HE models in absence of their full underlying code are therefore urgently needed to improve health policy making.
This study aimed to develop and test a generic dashboard to systematically explore the workings of HE models and validate their model parameters and outcomes.

Methods: The Probabilistic Analysis Check dashBOARD (PACBOARD) was developed using insights from literature, health economists, and a data scientist.
Functionalities of PACBOARD are (1) exploring and validating model parameters and outcomes using standardized validation tests and interactive plots, (2) visualizing and investigating the relationship between model parameters and outcomes using metamodeling, and (3) predicting HE outcomes using the fitted metamodel.
To test PACBOARD, 2 mock HE models were developed, and errors were introduced in these models, eg, negative costs inputs, utility values exceeding 1. PACBOARD metamodeling predictions of incremental net monetary benefit were validated against the original model's outcomes.

Results: PACBOARD automatically identified all errors introduced in the erroneous HE models. Metamodel predictions were accurate compared with the original model outcomes.

Conclusions: PACBOARD is a unique dashboard aiming at improving the feasibility and transparency of validation efforts of HE models. PACBOARD allows users to explore the working of HE models using metamodeling based on HE models' parameters and outcomes.

Keywords: health economic model, metamodels/emulators, probabilistic sensitivity analysis, sensitivity analysis, validation.

VALUE HEALTH. 2024; 27(8):1073–1084

Highlights

- Health economic evaluation guidelines emphasize the importance of health economic model validation, but there is currently no easy-to-use tool available to systematically validate (parts of) health economic models.
- This article presents the development and testing of the Probabilistic Analysis Check dashBOARD, which is a unique online tool aimed at validating the plausibility of parameters and outcomes of health economic models and at exploring the working of health economic models, using interactive plots and metamodeling methods.
- The Probabilistic Analysis Check dashBOARD is a useful tool for model developers to validate health economic models at different stages of model development, through uploading probabilistic parameters and outcomes of the health economic model. Also, it enables reviewers to efficiently evaluate health economic models.

Introduction

Health economic (HE) models are routinely developed to support health policy decisions but are not often publicly available. Consequently, HE models cannot be independently validated, which may increase the risk of errors in HE models informing health policy decisions. Additionally, HE models are usually not extensively validated, or the extent of HE model validation is not always systematically reported,[1] although checklists have been developed to this end.[2,3] This lack of validation may undermine HE models' credibility and increases the risk of making wrong decisions, potentially leading to health losses and inefficient healthcare spending.

There are several explanations for this limited transparency and (reporting of) validation of HE models. First, HE model developers from academia and from pharmaceutical and consultancy companies do not have incentives to transparently report their HE models or to make them publicly available,[4] which can partly be explained by the lack of reward system for performing transparent research within health economics.[5] Second, HE models' assumptions, parameters, and undertaken validation efforts are often not fully disclosed, partly because HE models are reported in scientific journals with limited word counts (although many scientific journals accept extensive appendices) and partly because of the nondisclosure of privacy-sensitive and confidential data informing HE models.[6] Third, proper model validation and engaging with stakeholders during development and validation of HE models requires substantial time and budget, limiting validation feasibility in many practical settings.[7]

Current guidelines for HE models' validation[2,3,8] advise on a series of validation tests that can be performed to assess HE models' validity but leave model developers free regarding which validation tests to perform and how to perform and report these. Additionally, they do not facilitate independent scrutiny and

interaction with the HE models. Therefore, tools to support the systematic validation of HE models in a standardized manner, without requiring the full model code, and within limited time are urgently needed. Reducing the complexity and time-investment needed for model validation may improve the implementation and reporting of validation efforts and increase the feasibility of stakeholders' engagement during validation processes. Metamodeling methods may contribute to the development of such validation tools because they can rapidly create insights in the relationships between HE model parameters and outcomes, and almost instantaneously approximate HE model outcomes without having access to the full model code. Metamodeling consists of fitting a statistical model (eg, a linear regression) to HE model parameters and their corresponding outcomes. These parameters and outcomes can be obtained from the standard probabilistic analysis, which is mandatory in most jurisdictions.

This study aimed to develop a generic interactive dashboard that supports the systematic exploration and validation of HE models' parameters and outcomes when the underlying HE model is unavailable, using standard validation tests, interactive plots, and metamodeling.

Methods

In the following sections, we describe the development of the R shiny Probabilistic Analysis Check dashBOARD (PACBOARD) and the literature reviews undertaken to inform the functionalities of PACBOARD. We then describe how we tested PACBOARD.

Development of PACBOARD

A requirement of PACBOARD was that users would only need to upload the model parameter values and corresponding outcome values of a probabilistic analysis to be able to use it. In PACBOARD, these parameters and outcomes should be uploaded as a comma separated value file (.csv) with variable names in the first row.

The primary focus of PACBOARD and its potential functionalities were first discussed among the team of health economists who initiated this project (X.P., M.K., N.v.d.L., and H.K.). Afterward, X.P. performed literature reviews to complete the list of potential functionalities. The literature reviews focused on identifying standard validation efforts that could be performed on any probabilistic parameters and outcomes of a HE model and on identifying metamodeling methods to explore and validate simulation models. X.P. and K.K. then developed a draft version of PACBOARD. This version was reviewed by M.K., N.v.d.L., and H.K. and was discussed during the LowLands Health Economics Study Group conference (LOLA HESG) 2022, a yearly health economics conference attended by approximately 150 participants. The feedback obtained from this discussion was used to finalize version 1.0 of PACBOARD.

Literature reviews

Two literature reviews were performed on February 23, 2023 (update of the original literature search) using one round of citation searching.[9] The first literature search focused on identifying studies describing simulation model validation efforts. The inclusion criteria were that the articles mentioned and described at least 1 method to validate simulation models using the original model or a metamodel. The seed article used for this literature search was the article describing the development of the Technical Verification (TECH-VER) checklist.[2] This article was chosen because of its recent publication and because it contains a list of verification efforts for HE models, which is derived from a literature review on model validation and verification. X.P. also screened the references cited in the TECH-VER publication and the literature referencing to TECH-VER. All publications that seemed to describe or apply validation efforts for simulation models based on their title and abstract were included. All validation efforts mentioned and applied in the included articles were extracted. Validation efforts for value-of-information analyses were beyond the scope of this study.

