Journal of Clinical Epidemiology 115 (2019) 106−115

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Decision analytic modeling was useful to assess the impact of a prediction model on health outcomes before a randomized trial

Kevin Jenniskens[a,*,1], Ghizelda R. Lagerweij[a,1], Christiana A. Naaktgeboren[a], Lotty Hooft[a,b], Karel G.M. Moons[a,b], Judith M. Poldervaart[a], Hendrik Koffijberg[c], Johannes B. Reitsma[a,b]

[a]*Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, University Utrecht, Utrecht, The Netherlands*
[b]*Cochrane Netherlands, University Medical Center Utrecht, University Utrecht, Utrecht, The Netherlands*
[c]*Department of Health Technology and Services Research, Technical Medical Centre, University of Twente, Enschede, The Netherlands*

Accepted 16 July 2019; Published online 19 July 2019

Abstract

Objective: To demonstrate how decision analytic models (DAMs) can be used to quantify impact of using a (diagnostic or prognostic) prediction model in clinical practice and provide general guidance on how to perform such assessments.

Study Design and Setting: A DAM was developed to assess the impact of using the HEART score for predicting major adverse cardiac events (MACE). Impact on patient health outcomes and health care costs was assessed in scenarios by varying compliance with and informed deviation (ID) (using additional clinical knowledge) from HEART score management recommendations. Probabilistic sensitivity analysis was used to assess estimated impact robustness.

Results: Impact of using the HEART score on health outcomes and health care costs was influenced by an interplay of compliance with and ID from HEART score management recommendations. Compliance of 50% (with 0% ID) resulted in increased missed MACE and costs compared with usual care. Any compliance combined with at least 50% ID reduced both costs and missed MACE. Other scenarios yielded a reduction in missed MACE at higher costs.

Conclusion: Decision analytic modeling is a useful approach to assess impact of using a prediction model in practice on health outcomes and health care costs. This approach is recommended before conducting an impact trial to improve its design and conduct.

Keywords: Decision analysis; Prediction model; Impact study; HEART score; Cost-effectiveness; Research waste

1. Introduction

Diagnostic or prognostic prediction models can be used to support management decisions such as subsequent testing, treatment, or lifestyle changes. Developed prediction models require external validation to ensure they have adequate predictive performance [1−4]. However, good predictive performance does not imply that implementation in clinical practice will improve health outcomes or reduce health care costs. The impact of using risk prediction models in clinical practice on patient health and monetary outcomes can be evaluated in impact studies, such as comparative longitudinal (ideally (cluster) randomized) trials, in which care directed by the prediction model is compared with usual care [5−10].

Impact studies for prediction models are infrequent, most likely due to their complexity, long follow-up, associated high costs, and lack of regulatory requirements [7−9,11−13]. In addition, the benefits observed in such impact studies have typically been smaller than expected or even lacking [14−16]. An approach using a decision analytic model (DAM) may prove useful, making use of evidence available at the time an impact study is being considered. A DAM could provide insight in the conditions under which a prediction model is likely to result in favorable health outcomes or costs when implemented in clinical practice.

Decision analytic modeling is a method that integrates multiple sources of evidence to assess the downstream cost-effectiveness of applying a prediction model in daily practice [7−9,17,18]. Constructing a DAM forces researchers to think about the pathway through which

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

[1] Shared first author.

* Corresponding author. Tel.: +31 88 75 680 50.
E-mail address: K.Jenniskens@umcutrecht.nl (K. Jenniskens).

https://doi.org/10.1016/j.jclinepi.2019.07.010
0895-4356/

What is new?

Key findings

- Decision analytic models are a valuable tool for assessing the potential impact of a diagnostic or prognostic prediction model on actual health outcomes and costs, before implementing the model, either in clinical practice or in the context of an empirical study.

What this adds to what was known?

- An illustrative example of the HEART score, a prediction model used for diagnosing major adverse cardiac events, shows it is feasible to perform a decision analytic modeling assessment without requiring data from an empirical longitudinal randomized impact study.
- Decision analytic models provide insight in the impact of compliance with and informed deviation from HEART score predictions and management recommendations on patient health outcomes and health care costs.
- General guidance is given on how to perform impact assessments of the use of diagnostic or prognostic prediction models through decision analytic modeling approaches, which methods can be applied, and what data sources can be used.

What is the implication and what should change now?

- Researchers contemplating an empirical longitudinal randomized impact study for a risk prediction model should first perform a decision analytic modeling assessment. Based on the outcome of this assessment, researchers can decide if, and under what conditions, an empirical impact study is warranted.

(multiple alternative) complex interventions can lead to health and monetary benefits, such as variation in the interplay between the model predictions and subsequent patient management based on these predicted risks. DAMs also allow for uncertainty on parameters, such as distribution of predicted probabilities or effectiveness of treatment, to be taken into account. In addition, downstream effects of hypothetical scenarios can be analyzed, by varying values of parameters for which there is little or no evidence. The results are then used to inform decisions for an individual patient or health care policy. DAMs have also been proposed and performed before conducting longitudinal comparative trials to assess impact of (complex) therapeutic interventions and diagnostic tests [19–21], although they are still rare for diagnostic or prognostic prediction models. An explanation for this could be that using DAMs to assess impact is more complex for prediction models than for interventions, as the former would not only need to include accuracy of predictions but also downstream effects of, for example, benefits and harms of subsequent tests. In addition, lack of available evidence on compliance with management recommendations from a prediction model based on the predicted risk, and informed deviation (ID) from that compliance (i.e., whether there is incremental value of a clinician's experience on top of predictions provided by a model) may also explain the limited number of DAMs assessing impact of prediction models before conducting a formal large-scale, long-term, costly, empirical impact study. Although DAMs are particularly ideal to estimate the impact when evidence is lacking, namely by simulating multiple (hypothetical) scenarios.

In this article, we demonstrate how to assess the potential impact of a prediction model on patient health outcomes and health care costs using a DAM approach, specifically focusing on the effect of compliance with management recommendations. We will use the HEART score prediction model for diagnosis of major adverse cardiac events (MACE) in patients with chest pain as a case study [22]. This article will conclude by providing generic guidance on how to perform a DAM-based assessment to estimate the impact of using a prediction model in daily practice and elaborate on how the results of such DAM can inform the design and conduct of a subsequent prospective comparative prediction model impact study.

