PharmacoEconomics
https://doi.org/10.1007/s40273-025-01491-2

SYSTEMATIC REVIEW

Under-reporting of Validation Efforts for Health Economic Models Persists Despite the Availability of Validation Tools: A Systematic Review

Katharina Abraham[1] · Isaac Corro Ramos[2] · C. Louwrens Braal[3] · Talitha Feenstra[4,5] · Anne Kleijburg[6,7] · George A. K. van Voorn[8] · Carin Uyl-de Groot[9]

Accepted: 23 March 2025

Abstract

Objective In this study we aimed to identify possible changes over time in validation efforts and the way in which they are reported for model-based health economic (HE) evaluations, given the introduction of several new validation tools and methods in the past decade.

Methods A systematic review was conducted using PubMed and Embase on published HE models for early breast cancer (EBC) for the period 2016 to 2024. AdViSHE—consisting of four validation categories that cover the main HE model aspects—was utilized to systematically evaluate the reported evaluation efforts. The percentage of studies reporting validation per category was compared with the review by de Boer et al. that covers the years 2008 to 2015.

Results Of the 2199 records, 78 studies fulfilled the eligibility criteria. Reported validation efforts did not significantly improve compared with the previous period, except for the validation of input data by experts. Reporting on the validation of the conceptual model remained low with around 10% of the studies providing validation. Validation of the computerized model and validation against outcomes using alternative input data were the most underreported validation categories with <4% of the studies. The validation of model outcomes, specifically cross validity and the comparison with empirical data, remained the most reported categories in this review also, with 52% and 36%, respectively. When validation efforts were reported, this was done in a non-systematic manner, and the tests and results were rarely detailed.

Conclusion Overall reporting of validation efforts for model-based HE evaluations in the past decade did not significantly change compared with the previous decade.

Key Points for Decision Maker

Reporting of validation efforts for model-based health economic analyses has remained unchanged for the past two decades.

Studies originating from the United States were found to report validation more frequently in a systematic manner than studies originating from other countries.

1 Introduction

Model-based health economic (HE) evaluations aim to estimate the trade-off between costs and health effects of health intervention options for a specific population and time frame (‘the decision problem’) by synthesizing clinical, epidemiological, and economic evidence, often from different sources. The development of models for HE evaluations can be categorized into four main aspects: (1) the development of the conceptual framework, (2) its implementation as a computerized model, (3) the population of the model with clinical and resource data, and (4) judging the outputs of the model for plausibility. The main health outcomes on the patients’ disease trajectory and their relationships provide the premise for the evaluation and are commonly visualized as a conceptual framework. Extensive clinical subject knowledge is required to appropriately reflect the disease trajectory in the conceptual framework. The framework is implemented as a computerized model, commonly in code language, where the interdependencies between health outcomes are translated into the appropriate mathematical operations. Considering the sometimes-intricate relationships between health outcomes, numerous (complex) mathematical operations need to be coded for the conceptual

Extended author information available on the last page of the article

model to be fully implemented as a mathematical model. The computerized model is populated with data on (a) the clinical relationships between the main health outcomes, (b) other health consequences, such as adverse events, and (c) associated resource use and unit costs. Given the range of input data required, as well as the need for long-term follow-up data, usually beyond the duration of clinical trials, multiple data sources are commonly combined. This requires subject knowledge for the selection of appropriate data (sources), and the appropriate manipulation of the data for use in the computerized model (e.g., re-calculation to a particular cycle length, extrapolation of clinical outcomes, such as progression-free survival [PFS] and overall survival [OS], to the time horizon). Lastly, the outputs of the model, such as the estimated survival, need to be judged for clinical plausibility. Appreciating the complexity involved in modelling, each of these main model development aspects are expected to be subjected to validation to ensure trust in the results of the model by decision makers as well as researchers who may utilize published HE models in their work, for example when developing global HE models.

In 2016, de Boer and colleagues systematically reviewed the reporting of validation efforts for model-based HE evaluations on seasonal influenza (SI) and early breast cancer (EBC), to gain insight into the reporting of model validation efforts in the overall health economic literature; however, their review showed that the reporting of validation efforts is not standard practice [1]. Between 2008 and 2015, about a third of the reviewed publications did not report validation efforts on any of the main economic modelling aspects. Only 9% of the studies on SI and 37% on EBC reported on two or more validation aspects. While follow-up with the authors revealed that more validation efforts were performed than those reported in the publications, de Boer and colleagues concluded that the systematic reporting of validation efforts as standard practice was desirable to enhance decision makers' confidence in HE models and their outcomes [1]. Similar conclusions on the lack of reporting were drawn in 2020 by Zhang and colleagues, who evaluated the impact on reporting quality including validation efforts, before (1997–2012) and after (2013–2016) the introduction of the consensus-based guidelines on modelling practices for discrete event simulations by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR)-Society for Medical Decision Making (SMDM) [2, 3]. No significant improvement was found on the reporting quality with reporting of predictive validation (2.8%), cross validation (8.5%), and face validation (26.5%) showing the lowest performance.

Several reasons may have played a role in the lack of reporting validation efforts, such as a general unawareness among model users and modellers on the importance of reporting validation efforts, the lack of time and budget to facilitate the validation of models, or researchers not being sufficiently familiar with the term 'validation', mistakenly referring to one-way and probabilistic sensitivity analyses (OWSA and PSA) as an approach for validating models [4, 5]. Since the publication of the ISPOR-SMDM best practice guidelines in 2012 [3, 6], several new validation tools have been developed ranging from generic tools that cover validation of multiple model aspects [2, 7–9] to tools that specifically focus on the verification of the computerized model [10, 11], or specific validation aspects, such as face or external validity [12, 13], and tools for collaboration on the validation of models [14, 15].

