ScienceDirect

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

Economic Evaluation

Using Metamodeling to Identify the Optimal Strategy for Colorectal Cancer Screening

Hendrik Koffijberg, PhD, Koen Degeling, PhD, Maarten J. IJzerman, PhD, Veerle M.H. Coupé, PhD, Marjolein J.E. Greuter, PhD

ABSTRACT

Objectives: Metamodeling can address computational challenges within decision-analytic modeling studies evaluating many strategies. This article illustrates the value of metamodeling for evaluating colorectal cancer screening strategies while accounting for colonoscopy capacity constraints.

Methods: In a traditional approach, the best screening strategy was identified from a limited subset of strategies evaluated with the validated Adenoma and Serrated pathway to Colorectal CAncer model. In a metamodeling approach, metamodels were fitted to this limited subset to evaluate all potentially plausible strategies and determine the best overall screening strategy. Approaches were compared based on the best screening strategy in life-years gained compared with no screening. Metamodel runtime and accuracy was assessed.

Results: The metamodeling approach evaluated >40 000 strategies in <1 minute with high accuracy after 1 adaptive sampling step (mean absolute error: 0.0002 life-years) using 300 samples in total (generation time: 8 days). Findings indicated that health outcomes could be improved without requiring additional colonoscopy capacity. Obtaining similar insights using the traditional approach could require at least 1000 samples (generation time: 28 days). Suggested benefits from screening at ages <40 years require adequate validation of the underlying Adenoma and Serrated pathway to Colorectal CAncer model before making policy recommendations.

Conclusions: Metamodeling allows rapid assessment of a vast set of strategies, which may lead to identification of more favorable strategies compared to a traditional approach. Nevertheless, metamodel validation and identifying extrapolation beyond the support of the original decision-analytic model are critical to the interpretation of results. The screening strategies identified with metamodeling support ongoing discussions on decreasing the starting age of colorectal cancer screening.

Keywords: colorectal cancer, screening, simulation, metamodel, optimization.

VALUE HEALTH. 2021; 24(2):206–215

Introduction

Health policy recommendations regarding the implementation of screening or reimbursement of diagnostic tests or treatments for a specific condition follow from a thorough assessment of its expected impact in patient outcomes, healthcare outcomes, and costs. Such analyses ensure effective and efficient care to be delivered, and thereby support optimal use of healthcare resources. Although impact assessments traditionally have been performed based on randomized controlled trials (RCTs), such trials are not always feasible given their typically high cost, short duration (ie, inability to observe long-term outcomes), and limited number of considered strategies (ie, study arms).

In particular, when considering multiple diagnostic or screening strategies involving combinations of tests, procedures, and their timing, a large number of different strategies can be distinguished. Ideally, all relevant strategies would be compared to determine the best strategy. Nevertheless, this is clearly not feasible in RCTs, as demonstrated by the rarity of even simple test + treatment RCTs.[1] Therefore, in silico decision-analytic impact studies are increasingly applied as a cheaper and more efficient method to evaluate strategies, and also to guide the design of RCTs.[2,3] Such modeling studies can easily incorporate multiple screening or diagnostic strategies for comparison. In doing so, in silico models should adequately represent clinical practice and disease progression. These models should also properly reflect the (indirect) impact of test outcomes on treatment decisions and resulting health outcomes and costs, for incorrect test results, for different test cutoff values, for patient stratification based on test outcomes, and other relevant aspects.[4] To reflect these aspects, patient-level simulation models (ie, microsimulations) are typically required, which may be more computationally demanding compared to conventional cohort-level models.[5]

Even with computationally demanding simulation models, performing traditional model analyses, including a probabilistic

analysis to reflect uncertainty in the modeling outcomes,[6] is often feasible within acceptable time frames. Performing more advanced analyses like Value of Information analysis, for example, or applying optimization approaches with such models, however, may not be feasible unless simulation code is parallelized and run on high performance computing clusters (see Appendix A in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.08.2099 for a numerical example). Alternatively, approximation methods can sometimes be applied to estimate outcomes of interest within a feasible time frame, for example, in the context of Value of Information analysis.[7,8] Furthermore, specific algorithms and operation research methods can be directly applied to mathematical models to optimize health services accounting for constraints (eg, in healthcare system characteristics or budgets).[9,10]

For screening models, which are often complex microsimulation models, it may prove computationally challenging to evaluate a large number of strategies to identify the best screening strategy for a specific condition. This explains why typically only a limited number of screening strategies are evaluated, for example, between 2 and 25 strategies, and compared with a reference strategy.[11,12] Although limiting the number of strategies to be evaluated improves the feasibility of the analysis, a direct consequence is that the best screening strategy may not be included in the set of evaluated strategies and, hence, will *not* be identified. Given that typically tens of thousands to millions of strategies can be identified (see Appendix B in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.08.2099 for a numerical example), the likelihood that the best screening strategy is actually identified may be very small.

Estimating the impact of all potential strategies or applying optimization approaches may become feasible, however, when approximating rather than evaluating the outcomes of a computationally demanding simulation model. Metamodels, also known as surrogate models or emulators, can provide such approximations of (complex) simulation models, and can be used to perform complex analyses, instead of using the original simulation model.[13,14] Metamodels have been used in different research contexts, but their application in health economics is still limited.[15]

In this article, metamodeling methods are used to address computational challenges associated with a complex simulation model, to enable optimization in a case study on colorectal cancer (CRC) screening in The Netherlands, while accounting for colonoscopy capacity constraints.

Methods

Case Study on Colorectal Cancer Screening

Colorectal cancer accounts for 1.8 million cases and over 850 000 deaths worldwide, making it a major public health issue.[16] Colorectal cancer originates from colorectal polyps, that is, adenomas and serrated lesions. These benign precursor lesions and the long preclinical phase make CRC an excellent target for screening. Indeed, several long-term RCTs have shown that CRC screening can considerably decrease CRC mortality.[17,18] As the preventive impact of CRC screening is widely recognized, CRC screening is now implemented in almost 60 countries.[19] These screening efforts vary widely in, among others, the type of screening program (ie, organized population-based screening versus opportunistic screening), type of test used, the age ranges at which screening is offered, and the screening interval. This diversity is mainly owing to differences in available healthcare resources, with colonoscopy capacity being the most relevant constraint. In The Netherlands, a population-based CRC screening program was implemented in 2014.[20] This program consists of biennial fecal immunochemical testing (FIT) in individuals aged 55 to 75 years. Individuals with a positive FIT result are referred to get a colonoscopy during which all detected colorectal polyps are removed. This program was defined based on a maximum capacity of 550 colonoscopies over the lifetime of 1000 patients, and is expected to reduce CRC mortality with at least 40% compared to no screening in 2040.[12]