2. Methods

2.1. Case study

We compared implementation of the HEART score prediction model to usual care in a DAM as an example of how compliance with management recommendations from a prediction model influences the impact of that model on patients' health outcomes, health care costs, and cost-effectiveness. The HEART score provides an excellent example for illustrating the usefulness of a DAM, as model development [22] and several external validations have shown that the HEART score can correctly predict and stratify patients according to their risk of having MACE [23–26], and HEART score predictions were categorized and linked to management recommendations (Table 1). Although a randomized impact trial has recently been conducted for the HEART score prediction model [27], the DAM only used information from studies and data sources available before this trial was conducted. Note that the main aim of this article was not to replicate the results from this impact trial [27] but rather to illustrate how a DAM can be used to assess the impact of a prediction model on patient health outcomes and health care costs.

Table 1. Overview of the HEART score predictions, categories, and their associated risk-based management recommendations

HEART score	HEART score category	Management recommendation
0–3	Low	Discharge home
4–6	Intermediate	Noninvasive testing
7–10	High	Invasive testing

The HEART score is a prediction model that uses routinely collected information from patient history and blood tests to predict MACE in patients presenting with chest pain at the emergency room, to generate a risk score ranging from 0 to 10 (Table 1) [22]. The potential benefit of using the HEART score lies in its ability to stratify patients according to their risk of MACE and provide risk-based management recommendations. Physicians are advised to promptly discharge low-risk patients (i.e., HEART score $\leq$ 3), reducing utilization of health care resources, and providing additional diagnostic testing in higher-risk patients (i.e., HEART score $\geq$ 4), to prevent unnecessary delay in treatment initiation. Noninvasive diagnostic testing for the intermediate HEART score category consisted of stress bicycle ECG, myocardial scintigraphy, coronary CT angiography, and cardiac MRI. Invasive diagnostic testing for the high HEART score category consisted of coronary angiography, in combination with any of the noninvasive tests.

We evaluated the HEART score purely as a diagnostic instrument for MACE, meaning that in our model, the HEART score and any subsequent actions do not have an impact on the total number of MACE. MACE found during diagnostic workup (detected MACE) were considered a favorable outcome, whereas MACE in discharged patients (missed MACE) were considered an unfavorable outcome.

2.2. *Structure of the decision analytic model*

Fig. 1 and Appendix A display the DAM comparing usual care to the HEART score strategy. In the usual care strategy, HEART scores are not available to clinicians and are therefore not used to guide subsequent patient management decisions. In the HEART score strategy, we mimicked that clinicians at the emergency department (ED) would calculate the HEART score and they would be given clear guidance on subsequent risk-based patient management recommendations (see Table 1).

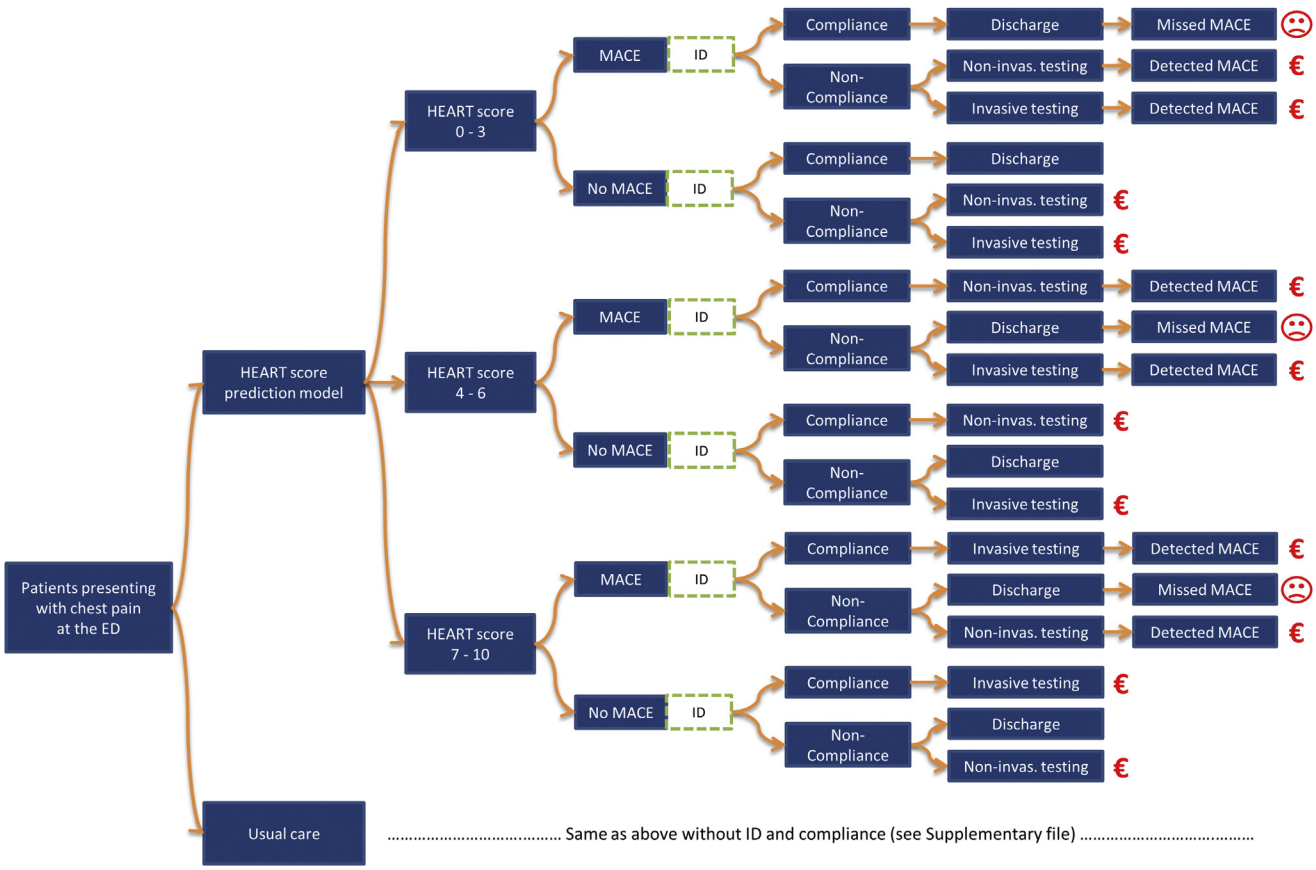

Fig. 1. Decision tree for using the HEART score prediction model for management decisions in patients presenting with chest pain at the ED. Euro signs and emoticons represent negative effects on costs and health outcomes, respectively. *Abbreviations:* ED, emergency department; ID, informed deviation from management recommendations corresponding to HEART score predictions; MACE, major adverse cardiac events.