In this study, we aimed to quantify the change in reporting validation efforts for model-based HE evaluations since the introduction of new validation tools. Specifically, our review aimed to clarify whether validation has become an integral part of reporting given the availability of time-efficient validation tools, and if so, in what manner validation efforts of economic models are currently conducted and reported.

2 Methods

A systematic literature search was performed in PubMed and Embase to retrieve HE modelling studies on EBC published between January 2015 and February 2024. In this follow-up on the review by de Boer and colleagues, who evaluated reporting efforts between January 2008 and December 2014, we focused on EBC only to keep the review manageable. To allow comparison of validation efforts with the review of de Boer et al., the same search strategy (Supplementary Data 1, see electronic supplementary material [ESM]) and inclusion criteria were used, except for the publication date: (1) HE evaluations published between January 2015 and February 2024; (2) presentation of costs and health effects; (3) use of a model to generate results; (4) full-text available, and (5) English language. Review papers, meta-analyses, and letters were excluded. Studies on metastatic breast cancer, screening, and diagnostic systems to stage breast cancer were excluded.

The records were deduplicated and abstracts of the unique records were screened for inclusion by KA and ICR in Rayyan to support the deduplication and comparison of the results from the title/abstract screening [16]. Records with disagreement were solved by consensus. The included records were then exported to Zotero for review in full text [17]. Any wrongly included records were excluded at this point. Model validation efforts were evaluated in the same manner as in de Boer et al. by categorizing the reported validation efforts based on the AdViSHE tool (Table 1) [1, 7]. In addition to comparing the proportion of studies reporting validation efforts with the review by de Boer et al., we evaluated whether validation efforts were reported in a

Table 1 Validation aspects according to AdViSHE [1]

Category	Model aspect	Validity test	Question to be answered with the validity test
A	Conceptual model	Face validity (A1)	Have experts been asked to judge the appropriateness of the conceptual model?
		Cross validity (A2)	Has this model been compared with other conceptual models found in the literature or clinical textbooks?
B	Input data	Face validity (B1)	Have experts been asked to judge the appropriateness of the input data?
		Model fit (B2)	When input parameters are based on regression models, have statistical tests been performed?
C	Computerized model	External review (C1)	Has the computerized model been examined by modelling experts?
		Extreme value testing (C2)	Has the model been run for specific, extreme sets of parameter values in order to detect any coding errors?
		Testing of traces (C3)	Have patients been tracked through the model to determine whether its logic is correct?
		Unit testing (C4)	Have individual submodules of the computerized model been tested?
D	Operational	Face validity (D1)	Have experts been asked to judge the appropriateness of the model outcomes?
		Cross validity (D2)	Have the model outcomes been compared with the outcomes of other models that address similar problems?
		Alternative input (D3)	Have the model outcomes been compared with the outcomes obtained when using alternative input data?
		Empirical data (D4)	Have the model outcomes been compared with empirical data? (A) Comparison against the data sources on which the model is based (dependent validation) (B) Comparison against a data source that was not used to build the model (independent validation)
E	Other	Other techniques	Have any other validation techniques been performed?

systematic manner and a validation tool was used. Reporting validation efforts in a systematic manner was defined as either (a) a separate section or paragraph on model validation specifically or (b) stating that a validation tool was used. Lastly, we described whether validation of the model was reported in general statement(s) or in detail, and whether any information on adjustments after validation were provided. The reporting of validation efforts was described as absolute numbers of studies fulfilling a validation category and relative to the total number of studies included. The difference in the proportion of reporting efforts between the two time periods were compared for statistical significance for each category of the AdViSHE tool with the two proportions z-test in R version 4.3.2 (2023 - 10- 31). To account for the potential impact of new validation tools on reporting, eligible records between January and December 2015, before the publication of AdViSHE, were added to the validation categories in the de Boer et al. review. All eligible studies published from January 2016 onwards were included in this follow-up review to allow a before–after comparison.

3 Results

The systematic literature review resulted in 85 relevant studies in the period January 2015 to February 2024 (Fig. 1).

The general characteristics of the included studies can be found in Table 2. Almost all studies ($N = 80$) were cost–utility analyses reporting on quality-adjusted life years (QALYs) and life years (LYs) as their main outcomes. The remaining studies were cost-effectiveness ($N = 4$) and cost–benefit ($N = 1$) analyses. The type of models utilized, as characterized in the papers, were Markov models ($N = 55$), microsimulation models ($N = 10$), a combination of a Markov model and a decision tree ($N = 7$), decision trees ($N = 5$), semi-Markov models ($N = 3$), partitioned survival models ($N = 3$), and a Markov macro-simulation ($N = 1$). One study did not specify the type of model. Seven of the 85 studies had a publication date before December 2015 (the publication date of the first validation tool [7] after the review by de Boer et al.) and were consequently included in the published numbers by de Boer et al., per validation category for the comparison. As a result, 78 studies represent the validation efforts in the publication period of the new validation tools (2016–2024). In the review by de Boer et al., originally 41 studies were included, which, in combination with the seven studies found in this review, represent the validation efforts for the period prior to the publication of the new validation tools (2008–2015).