The second literature review aimed to identify metamodeling methods to explore and validate the working of simulation models. The seed article for this literature review was the scoping review of Degeling et al[10] because it provides a recent overview of metamodeling applications within HE modeling research and refers to literature from other research fields on metamodels in simulation modeling.[10] XP also screened the references cited in Degeling et al[10] and the literature referencing to Degeling et al.[10] Based on titles and abstracts, articles describing the use of metamodels to explore or validate the original simulation model were included. All metamodel methods for verification and validation were extracted. Articles describing the use of metamodels for simulation, calibration, or optimization were excluded.

Both literature reviews were limited to articles referring to the seed articles and articles cited in the seed articles. We decided to limit the literature reviews to these articles because both seed articles were based on recent scoping reviews, and more extensive literature searches were unlikely to yield additional validation efforts and metamodel applications. Similar validation efforts and metamodeling methods (eg, different names but same validation/metamodel methods) were clustered. Validation efforts and metamodeling methods were then classified into validation efforts that could and could not be performed using solely the probabilistic parameters and outcomes of a HE model. Validation efforts that could be performed using the probabilistic parameters and outcomes of a HE model were considered for inclusion in PACBOARD. Validation efforts and metamodeling features initially included in PACBOARD were selected based on their feasibility. Simple and easy to implement tests were included first. More complex tests may be added at a later stage. These decisions were made by X.P. and K.K.

We excluded abstracts without full-text articles and articles and books of which we could not access the full text ($N = 1$). We did not restrict our inclusion to a specific type of journal article but restricted inclusion to English, Dutch, and French articles.

Testing PACBOARD

Two HE models were developed to test PACBOARD (Appendix 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.04.008).

HE models description

The first HE model was a cohort state transition model (STM) and the second a partitioned survival model (PSM). The model structure was similar for both HE models and was typical for advanced cancer with 3 health states: progression free (PF), progressed disease (PD), and dead. The cohort started in the PF health state and could transit to the PD and dead health states. Once individuals progressed, they could not transit back to the PF health state but could die. Dead was the absorbing health state. Each health state was assigned a utility value and costs. Both HE models compared a new treatment (the intervention) with usual care, which was "do nothing" (the comparator) for treating advanced breast cancer. In the STM, the intervention reduced the probability to progress from PF to PD, whereas it reduced the probability to progress and die in the PSM. The intervention was associated with additional treatment costs and a probability of experiencing

adverse events, leading to quality-of-life decrements and costs once at the start of the HE model. The STM contained 12 input parameters: 3 transition probabilities, utility values and costs of PF and PD health states, costs and relative effectiveness of the intervention, and the probability, utility decrement, and costs associated with adverse events. The PSM contained the same input parameters, except that the transition probabilities and relative risk were replaced by parameters of the survival curves for PF (exponential) and overall survival (Weibull) for the comparator and the intervention. The following intermediate HE model outcomes were calculated per health state per strategy: total discounted life years (LY), total discounted quality-adjusted LY (QALY), and total discounted costs, total costs, and QALY decrements associated with the occurrence of adverse events. The following outcomes were calculated for each strategy: total (un) discounted LYs, QALYs, and costs per strategy. Also, the incremental LYs, QALYs, and costs of the intervention versus the comparator were calculated (N = 29 outcome variables). Probabilistic analyses containing 10 000 iterations were performed using the developed HE models.

Testing verification functionalities of PACBOARD

The following errors were incorporated in the STM to test whether PACBOARD would identify them automatically:

- Negative utility values ("u_pfs" model parameter)
- Utility values above 1 ("u_pd" model parameter)
- Negative probabilities ("p_pfspd" model parameter)
- Probabilities above 1 ("p_pdd" model parameter)
- Negative costs ("c_pd" model parameter)
- Negative relative risks ("rr" model parameter)
- Negative total outcomes for a strategy ("t_qaly_d_comp" model outcome)
- Discounted outcomes higher than undiscounted outcomes for a strategy ("t_qaly_d_int" higher than "t_qaly_int")

To test whether PACBOARD could identify 2 crossing survival curves, an extremely low rate (0.01) for the exponential PF survival curve of the comparator ("r_exp_pfs_comp") was implemented in several iterations, which led to a crossing of the PF survival curve with the overall survival curve.

Metamodel development and prediction accuracy

Because an aim of PACBOARD was to investigate the possibility to use metamodeling to validate HE models, we decided to implement a linear regression model based on the literature review findings. A linear regression models was selected because of its easy interpretation, its low computational burden, and its ease of implementation in an online application. Previous literature further demonstrates that linear regression models are suitable for verification purposes of simulation models[11] (which is a goal of PACBOARD) and for the approximation of simple HE models.[12]

The outcome for the development and prediction of the metamodel was the incremental net monetary benefit (iNMB) of the intervention versus the comparator, which was calculated using PACBOARD. For each iteration, the iNMB of the intervention versus the comparator was calculated using a willingness-to-pay threshold of €80 000 per QALY. The 10 000 original probabilistic parameters and outcomes of the STM were divided in a training (N = 7500) and validation (N = 2500) set to validate the linear regression metamodel.

First, a linear regression metamodel including all 12 parameters of the HE model was fitted on the training set. Second, backward variable selection was used to determine the final selection of variables for the metamodel. Variables with a P value higher than .05 were discarded one at a time, beginning with the variable with the highest P value. This process was continued until only statistically significant variables remained in the metamodel. The metamodel was then used to predict the iNMBs in the validation set, and the predicted versus observed values in the validation set were plotted. The mean and standard deviation of the iNMBs and the probability that the intervention was cost effective versus the comparator at a willingness-to-pay threshold of €80 000 per QALY were calculated using the validation set and using the metamodel predictions. In addition, the mean absolute and relative errors and R^2 were calculated. To ensure reproducibility of the results, these analyses were performed outside PACBOARD (although PACBOARD provides the calibration plot and the prediction accuracy measures in the validation set).

Results

PACBOARD is available at https://bdsi.shinyapps.io/pacboard/ (source code PACBOARD available at: https://github.com/BDSI-Utwente/shiny-meta-models). PACBOARD is partly based on functions contained in the *pacheck* R package (development version of pacheck available at https://github.com/Xa4P/pacheck). The online version of PACBOARD is hosted on a R Shiny server and does not store any data after a user session has been completed. We rely on industry standard open-source software and best practices to secure the data uploaded in PACBOARD but recommend downloading PACBOARD using the Github link and running PACBOARD locally for use cases that require a higher level of security and/or certification.