The DAM used an assistive (as opposed to a directive) prediction model approach, meaning physicians were not forced to comply with management recommendations [8,28]. This allows for better mimicking actual implementation of the prediction model in clinical practice, and thus provides more realistic and generalizable quantification of impact. The focus of this study is to quantify the impact of compliance with the HEART score predictions and subsequent management recommendations on patient-relevant health outcomes and monetary outcomes. Accordingly, we varied the amount of compliance to the prediction model's management recommendations (i.e., the percentage of patients in whom the specified management recommendation was followed) in several scenarios (see "Scenario analysis" paragraph).

In the DAM, clinicians could deviate from recommended management based on additional patient information (e.g., signs and symptoms) or clinical expertise, leading to more appropriate stratification of management given to patients, so-called ID. This ID was included as a variable in the DAM, defined as the proportion of patients for whom the initial management recommendations according to the prediction model were incorrect, in which physicians—informed by additional knowledge—correctly deviate from those recommendations. ID ranged from 0% (uninformative compliance; compliance is equal in patients with and without MACE) to 100% (fully informative compliance; patients with MACE follow a diagnostic pathway, and patients without MACE are discharged). For an example of how ID influences management recommendations in the low–HEART score category, see Table 2. Introducing 50% ID to a scenario in which there is 80% compliance to management recommendations would lead to an additional 40% of patients with MACE receiving testing and an additional 10% of individuals without MACE being discharged.

2.3. Input parameters for the decision analytic model

To operationalize the DAM, each parameter requires an input value. Three types of input parameters are considered. First, transition probabilities, which are the probabilities for transitioning from one (health) state to the next, are defined (marked in Fig. 1 by the orange arrows). Second, we defined the main and other health outcomes. Finally, input values for the intended and unintended effects and costs of any subsequent tests, treatments, and conditions need to be determined. Input for most of these parameters in the "usual care" strategy was based on the observational data from the study by Nieuwets et al. [29]. See Appendix B for an overview of all input parameters.

2.3.1. Transition probabilities

The distribution of the target patient population across HEART score categories and MACE rates per HEART score category were derived from development [22] and multiple external validation studies of the HEART score prediction model [23,24,30]. Values for compliance and ID were not available and are further described in the "Scenario Analysis" Section. Transition probabilities and likelihood of receiving specific diagnostic tests (e.g., a stress bicycle ECG) in noninvasive and invasive diagnostic testing pathways were derived from a study measuring consumption of health care resources in usual care [29].

2.3.2. Health outcomes

The health outcome of interest was defined as the proportion of missed MACE, that is, patients with MACE at 6 weeks who were (initially) discharged without any subsequent diagnostic workup. MACE detected during or occurring after diagnostic workup was not included as adverse outcome, as this would have been found and managed accordingly in clinical practice. MACE was defined as occurrence of one or more of the following events or interventions: acute myocardial infarction (both ST-segment and non-ST-segment elevation), unstable angina, percutaneous coronary intervention, coronary artery bypass grafting, significant stenosis (>50%) managed conservatively, and death due to any cause [31].

2.3.3. Health care costs

Calculation of the HEART score relies on readily available predictors; hence, no extra costs are associated with

Table 2. Illustration of how compliance with and informed deviation from HEART score management recommendations for the low–HEART score category (0–3) influence the proportion of patients being discharged vs. receiving additional (non-)invasive testing

	Scenario (HEART score 0–3)		
Course of action	**Compliance 80% ID 0%**	**ID 50%**	**Compliance 80% ID 50%**
MACE			
Discharged	80%	−40% (80% × 0.5)	40%
Additional testing*	20%	+40% (80% × 0.5)	60%
No MACE			
Discharged*	80%	+10% (20% × 0.5)	90%
Additional testing	20%	−10% (20% × 0.5)	10%

Abbreviations: ID, informed deviation; MACE, major adverse cardiac events.
* Preferred course of action for patients with and without MACE.

collection of these predictors when compared with usual care. Costs of MACE were calculated based on a weighted average of costs and probability of each individual MACE component, derived from scientific literature [29,32–35]. Costs of noninvasive and invasive testing pathways in a specific HEART score category were calculated by taking the average number of times a specific diagnostic test was used per patient in that pathway and multiplying it by its unit costs [29]. Summing the average cost for all diagnostic tests in each of the pathways yielded the total costs of diagnostic testing. Similarly, the average number of admission and readmission days were calculated for each of the diagnostic pathways. Complication rates in noninvasive and invasive testing pathways were not explicitly included in the model; however, the expected frequency of severe complications for procedures included in the DAM is low [36–38], and expected costs of complications are largely captured by the number of (re)admission days.

2.4. Analyses

Scenario analysis was performed, comparing hypothetical scenarios in which compliance and ID were varied. Furthermore, a probabilistic sensitivity analysis was performed, in which a cohort was run through a series of simulations to take into account uncertainty surrounding the parameters in the DAM. A time horizon of 6 weeks was taken for the analyses, for which discounting was not deemed necessary.

2.4.1. Scenario analysis

Scenario analysis focused on comparing different compliances to HEART score predictions and corresponding management recommendations, combined with varying degrees of ID from those compliances. The influence of compliance on missed MACE and costs was investigated in three different scenarios: low (50%), medium (75%), and full (100%) compliance. Furthermore, four scenarios were defined for ID: no (0%), low (25%), medium (50%), and high (75%) ID.