Overall, the number of studies reporting across the different categories of validation did not significantly change since the introduction of the new validation tools, except for the face validation of input data (Supplementary Data 2, see

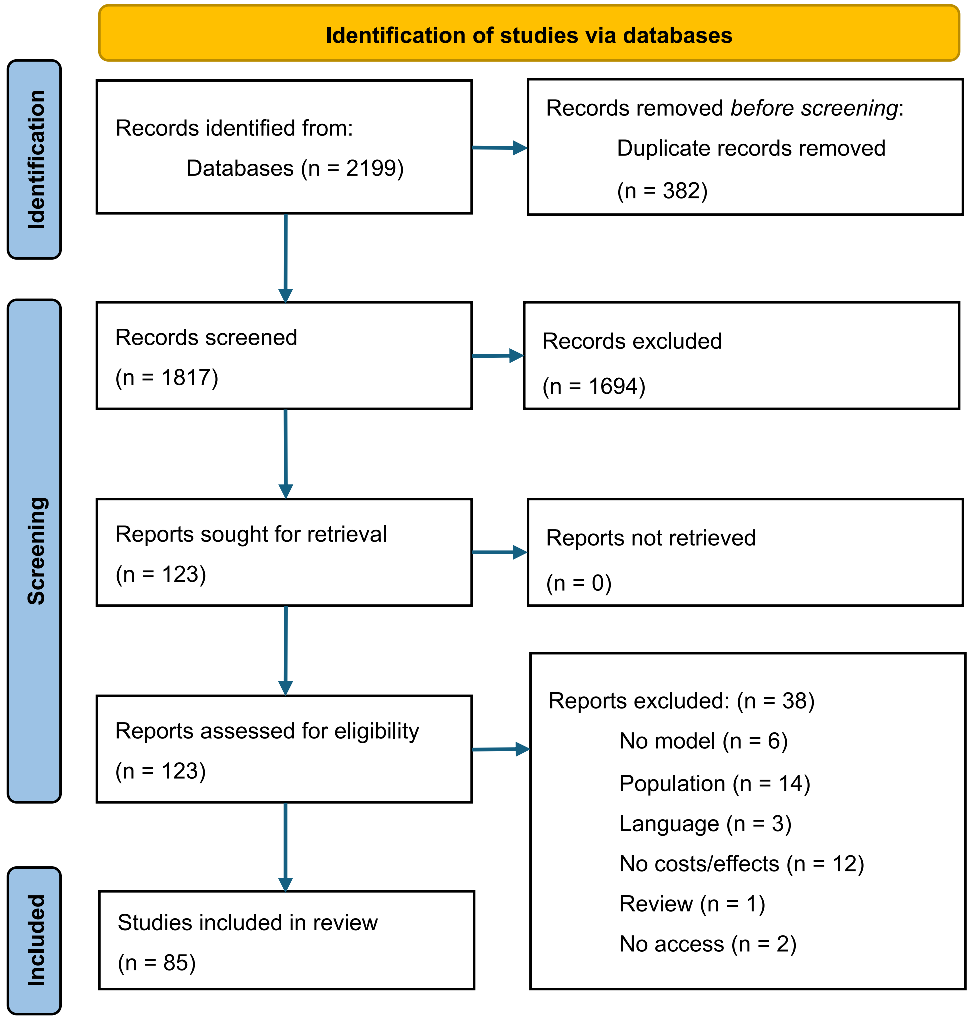

Fig. 1 PRISMA flow diagram

ESM). Before the introduction of the new tools and methods, 2% (1/48) of the studies reported that experts had validated the model input data, which increased to 20% (16/78) during this past decade (B1 in Fig. 2). A non-significant negative change was found for the reporting of statistical tests to validate input data for the model, such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) for selecting extrapolated survival curves. Before the introduction of new validation tools, de Boer et al. found that 13% of the studies reported statistical tests for input data while in this review only 3% (2/78) of the studies [18, 19] reported on statistical tests for input data (B2 in Fig. 2). Davie et al. [19], for example, reported the use of the AIC and the BIC to select the best fitting distributions for the extrapolated survival curves beyond the trial period. Statistical testing for input data may not be appropriate in all the model-based HE evaluations included in this review, or the review conducted by de Boer et al.

Face and cross validity of the conceptual model showed non-significant improvements since the introduction of new validation tools with 10% (8/78) versus 8% (4/48) of the studies reporting that experts were asked to judge the appropriateness of the conceptual model [20–26], and 12% (9/78) vs 6% (3/48) of the studies reporting that the conceptual model was compared with literature [20, 24–31] (A1 and A2 in Fig. 2). For example, Hajjar et al. [21] reported for the validation of the conceptual model that their co-author shared his clinical expertise on the model's structure and assumptions. Alshreef et al. [20] and Younis et al. [24] both reported the utilization of conceptual models validated by the NIHR (National Institute for Health and Care Research), and the Health Technology Assessment (HTA) agencies National Institute for Health and Care Excellence (NICE) and Canadian Agency for Drugs and Technologies in Health (CDA-AMC), respectively. The details of the comparison of the conceptual model with literature was, however, not described in any of the studies. The verification of the computerized model also remained underreported in this review (C1 to C4 in Fig. 2). While between 2008 and 2015, none of the studies (0/48) reported on the verification of the computerized model, in this review two of the 78 studies [25, 32] explicitly stated an external review of the computerized

Table 2 Study characteristics ($N = 85$)

Region	N **= 88***
Europe	18
North America	36
South America	2
Oceania	5
Asia	23
Africa	4
Income	N **= 88***
High	66
Middle	22
Low	0
Funding source	N **= 85**
Public sources	40
Industry	17
Public and industry	1
No external funding	17
Not stated	10
Type of intervention	N **= 91***
Pharmaceutical	61
Non-pharmaceutical	30
Type of evaluation	N **= 85**
Cost–benefit	1
Cost-effectiveness	4
Cost–utility	80
Model characteristic*	N **= 85**
Markov	55
Microsimulation (Markov, discrete event simulation (DES), continuous time)	10
Decision tree + Markov	7
Decision tree	5
Semi-Markov	3
Partitioned survival model	3
Markov macro-simulation	1
Not specified	1

*Some studies report for multiple countries, interventions, and/or model types. As described in the study

model (C1 in Fig. 2), and two studies [33, 34] reported on extreme value testing (C2 in Fig. 2). Earl et al. [25], for example, reported that two health economists independently validated the code of the computerized model by working on different parts of the code and subsequently explaining the code to each other to spot errors. Chandler et al. [34] reported that extreme parameter values were entered in the model to ensure that the output varied in the expected directions. One study [35] referred to the use of TECH-VER [10] but did not further specify whether all tests were conducted nor the results of those tests. Model outcomes were compared with outcomes from other models addressing a similar decision problem and with empirical data in 53% (41/78) and 36% (28/78) of the reviewed studies versus 50% (24/48) and 42% (20/48) in the previous period (D2 and D4 in Fig. 2). Like in the previous period, none of the studies reported on the validation of model outcomes with alternative input (D3 in Fig. 2). Also, none of the studies in either of the two time periods reported on all four main modelling aspects (A to D).