The original ASCCA simulation model

The Adenoma and Serrated pathway to Colorectal CAncer (ASCCA) simulation model was designed and implemented to reflect the development of CRC. The ASCCA model is calibrated to Dutch adenoma and serrated polyp prevalence rates as well as Dutch CRC incidence and mortality rates,[21] and uses screening adherence rates based on achieved adherence rates in The Netherlands, as reported by the national monitor.[20] Details of the ASCCA model are presented in Appendix C (see Appendix C in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.08.2099), and extensive model descriptions have previously been published.[22] ASCCA has been used, among others, to evaluate the long-term impact of the Dutch national colorectal screening program in cancer incidence, mortality, and colonoscopy capacity requirement,[12] and to assess the effectiveness and cost-effectiveness of colonoscopy surveillance in FIT-based screening.[23] Although the ASCCA model is programmed in C++, one of the fastest general programming environments,[24] a single simulation run to evaluate 1 strategy takes about 40 minutes on a modern desktop computer. This computational burden is owing to the fact that (1) 10 million men and 10 million women need to be simulated to obtain stable results, and (2) previous screening and surveillance results of individuals are stored and used in the model to simulate future events. Despite the model's valuable ability to evaluate different screening strategies, its computational cost makes it infeasible to identify the optimal screening strategy by evaluating a comprehensive set of thousands or more screening strategies within a reasonable time frame (eg, a few days).

Approaches to optimize screening strategies

Potential FIT-based screening strategies were defined based on a limited set of 4 parameters: start age (in years, range 30-90), screening interval (in years, range 1-60), number of screening rounds (range 1-31), and FIT cutoff for referral (discrete values: 50, 75, 100, or 150 ng/mL). When limiting the maximum age of screening to 90 years, these combined ranges lead to a set comprising 40 864 unique screening strategies (see Appendix B in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.08.2099 for calculation). In all strategies, overall adherence to FIT screening was set at 73%. Screen-positive individuals are referred to diagnostic colonoscopy, for which compliance was set at 92%. In individuals referred to the surveillance program based on findings at the diagnostic colonoscopy, compliance was assumed to be 92% as well.

We compared 2 approaches to identify the best screening strategy from all potential strategies, that is, the traditional and the metamodeling approach. Both approaches were used to identify a best strategy, compared to no screening, in life-years gained (LYG) and a best strategy in net monetary benefit (NMB), satisfying a predefined maximum colonoscopy capacity constraint. For the NMB a willingness-to-pay of €20 000 per life-year gained

was used. Colonoscopy capacity was defined as the number of colonoscopies available in the lifetime of 1000 patients.

In the *traditional approach,* the optimal strategy was determined by evaluating a fixed set of strategies using the original ASCCA model and selecting the best strategy in LYG or NMB that does not exceed the maximum colonoscopy capacity.

In the *metamodeling approach,* the exact same set of strategies evaluated in the traditional approach was used to fit Gaussian process metamodels. Gaussian process regression is a nonparametric regression method also known as kriging, which uses information on neighbor observations for new predictions.[13,25] This metamodeling technique was selected because it is particularly suited for scenarios involving low number of observations and input parameters, because the computational burden, both in fitting and predicting, increases dramatically with increasing numbers of observations and parameters. Based on the screening parameters defined earlier, metamodels were fitted to estimate (1) the number of LYG (primary outcome), (2) incremental costs, (3) NMB of a screening strategy compared to no screening, and (4) the number of colonoscopies required for a screening strategy. These metamodels were then used to evaluate all possible screening strategies to determine the overall best strategy in predicted LYG or NMB that meets the colonoscopy capacity constraint. The main benefit of the metamodeling approach is that it enables a large number of screening strategies (ie, search space) to be explored, rather than the limited set of strategies that can be evaluated using the traditional approach, and thereby may identify a better best screening strategy. As the metamodeling approach can be considered an approximation to the ASCCA model, a feedback step (or iterative loop) to this original model was implemented to check metamodel performance. Finally, because searching through all possible screening strategies might result in the evaluation of strategies beyond the boundaries of the evidence supporting the ASCCA model, we checked to what extent the identified best strategies were based on extrapolation.

Analysis and outcomes of the approaches

We performed 3 steps to illustrate the use and advantages of the metamodeling approach, and assessed how its performance varies with the size of the fixed set of strategies evaluated with the ASCCA model. An overview of all steps is provided in Figure 1. In the first step, the potential benefit of optimization using a metamodel in LYG (ie, ignoring the NMB and costs) across a range of colonoscopy capacity constraints was illustrated. Hence, the optimization problem addressed was to maximize the LYG by screening for CRC given a colonoscopy capacity of 550 in the lifetime of 1000 patients. The process was as follows:

1.1. Different sets of 25, 50, 100, and 150 screening strategies that each cover the search space as much as possible were generated by sampling according to a Latin hypercube design. A constraint to the sampling was applied to ensure only strategies that do not exceed the maximum age of screening of 90 years were included. A Latin hypercube design was selected, because it has been used often for designing computer experiments following its ability to efficiently cover the full parameter space.[26,27]
1.2. Each of the 4 sets of screening strategies was evaluated with the ASCCA model to obtain their corresponding outcomes in LYG and number of colonoscopies required for each screening strategy. For the traditional approach, the best strategy given a certain colonoscopy capacity was identified within these (limited) evaluated sets.

The traditional approach stopped here—metamodeling approach continued.

1.3. Metamodels for the LYG and number of colonoscopies required were fitted on each set of screening strategies separately, resulting in 4 groups of 2 metamodels (1 group for each set of 25, 50, 100 and 150 screening strategies).
1.4. For each group of metamodels separately, the metamodels were used to estimate the LYG and number of colonoscopies required for all 40 864 potential screening strategies.
1.5. For each group of metamodels separately, the best outcomes in LYG from the traditional and metamodeling approach were presented in an optimization frontier for different colonoscopy capacities ranging from 25 to 900 (in steps of 25).