For each scenario, the incremental proportion of missed MACE, health care costs, and cost per missed MACE was given per HEART score category and for all HEART score categories combined, as compared with usual care. Cost-effectiveness planes were provided to give insight in the distribution of missed MACE and health care costs in the presence of parameter uncertainty.

2.4.2. Probabilistic sensitivity analysis

Monte Carlo simulation was used to assess the robustness of expected health outcomes and health care costs based on uncertainty surrounding the different parameters. A series of 10,000 simulations was run per scenario, each with a patient population of 200,000, reflective of the annual Dutch population visiting the ED with chest pain [39]. Parameter uncertainty was reflected by calculating standard errors and defining appropriate statistical distributions for each parameter. Beta and Dirichlet distributions were used to account for uncertainty in transition probabilities. Gamma distributions were used for uncertainty surrounding costs (see Appendix B).

3. Results

In usual care, the average proportion of patients with missed MACE was estimated at 0.016 (95% confidence interval 0.007–0.027) or an average of 16 MACE in discharged patients per 1,000 individuals presenting with chest pain at the ED. The average cost per patient in usual care was €2,870 [29].

3.1. Scenario analysis

The impact of compliance and ID on the number of missed MACE (i.e., effects), costs, and cost-effectiveness was investigated in various scenarios. Negative values for missed MACE indicate a decrease, and positive numbers, an increase in missed MACE, compared with usual care. The values in Tables 3–5 are marked in bold to indicate a beneficial effect or underlined to indicate an unbeneficial effect of using the HEART score in practice.

Table 3. Average difference in missed MACE per person between the HEART score strategy and usual care

		ID			
	Compliance	**0%**	**25%**	**50%**	**75%**
HEART score 0–3	50%	0.006	0.004	0.001	**−0.001**
	75%	0.011	0.007	0.004	0.000
	100%	0.016	0.011	0.006	0.001
HEART score 4–6	50%	0.002	**−0.005**	**−0.011**	**−0.018**
	75%	**−0.011**	**−0.015**	**−0.018**	**−0.021**
	100%	**−0.025**	**−0.025**	**−0.025**	**−0.025**
HEART score 7–10	50%	0.003	**−0.003**	**−0.009**	**−0.014**
	75%	**−0.009**	**−0.011**	**−0.014**	**−0.017**
	100%	**−0.020**	**−0.020**	**−0.020**	**−0.020**
Total	50%	0.004	**−0.001**	**−0.006**	**−0.011**
	75%	**−0.002**	**−0.005**	**−0.009**	**−0.012**
	100%	**−0.008**	**−0.010**	**−0.012**	**−0.014**

Abbreviations: ID, informed deviation; MACE, major adverse cardiac events.

A negative number represents a reduction in missed MACE. The average number of missed MACE per patient in usual care was 0.016. Values marked in bold indicate a beneficial effect and underlined values indicate an unbeneficial effect of using the HEART score in practice.

Table 4. Average difference in costs per patient between the HEART score strategy and usual care

		ID			
	Compliance	**0%**	**25%**	**50%**	**75%**
HEART score 0–3	50%	€26	**−€55**	**−€137**	**−€219**
	75%	**−€148**	**−€186**	**−€224**	**−€262**
	100%	**−€323**	**−€317**	**−€312**	**−€306**
HEART score 4–6	50%	€7	**−€114**	**−€235**	**−€356**
	75%	**−€102**	**−€196**	**−€289**	**−€383**
	100%	**−€211**	**−€278**	**−€344**	**−€410**
HEART score 7–10	50%	€537	€613	€689	€764
	75%	€1,465	€1,309	€1,153	€996
	100%	€2,393	€2,005	€1,617	€1,228
Total	50%	€98	€23	**−€51**	**−€126**
	75%	€125	€43	**−€38**	**−€119**
	100%	€151	€64	**−€24**	**−€112**

Abbreviations: ID, informed deviation; MACE, major adverse cardiac events.

A negative number represents a reduction in costs. The average cost per patient in usual care was €2,870. Values marked in bold indicate a beneficial effect and underlined values indicate an unbeneficial effect of using the HEART score in practice.

3.1.1. Missed MACE

Table 3 shows the average difference in missed MACE (per person) for each of the HEART score categories and for the total patient population, as compared with usual care. Maybe somewhat surprisingly at first sight, the low–HEART score category shows an increase in the proportion of missed MACE as compliance increases, whereas in the intermediate– and high–HEART score categories there is an inverse relation. This can be explained by the different management recommendations associated with each HEART score category. Higher compliance in the low–HEART score category obviously leads to more patients being discharged, running the risk of missing more MACE in these patients. On the other hand, compliance in the intermediate– and high–HEART score categories automatically implies more diagnostic testing, reducing the risk of missing MACE. ID counteracts the higher proportion of missed MACE in the low–HEART score category and further reduces missed MACE in the intermediate and high categories.

Table 5. Ratios of the average difference in cost and missed MACE between the HEART score strategy and usual care

		ID			
	Compliance	**0%**	**25%**	**50%**	**75%**
Total	50%	€25,946	€21,749	**€8,614**	**€11,648**
	75%	€64,113	€8,099	**€4,292**	**€9,738**
	100%	€19,751	€6,576	**€2,092**	**€8,228**

Abbreviations: ID, informed deviation; MACE, major adverse cardiac events.

Cost-effective scenarios are marked in bold, where numbers represent the reduction in costs to prevent one missed MACE. Not cost-effective scenarios are underlined, where numbers represent the increase in costs for one extra missed MACE. Other scenarios are unmarked (plain black), where cost-effectiveness depends on the willingness to pay for preventing missed MACE. Numbers represent the increase in costs to prevent one missed MACE.

3.1.2. Costs

Table 4 shows the average difference in costs per patient between the HEART score strategy and usual care. Costs declined for the low– and intermediate–HEART score category when compliance and ID increased. A different pattern is observed when the high–HEART score category is taken into consideration, where a higher compliance led to higher costs. In the total patient population, an ID of at least 50% reduced costs of the HEART score strategy compared with usual care.