Development of PACBOARD

Literature reviews

Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.04.008 provides an overview of the scientific articles included in the literature reviews and a list of all validation efforts and metamodeling methods that were identified in the included articles. It further contains the complete lists of excluded validation efforts and metamodeling methods.

LOLA HESG 2022

During LOLA HESG 2022, the following additional functionalities were suggested by the audience: adding skewness to the summary statistics of the variables, implementing survival modeling validation, and adding explanations to improve user friendliness. All these suggestions were implemented in the version of PACBOARD described in this manuscript (v1.0).

PACBOARD functionalities

After clustering the validation efforts identified in the literature, the identified validation efforts (1-3) and metamodeling methods (4 and 5) could be classified in these 5 categories:

1. Investigate (the plausibility of) values of (groups of) model parameters and outcomes using statistical criteria and interactive plots
2. Investigate the relationships between model parameters and outcomes using statistical criteria and interactive plots
3. Investigate survival models
4. Perform metamodeling to explore the relation between model parameters and outcomes
5. Perform sensitivity analyses using metamodeling

Tables 1 and 2 respectively provide an overview of the validation efforts and of the metamodeling methods that can be

Table 1. Validation efforts which can be undertaken using PACBOARD or pacheck.

Validation effort category	Examples validation effort described in the literature	Source	Functionality of PACBOARD	Function from pacheck R package
Investigate (the plausibility of) single or groups of model parameters and outcomes	For the treatment effect inputs, if the model uses outputs from Windows Bayesian inference Using Gibbs Sampling (WINBUGS), are the OR, HR, and RR values all within plausible ranges? (Should all be nonnegative and the average of these WINBUGS outputs should give the mean treatment effect).	2	Summary statistics-tab; Summary statistics box	do_quick_check; check_positive
	Check if all probabilities and number of patients in a state are greater than or equal to 0.	2	Prepare data-tab; Quick checks box	do_quick_check; check_positive
	Check if all probabilities are smaller than or equal to 1	2	Prepare data-tab; Quick checks box	do_quick_check
	Decision tree specific: Calculate the sum of the expected probabilities of the terminal nodes: Should sum up to 1	2	Data inspection-tab; Check sum of probabilities box	check_sum_probs
	Total undiscounted results greater than the discounted results for each comparator.	2	Prepare data-tab; Quick checks box	do_quick_check; do_discount_check
	Divide undiscounted total QALYs by undiscounted life years: This value should be within the outer ranges (maximum and minimum) of all the utility value inputs.	2	Not implemented in PACBOARD yet	check_mean_qol
	Could you generate all the results in the report from the model (including the uncertainty analysis results)?	2	Model outcomes-tab; Incremental cost-effectiveness plane, Cost-effectiveness plane, Net benefits plane, Cost-effectiveness acceptability curve boxes	generate_sum_stats; summary_ice; plot_ice; calculate_ceac; plot_ceac; calculate_nb
	If disentangled results are presented, do they sum up to the total results (eg, different cost types sum up to the total costs estimate)?	2	Not implemented in PACBOARD yet	check_sum_vars
	Are the upper and lower bounds used in the one-way sensitivity analysis using confidence intervals based on the statistical distribution assumed for that parameter?	2	Summary statistics-tab; Summary statistics box	generate_sum_stats; vis_1_param; fit_dist
	Check that all parameters used in the sensitivity analysis have appropriate associated distributions – upper and lower bounds should surround the deterministic value (ie, upper bound $\geq$ mean $\geq$ lower bound).	2	Summary statistics-tab; Summary statistics, Univariate distributions boxes	generate_sum_stats; vis_1_param; vis_2_params; fit_dist
	Strictly positive (eg, lognormal/ gamma*) distribution for HRs and costs/resource use.	2	Summary statistics-tab; Summary statistics, Univariate distributions boxes	generate_sum_stats; vis_1_param; vis_2_params; fit_dist
	Distribution resulting in values between 0 and 1 (eg, beta*) for utilities and proportions/ probabilities.	2	Summary statistics-tab; Summary statistics, Univariate distributions boxes	generate_sum_stats; vis_1_param; vis_2_params; fit_dist

continued on next page

Table 1. Continued

Validation effort category	Examples validation effort described in the literature	Source	Functionality of PACBOARD	Function from pacheck R package
	Normal for other variables as long as samples do not violate the requirement to remain positive when appropriate.	2	Summary statistics-tab; Summary statistics, Univariate distributions boxes	generate_sum_stats; vis_1_param; vis_2_params; fit_dist
	Check PSA output mean costs, QALYs, and ICER compared with the deterministic results. Is there a large discrepancy?	2	Summary statistics-tab; Summary statistics box	generate_sum_stats; vis_1_param; vis_2_params; fit_dist; summary_ice
	Does the PSA cloud, based on the original probabilistic analysis, demonstrate an unexpected behavior or have an unusual shape? [No].	2	Model outcomes-tab; Incremental cost-effectiveness plane, Cost-effectiveness plane, Net benefits plane boxes	generate_sum_stats; vis_1_param; vis_2_params; summary_ice; plot_ice; calculate_ceac; plot_ceac; calculate_nb
	Is the sum of all CEAC lines equal to 1 for all WTP values? (Yes).	2	Model outcomes-tab; Cost-effectiveness acceptability curve boxes	calculate_ceac; plot_ceac; calculate_nb
	Compare the mean of the parameter samples generated by the model with the point estimate for that parameter; use graphical methods to examine distributions, functions (the sample means and the point estimates will overlap, the graphs will be similar to the corresponding distribution functions [normal, gamma, etc]).	2	Summary statistics-tab; Summary statistics, Univariate distributions box	generate_sum_stats; vis_1_param; fit_dist
	Check if sensitivity analyses include any parameters associated with methodological/structural uncertainty (eg, annual discount rates, time horizon) (No).	2	Summary statistics-tab; Summary statistics box	generate_sum_stats
	Did the electronic model pass the black-box tests of the previous verification stages in all PSA iterations and in all scenario analysis settings? (Additional macro can be embedded to the PSA code, which stops the PSA when an error such as negative transition probability is detected) (Yes).	2	Prepare data-tab; Quick checks box	do_quick_check; generate_sum_stats; vis_1_param; vis_2_params fit_dist
	Testing that all parameter and constant values have been assigned correctly for the base-case scenario (ie, preventing mechanical errors).	13	Summary statistics-tab; Summary statistics box	generate_sum_stats; vis_1_param; vis_2_params fit_dist
Investigate the relation between model parameters and outcomes	Does the technology (drug/device, etc) acquisition cost increase with higher prices?	2	Metamodeling, Metamodel predictions-tabs; Summary statistics-tab; Correlation matrix, Bivariate distributions boxes	vis_2_params; generate_cor; fit_lm_metamodel
	Does the drug acquisition cost increase for higher weight or body surface area?	2	Metamodeling, Metamodel predictions-tabs; Summary statistics-tab; Correlation matrix, Bivariate distributions boxes	vis_2_params; generate_cor; fit_lm_metamodel