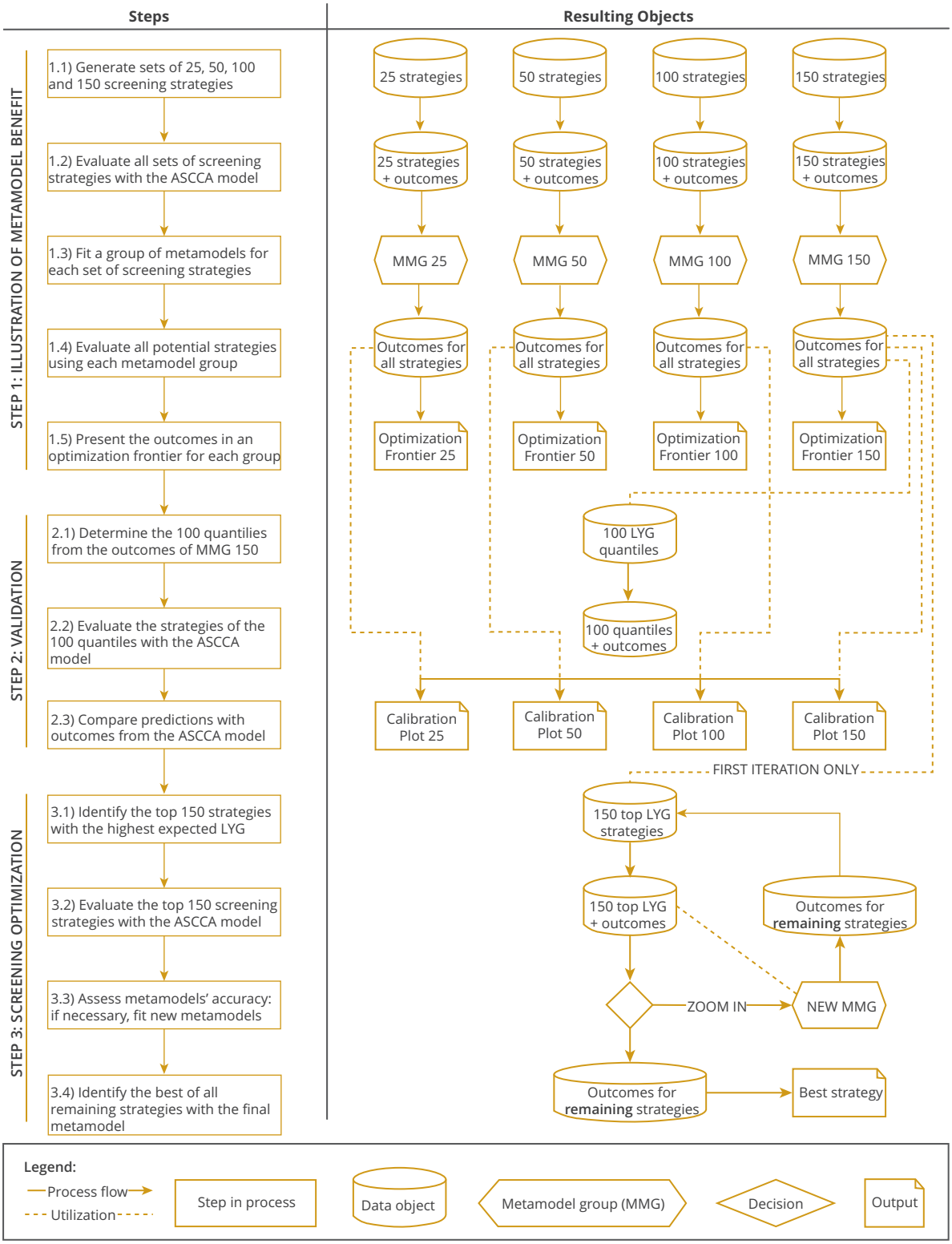
The second step investigated how accurate, and therefore meaningful, outcomes of the LYG metamodels were, to illustrate the validation process:

2.1. Based on the outcomes of the 40 864 unique screening strategies estimated in step 1.4 and using the metamodel for LYG fitted to 150 experiments in step 1.3, the 100 strategies with outcomes corresponding to the 100 quantiles of LYG were identified. Although LYG outcomes predicted by one of the metamodels fitted to other (smaller) sets of strategies could also have been used, the metamodel fitted to the largest number of strategies (ie, 150) was expected to perform best and, hence, the outcomes as predicted by that metamodel were used to identify the quantiles.
2.2. The set of 100 strategies corresponding to the identified quantiles were evaluated with the ASCCA model to obtain the 'true' outcomes.
2.3. For all LYG metamodels fitted to the differently sized sets of strategies (step 1.3), outcomes predicted by the LYG metamodels were compared with those from the ASCCA model in calibration plots, and error values were calculated, to assess the accuracy of the metamodels.

The third step concerned the identification of the CRC screening strategy expected to be optimal for The Netherlands. For this purpose, we used the group of metamodels fitted based on the largest set of 150 experiments, because these metamodels were most accurate based on the mean (absolute) relative error.

3.1. Based on the outcomes of all (remaining) screening strategies predicted in step 1.4 using the group of metamodels fitted to 150 experiments in step 1.3, the top 150 strategies with the highest predicted LYG were identified, across all strategies complying with the capacity constraint. As will be discussed below, outcomes of all screening strategies from step 1.4 are used for the first iteration of this step 3.1.
3.2. These identified top 150 strategies were evaluated with the ASCCA model to obtain their true outcomes.
3.3. The outcomes predicted by the metamodels were compared to those from the ASCCA model to check for discrepancies. If necessary, new metamodels were fitted on the ASCCA outcomes of the top 150 strategies and used to repeat steps 3.1 and 3.2. This was an adaptive sampling step, to zoom in on the region with strategies of interest and, thereby, obtain more accurate metamodel predictions in that region. In our illustration, not aimed to directly inform policy decisions, a mean absolute error of 0.01 in LYG was assumed to be acceptable; larger discrepancies would require zooming in and fitting new metamodels. Hence, when repeating step 3.1 based on the outcomes of all potential screening strategies as predicted by the newly developed metamodels, the strategies considered were limited to those

Figure 1. Overview of the steps performed and resulting objects.



covered by the parameter ranges as observed in the top 150 strategies used to fit the new metamodels.

3.4. When the metamodels were considered sufficiently accurate based on the calibration plot and error measures, they were used to evaluate all remaining strategies (ie, those covered by the parameter ranges on which the metamodels were fitted) and, thereby, identify the best strategy given a colonoscopy capacity constraint.

Step 3 was also performed with NMB rather than LYG as primary outcome to address the optimization problem of maximizing the NMB of screening for CRC given a colonoscopy capacity of 550 in the lifetime of 1000 patients. An overview of the analysis, conforming to the The Professional Society for Health Economics and Outcomes Research guidance on optimization methods,[10] is provided in Appendix E in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.08.2099.

Results

Figure 2 shows the LYG optimization frontiers as a function of the maximum colonoscopy capacity, for each group of metamodels fitted based on screening strategy sets of different sizes (step 1). Regardless of the number of included strategies, it is clear that most strategies sampled for the *traditional approach* according to a Latin hypercube design require limited colonoscopy capacity and yield only small health benefits. Most strategies are located in the lower left corner, yielding between 0 and 0.03 LYG per individual and requiring less than 250 colonoscopies lifelong per 1000 individuals.