3.1.3. Costs/missed MACE ratio

To gain insight in the monetary investment required to reduce missed MACE, the ratio for the difference in costs and missed MACE between the HEART strategy and usual care is calculated. Table 5 shows the results for the different scenarios of compliance and ID, exhibited for the total patient population. HEART score strategy is considered cost-effective when there are less costs and fewer missed MACE compared with usual care (marked in bold in Table 5). HEART score strategy is considered not cost-effective when there are both extra costs and more missed MACE compared with usual care (underlined in Table 5). When missed MACE could be reduced at higher costs, cost-effectiveness depends on the willingness to pay for reducing missed MACE (unmarked in Table 5).

The impact of introducing the HEART score strategy on cost per missed MACE depended greatly on the interplay between compliance and ID. For scenarios with a compliance of at least 50% combined with at least 50% ID, costs and missed MACE were both reduced, resulting in a promising (i.e., cost-effective) strategy.

3.2. Probabilistic sensitivity analysis

Fig. 2 shows the incremental cost-effectiveness plane of four scenarios (compliance of 50%/100% and ID of 0%/75%) of the HEART score strategy compared with usual care. In the scenario with 50% compliance with (and 0% ID from) management recommendations, 68% of simulations resulted in an outcome that would be considered not cost-effective. This means that there are both more missed MACE and higher costs for the HEART score strategy compared with usual care. When 100% compliance (and 0% ID) was assumed, there was a reduction in missed MACE, but in all simulations, costs per patient were higher. For both the 50% compliance/75% ID and 100%

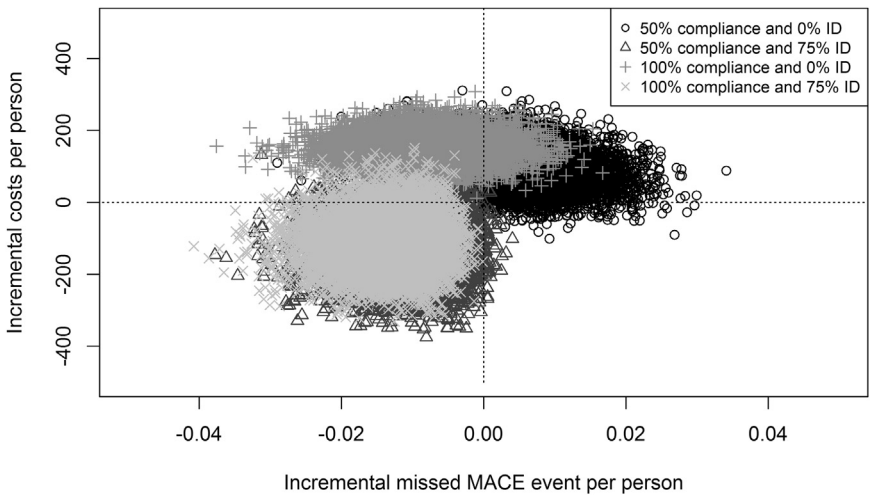

Fig. 2. Incremental cost-effectiveness planes for 10,000 simulations (each symbol represents one simulation) comparing the HEART score strategy with usual care for the following scenarios: 50% compliance with 0% ID; 50% compliance with 75% ID; 100% compliance with 0% ID; 100% compliance with 75% ID. Note that negative numbers indicate a more desirable outcome (less missed MACE and/or reduction in costs). *Abbreviations*: ID, informed deviation; MACE, major adverse cardiac events.

compliance/75% ID scenarios, 94% of the simulations resulted in a reduction in missed MACE as well as cost savings.

4. Discussion

We have shown how a DAM can be used to estimate the potential health-economic impact of using a diagnostic or prognostic prediction model in practice, using only data and information available before performing a costly, long-term, randomized impact trial. We illustrated this for various hypothetical scenarios if the HEART score prediction model were to be implemented in clinical practice.

Generating a DAM for impact assessment of a prediction model forces researchers to think about its main goals (e.g., reducing the primary outcome, reducing side effects, and optimizing diagnostic and treatment pathways) and how it aims to achieve these goals. DAMs can help demonstrate under what conditions (e.g., amount of required compliance and deviation of model adherence) a prediction model is likely to have the desired impact on health outcomes and/or costs. If it is unlikely that these conditions are going to be satisfied, then one should consider whether investment in a large-scale prediction model impact trial is justified [7–9,40]. Should those conditions be deemed plausible, a pilot study or qualitative assessments with experts in the field might be considered to gain more insight and reduce parameter uncertainty. This information can then be used to update the DAM, allowing researchers to reassess the prediction model's expected impact. Researchers should ensure using representative and valid input parameters for their DAM, preventing goal-oriented model construction and assessment. In general, DAMs should be used for optimizing the design and conduct of an upcoming impact study [9,41].

A DAM can be developed for any type of prediction model to evaluate its potential impact. Fig. 3 provides a concise overview on how to conduct a model-based impact assessment of a prediction model. The first step is designing the DAM, for which different structures can be chosen, such as a decision tree, Markov model, or microsimulation model. Next, parameter estimates should be collected, such as probabilities (e.g., transition probability between [health] states), health outcomes (e.g., quality of life), and costs (e.g., cost of diagnostic tests). Feasibility of creating a DAM depends on availability of these data. Analyses can then be run for different scenarios, typically by varying parameters with the greatest uncertainty surrounding them (e.g., compliance and ID for our case study). Alternatively, scenarios can look at other cutoffs for stratifying patients into risk categories. Robustness of outcome measures can be assessed by using Monte Carlo simulation, varying parameter estimates based on the uncertainty surrounding them. In the final step, the results of the DAM can guide the decision on whether a trial to study the impact of a prediction model is warranted. If so, a DAM could provide directions for a pilot study or qualitative assessment before the trial to help optimize its design and conduct. More details on how to develop and analyze DAMs can be found in literature [17,47].