A systematic validation approach, defined as the utilization of validation tools and/or reporting of validation efforts within specified sections of the publication, was found in 12 of the 78 studies (15%) [21, 28, 32, 34–42]. Of these 12 studies, five reported the use of a validation tool or guideline (ISPOR-SMDM [40, 42], TECH-VER [35], and AdViSHE [36], and an adapted version of AdViSHE [38]). While these systematic validation tools were named in the publications, their results were reported only in one study and only about one validation category [36]. When we compared the proportions of systematic reporting between studies originating from the United States (US) and non-US studies (two proportions z-test), we found that studies originating from the US reported validation more frequently in a systematic manner than non-US studies (Supplementary Data 3, see ESM). No country-specific comparisons (e.g. US vs Germany) could be done due to small sample sizes. Overall, 9 of the 78 studies (12%) included the results of their validation efforts in their publication [28, 30, 36, 39, 41, 43–46]. A specific validation section was included in 7 of the 78 studies (9%) [21, 28, 32, 34, 37, 39, 41]. When included, the validation section was part of either the 'Methods' or the 'Results' section of the publication. Only one study reported validation in both sections [28]. When operational validation was reported, specifically the comparison with other (empirical) studies (D2, D4), this was mainly found in the 'Discussion' section of the publication.

The type of model did not seem to influence reporting efforts when comparing presumably simpler models, such as Markov models, Markov model + decision tree, and decision trees with presumably more complex model types, such as partitioned survival model, microsimulation and semi-Markov model. No other model type-specific comparisons could be made due to small sample sizes.

4 Discussion

Since the introduction of new validation tools and methods, significant improvements were found only for the reporting of face validity of input data, which increased from 2% of all included studies to 20%. Improvements in the validation of the conceptual model and the computerized model were found to be non-significant and remained low with 12% and 3% of the studies reporting, respectively. Thus, while more than 50% of the authors reported that they had compared the results between models, they often did so without comparing

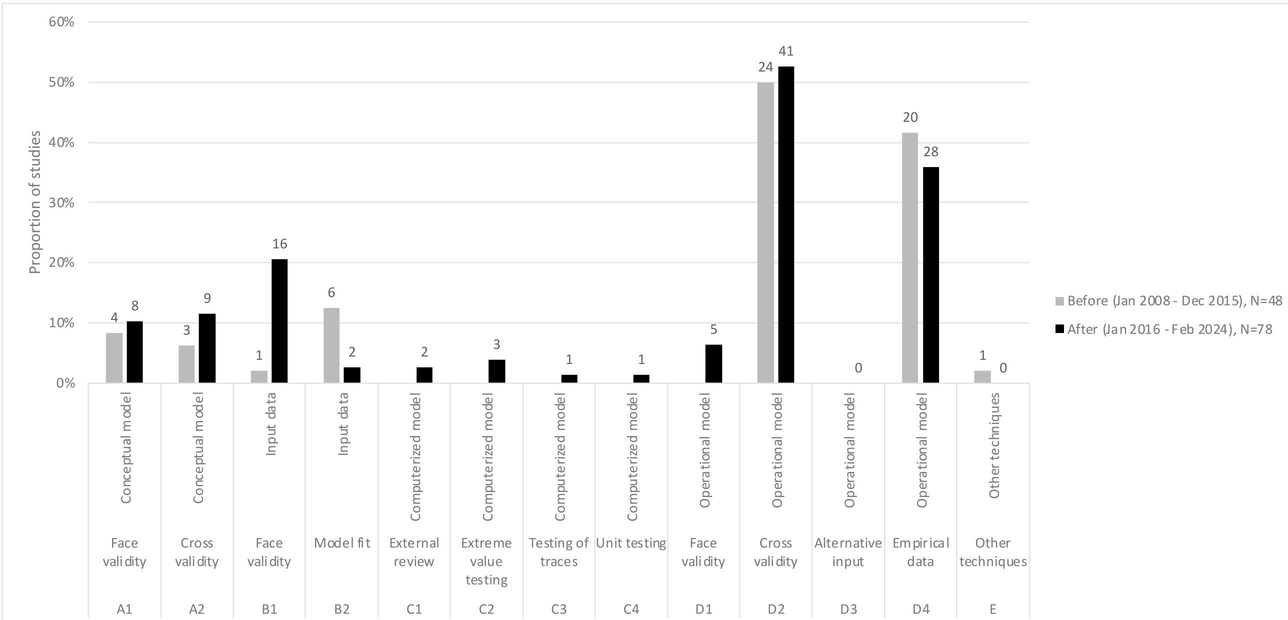

Fig. 2 Comparison of model validation efforts between the period 2016–2024 and the period 2008–2015

the models themselves. The low reporting on the validation of the conceptual model could be due to EBC being a commonly studied disease area, given its prevalence and known relevant disease outcomes. Therefore, authors may erroneously consider the reporting of the conceptual model validation as superfluous.