The potential benefit of using the metamodeling approach is visible as the area between the metamodel frontier and the traditional frontier. The latter represents experiments corresponding to the best outcome as a function of the number of colonoscopies. The benefit is observed across the entire range of maximum colonoscopy capacity but varies in size, being highest in the region where few strategies have been evaluated (ie, top right). As expected, the benefit of using metamodels is smallest for a set of 150 strategies. When the number of strategies evaluated with the ASCCA model and used to fit metamodels increases, the best strategies from both approaches will converge and, hence, the added value of the metamodeling approach becomes smaller.

Looking only at the difference between the frontiers, however, would ignore that the metamodels are approximations of the ASCCA model and that predicted outcomes from the metamodels may not accurately reflect outcomes of the ASCCA model. Figure 3 shows the results of assessing the accuracy of the LYG metamodel (step 2): calibration plots for the different number of screening strategies used to fit metamodels. The corresponding error values are presented in Table 1. Clearly, calibration is relatively good for low health benefits and relatively poor for high health benefits, owing to the aforementioned distribution of strategies in the lower-left corners of the plots in Figure 2. Although calibration improves with an increasing number of strategies considered, calibration for large health benefit values is still poor, even when considering 150 strategies.

To identify the optimal screening strategy (step 3), the performance of the LYG metamodel developed based on 150 strategies (step 1.3) regarding the 150 best strategies (step 1.4) was assessed. The mean absolute error of 0.0245 and mean relative absolute error of 30.7% were deemed too large to accurately identify the best strategy with the metamodel. Hence, an adaptive sampling step to zoom in was performed: a new Gaussian process metamodel was fitted on the ASCCA outcomes of these top 150 strategies. Figure 4A shows the calibration plot for the new metamodel for the 150 best strategies according this new LYG metamodel (step 3.3, second iteration). Following this adaptive sampling step, performance of this new metamodel was considered sufficient (mean absolute error 0.0002; mean relative absolute error 0.2%).

The new metamodel was used to evaluate all scenarios within the parameter ranges as covered in the top 150 strategies identified in step 3.1 (first iteration), because these were used to fit the new metamodel in step 3.3. These parameter ranges after zooming in were start age (range 30-43), screening interval (range 1-3), number of screening rounds (range 18-21), and FIT cutoff values (50, 75, 100, and 150 ng/mL), resulting in 536 remaining feasible strategies. The optimization frontier of these strategies is presented in Figure 4B.

All possible screening strategies (ie, 40 864 strategies) could be evaluated with the metamodels in less than 1 minute, based on 150 + 150 generated samples (generation time ~8 days). Conversely, the traditional approach required ~4 days of computation time to evaluate just 150 (predefined) screening strategies, and is expected to require generation of at least 1000 samples (generation time ~28 days) to provide similar insights.

Table 2 shows that for a maximum of 550 colonoscopies lifelong per 1000 individuals, which reflects current colonoscopy capacity, the optimal screening strategy in LYG yields an expected 0.092 LYG per individual and starts at age 33 years, and has a screening interval of 2 years, 21 screening rounds, and a FIT cutoff of 150 ng/mL. In comparison, the best screening strategy identified by the traditional approach yields 0.081 LYG and starts screening at age 32, and has an interval of 3 years and 14 screening rounds and also uses FIT cutoff of 150 ng/mL. The definition and outcomes for the top 10 strategies according to the metamodel are shown in Table 3. Clearly, both strategy definitions and outcomes are quite similar for these top 10 strategies. When colonoscopy capacity would be substantially lower, the screening interval decreases to 1 year (other aspects remain comparable). Conversely, when colonoscopy capacity would be substantially higher, the FIT cutoff decreases to 75 ng/mL (other aspects remains comparable) (Table 2). Results found when optimizing the NMB rather than LYG did not provide additional insights, because the impact of costs is limited compared to the impact of LYG (see Appendix D in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.08.2099).

These optimal strategies, however, are based on extrapolations beyond the scope of the original ASCCA model. Colorectal polyp incidence rates in the ASCCA model were calibrated against polyp prevalence rates as reported by a Dutch study in which 1420 screening-naïve individuals aged 50 to 75 years underwent colonoscopy.[28] For the ages 20 to 50 and 75 to 90, prevalence rates reported by Rutter et al[29] were used. As the confidence intervals in this latter study were extremely wide, there is considerable uncertainty regarding the ASCCA model predictions for screening at ages below 50 and over 75 years. Consequently, even though the metamodel is an accurate approximation of the ASCCA model, the ASCCA model itself requires validation for younger ages before its results can be used for policy recommendations concerning a decrease in the starting age of screening.

Discussion

The computational burden of using complex, patient-level simulation models for health economic analyses can be problematic. In this case study, metamodeling techniques were used to negate computational challenges with the original ASCCA model,

Figure 2. Optimization frontier for the LYG as function of maximum colonoscopy capacity. LYG, life-years gained.