To provide insight in the validity of our DAM, it is worthwhile to compare its results to those of the impact trial that was performed by Poldervaart et al. [27]. Unfortunately, health outcomes could not be compared because a different health outcome was used in the impact trial compared with the DAM (any MACE vs. missed MACE). Furthermore, cost data in the trial were collected over a 3-month time horizon, different from the 6-week time horizon used in literature available before the trial. Still, the impact of noncompliance in our DAM study can be translated to the actual HEART impact trial. The DAM showed that noncompliance without ID in patients with a low

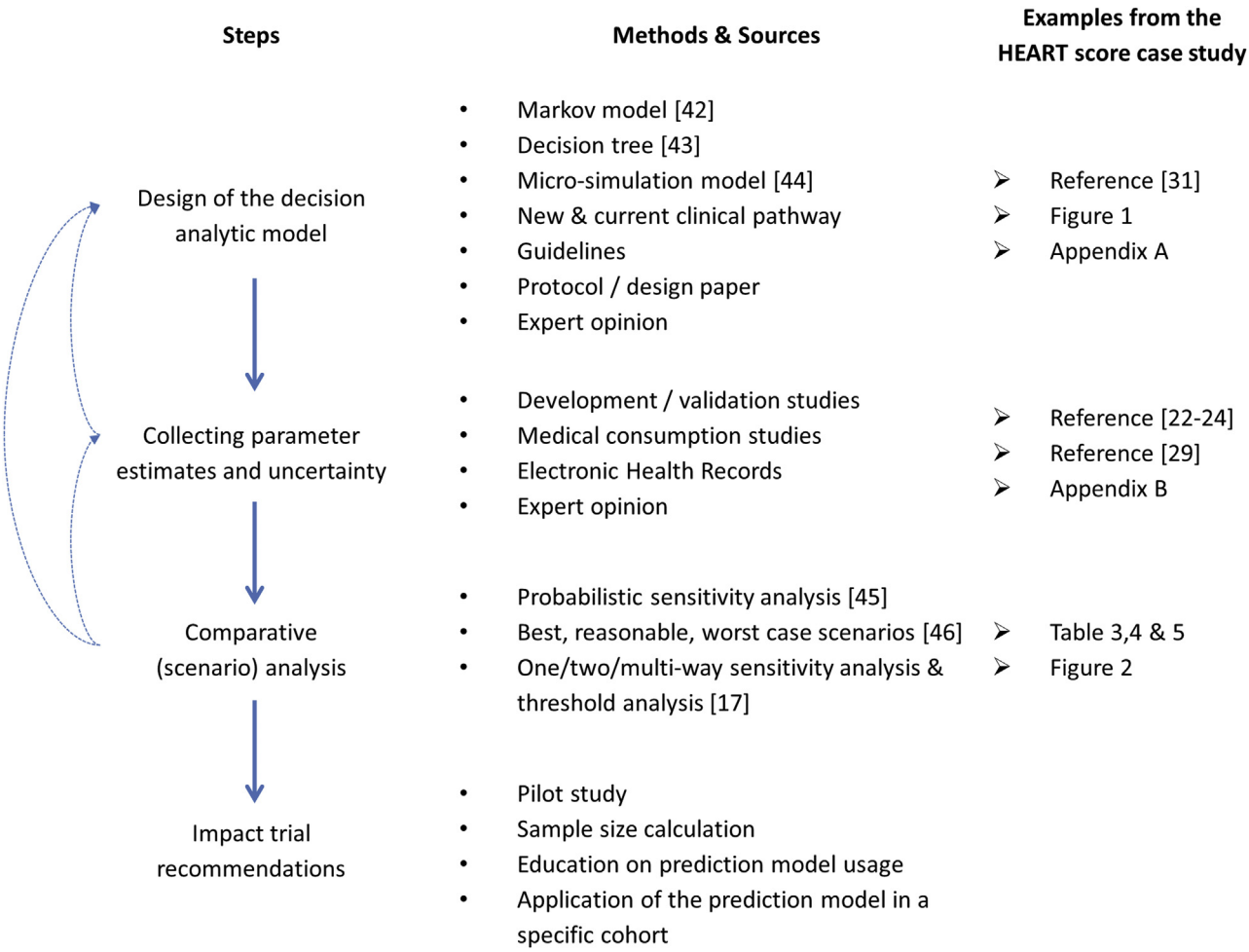

Fig. 3. Guidance for a model-based impact assessment of prediction models, before data on clinical impact have become available [42–46]. Solid arrows mark the logical sequence in which the steps should be taken. Dotted arrows allow researchers to adapt and adjust decisions in previous steps, based on newly available information.

HEART score had a detrimental effect on potential cost savings, which was also the main finding of the HEART impact trial: substantial noncompliance in the low–HEART score category led to small differences in total cost reduction. This information could have been known before the impact trial and hence could have been used to support a more efficient design and conduct by, for example, assessing potential compliance of physicians beforehand using interviews or performing a pilot study.

Few other DAM-based assessments have been previously performed that assess the potential impact of prediction models before an empirical impact study has been executed. One study assessed the value of a prediction model for predicting shoulder pain in patients with early-stage oral cavity squamous cell carcinoma after surgical removal of lymph nodes [48]. Although the analysis did focus on specific scenarios regarding the accuracy of predictions, compliance or additional clinical expertise on top of the prediction model were not evaluated. DAM assessments have also been used for headroom analysis, a method that is used to assess the likelihood of potential cost-effectiveness of an intervention, often at very early stage of development, for a given willingness to pay threshold [19,49–53]. These analyses also make use of data before implementation of an innovation to assess potential benefit. Although a headroom approach is feasible for prediction models, to our knowledge, there are no articles on this topic described in literature.

This is one of the first examples in which a DAM was applied for impact assessment of implementing a prediction model in daily practice, using solely data available before conducting a trial. This method can be applied using data that is commonly available after prediction model development and validation or can be retrieved from (hospital) databases. Compared with a clinical trial, DAM assessments require a fraction of the time and cost and could help improve design and conduct of an impact trial, reducing research waste.

There are a few considerations to fully appreciate the findings of the impact assessment in this article. Use of health care resources in our model was based on the first 6 weeks of medical consumption [29]. It is likely that negative consequences from MACE will last beyond this

timeframe. Markov chain modeling could account for these long-term effects; however, reliable data for these effects were lacking [42]. Out-of-hospital costs, such as general practitioner visits, medication usage, and nonmedical costs (e.g., labor productivity losses, traveling expenses), were not included in the assessment. Although these are likely to influence the incremental costs and health from a societal perspective, the general conclusions will likely be similar.