continued on next page

Table 1. Continued

Validation effort category	Examples validation effort described in the literature	Source	Functionality of PACBOARD	Function from pacheck R package
	Does the probability of an event, derived from an OR/RR/HR and baseline probability, increase with higher OR/RR/HR?	2	Metamodeling, Metamodel predictions-tabs; Summary statistics-tab; Correlation matrix, Bivariate distributions boxes	vis_2_params; generate_cor; fit_lm_metamodel
	Check the incremental life years and QALYs gained results. Are they in line with the comparative clinical effectiveness evidence of the treatments involved? (If a treatment is more effective, it generally results in positive incremental LYs and QALYs in comparison with the less-effective treatments).	2	Metamodeling, Metamodel predictions-tabs	generate_cor; vis_2_params; fit_lm_metamodel
	Check the incremental cost results. Are they in line with the treatment costs? (If a treatment is more expensive, and if it does not have much effect on other costs, it generally results in positive incremental costs).	2	Metamodeling, Metamodel predictions-tabs; Summary statistics-tab; Correlation matrix, Bivariate distributions boxes	generate_cor; vis_2_params; fit_lm_metamodel
	Total life years greater than the total QALYs for each comparator.	2	Summary statistics-tab; Bivariate distributions box	generate_cor; vis_2_params
	Subgroup analysis results: How do the outcomes change if the characteristics of the baseline change? (Better total health outcomes for better baseline health conditions, and worse total health outcomes for worse baseline health conditions for each comparator. These better health outcomes may be achieved through [1] better quality of life [eg, individuals with diabetes may have lower quality of life compared with healthy individuals without diabetes] and [2] lower risks of experiencing negative health outcomes [eg, individuals with diabetes are on average at higher risk of experiencing a cardiovascular event compared with healthy individuals]).	2	Prepare data-tab; Categorize variables box, Scenario field; Metamodeling, Metamodel predictions-tabs	fit_lm_metamodel
	Check the correlation between 2 PSA results (ie, costs/QALYs under the SoC and costs/QALYs under the comparator) (should be very low [very high] if different (same) random streams are used for different arms).	2	Summary statistics-tab; Correlation matrix, Bivariate distributions box	generate_cor
Investigate survival models	In a partitioned survival model, does the progression-free survival curve or the time on treatment curve cross the overall survival curve?	2	Survival analysis-tab	check_surv_mod

continued on next page

Table 1. Continued

Validation effort category	Examples validation effort described in the literature	Source	Functionality of PACBOARD	Function from pacheck R package
Perform sensitivity analyses (approximations using metamodeling)	Decrease all state utilities simultaneously (but keep event-based utility decrements constant): lower utilities will be accumulated each time.	2	Metamodeling, Metamodel predictions-tabs	fit_lm_metamodel; predict_lm_metamodel
	Put adverse event/ discontinuation rates to 0 and then to an extremely high level (less costs and higher QALYS/ LYs when adverse event rates are 0, higher costs and lower QALYS/LYs when adverse event rates are extreme).	2	Metamodeling, Metamodel predictions-tabs	fit_lm_metamodel; predict_lm_metamodel
	Double the difference in efficacy and safety between the new intervention and comparator and report the incremental results (approximately twice the incremental effect results of the base case. If this is not the case, report and explain the underlying reason/mechanism).	2	Metamodeling, Metamodel predictions-tabs	fit_lm_metamodel; predict_lm_metamodel
	Half the difference in efficacy and safety (approximately half of the incremental effect results of the base case. If this is not the case, report and explain the underlying reason/mechanism).	2	Metamodeling, Metamodel predictions-tabs	fit_lm_metamodel; predict_lm_metamodel
	Are the resulting ICER, incremental costs/QALYs with upper and lower bound of a parameter plausible and in line with a priori expectations?	2	Model outcomes-tab; Incremental cost-effectiveness plane, Cost-effectiveness plane, Net benefits plane boxes	generate_sum_stats; vis_1_param; vis_2_params; summary_ice; plot_ice; calculate_ceac; plot_ceac; calculate_nb; fit_lm_metamodel; predict_lm_metamodel
	Ad hoc testing: unstructured testing of the model to uncover any potential logical errors.	13	Model outcomes-tab; Incremental cost-effectiveness plane, Cost-effectiveness plane, Net benefits plane boxes	fit_lm_metamodel; predict_lm_metamodel
	Fuzzy testing: inspecting probabilistic sensitivity analysis output outliers would be a form a fuzzy testing. However, the spirit of fuzzy testing would introduce other distributions for the selected distributions to pressure test a model.	13	Metamodeling, Metamodel predictions-tabs	vis_1_params; vis_2_params; ice_plot
	Extreme condition test: the model structure and outputs should be plausible for any extreme and unlikely combination of levels of factors in the system.	14	Metamodeling, Metamodel predictions-tabs	fit_lm_metamodel; predict_lm_metamodel

*These are the "standard" distributions that are used for these.
CEAC indicates cost-effectiveness acceptability curve; HR, hazard ratio; ICER, incremental cost-effectiveness ratio; LY, life year; OR, odds ratio; PACBOARD, Probabilistic Analysis Check dashBOARD; PSA, probabilistic sensitivity analysis; QALY, quality-adjusted life year; RR, relative risk; SoC, standard of care; WTP, willingness to pay.

performed using PACBOARD. The summary statistics of the data set uploaded in PACBOARD, the results of the validation efforts performed through PACBOARD, and different plots implemented in PACBOARD can be downloaded from the "Download report"-tab of PACBOARD.

Testing PACBOARD

Testing the verification functionalities of PACBOARD

PACBOARD automatically identified all errors introduced in the STM, which is indicated in the "Prepare data"-tab under the

Table 2. Metamodeling methods that can be undertaken using PACBOARD or pacheck.