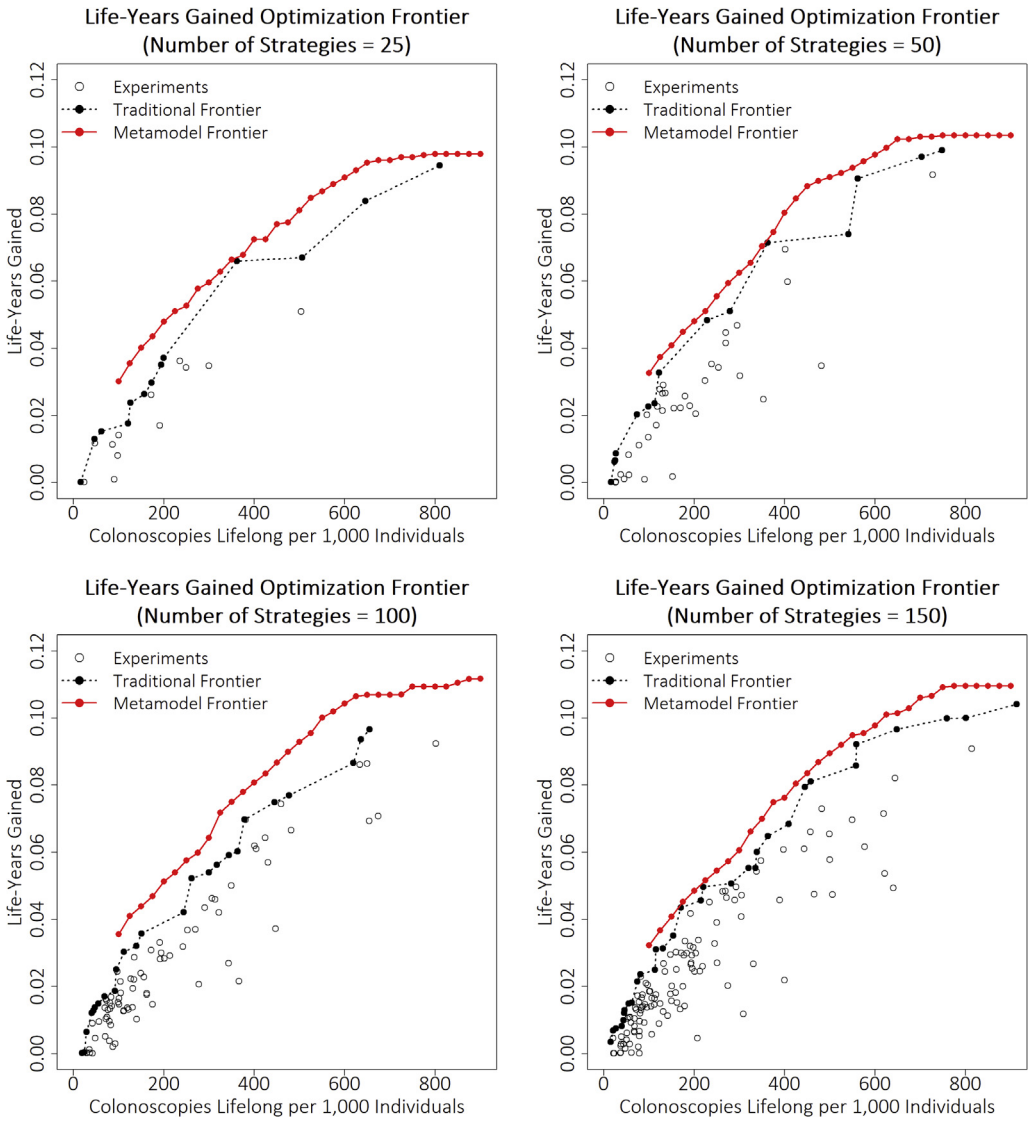


Table 1. Mean absolute and mean relative absolute errors for the LYG metamodels.

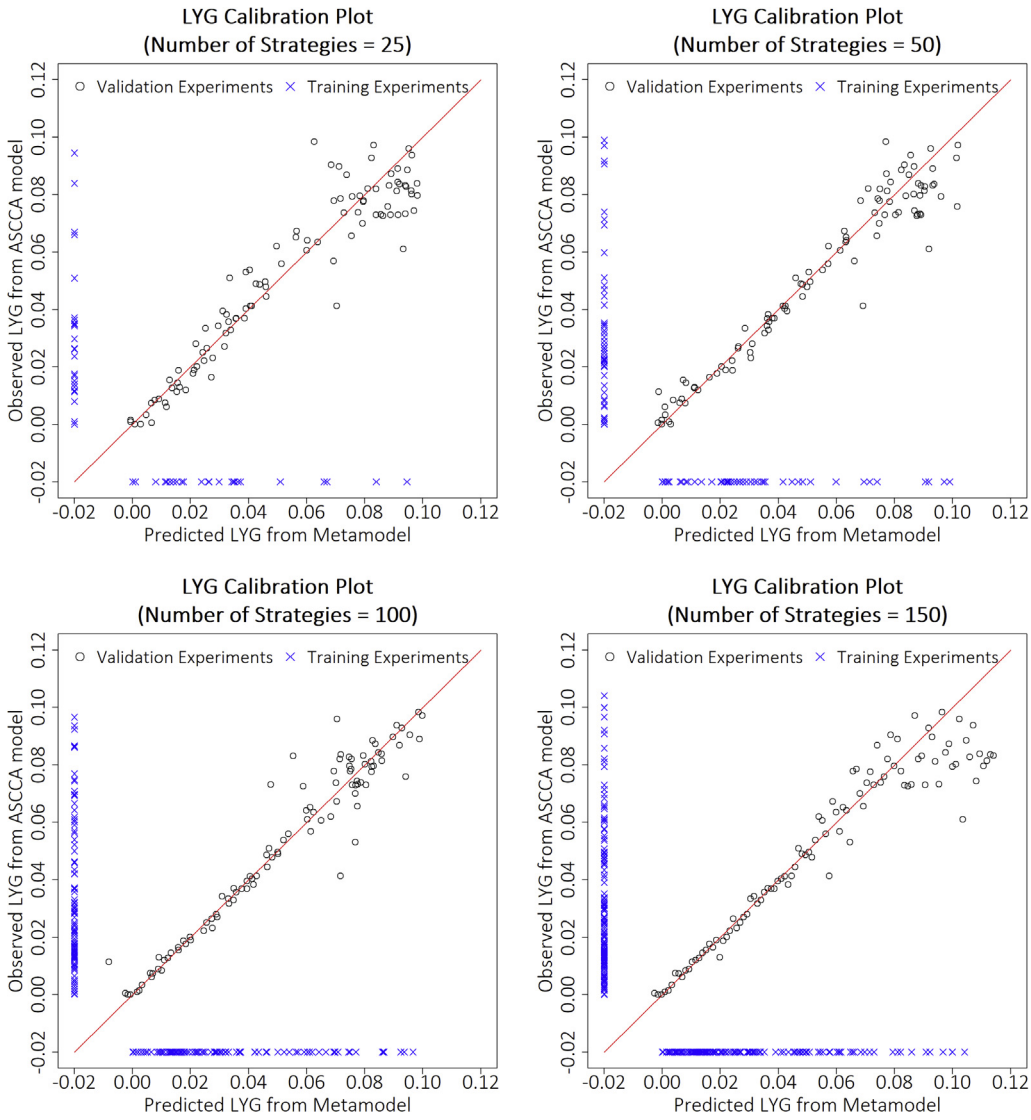
	100 quantiles (step 2.3)			
	ME	MAE	MRE (%)	MRAE (%)
MM25	0.0014	0.0072	84.8	101.2
MM50	0.0018	0.0055	48.3	78.9
MM100	−0.0002	0.0044	−53.9	64.5
MM150	0.0039	0.0068	−37.2	55.0

LYG indicates life-years gained; MAE mean absolute error; ME, mean error; MRAE, mean relative absolute error; MRE, mean relative error.

which allowed for identification of the optimal screening strategy for The Netherlands given a maximum colonoscopy capacity.