We viewed the HEART score purely as a diagnostic tool, which implies that using the HEART score cannot prevent MACE. It can only optimize correct stratification of patients and streamline subsequent management. Because MACE is not prevented, the natural outcome is missed MACE, associated with poorer outcome and additional costs. Others have argued that the HEART score can also be used to predict future MACE, opening the opportunity to prevent it. This would of course lead to a rather different DAM. We chose not to do this because HEART was designed for use in an acute care setting, where patients present with chest pain, which is clearly a diagnostic setting.

A DAM is ideal for assessing the expected impact of using a prediction model in clinical practice on patient health outcomes and health care costs, using solely data available before conducting an empirical long-term randomized impact study. With the results of such DAMs, one can decide whether an empirical impact trial is still deemed necessary, and if so, under what conditions such prediction model is likely to show favorable results. Efforts can then be directed at improving the use of the prediction model by clinicians and on improving the trial design. In general, DAMs can provide insight in the mechanism through which a prediction model and its risk-based management recommendations can lead to desired results and expose potential flaws in mechanistic pathways, allowing researchers to adapt the design of an empirical trial beforehand. Ultimately, model-based impact assessments have the potential to reduce research waste, by more efficient selection of prediction models in which an empirical impact trial is warranted.

CRediT authorship contribution statement

Kevin Jenniskens: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Visualization, Writing - original draft, Writing - review & editing. **Ghizelda R. Lagerweij:** Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing - original draft, Writing - review & editing. **Christiana A. Naaktgeboren:** Conceptualization, Investigation, Supervision, Writing - review & editing. **Lotty Hooft:** Writing - review & editing. **Karel G.M. Moons:** Conceptualization, Supervision, Writing - review & editing. **Judith M. Poldervaart:** Resources, Validation, Writing - review & editing. **Hendrik Koffijberg:** Supervision, Validation, Writing - review & editing. **Johannes B. Reitsma:** Conceptualization, Resources, Supervision, Validation, Writing - review & editing.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2019.07.010.

References

[1] Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol 2016;69:245−7.
[2] Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med 2013;10(2):e1001381.
[3] Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research: a clinical example. J Clin Epidemiol 2003;56:826−32.
[4] Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. BMJ 2009;338:b605.
[5] Dharmarajah B, Thapar A, Salem J, Lane TR, Leen EL, Davies AH. Impact of risk scoring on decision-making in symptomatic moderate carotid atherosclerosis. Br J Surg 2014;101(5):475−80.
[6] van Montfort P, Willemse JP, Dirksen CD, van Dooren IM, Meertens LJ, Spaanderman ME, et al. Implementation and effects of risk-dependent obstetric care in The Netherlands (expect study II): protocol for an impact study. JMIR Res Protoc 2018;7(5):e10066.
[7] Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ 2009;338:b606.
[8] Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart 2012;98:691−8.
[9] Kappen TH, van Klei WA, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, Moons KGM. Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. Diagn Progn Res 2018;2(1):11.
[10] Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1−73.
[11] Ferrante di Ruffano L, Davenport C, Eisinga A, Hyde C, Deeks JJ. A capture-recapture analysis demonstrated that randomized controlled trials evaluating the impact of diagnostic tests on patient outcomes are rare. J Clin Epidemiol 2012;65:282−7.
[12] van Giessen A, Peters J, Wilcher B, Hyde C, Moons C, de Wit A, et al. Systematic review of health economic impact evaluations of risk prediction models: stop developing, start evaluating. Value Health 2017;20(4):718−26.
[13] Wallace E, Uijen MJ, Clyne B, Zarabzadeh A, Keogh C, Galvin R, et al. Impact analysis studies of clinical prediction rules relevant to primary care: a systematic review. BMJ Open 2016;6(3):e009957.
[14] Kappen TH, Moons KG, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, van Klei WA. Impact of risk assessments on prophylactic antiemetic prescription and the incidence of postoperative nausea and vomiting: a cluster-randomized trial. Anesthesiology 2014;120:343−54.
[15] Huang DT, Yealy DM, Filbin MR, Brown AM, Chang CH, Doi Y, et al. Procalcitonin-guided use of antibiotics for lower respiratory tract infection. N Engl J Med 2018;379:236−49.
[16] Snooks H, Bailey-Jones K, Burge-Jones D, Dale J, Davies J, Evans B, et al. Predictive risk stratification model: a randomised stepped-