The very low reporting on the validation of the computerized model may be indicative of the lack of conducting validation tests on the programmed model. Considering the common occurrence of technical errors, as exemplified in a study on technical errors in company submissions to NICE [47], validation and the reporting thereof should be highly encouraged by journals and HTA agencies alike, and facilitated in project budgets and timelines. Reporting on technical model validation may be also lagging since specific tools for programming validation have been introduced only relatively recently, with TECH-VER [10] being one of the first. While TECH-VER [10] has been around since 2019, model developers may need to allocate more time to familiarize themselves with and to incorporate this extensive validation exercise within their workflow. More time-efficient tools, such as the online tool PACBOARD [11] and the CDA-AMC Model Validation Tool [9], have only been introduced in 2022 and 2024, and therefore uptake may still need time. PACBOARD [11], for example, has the advantage of requiring only the inputs and corresponding model outputs of a probabilistic analysis to check the validity of inputs and outputs and to investigate their relationships. The CDA-AMC Model Validation Tool [9] also facilitates time-efficient model verification requiring only a third of the black box tests outlined in the TECH-VER [10].

In addition to the aforementioned reasons, several other reasons may have played a role in the limited reporting of validation efforts, which are also recognized by the authors' own experiences with model-based HE. For example, confusion with the term 'validation' may have resulted in lower overall validation efforts. Sensitivity analyses such as OWSA and PSA were still mistaken for the validation of the conceptual model and inputs in this review [48–50]. While sensitivity analyses are important to illustrate (un)certainty of the model outcomes, their purpose is not to evaluate the validity of the (computerized) model, inputs or outcomes. In other words, whether the source selected for an input parameter is appropriate cannot be judged by running a PSA.

Furthermore, practical questions or lack of understanding when utilizing validation tools may have also contributed to the under-reporting of validation efforts. For example, it may not always be clear which model components fall under input versus outcome validation, as this may differ across different model types, impacting how and where to report validation efforts. This can be illustrated with the study by Braal et al. [51], where fitting the curves to the data falls under the validation of input data by testing for model fit, after which the resulting extrapolated and adjusted overall survival and recurrence-free curves are validated by experts as outcome. In contrast, for Markov models, transition probabilities are being validated as inputs, whereas the Markov traces adjusted for overall survival being validated as outcome. In particular, researchers new to the field may not be fully aware and/or understand the methods and tools for validation. Another practical question may arise regarding where to report the conducted validation tests and results. In the AdViSHE tool the recommendation is to fill in the form and indicate where the tests and outcomes were reported. However for a more homogenous way of reporting, the authors suggest either including a link in the publication to a repository that contains all validation efforts or to include the validation tests in a separate paragraph in the Methods section of the publication with the results of the tests in a separate paragraph in the Result section, as done by Ward et al. [28, 41].

Another explanation for the overall limited reporting of validation efforts since 2008 could be the lack of requirements for reporting validation of models when publishing model-based HE studies. Although the consolidated CHEERS 2022 checklist contains an item to describe the methods or approaches for transformed input data and any model used, model validation is not mandatory, nor does the item cover all the main aspects of model validation [52]. Journals could support the reporting of validation efforts by setting systematic reporting requirements for model validation or by requesting a reference to the publicly stored validation efforts, systematically conducted as described with the help of one of the validation tools. HTA agencies could also take a role in supporting the reporting of validation by demanding systematic validation of all submitted model-based HEs and by recommending the appropriate tools. This has been done recently by the Dutch HTA agency in their latest guidelines on cost effectiveness [53], in which the reporting of validation efforts for model-based HE analyses with the AdViSHE checklist is required. This could also help new model developers to select and train on the right tools for conducting validation of model-based HE analyses and the systematic reporting of the conducted tests and results.

Considering the overall agreement on the importance of model validation but the lack thereof for the past two decades, journals, HTA agencies, and model developers need to step up. While agencies and journals need to set the requirements and standards, model developers are encouraged to independently start with systematically reporting validation efforts. In addition, further research could be conducted on the perspectives of model developers regarding the reporting of validation and their choice of tools. These insights could help to tackle the perceived barriers on the model developers’ end.

In this review, we found that the reporting of validation efforts was low for the four main modelling aspects: (1) the development of the conceptual framework, (2) its implementation as a computerized model, (3) the population of the model, and (4) judging the outputs of the model, based on HE models for EBC. We cannot establish whether this means that models were not appropriately validated, since the focus of this review was on the change in validation efforts reported in published articles.

A limitation of this study could be the lack of other disease areas reviewed. While we think that EBC is a representative disease area for this review given its prevalence and thus the efforts of pharmaceutical companies to develop treatments and consequently the need for payers to make informed decisions, the results and conclusions should be interpreted with caution when generalizing the reporting of validation efforts for model-based HE evaluations. Furthermore, the reporting of model fit testing was found to be lower in this review compared with the review by de Boer et al. Since regression analysis may not be applicable to all the included HE evaluations in this or the previous review, the change in reporting on model fit needs to be interpreted with caution. Lastly, while the AdViSHE tool utilized in this study for reviewing the systematic reporting of validation covers the four main areas of model development, there are other aspects that contribute to the credibility of a model, such as the reporting of conflicts of interest or on the relevance of the PICO for the decision problem, which could be reported alongside model validation.

5 Conclusion

The introduction of additional validation tools since 2016 has not led to a statistically significant increase in the reporting of validation efforts in model-based HE evaluations for EBC. Transparent and systematic reporting of validation practices are essential to ensure the reliability of these models and should become standard practice. Journals and HTA agencies may need to take an active role by setting clear requirements and standards for the reporting of validation efforts.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40273-025-01491-2.

Declarations

Funding None of the authors received support from any organization for the submitted work.

Conflicts of Interest K.A., G.vV., T.F., A.K., C.UdG. and I.C.R. have no relevant financial or non-financial interests to disclose. C.L.B.: consultant for Eli Lilly, Daiichi Sankyo, CLB pharma consultancy (not during his work for this manuscript). All competing interest are outside the scope of this manuscript.

Availability of Data and Material Not applicable.

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Code Availability Not applicable.

Authors' Contributions K.A. and I.C.R. conceived the objective of the work. K.A. and I.C.R. performed the literature review. K.A. performed the data extraction, the analysis and drafted the article. I.C.R., G.vV., T.F., M.A., C.UdG., A.K. and C.L.B. critically and extensively revised the work. C.UdG. and I.C.R. supervised the work.