In our metamodeling application, the feedback loop with the ASCCA model to ensure accuracy of the predicted outcomes was crucial, because the validation showed an adaptive sampling step was necessary to improve metamodel performance. Clearly, proper validation should be part of every metamodeling study, which is straightforward when access to the original model is available. Without metamodeling, only a potential set of dozens or hundreds of strategies could have been evaluated with the ASCCA model. It is highly unlikely that the optimal screening would have been identified by such a traditional approach. The additional insight gained from evaluating a much more comprehensive set of strategies, within a single minute, is a clear advantage of

Figure 3. Calibration plots for the LYG metamodel. LYG, life-years gained.



metamodeling, and enables better-informed decision making. Although many strategies in this comprehensive set ultimately may not be feasible or acceptable in practice, excluding them beforehand (in the traditional approach, out of necessity) keeps potentially relevant new insights from researchers and policy makers. Beyond optimization based on population-level parameters, there is the potential for metamodels to enable further research into more personalized screening, for which even more potential strategies exist. Such screening strategies may be, for example, sex- and/or age-specific, or strategies in which the frequency of screening is adapted on the basis of previous test results.

Next to optimization, other applications of metamodeling in a health economic context consist of performing a probabilistic analysis,[30] a value of information analysis,[31,32] or a model calibration procedure.[33] Although different in goal, these applications have in common that metamodeling is used to perform extensive (repeated) analyses in the context of computationally demanding simulation models. In general, applications of metamodeling in which fewer associations need to be reflected, that is, the number of relevant input parameters is limited, may be easier to perform. For example, in our case study 4 input parameters were used, whereas a probabilistic analysis of the ASCCA model would require the use of more than 50 input parameters. Fitting metamodels with many (eg, >25) input parameters is not feasible using Gaussian processes, for which both the fitting time and calculation time (for prediction) increase exponentially with the number of input parameters. Fitting such metamodels with alternative techniques, for example, generalized additive models or neural networks, would be feasible in limited time, and predictions are likely to be very fast, which may enable performing probabilistic analysis in a matter of minutes.[34] The main challenge would be to obtain sufficient samples for fitting the metamodels from the original model, because hundreds of samples may be required to reach acceptable metamodel accuracy using neural networks, for example. The number of samples required can be decreased by performing parameter importance analysis,[30] but a discussion of those methods is beyond the scope of this study.

In our specific case study, the metamodeling approach suggests that optimal screening under a colonoscopy capacity

Figure 4. Calibration plot* (A) and optimization frontier (B) for the LYG metamodel after 1 adaptive sampling step. *Note: range of predicted LYG in this plot has been restricted to higher LYG values, as calibration was based on the top 150 strategies with highest expected LYG. LYG, life-years gained.

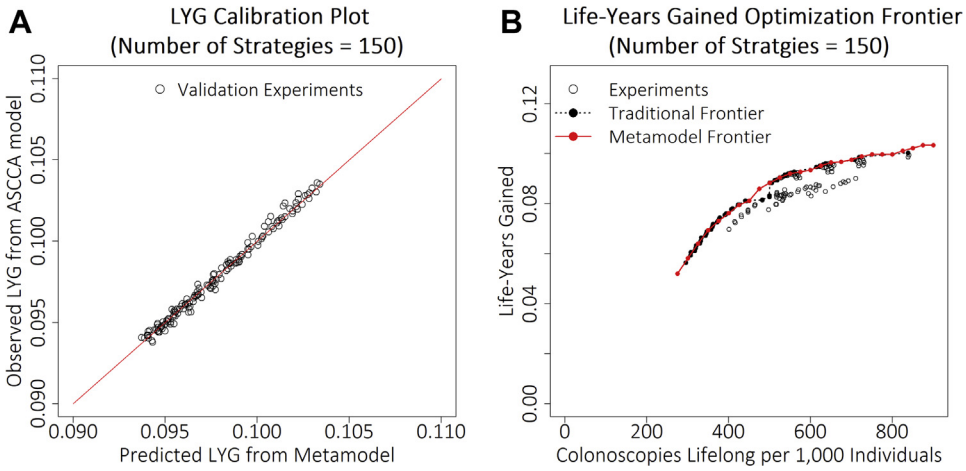


Table 2. Definitions and outcomes of screening strategies expected to be optimal in LYG.

Constraint	Screening strategy identified as optimal				Predicted strategy outcomes*			
Maximum colonoscopy capacity†	Start age screening (years)	Screening interval (years)	Number of screening rounds	FIT cutoff‡ for referral (ng/mL)	Number of colonoscopies†	LYG	Incremental costs (€)	NMB§
300	32	1	18	150	296	0.058	-266	1430
450	40	1	20	150	441	0.081	-316	1941
550	33	2	21	150	546	0.092	-361	2202
650	34	2	21	100	647	0.097	-363	2294
800	35	2	21	75	738	0.100	-380	2374

FIT indicates fecal immunochemical test; LYG, life-year gained; NMB, net monetary benefit.
*For the optimal screening strategy compared with no screening. No comparison between screening strategies was made, because their requirements in colonoscopy capacity, which will determine feasibility in clinical practice, varies widely.
†Colonoscopy capacity in lifelong colonoscopies, per 1000 individuals.
‡Only discret-e cutoff values of 50, 75, 100, and 150 ng/mL were considered in the analysis.
§NMB determined using a willingness-to-pay of €20 000 per life-year gained.

constraint of 550 colonoscopies lifelong per 1000 individuals, conforming to current Dutch colonoscopy requirements, should start at much younger age than in the current Dutch screening program with starting age 55 years (and 11 biennial screening rounds, using a 75 ng/mL FIT cutoff). Although this might imply that earlier screening potentially could save lives and decrease costs, without requiring additional colonoscopy capacity, the best-performing screening strategies identified in this study are large extrapolations beyond the scope of the original ASCCA model. Clearly, the ASCCA model itself requires validation for younger ages, an extensive comparison of predicted and observed adenoma prevalence rates in younger individuals, before its results can be used for policy recommendations concerning a decrease in the starting age of screening. To enable such a validation, future CRC screening studies should also include individuals aged <50 years and report age-specific adenoma prevalence rates. Current case study results may, therefore, stimulate further research on CRC screening in The Netherlands. CRC incidence is rising in young adults,[35] and there is increasing interest in the optimal start age of screening. This is underpinned by 2 recent studies evaluating screening strategies starting at age 40.[36,37]

Our study has certain limitations. Results shown are, of course, specific to the optimization problem addressed and the original simulation model used. When other CRC screening simulation models would have been used, results could have been different.[38] Furthermore, we applied a single metamodeling technique, Gaussian processes, to illustrate the potential advantages of metamodeling. Many alternative metamodeling techniques are available, for example, linear regression, artificial neural networks, and multivariate adaptive regression splines, each with their own (dis) advantages.[39,40] Nevertheless, rather than comparing these techniques, this study focused on increasing general awareness of the potential added value of metamodeling in a health economic context. For the same reason, validation efforts and performance checks for the cost and colonoscopy metamodels were limited, even though assessing metamodel accuracy is crucial and a range of resampling strategies is available.[41] The impact of varying the sample size for adaptive sampling was not studied, because this would require a comprehensive simulation study to generate generalizable results, which was beyond the scope of this article. Although metamodels are ideally fitted to samples generated through a probabilistic analysis for each strategy, this is currently

Table 3. Top 10 screening strategies in terms of LYG for a colonoscopy capacity of 550.