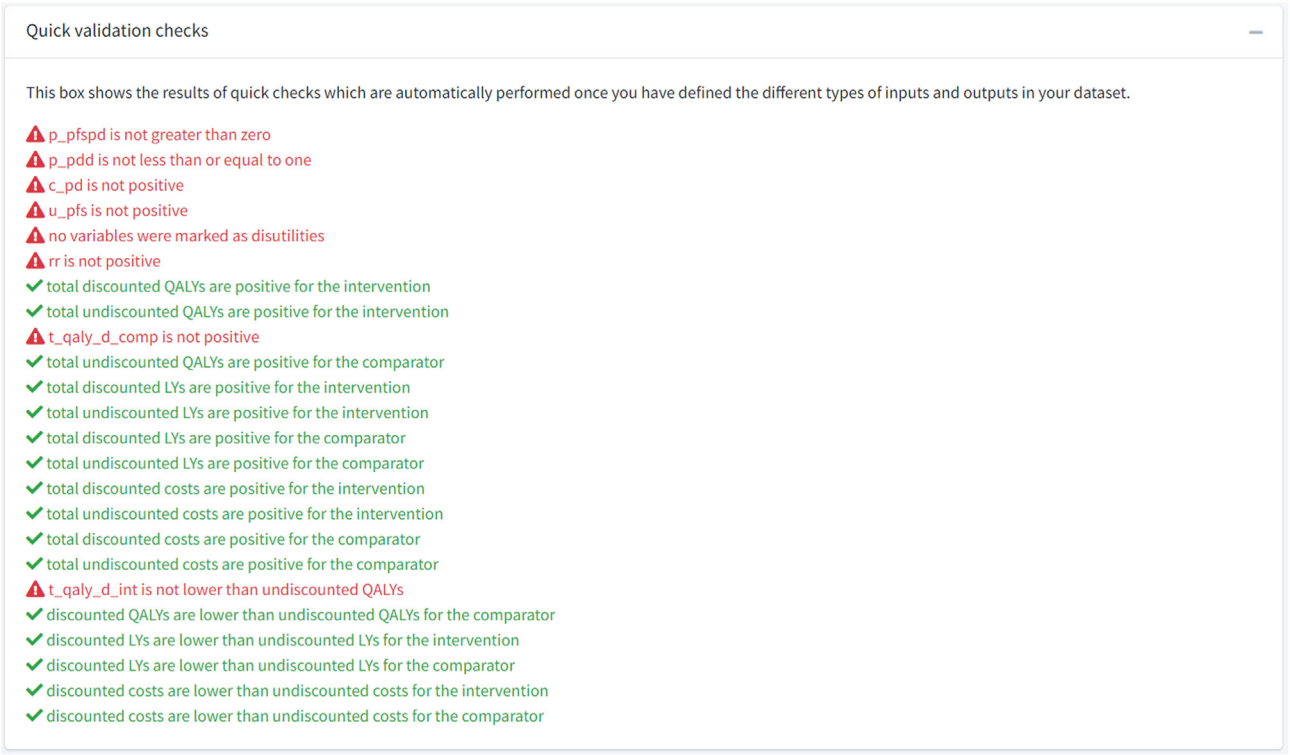
Metamodel methods for validation or exploration category	Function from pacheck R package	Examples validation or exploration effort described in the literature	Source	Included in current version of PACBOARD
Fitting linear regression metamodel	fit_lm_metamodel	Use of regression analysis and design of experiment.	15	Yes
Determining the influence of parameters on the result	fit_lm_metamodel	Normalizing input parameters using mean and standard deviation and then linear regression metamodeling on all separate outcomes (Effects, Costs, and net health benefits). Parameters of the regression on effects and costs are then used to compute the effects of the model parameters on the iNMB.	16	No, normalizing has to be performed outside PACBOARD. Normalizing can be performed when using the *pacheck* package
Perform scenario analysis	fit_lm_metamodel; predict_lm_metamodel	Threshold analysis using the linear regression metamodel.	16	Yes
Determining decision sensitivity	estimate_decision_sensitivity	Decision sensitivity to the input parameters assessed by logistic regression analysis. (…) The change in log odds attributable to a change in each input variable provides a measure of influence of each (independent) variable on the dependent variable, here, a decision favoring (a certain intervention). From total variation of all of the input variables over reasonable ranges, we can calculate a total change of the log odds attributable to the independent variables. Comparison of the relative change of the log odds due to variation of each input variable with the total log odds from all variables gives us a direct measure of the percentage contribution of each input variable to the distribution of preferred decisions.	17	No, beta version included in pacheck package (estimate_decision_sensitivity function)
Examining the direction of outcomes based on model parameters	fit_lm_metamodel	(After fitting a metamodel, investigate whether) the output increase or decrease for an increase in a specific input(.) This goal requires at least ordinal measurement scales for input and output. (…) The signs of the individual parameters of the metamodel should support prior knowledge about the problem entity." For instance, the costs of a healthcare intervention should have a positive relation with the total incremental costs of this intervention versus usual care.	11	Yes
Identifying important model parameters influencing the outcome	fit_lm_metamodel	Give a "short list" of the most important factors. For this goal, a rather crude metamodel may suffice (of course, there is always the danger that such a crude metamodel is misleading).	11	Yes
Validation of the metamodel itself: coefficient of determination of the metamodel	fit_lm_metamodel	Fit the metamodel, determine fit using : "The classic measure is the coefficient of determination (R^2 or adjusted R^2)."	11	Yes

iNMB indicates incremental net monetary benefit; PACBOARD, Probabilistic Analysis Check dashBOARD.

"Quick Check"-box of PACBOARD. For example, PACBOARD indicated the negative transition probabilities included in "p_pfspd" as "p_pfspd is not greater than 0." PACBOARD's warning messages are shown in Figure 1.

Concerning the crossing survival curves, PACBOARD identified that the PF curve crossed the overall survival curve in the comparator arm for the iterations containing a low rate for the exponential PF survival curve (Fig. 2).

Figure 1. Print screen of PACBOARD showing the identification of errors ("Quick validation checks"-box).