References

1. de Boer PT, Frederix GWJ, 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;34:833–45. https://doi.org/10.1007/s40273-016-0410-3.
2. Zhang X, Lhachimi SK, Rogowski WH. Reporting quality of discrete event simulations in healthcare—results from a generic reporting checklist. Value Health. 2020;23:506–14. https://doi.org/10.1016/j.jval.2020.01.005.
3. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling good research practices—overview: a report of the ISPOR-SMDM modeling good research practices task force-1. Value Health. 2012;15:796–803. https://doi.org/10.1016/j.jval.2012.06.012.
4. Bu DD, Schwam ZG, Kaul VF, Wong K, Fan C, Wanna GB, et al. Cost-effectiveness of canal wall-up vs canal wall-down mastoidectomy: a modeling study. Otolaryngol-Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg. 2022;167:552–9. https://doi.org/10.1177/01945998221076051.
5. Naved N, Umer F, Khowaja AR. Cost-effectiveness analysis of regenerative endodontics versus MTA apexification. JDR Clin Transl Res. 2023. https://doi.org/10.1177/23800844231191515.
6. Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, Möller J. Modeling using discrete event simulation: a report of the ISPOR-SMDM modeling good research practices task force–4. Med Decis Making. 2012;32:701–11. https://doi.org/10.1177/0272989X12455462.
7. 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;34:349–61. https://doi.org/10.1007/s40273-015-0327-2.
8. Jaime Caro J, Eddy DM, Kan H, Kaltz C, Patel B, Eldessouki R, et al. 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;17:174–82. https://doi.org/10.1016/j.jval.2014.01.003.
9. Coyle D, Haines A, Lee K. Model validation tool to assist in the conduct of economic evaluations. Can J Health Technol. 2024. https://doi.org/10.51731/cjht.2024.862.
10. Büyükkaramikli NC, Rutten-van Mölken MPMH, Severens JL, Al M. TECH-VER: a verification checklist to reduce errors in models and improve their credibility. Pharmacoeconomics. 2019;37:1391–408. https://doi.org/10.1007/s40273-019-00844-y.
11. Pouwels X, Kroeze K, van der Linden N, Kip M, Koffijberg H. MSR35 systematically validating health economic models using the probabilistic analysis check dashboard (PACBOARD). Value Health. 2022;25:S356. https://doi.org/10.1016/j.jval.2022.09.1767.
12. Tremblay G, Humphries B. A practical guide for clinical and key opinion leader validation of health economic models. J Med Econ. 2023;26:473–6. https://doi.org/10.1080/13696998.2023.2193112.
13. Corro Ramos I, van Voorn GAK, Vemer P, Feenstra TL, Al MJ. A new statistical method to determine the degree of validity of health economic model outcomes against empirical data. Value Health. 2017;20:1041–7. https://doi.org/10.1016/j.jval.2017.04.016.
14. Adibi A, Harvard S, Sadatsafavi M. Programmable interface for statistical & simulation models (PRISM): towards greater accessibility of clinical and healthcare decision models 2022. https://doi.org/10.48550/arXiv.2202.08358.
15. Harvard S, Adibi A, Easterbrook A, Werker GR, Murphy D, Grant D, et al. Developing an online infrastructure to enhance model accessibility and validation: the peer models network. Pharmacoeconomics. 2022;40:1005–9. https://doi.org/10.1007/s40273-022-01179-x.
16. Rayyan – Intelligent Systematic Review - Rayyan 2021. https://www.rayyan.ai/ (**accessed Jun 11, 2024**).
17. Zotero | Your personal research assistant n.d. https://www.zotero.org/ (**accessed Jun 11, 2024**).
18. Clarke CS, Hunter RM, Shemilt I, Serra-Sastre V. Multi-arm Cost-Effectiveness Analysis (CEA) comparing different durations of adjuvant trastuzumab in early breast cancer, from the English NHS payer perspective. PLoS One. 2017;12: e0172731. https://doi.org/10.1371/journal.pone.0172731.
19. Davie A, Traoré S, Giovannitti M, Pompilio G, Lambton M, Cakar E, et al. Cost-effectiveness analysis of abemaciclib with endocrine therapy (ET) versus ET alone for HR+, HER2-, node-positive, high-risk early breast cancer in Italy. Glob Reg Health Technol Assess. 2023;10:62–9. https://doi.org/10.33393/grhta.2023.2561.
20. Alshreef A, MacQuilkan K, Dawkins B, Riddin J, Ward S, Meads D, et al. Cost-effectiveness of docetaxel and paclitaxel for adjuvant treatment of early breast cancer: adaptation of a model-based economic evaluation from the United Kingdom to South Africa. Value Health Reg Issues. 2019;19:65–74. https://doi.org/10.1016/j.vhri.2019.03.001.
21. Hajjar A, Ergun MA, Alagoz O, Rampurwala M. Cost-effectiveness of adjuvant paclitaxel and trastuzumab for early-stage node-negative, HER2-positive breast cancer. PLoS One. 2019;14: e0217778. https://doi.org/10.1371/journal.pone.0217778.
22. Kongsakon R, Lochid-amnuay S, Kapol N, Pattanaprateep O. From research to policy implementation: trastuzumab in early-stage breast cancer treatment in Thailand. Value Health Reg Issues. 2019;18:47–53. https://doi.org/10.1016/j.vhri.2018.08.004.
23. Elsisi GH, Nada Y, Rashad N, Carapinha J, Noor AO, Almasri DM, et al. Cost-effectiveness of six months versus 1-year adjuvant trastuzumab in HER2 positive early breast cancer in Egypt. J Med Econ. 2020;23:575–80. https://doi.org/10.1080/13696998.2020.1724682.
24. Younis T, Lee A, Coombes ME, Bouganim N, Becker D, Revil C, et al. Economic evaluation of adjuvant trastuzumab emtansine

in patients with her2-positive early breast cancer and residual invasive disease after neoadjuvant taxane and trastuzumab–based treatment in Canada. Curr Oncol. 2020;27:578–89. https://doi.org/10.3747/co.27.6517.