Screening strategy identified as optimal				Predicted strategy outcomes*			
Start age screening (years)	Screening interval (years)	Number of screening rounds	FIT cutoff‡ for referral (ng/mL)	Number of colonoscopies†	LYG	Incremental costs (€)	NMB§
33	2	21	150	546	0.0920	-361	2202
34	2	20	150	545	0.0919	-348	2185
36	2	19	150	548	0.0917	-348	2183
35	2	19	150	541	0.0916	-361	2193
33	2	20	150	537	0.0914	-359	2186
32	2	21	150	538	0.0912	-343	2166
34	2	19	150	533	0.0911	-343	2166
37	2	18	150	543	0.0911	-361	2183
36	2	18	150	536	0.0910	-343	2163
38	2	18	150	549	0.0909	-347	2165

FIT indicates fecal immunochemical test; LYG, life-year gained; NMB, net monetary benefit.
*For the optimal screening strategy compared with no screening. No comparison between screening strategies was made, because their requirements in colonoscopy capacity, which will determine feasibility in clinical practice, varies widely.
†Colonoscopy capacity in lifelong colonoscopies, per 1000 individuals.
‡Only discrete cutoff values of 50, 75, 100, and 150 ng/mL were considered in the analysis.
§NMB determined using a willingness-to-pay of €20 000 per life-year gained.

not supported by the ASCCA model. As discussed previously, metamodels can also be used to enable probabilistic analysis of computationally demanding models like the ASCCA model,[30] but this is part of further research. Also, the number of screening strategies considered to illustrate the benefit of the metamodeling approach over the traditional approach and its impact on metamodel accuracy is an arbitrary decision. Nevertheless, although using different numbers of strategies might have resulted in different optimization frontiers and accuracy estimates, this is not expected to influence the final outcomes and conclusions owing to the high accuracy of the final metamodel. Finally, the comparison of the traditional and metamodeling approach was based on use of the same limited set of strategies generated using a Latin hypercube design. Using such a design to cover the entire space of strategies for the metamodeling approach is sensible. Nevertheless, when only a very limited set of diagnostic strategies (eg, in range of 2 to 25) is evaluated with the traditional approach, strategies are often not generated by design but defined by researchers. Such a manual selection potentially could either increase or decrease the benefits of metamodeling, depending on how close the strategies selected for evaluation would be to the optimal strategy.

Conclusions

Metamodels can be used to address computational challenges in complex, model-based health economic analyses and allow for optimization rather than evaluation of strategies. This was illustrated for a CRC screening context, while accounting for practical constraints regarding the maximum number of colonoscopies to be performed. When applying metamodeling methods, validation showed to be crucial, and an adaptive sampling step was necessary to improve metamodel accuracy. Outcomes from the metamodel indicated potential benefits of starting CRC screening at a younger age than is currently considered. This raised questions concerning the evidence base (ie, generalizability) of the original model, may lead to additional validation efforts, and provides insights beyond those that would have been obtained using only the original simulation model.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2020.08.2099.

Article and Author Information

Accepted for Publication: August 18, 2020

Published Online: October 27, 2020

doi: https://doi.org/10.1016/j.jval.2020.08.2099

Author Affiliations: Health Technology and Services Research Department, Technical Medical Centre, University of Twente, Enschede, The Netherlands (Koffijberg, Degeling, IJzerman); Centre for Cancer Research and Centre for Health Policy, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne, Australia (IJzerman); Melbourne School of Population and Global Health, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne, Australia (IJzerman); Department of Cancer Research, Peter MacCallum Cancer Centre, Melbourne, Australia (IJzerman); Decision Modeling Center, Department of Epidemiology and Data Science, Amsterdam UMC – location VUmc, Amsterdam, The Netherlands (Coupé, Greuter).

Correspondence: Hendrik Koffijberg, PhD, Health Technology & Services Research Department, University of Twente, PO Box 217, 7500 AE, Enschede, The Netherlands. Email: h.koffijberg@utwente.nl

Author Contributions: *Concept and design*: Koffijberg, Degeling, IJzerman, Coupé
Acquisition of data: Greuter
Analysis and interpretation of data: Koffijberg, Degeling, IJzerman, Coupé, Greuter
Drafting of the manuscript: Koffijberg, Degeling, Coupé
Critical revision of the paper for important intellectual content: IJzerman, Coupé, Greuter
Statistical analysis: Koffijberg, Degeling
Administrative, technical, or logistic support: Degeling
Supervision: Koffijberg, IJzerman

Conflict of Interest Disclosures: Dr IJzerman is an editor for *Value in Health* and had no role in the peer review process of this article. No other disclosures were reported.