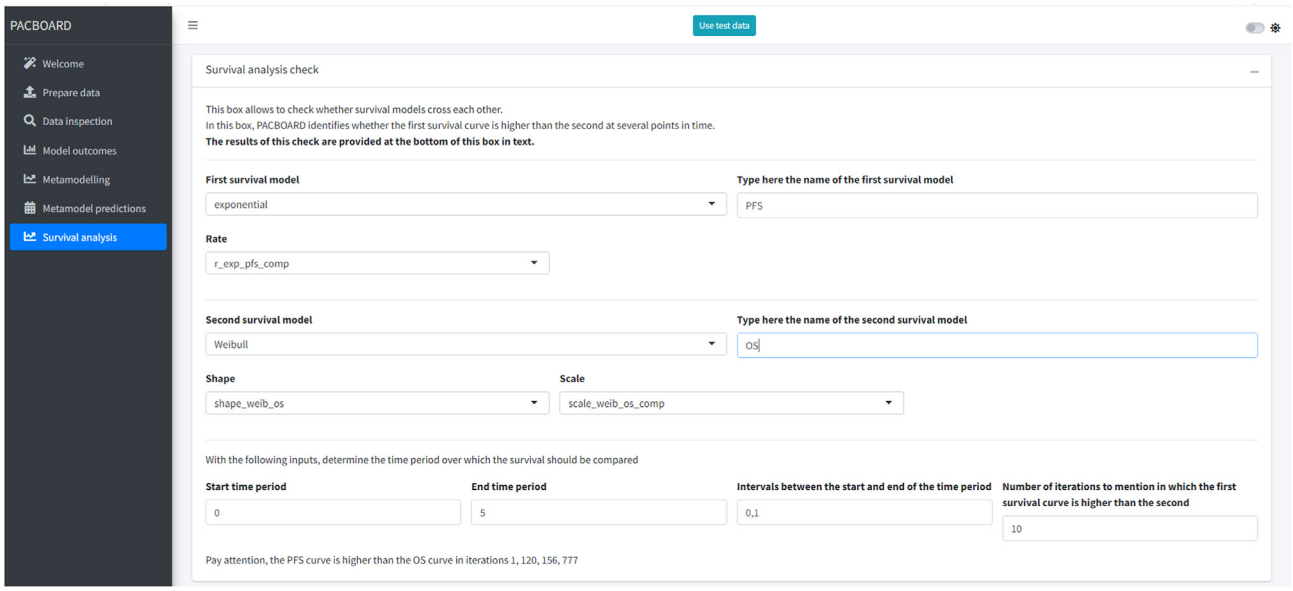


PACBOARD indicates The Probabilistic Analysis Check dashboard.

Metamodel development and prediction accuracy

Two parameters "c_ae" (costs of treating an adverse event) and "c_pd" (costs of progressed-disease health state) were removed from the definitive metamodel because they both had *P* values above .05. The mean observed iNMB in the validation set was −€10 225 (standard deviation: €9792), and the mean predicted iNMB by the metamodel was −€10 159 (standard deviation: €8841). The probabilities of the intervention being cost

Figure 2. Print screen of PACBOARD showing the identification of crossing survival curves.



PACBOARD indicates The Probabilistic Analysis Check dashboard.

Table 3. Validation criteria of the iNMB metamodel predictions versus observed iNMB in the validation set.

Criterion	Minimum	First quartile	Median	Mean	Third quartile	Maximum
Absolute error	€ 2.61	€ 656.01	€ 1561.28	€ 2469.87	€ 3250.98	€ 27 776.86
Relative error	0%	6%	15%	60%	34%	23 709%
R^2	-	-	-	0.858	-	-

iNMB indicates incremental net monetary benefit.
NOTE: The performed analyses can be found in the development version of the R package pacheck: https://github.com/Xa4P/pacheck. PACBOARD can be accessed at https://bdsi.shinyapps.io/pacboard/. The source code of PACBOARD (GPL-3 license) can be accessed at https://github.com/BDSI-Utwente/shiny-meta-models.

effective versus the comparator at a willingness-to-pay threshold of €80 000 per QALY were 13.2% in the original validation set and 12.9% when estimated using the metamodel.

Table 3 provides an overview of the absolute and relative error statistics and of the R^2 of the metamodel predictions in the validation set. The R^2 of 0.86 indicates a relatively satisfactory prediction accuracy, whereas the mean absolute and relative error of, respectively, €2470 and 60% indicated that the metamodel prediction could be improved, especially in the tails of the iNMB distribution. iNMBs above €0 were less accurately estimated by the metamodel. This can be observed in Figure 3 in which the distance between the dots and the 45 degrees red line became greater than for values below €0.

Discussion

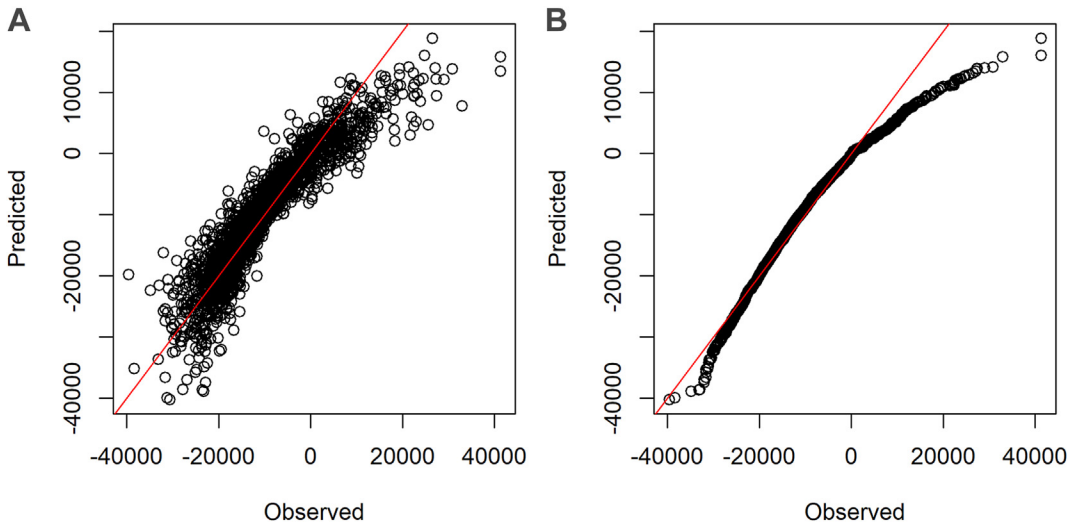
The current manuscript describes the development and testing of PACBOARD v1.0. PACBOARD is a unique tool to validate and explore HE models, which requires users to solely upload probabilistic parameters and outcomes of HE models. PACBOARD aims at improving the feasibility and transparency of HE model validation efforts. It contains a suite of validation functions and allows users to explore the relationships between HE models' parameters and outcomes using interactive visualization and linear regression metamodeling.