25. Earl H, Hiller L, Vallier A-L, Loi S, McAdam K, Hughes-Davies L, et al. Six versus 12 months' adjuvant trastuzumab in patients with HER2-positive early breast cancer: the PERSEPHONE non-inferiority RCT. Health Technol Assess. 2020;24:1–230. https://doi.org/10.3310/hta24400.
26. Sussell J, Singh Jhuti G, Antao V, Herrera-Restrepo O, Wehler E, Bilir SP. Cost-effectiveness analysis of Ado-trastuzumab Emtansine (T-DM1) for the adjuvant treatment of patients with residual invasive HER2+ early breast cancer in the United States. Am J Clin Oncol. 2021;44:340–9. https://doi.org/10.1097/COC.0000000000000816.
27. Vaidya A, Vaidya P, Both B, Brew-Graves C, Bulsara M, Vaidya JS. Health economics of targeted intraoperative radiotherapy (TARGIT-IORT) for early breast cancer: a cost-effectiveness analysis in the United Kingdom. BMJ Open. 2017;7: e014944. https://doi.org/10.1136/bmjopen-2016-014944.
28. Ward MC, Vicini F, Chadha M, Pierce L, Recht A, Hayman J, et al. Radiation therapy without hormone therapy for women age 70 or above with low-risk early breast cancer: a microsimulation. Int J Radiat Oncol Biol Phys. 2019;105:296–306. https://doi.org/10.1016/j.ijrobp.2019.06.014.
29. Botelho CH, Estevez-Diz MDP, Campolina AG. Cost-effectiveness analysis of trastuzumab for early breast cancer in Brazil. Expert Rev Pharmacoecon Outcomes Res. 2022;22:63–72. https://doi.org/10.1080/14737167.2021.1909478.
30. Pham PD, Muchadeyi MT, Lindholm L. In situ breast cancer surgeries in Sweden: lumpectomy or mastectomy?—A cost-effectiveness analysis over a 30-Year period using Markov model. Cost Eff Resour Alloc. 2023. https://doi.org/10.1186/s12962-023-00495-z.
31. Zambelli A, Cazzaniga M, La Verde N, Munzone E, Antonazzo IC, Mantovani LG, et al. A cost-consequence analysis of adding pertuzumab to the neoadjuvant combination therapy in HER2-positive high-risk early breast cancer in Italy. Breast Edinb Scotl. 2023;71:113–21. https://doi.org/10.1016/j.breast.2023.08.005.
32. Huang M, Fasching PA, Haiderali A, Xue W, Yang C, Pan W, et al. Cost-effectiveness of neoadjuvant pembrolizumab plus chemotherapy followed by adjuvant single-agent pembrolizumab for high-risk early-stage triple-negative breast cancer in the United States. Adv Ther. 2023;40:1153–70. https://doi.org/10.1007/s12325-022-02365-1.
33. Corkum JP, Bezuhly M. Combining vascularized lymph node transfer with autologous breast reconstruction: a Surveillance, Epidemiology and End Results (SEER) Database cost-utility analysis. J Plast Reconstr Aesthet Surg. 2020;73:1879–88. https://doi.org/10.1016/j.bjps.2020.05.027.
34. Chandler Y, Schechter C, Jayasekera J, Isaacs C, Kurian AW, Cadham C, et al. Simulation modeling of breast cancer endocrine therapy duration by patient and tumor characteristics. Cancer Med. 2022;11:297–307. https://doi.org/10.1002/cam4.4084.
35. Glynn D, Bliss J, Brunt AM, Coles CE, Wheatley D, Haviland JS, et al. Cost-effectiveness of 5 fraction and partial breast radiotherapy for early breast cancer in the UK: model-based multi-trial analysis. Breast Cancer Res Treat. 2023;197:405–16. https://doi.org/10.1007/s10549-022-06802-1.
36. Deshmukh AA, Shirvani SM, Lal L, Swint JM, Cantor SB, Smith BD, et al. Cost-effectiveness analysis comparing conventional, hypofractionated, and intraoperative radiotherapy for early-stage breast cancer. J Natl Cancer Inst. 2017. https://doi.org/10.1093/jnci/djx068.
37. Garrison LPJ, Babigumira J, Tournier C, Goertz H-P, Lubinga SJ, Perez EA. Cost-effectiveness analysis of pertuzumab with trastuzumab and chemotherapy compared to trastuzumab and chemotherapy in the adjuvant treatment of HER2-positive breast cancer in the United States. Value Health J Int Soc Pharmacoeconomics Outcomes Res. 2019;22:408–15. https://doi.org/10.1016/j.jval.2018.11.014.
38. Hassett MJ, Li H, Burstein HJ, Punglia RS. Neoadjuvant treatment strategies for HER2-positive breast cancer: cost-effectiveness and quality of life outcomes. Breast Cancer Res Treat. 2020;181:43–51. https://doi.org/10.1007/s10549-020-05587-5.
39. Favre-Bulle A, Huang M, Haiderali A, Bhadhuri A. Cost-effectiveness of neoadjuvant pembrolizumab plus chemotherapy followed by adjuvant pembrolizumab in patients with high-risk, early-stage, triple-negative breast cancer in Switzerland. PharmacoEconomics - Open. 2024;8:91–101. https://doi.org/10.1007/s41669-023-00445-8.
40. Nguyen HT, De Allegri M, Heil J, Hennigs A. Population-level impact of omitting axillary lymph node dissection in early breast cancer women: evidence from an economic evaluation in Germany. Appl Health Econ Health Policy. 2023;21:275–87. https://doi.org/10.1007/s40258-022-00771-8.
41. Ward MC, Vicini F, Al-Hilli Z, Chadha M, Abraham A, Recht A, et al. Cost-effectiveness analysis of no adjuvant therapy versus partial breast irradiation alone versus combined treatment for treatment of low-risk DCIS: a microsimulation. JCO Oncol Pract. 2021;17:e1055–74. https://doi.org/10.1200/OP.20.00992.
42. Le V, Zhong L, Narsipur N, Hays E, Tran DK, Rosario K, et al. Cost-effectiveness of ribociclib plus endocrine therapy versus placebo plus endocrine therapy in HR-positive, HER2-negative breast cancer. J Manag Care Spec Pharm. 2021;27:327–38. https://doi.org/10.18553/JMCP.2021.27.3.327.
43. Lang H-C, Chen H-W, Chiou T-J, Chan ALF. The real-world cost-effectiveness of adjuvant trastuzumab in HER-2/neu-positive early breast cancer in Taiwan. J Med Econ. 2016;19:923–7. https://doi.org/10.1080/13696998.2016.1185013.
44. Ward MC, Vicini F, Al-Hilli Z, Chadha M, Pierce L, Recht A, et al. Cost-effectiveness analysis of endocrine therapy alone versus partial-breast irradiation alone versus combined treatment for low-risk hormone-positive early-stage breast cancer in women aged 70 years or older. Breast Cancer Res Treat. 2020;182:355–65. https://doi.org/10.1007/s10549-020-05706-2.
45. Xie Y, Guo B, Zhang R. Cost-effectiveness analysis of radiotherapy techniques for whole breast irradiation. PLoS One. 2021;16: e0248220. https://doi.org/10.1371/journal.pone.0248220.
46. Ward MC, Recht A, Vicini F, Al-Hilli Z, Asha W, Chadha M, et al. Cost-effectiveness analysis of ultra-hypofractionated whole breast radiation therapy alone versus hormone therapy alone or combined treatment for low-risk ER-positive early stage breast cancer in women aged 65 years and older. Int J Radiat Oncol Biol Phys. 2023;116:617–26. https://doi.org/10.1016/j.ijrobp.2022.12.028.
47. Radeva D, Hopkin G, Mossialos E, Borrill J, Osipenko L, Naci H. Assessment of technical errors and validation processes in economic models submitted by the company for NICE technology appraisals. Int J Technol Assess Health Care. 2020;36:311–6. https://doi.org/10.1017/S0266462320000422.
48. Huang X, Liu Y, Lin S, Wang H, Deng Y, Rao X, et al. Adjuvant zoledronic acid therapy for postmenopausal women with early breast cancer in China: a cost-effectiveness analysis. Int J Qual Health Care J Int Soc Qual Health Care. 2023;35: mzad016. https://doi.org/10.1093/intqhc/mzad016.
49. Asban A, Homsy C, Chen L, Fisher C, Losken A, Chatterjee A. A cost-utility analysis comparing large volume displacement oncoplastic surgery to mastectomy with single stage implant reconstruction in the treatment of breast cancer. Breast Edinb Scotl. 2018;41:159–64. https://doi.org/10.1016/j.breast.2018.07.012.