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

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

REFERENCES

1. Ferrante di Ruffano L, Davenport C, Eisinga A, et al. A capture-recapture analysis demonstrated that randomized controlled trials evaluating the impact of diagnostic tests on patient outcomes are rare. *J Clin Epidemiol*. 2012;65:282–287.
2. Buxton MJ, Drummond MF, Van Hout BA, et al. Modelling in economic evaluation: an unavoidable fact of life. *Health Econ*. 1997;6:217–227.
3. McKenna C, Claxton K. Addressing adoption and research design decisions simultaneously: the role of value of sample information analysis. *Med Decis Making*. 2011;31:853–865.
4. Kip MMA, IJzerman MJ, Henriksson M, et al. Toward Alignment in the Reporting of Economic Evaluations of Diagnostic Tests and Biomarkers: The AGREEDT Checklist. *Med Decis Making*. 2018;38:778–788.
5. Degeling K, Koffijberg H, IJzerman MJ. A systematic review and checklist presenting the main challenges for health economic modeling in personalized medicine: towards implementing patient-level models. *Expert Rev Pharmacoecon Outcomes Res*. 2017;17:17–25.
6. Briggs AH, Weinstein MC, Fenwick EA, et al. Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–6. *Value Health*. 2012;15:835–842.
7. Strong M, Oakley JE, Brennan A. Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: a nonparametric regression approach. *Med Decis Making*. 2014;34:311–326.
8. Baio G. BCEA: A R Package to Perform Bayesian Cost-Effectiveness Analysis. *Value Health*. 2014;17:A550.
9. Crown W, Buyukkaramikli N, Sir MY, et al. Application of Constrained Optimization Methods in Health Services Research: Report 2 of the ISPOR Optimization Methods Emerging Good Practices Task Force. *Value Health*. 2018;21:1019–1028.
10. Crown W, Buyukkaramikli N, Thokala P, et al. Constrained Optimization Methods in Health Services Research-An Introduction: Report 1 of the ISPOR Optimization Methods Emerging Good Practices Task Force. *Value Health*. 2017;20:310–319.
11. Bor AS, Koffijberg H, Wermer MJ, et al. Optimal screening strategy for familial intracranial aneurysms: a cost-effectiveness analysis. *Neurology*. 2010;74:1671–1679.
12. Greuter MJ, Berkhof J, Fijneman RJ, et al. The potential of imaging techniques as a screening tool for colorectal cancer: a cost-effectiveness analysis. *Br J Radiol*. 2016;89:20150910.
13. Barton RR. Simulation metamodels. Proceedings of the 30th conference on Winter simulation. Washington, D.C., USA: IEEE Computer Society Press, 1998.
14. Simpson TW, Poplinski JD, Koch PN, et al. Metamodels for Computer-based Engineering Design: Survey and recommendations. *Eng Computers*. 2001;17:129–150.
15. Degeling K, IJzerman MJ, Koffijberg H. A scoping review of metamodeling applications and opportunities for advanced health economic analyses. *Expert Rev Pharmacoecon Outcomes Res*. 2019;19:181–187.
16. International Agency for Research on Cancer. *Global Cancer Observatory*. World Health Organization; 2018.
17. Shaukat A, Mongin SJ, Geisser MS, et al. Long-term mortality after screening for colorectal cancer. *N Engl J Med*. 2013;369:1106–1114.
18. Scholefield JH, Moss SM, Mangham CM, et al. Nottingham trial of faecal occult blood testing for colorectal cancer: a 20-year follow-up. *Gut*. 2012;61:1036–1040.
19. Young GP, Rabeneck L, Winawer SJ. The Global Paradigm Shift in Screening for Colorectal Cancer. *Gastroenterology*. 2019;156:843–851.e2.
20. IKNL. Monitor Bevolkingsonderzoek Darmkanker 2018. Integraal Kankercentrum Nederland. https://www.iknl.nl/kankersoorten/darmkanker/onderzoek/monitor-bevolkingsonderzoek-darmkanker. Accessed November 22, 2019. Accessed.
21. IKNL. Cijfers over kanker 2019. Integraal Kankercentrum Nederland. http://www.cijfersoverkanker.nl/. Accessed November 22, 2019.
22. Greuter MJ, Xu XM, Lew JB, et al. Modeling the Adenoma and Serrated pathway to Colorectal CAncer (ASCCA). *Risk Anal*. 2014;34:889–910.
23. Greuter MJE, de Klerk CM, Meijer GA, et al. Screening for Colorectal Cancer With Fecal Immunochemical Testing With and Without Postpolypectomy Surveillance Colonoscopy: A Cost-Effectiveness Analysis. *Ann Intern Med*. 2017;167:544–554.
24. Aruoba SB, Fernandez-Villaverde J. A comparison of programming languages in macroeconomics. *J Econ Dyn Control*. 2015;58:265–273.
25. O'Hagan A. Curve Fitting and Optimal Design for Prediction (with discussion). *J Royal Stat Soc Series B*. 1978;40:1–42.
26. Garud SS, Karimi IA, Kraft M. Design of computer experiments: a review. *Comput Chem Eng*. 2017;106:71–95.
27. Husslage BGM, Rennen G, van Dam ER, et al. Space-filling Latin hypercube designs for computer experiments. *Optim Eng*. 2011;12:611–630.
28. Stoop EM, de Haan MC, de Wijkerslooth TR, et al. Participation and yield of colonoscopy versus non-cathartic CT colonography in population-based screening for colorectal cancer: a randomised controlled trial. *Lancet Oncol*. 2012;13:55–64.
29. Rutter CM, Yu O, Miglioretti DL. A hierarchical non-homogenous Poisson model for meta-analysis of adenoma counts. *Stat Med*. 2007;26:98–109.
30. de Carvalho TM, Heijnsdijk EAM, Coffeng L, et al. Evaluating parameter uncertainty in a simulation model of cancer using emulators. *Med Decis Making*. 2019;39:405–413.
31. Jalal H, Alarid-Escudero F. A Gaussian approximation approach for value of information analysis. *Med Decis Making*. 2018;38:174–188.
32. Jutkowitz E, Alarid-Escudero F, Kuntz KM, et al. The curve of optimal sample size (COSS): a graphical representation of the optimal sample size from a value of information analysis. *Pharmacoeconomics*. 2019;37:871–877.
33. Andrianakis I, Vernon IR, McCreesh N, et al. Bayesian history matching of complex infectious disease models using emulation: a tutorial and a case study on HIV in Uganda. *PLoS Comput Biol*. 2015;11:e1003968.
34. Degeling K, IJzerman MJ, Lavieri MS, et al. An introduction to metamodeling for reducing computational burden of advanced analyses with health economic models: a structured overview of metamodeling methods in a six-step application process. *Med Decis Making*. 2020;40(3):348–363.
35. O'Connell JB, Maggard MA, Liu JH, et al. Rates of colon and rectal cancers are increasing in young adults. *Am Surg*. 2003;69:866–872.
36. Peterse EFP, Meester RGS, Siegel RL, et al. The impact of the rising colorectal cancer incidence in young adults on the optimal age to start screening: microsimulation analysis I to inform the American Cancer Society colorectal cancer screening guideline. *Cancer*. 2018;124:2964–2973.
37. Lew JB, St John DJB, Macrae FA, et al. Benefits, harms, and cost-effectiveness of potential age extensions to the ational Bowel Cancer Screening Program in Australia. *Cancer Epidemiol Biomarkers Prev*. 2018;27:1450–1461.
38. Silva-Illanes N, Espinoza M. Critical analysis of Markov Models used for the economic evaluation of colorectal cancer screening: a systematic review. *Value Health*. 2018;21:858–873.
39. Tappenden P, Chilcott JB, Eggington S, et al. Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon-beta and glatiramer acetate for multiple sclerosis. *Health Technol Assess*. 2004;8(iii):1–78.
40. Li YF, Ng SH, Xie M, et al. A systematic comparison of metamodeling techniques for simulation optimization in Decision Support Systems. *Appl Soft Comput*. 2010;10:1257–1273.
41. Bischl B, Mersmann O, Trautmann H, et al. Resampling methods for meta-model validation with recommendations for evolutionary computation. *Evol Comput*. 2012;20:249–275.