In this illustration, the linear regression metamodel developed through PACBOARD provided satisfactory predictions for the purpose of validating and verifying the developed HE model. The high mean absolute and relative error are mostly driven by high maximum absolute and relative error. Prediction accuracy decreased for the iNMB above €0, which is expected because results in this range may be caused by more unlikely and extreme combinations of model parameter values. Readers should be aware that these results are case-specific and that the accuracy of the fitted metamodel will differ across use cases. Before using the metamodel to perform sensitivity analyses, one should validate the fitted metamodel using the functionalities included in PACBOARD. The advantage of performing sensitivity analyses using metamodels is that the results are obtained almost instantaneously compared with having to rerun the original HE model, which can be computationally intensive.

The primary purpose of PACBOARD being to explore and validate HE models, linear regression metamodels are useful since they are easily interpretable. Also, linear regressions have been deemed suitable metamodels for validation and verification purposes[11] and for approximating simple HE models.[12] Still, they have limited flexibility concerning the modeled relationships between (meta)model parameters and outcomes. We therefore do not recommend using PACBOARD to approximate HE model outcomes yet if users are aware that the relationships between HE model parameters and outcomes cannot be accurately estimated using a linear regression model.

Metamodeling methods included in PACBOARD focus on exploring the workings of HE models and on performing sensitivity analyses through metamodeling. Metamodels can also be used for calibration, simulation (value-of-information analyses), and optimization. These applications are beyond the scope of

Figure 3. Comparison of incremental net monetary benefits prediction and observation in the validation set. (A) Predicted versus observed value in the validation set and (B) quantile-quantile plot of these observed and predicted values.



model validation and are therefore not included in PACBOARD. Examples of these metamodel methods within the HE literature have been reviewed recently.[10]

PACBOARD distinguishes itself from other online tools by its standardized validation functions and by its flexible linear metamodeling functionalities. Other tools mostly focus on performing value-of-information analyses (Sheffield Accelerated Value of Information [SAVI], the web interface to the BCEA [Bayesian Cost-Effectiveness Analysis] R package [BCEAweb], Rapid Assessment of Need for Evidence [RANE], and Value of Information for Cardiovascular Trials and Other Comparative Research [VICTOR]), and drawing the "standard" HE plots (BCEAweb).[18] One recently developed tool, Model Examiner (ModEx),[16,19,20] also allows to explore the relationship of HE model parameters and outcomes using metamodeling but is less flexible compared with PACBOARD concerning which parameters and outcomes to include within the metamodel.

PACBOARD responds to the need for efficiently standardizing the validation of HE models to improve the feasibility of validation during HE models development.[21] We demonstrated the functionalities of PACBOARD using probabilistic parameters and outcomes of simple HE models, but PACBOARD could be used on parameters and outcomes obtained from other designs of experiments (DoE), such as Latin Hypercube.[22] Using other DoE may prevent users from investigating the distributions of the parameters used during the probabilistic analysis because these alternative DoE do not necessarily reflect the probability density of the parameters.

To fully profit from the functionalities of PACBOARD, we advise users to also upload intermediate outcomes from their HE models. These intermediate outcomes, such as the number of cardiovascular events (when evaluating intervention focusing on cardiovascular events prevention for instance), can provide valuable information for validating HE models with clinical experts because they are easier to interpret and validate compared with total or incremental costs and QALYs.

A strength of PACBOARD is its reliance on previous HE literature and its generic character. PACBOARD further demonstrates that parts of HE model validation may be standardized. Although PACBOARD supports the technical verification of HE models by assessing the plausibility of model parameters and outcomes and supports assessing the face validity of HE models' outcomes, PACBOARD does not support other important aspects of HE model validation, such as face validity of the conceptual model and the external validity of model outcomes.[3,8] In addition, the plots included in PACBOARD, such as the (incremental) cost-effectiveness plane and cost-effectiveness acceptability curves, are currently limited to pairwise comparisons. Another limitation is that PACBOARD requires model developers to share the parameters and (intermediate) outcomes of their HE model to allow validation by third parties. Although this is not standard practice yet, it is likely more acceptable compared with sharing full HE models. Besides, PACBOARD only contains the possibility to fit a linear regression metamodel and does not yet allow to include interaction terms. Finally, we did not perform systematic literature reviews to inform the functionalities of PACBOARD, and we prioritized the implementation of validation efforts which did not require heavy computations because of time constraints.

The current version of PACBOARD mainly focuses on HE models' validation and its primary audience is model developers and reviewers. We envision PACBOARD as a tool to support the iterative validation of HE models during their development[23] and as a companion of model reviewers (eg, from Health Technology Assessment agencies) to accelerate parts of HE models review processes because thoroughly reviewing a HE model is time consuming even when having access to the source code of the model. Besides, the included interactive elements of PACBOARD may provide other stakeholders, such as clinicians and patients, the opportunity to interact with HE models without requiring extensive technical skills. Direct and instantaneous interaction using the metamodel may increase their understanding of HE models, their participation in HE model development and validation, and empower them in health policy making.

Finally, PACBOARD has been tested using relatively simple HE models, and its user friendliness and usefulness for validating more complex HE models has not been investigated. Further research is warranted on the following topics: testing PACBOARD using more complex HE models, assessing the value of adding multiple functional forms of metamodels within PACBOARD and the possibility to fit multiple metamodels, investigating how PACBOARD may contribute to health policy making, and the communication of HE models to a broader audience using PACBOARD.

Conclusion

This article describes the development and testing of PACBOARD. PACBOARD is an interactive dashboard containing a suite of standard validation functions for HE models, allowing users to explore the relationships between parameters and outcomes of HE models using interactive visualization and metamodeling. PACBOARD was able to identify implausible model parameters and outcomes of a HE model. Metamodel predictions in PACBOARD were deemed suitable for validation purposes. PACBOARD is a unique interactive dashboard aiming at improving the feasibility and transparency of validation efforts of HE models.

Author Disclosures

Author disclosure forms can be accessed below in the Supplemental Material section.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2024.04.008.