wedge trial in primary care (PRISMATIC). Health Serv Deliv Res 2018;6.
[17] Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddart G. Methods for the economic evaluation of health care programme. 3rd ed. Oxford: Oxford University Press; 2005.
[18] IJzerman M, Koffijberg H, Fenwick E, Krahn M. Emerging use of early health technology assessment in medical product development: a scoping review of the literature. Pharmacoeconomics 2017;35:727−40.
[19] McAteer H, Cosh E, Freeman G, Pandit A, Wood P, Lilford R. Cost-effectiveness analysis at the development phase of a potential health technology: examples based on tissue engineering of bladder and urethra. J Tissue Eng Regen Med 2007;1(5):343−9.
[20] Greving JP, Buskens E, Koffijberg H, Algra A. Cost-effectiveness of aspirin treatment in the primary prevention of cardiovascular disease events in subgroups based on age, gender, and varying cardiovascular risk. Circulation 2008;117:2875−83.
[21] Miquel-Cases A, Steuten LM, Retel VP, van Harten WH. Early stage cost-effectiveness analysis of a BRCA1-like test to detect triple negative breast cancers responsive to high dose alkylating chemotherapy. Breast 2015;24(4):397−405.
[22] Six AJ, Backus BE, Kelder JC. Chest pain in the emergency room: value of the HEART score. Neth Heart J 2008;16(6):191−6.
[23] Backus BE, Six AJ, Kelder JC, Bosschaert MA, Mast EG, Mosterd A, et al. A prospective validation of the HEART score for chest pain patients at the emergency department. Int J Cardiol 2013;168:2153−8.
[24] Backus BE, Six AJ, Kelder JC, Mast TP, van den Akker F, Mast EG, et al. Chest pain in the emergency room: a multicenter validation of the HEART Score. Crit Pathw Cardiol 2010;9(3):164−9.
[25] Streitz MJ, Oliver JJ, Hyams JM, Wood RM, Maksimenko YM, Long B, et al. A retrospective external validation study of the HEART score among patients presenting to the emergency department with chest pain. Intern Emerg Med 2018;13(5):727−48.
[26] Oliver JJ, Streitz MJ, Hyams JM, Wood RM, Maksimenko YM, Long B, et al. An external validation of the HEART pathway among emergency department patients with chest pain. Intern Emerg Med 2018;13:1249−55.
[27] Poldervaart JM, Reitsma JB, Backus BE, Koffijberg H, Veldkamp RF, Ten Haaf ME, et al. Effect of using the HEART score in patients with chest pain in the emergency department: a stepped-wedge, cluster randomized trial. Ann Intern Med 2017;166:689−97.
[28] Kappen TH, van Loon K, Kappen MA, van Wolfswinkel L, Vergouwe Y, van Klei WA, et al. Barriers and facilitators perceived by physicians when using prediction models in practice. J Clin Epidemiol 2016;70:136−45.
[29] Nieuwets A, Poldervaart JM, Reitsma JB, Buitendijk S, Six AJ, Backus BE, et al. Medical consumption compared for TIMI and HEART score in chest pain patients at the emergency department: a retrospective cost analysis. BMJ Open 2016;6(6):e010694.
[30] Six AJ, Cullen L, Backus BE, Greenslade J, Parsonage W, Aldous S, et al. The HEART score for the assessment of patients with chest pain in the emergency department: a multinational validation study. Crit Pathw Cardiol 2013;12(3):121−6.
[31] Poldervaart JM, Reitsma JB, Koffijberg H, Backus BE, Six AJ, Doevendans PA, et al. The impact of the HEART risk score in the early assessment of patients with acute chest pain: design of a stepped wedge, cluster randomised trial. BMC Cardiovasc Disord 2013;13:77.
[32] Soekhlal RR, Burgers LT, Redekop WK, Tan SS. Treatment costs of acute myocardial infarction in The Netherlands. Neth Heart J 2013;21(5):230−5.
[33] Bekelman JE, Halpern SD, Blankart CR, Bynum JP, Cohen J, Fowler R, et al. Comparison of site of death, health care utilization, and hospital expenditures for patients dying with cancer in 7 developed countries. JAMA 2016;315:272−83.
[34] Mahoney EM, Wang K, Cohen DJ, Hirsch AT, Alberts MJ, Eagle K, et al. One-year costs in patients with a history of or at risk for atherothrombosis in the United States. Circ Cardiovasc Qual Outcomes 2008;1(1):38−45.
[35] Osnabrugge RL, Magnuson EA, Serruys PW, Campos CM, Wang K, van Klaveren D, et al. Cost-effectiveness of percutaneous coronary intervention versus bypass surgery from a Dutch perspective. Heart 2015;101:1980−8.
[36] Baim D, Grossman W. Complications of cardiac catheterization. Baltimore, MD: Williams & Wilkins; 1996.
[37] Johnson LW, Lozner EC, Johnson S, Krone R, Pichard AD, Vetrovec GW, et al. Coronary arteriography 1984-1987: a report of the registry of the society for cardiac angiography and interventions. I. Results and complications. Cathet Cardiovasc Diagn 1989;17(1):5−10.
[38] Wyman RM, Safian RD, Portway V, Skillman JJ, McKay RG, Baim DS. Current complications of diagnostic and therapeutic cardiac catheterization. J Am Coll Cardiol 1988;12:1400−6.
[39] van der Zalm P. HEART-score biedt arts extra handvat. Zorgvisie. 2016. Available at https://www.zorgvisie.nl/heart-score-biedt-arts-extra-handvat/. Accessed October 31, 2018.
[40] Lilford RJ, Chilton PJ, Hemming K, Girling AJ, Taylor CA, Barach P. Evaluating policy and service interventions: framework to guide selection and interpretation of study end points. BMJ 2010;341:c4413.
[41] Tuffaha HW, Gordon LG, Scuffham PA. Value of information analysis in healthcare: a review of principles and applications. J Med Econ 2014;17(6):377−83.
[42] Briggs A, Sculpher M. An introduction to Markov modelling for economic evaluation. Pharmacoeconomics 1998;13:397−409.
[43] Ryder HF, McDonough C, Tosteson AN, Lurie JD. Decision analysis and cost-effectiveness analysis. Semin Spine Surg 2009;21(4):216−22.
[44] Zucchelli E, Jones AM, Rice N. The evaluation of health policies through dynamic microsimulation methods. Int J Microsimul 2012;5(1):2−20.
[45] Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health Econ 2005;14:339−47.
[46] Briggs AH, Gray AM. Handling uncertainty when performing economic evaluation of healthcare interventions. Health Technol Assess 1999;3:1−134.
[47] ISPOR. Synthesizing health care evidence - modelling/decision analysis methods. 2018. [20-12-2018]; Available from https://www.ispor.org/heor-resources/more-heor-resources/outcomes-research-guidelines-index. Accessed December 20, 2018.
[48] Govers TM, Rovers MM, Brands MT, Dronkers EAC, Baatenburg de Jong RJ, Merkx MAW, et al. Integrated prediction and decision models are valuable in informing personalized decision making. J Clin Epidemiol 2018;104:73−83.
[49] Girling A, Lilford R, Cole A, Young T. Headroom approach to device development: current and future directions. Int J Technol Assess Health Care 2015;31(5):331−8.
[50] Girling A, Young T, Brown C, Lilford R. Early-stage valuation of medical devices: the role of developmental uncertainty. Value Health 2010;13(5):585−91.
[51] Van Nimwegen KJ. Feasibility of the headroom analysis in early economic evaluation of innovative diagnostic technologies with no immediate treatment implications. Value Health 2014;17(7):A550.
[52] Cosh E, Girling A, Lilford R, McAteer H, Young T. Investing in new medical technologies: a decision framework. J Commer Biotechnol 2007;13(4):263−71.
[53] Miquel-Cases A, Retel VP, Lederer B, von Minckwitz G, Steuten LM, van Harten WH. Exploratory cost-effectiveness analysis of response-guided neoadjuvant chemotherapy for hormone positive breast cancer patients. PLoS One 2016;11:e0154386.