50. Lester-Coll NH, Rutter CE, Evans SB. Cost-effectiveness assessment of lumpectomy cavity boost in elderly women with early stage estrogen receptor positive breast cancer receiving adjuvant radiotherapy. Radiother Oncol J Eur Soc Ther Radiol Oncol. 2016;119:52–6. https://doi.org/10.1016/j.radonc.2016.01.026.
51. Braal CL, Kleijburg A, Jager A, Koolen SLW, Mathijssen RHJ, Corro Ramos I, et al. Therapeutic drug monitoring-guided adjuvant tamoxifen dosing in patients with early breast cancer: a cost-effectiveness analysis from the prospective TOTAM trial. Clin Drug Investig. 2022;42:163–75. https://doi.org/10.1007/s40261-021-01114-6.
52. Husereau D, Drummond M, Augustovski F, de Bekker-Grob E, Briggs AH, Carswell C, et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. Pharmacoeconomics. 2022;40:601–9. https://doi.org/10.1007/s40273-021-01112-8.
53. Welzijn en Sport Ministerie van Volksgezondheid. Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg (versie 2024) - Publicatie - Zorginstituut Nederland 2024. https://www.zorginstituutnederland.nl/publicaties/publicatie/2024/01/16/richtlijn-voor-het-uitvoeren-van-economische-evaluaties-in-de-gezondheidszorg (**accessed Jan 18, 2025**).

Authors and Affiliations

Katharina Abraham[1] · Isaac Corro Ramos[2] · C. Louwrens Braal[3] · Talitha Feenstra[4,5] · Anne Kleijburg[6,7] · George A. K. van Voorn[8] · Carin Uyl-de Groot[9]

✉ Katharina Abraham
abraham@eshpm.eur.nl

1 Erasmus University Rotterdam (EUR), Erasmus School of Health Policy and Management, Rotterdam, South Holland, The Netherlands

2 Institute for Medical Technology Assessment (iMTA), EUR, Rotterdam, South Holland, The Netherlands

3 Department of Medical Oncology, Erasmus University MC Cancer Institute, Rotterdam, The Netherlands

4 Faculty of Science and Engineering, University of Groningen, Groningen Research Institute of Pharmacy, Groningen, The Netherlands

5 National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands

6 Department of Health Services Research, CAPHRI Care and Public Health Research Institute, Maastricht University, Maastricht, The Netherlands

7 Centre of Economic Evaluations and Machine Learning, Trimbos Institute, Netherlands Institute of Mental Health and Addiction, Utrecht, The Netherlands

8 Applied Mathematics and Statistics/Biometrics, Wageningen University and Research, Wageningen, The Netherlands

9 Erasmus School of Health Policy and Management, EUR, Rotterdam, South Holland, The Netherlands