Article and Author Information

Accepted for Publication: April 9, 2024

Published Online: June 4, 2024

doi: https://doi.org/10.1016/j.jval.2024.04.008

Author Affiliations: Section of Health Technology and Services Research, Technical Medical Centre, Faculty of Behavioural, Management, and Social Sciences, University of Twente, Enschede, Overijssel, The Netherlands (Pouwels, van der Linden, Kip, Koffijberg); Behavioural Data Science incubator, Faculty of Behavioural, Management, and Social Sciences, University of Twente, Enschede, Overijssel, The Netherlands (Kroeze); Institute for Health Systems Science, Faculty of Technology, Policy and Management, Delft University of Technology, Delft, South Holland, The Netherlands (van der Linden).

Correspondence: Xavier G.L.V. Pouwels, PhD, Section of Health Technology and Services Research, Technical Medical Centre, Faculty of Behavioural, Management, and Social Sciences, University of Twente, Xavier Pouwels, P.O. box 217, Enschede 7500 AE, Overijssel, The Netherlands. Email: x.g.l.v.pouwels@utwente.nl

Author Contributions: *Concept and design:* Pouwels, Linden, Kip, Koffijberg
Acquisition of data: Pouwels
Analysis and interpretation of data: Pouwels, Kroeze, Linden, Kip
Drafting of the manuscript: Pouwels, Linden
Critical revision of the paper for important intellectual content: Pouwels, Kroeze, Linden, Kip, Koffijberg
Statistical Analysis: Pouwels, Kroeze, Linden, Kip
Administrative, technical, or logistic support: Kroeze
Supervision: Koffijberg

Funding/Support: The authors received no financial support for this research.

Acknowledgment: The authors acknowledge Dr Frederick Thielen for his thoughtful feedback on this research during the 2022 Lowlands Health Economic Study Group conference. The source code of PACBOARD at the time of publication is available on ZENODO at https://doi.org/10.5281/zenodo.11279943.

REFERENCES

1. 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;34(9):833–845.
2. Büyükkaramikli NC, Rutten-van Mölken M, Severens JL, Al M. TECH-VER. A verification checklist to reduce errors in models and improve their credibility. *Pharmacoeconomics*. 2019;37(11):1391–1408.
3. Vemer P, Corro Ramos I, 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(4):349–361.
4. Tappenden P, Caro JJ. Improving transparency in decision models: current issues and potential solutions [Editorial]. *Pharmacoeconomics*. 2019;37(11):1303–1304.
5. Pouwels XGLV, Sampson CJ, Arnold RJG. Opportunities and barriers to the development and use of open source health economic models: a survey. *Value Health J Int Soc Pharmacoecon Outcomes Res*. 2022;25(4):473–479.
6. Bullement A, Taylor M, McMordie ST, Waters E, Hatswell AJ. NICE, in confidence: an assessment of redaction to obscure confidential information in single technology appraisals by the National Institute for Health and Care Excellence. *Pharmacoeconomics*. 2019;37(11):1383–1390, 2019.
7. Xie RZ, Malik ED, Linthicum MT, Bright JL. Putting stakeholder engagement at the Center of Health Economic Modeling for Health Technology Assessment in The United States. *Pharmacoeconomics*. 2021;39(6):631–638.
8. Eddy DM, Hollingworth W, Caro JJ, et al. Model transparency and validation: a report of the ISPOR-SMDM modeling good research practices task force-7. *Value Health*. 2012;15(6):843–850.
9. Hinde S, Spackman E. Bidirectional citation searching to completion: an exploration of literature searching methods. *Pharmacoeconomics*. 2015;33(1):5–11.
10. Degeling K,MJ, H Koffijberg IJ, Koffijberg H. A scoping review of metamodeling applications and opportunities for advanced health economic analyses. *Expert Rev Pharmacoecon Outcomes Res*. 2019;19(2):181–187.
11. Kleijnen JPC, Sargent RG. A methodology for fitting and validating metamodels in simulation. *Eur J Oper Res*. 2000;120(1):14.
12. de Carvalho TM, van Rosmalen J, Wolff HB, Koffijberg H, Coupé VMH. Choosing a metamodel of a simulation model for uncertainty quantification. *Med Decis Mak Int J Soc Med Decis Mak*. 2022;42(1):28–42.
13. Dasbach EJ, Elbasha EH. Verification of decision-analytic models for health economic evaluations: an overview. *Pharmacoeconomics*. 2017;35(7):673–683.
14. Sargent RG. *Verification and Validation of Simulation Models*. Presented at: Proceedings of the 2007 Winter Simulation Conference, Washington, DC, December 9-12, 2007.
15. Kleijnen JP. Verification and validation of simulation models. *Eur J Oper Res*. 1995;82(1):145–162.
16. Jalal H, Dowd B, Sainfort F, Kuntz KM. Linear regression metamodeling as a tool to summarize and present simulation model results. *Med Decis Mak Int J Soc Med Decis Mak*. 2013;33(7):880–890.
17. Merz JF, Small MJ, Fischbeck PS. Measuring decision sensitivity: a combined Monte Carlo-logistic regression approach. *Med Decis Mak*. 1992;12(3):189.
18. Tuffaha H, Rothery C, Kunst N, Jackson C, Strong M, Birch S. A review of web-based tools for value-of-information analysis. *Appl Health Econ Health Policy*. 2021;19(5):645–651.
19. Jalal H. Model examiner online R shiny application. https://metamodel.dashlab.ca/. Accessed February 2, 2024.
20. Jalal H, Alarid-Escudero F. A gaussian approximation approach for value of information analysis. *Med Decis Mak*. 2018;38(2):174.
21. Feenstra T, Corro-Ramos I, Hamerlijnck D, van Voorn G, Ghabri S. Four aspects affecting health economic decision models and their validation. *Pharmacoeconomics*. 2022;40(3):241–248.
22. Degeling K, Ijzerman MJ, Lavieri MS, Strong M, Koffijberg H. Introduction to Metamodeling for Reducing Computational Burden of Advanced Analyses with Health Economic Models: a Structured Overview of Metamodeling Methods in a 6-Step Application Process. *Medi Deci Making*. 2020;40(3):348–363.
23. Kunst N, Burger EA, Coupé VMH, Kuntz KM, Aas E. A guide to an iterative approach to model-based decision making in health and medicine: an iterative decision-making framework. *Pharmacoeconomics*. 2024;42(4):363–371.