Original Article

MDM Medical Decision Making

Medical Decision Making
2018, Vol. 38(7) 810–821

DOI: 10.1177/0272989X18792283
journals.sagepub.com/home/mdm
SAGE

Nonidentifiability in Model Calibration and Implications for Medical Decision Making

Fernando Alarid-Escudero, Richard F. MacLehose, Yadira Peralta, Karen M. Kuntz, and Eva A. Enns

Background. Calibration is the process of estimating parameters of a mathematical model by matching model outputs to calibration targets. In the presence of nonidentifiability, multiple parameter sets solve the calibration problem, which may have important implications for decision making. We evaluate the implications of nonidentifiability on the optimal strategy and provide methods to check for nonidentifiability. **Methods.** We illustrate nonidentifiability by calibrating a 3-state Markov model of cancer relative survival (RS). We performed 2 different calibration exercises: 1) only including RS as a calibration target and 2) adding the ratio between the 2 nondeath states over time as an additional target. We used the Nelder-Mead (NM) algorithm to identify parameter sets that best matched the calibration targets. We used collinearity and likelihood profile analyses to check for nonidentifiability. We then estimated the benefit of a hypothetical treatment in terms of life expectancy gains using different, but equally good-fitting, parameter sets. We also applied collinearity analysis to a realistic model of the natural history of colorectal cancer. **Results.** When only RS is used as the calibration target, 2 different parameter sets yield similar maximum likelihood values. The high collinearity index and the bimodal likelihood profile on both parameters demonstrated the presence of nonidentifiability. These different, equally good-fitting parameter sets produce different estimates of the treatment effectiveness (0.67 v. 0.31 years), which could influence the optimal decision. By incorporating the additional target, the model becomes identifiable with a collinearity index of 3.5 and a unimodal likelihood profile. **Conclusions.** In the presence of nonidentifiability, equally likely parameter estimates might yield different conclusions. Checking for the existence of nonidentifiability and its implications should be incorporated into standard model calibration procedures.

Keywords
calibration, decision-analytic models, estimation, likelihood function, nonidentifiability

Date received: January 18, 2017; accepted: July 6, 2018

Model calibration is the process of estimating values for unknown or uncertain parameters of a mathematical model by matching model outputs to observed clinical, epidemiological, or any other type of data arising from a physical or biological system (known as calibration targets).[1] The goal is to identify parameter values that maximize the fit between model outputs and the calibration targets.[2] Previous literature has shown how model calibration can be described as a statistical estimation problem.[1,3,4] A desirable property for a statistical model is that of identifiability, which requires that different sets of parameter estimates cannot lead to the same probability distribution of the data.[5,6] In the context of model calibration, this means that there exists a unique set of model parameter values that yields the best fit to the chosen calibration targets.[7,8]

Corresponding Author
Fernando Alarid-Escudero, Division of Health Policy and Management, University of Minnesota School of Public Health, 420 Delaware St SE (MMC 729), Minneapolis, MN 55455 (alari006@umn.edu).

With the increasing complexity of mathematical models parameterized with a large number of unknown inputs, concerns have been raised about calibrating to an insufficient number of targets relative to the number of parameters to be estimated, making the model nonidentifiable.[9] However, a model need not be of high complexity for its parameters to be nonidentifiable. Identifiability in the context of model calibration in medical decision making has been briefly discussed[10,11] but has not been formally described. In this article, we explicitly define nonidentifiability in the context of model calibration and demonstrate that even a simple disease simulation model can have nonidentifiable parameters. We start by defining the problem of nonidentifiability in a calibration framework. We then describe and calibrate a 2-parameter simulation model of cancer recurrence that is calibrated to relative survival. We show that for this simple case, calibration does not yield a unique solution, exhibiting problems of nonidentifiability. We demonstrate that this problem could potentially be addressed by incorporating additional information, and we assess the implications of nonidentifiability in the evaluation of the comparative effectiveness of a hypothetical treatment. In addition, we provide a description of methods to detect nonidentifiability and apply them to our illustrative example and a realistic model of the natural history of colorectal cancer. We conclude with suggestions on how to potentially achieve identifiability and methods to estimate parameters whenever nonidentifiability is not avoidable.

Division of Health Policy and Management, University of Minnesota School of Public Health, Minneapolis, MN (FA-E, KMK, EAE); Division of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, MN (RFM); and Department of Educational Psychology, University of Minnesota, Minneapolis, MN (YP). Drs. Alarid-Escudero and Kuntz were supported by a grant from the National Institutes of Health (award U01 CA 199335) as part of the Cancer Intervention and Surveillance Modeling Network (CISNET). Financial support for this study also was provided in part by a grant from the National Council of Science and Technology of Mexico (CONACYT) and a Doctoral Dissertation Fellowship from the Graduate School of the University of Minnesota as part of Dr. Alarid-Escudero's doctoral program. Dr. Enns was supported by a grant from the National Institute of Allergy and Infectious Disease of the National Institutes of Health under award no. K25AI118476. The funding agencies had no role in the design of the study, interpretation of results, or writing of the manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

The preliminary findings from this analysis were presented at the 39th annual meeting of the Society for Medical Decision Making in October 2017.

Methods

Calibration as an Estimation Problem

Let M denote a mathematical model that takes a set of parameters $\theta \in \Theta \subseteq \mathbb{R}^p$ as input and produces a set of outputs denoted as $\phi \in \Phi \subseteq \mathbb{R}^m$, that is $\phi = M(\theta)$. The last equality holds true if M is a deterministic model; for stochastic models, ϕ would be defined as the expected value of model outputs, $\phi = \mathbf{E}[M(\theta)]$. In the context of medical decision making, typical model types include Markov models, microsimulations, discrete-event simulations, and dynamic transmission models.[12,13] Input parameters θ might include transition probabilities or rates, while model outputs ϕ might be quantities such as prevalence, incidence, or survival curves.[14] There is typically a subset of parameters θ that are unobserved due to financial, practical, or ethical reasons.[15] Thus, $\theta = (\theta_u, \theta_k)$, where θ_u denotes unknown parameters that need to be estimated via calibration,[16,17] and θ_k denotes parameters that are either known a priori or that can be estimated directly from available data without the use of the mathematical model M.[18] θ_u are the parameters of interest in the calibration problem, often called calibrated input parameters,[19–21] whereas parameters θ_k are fixed for the purposes of calibration. In medical decision making, unknown parameters might be disease progression or regression rates, probability of symptom-based detection in natural history models of chronic diseases,[10,11,22,23] or the transmission probabilities in infectious disease dynamic models,[24] among others.

Let Y denote the clinical or epidemiological phenomenon in the population of interest. The empirical data, y, referred to as the calibration targets, are a realization of Y. The process of model calibration is an estimation problem where we seek to estimate θ_u by using a summary measure of the discrepancy between the corresponding model output, ϕ, and y.[4] In this article, we will use the likelihood function as the summary measure; however, other possible summary measures include the sum of squared difference (SSD), a weighted sum of squared difference (WSSD), absolute difference, and so on.[19,25–27] To define a likelihood, a probability distribution of the calibration targets, f, is specified as a function of model parameters, θ.[10] In the context of model calibration, $Y \sim f(y; \phi)$, where f denotes a probability distribution that is conditional on the calibration targets from the model outputs, ϕ, which in turn depend on θ. Given that $Y = y$ is observed and M is a deterministic model, the likelihood function can be defined as $L(\theta) \equiv L(\theta; y) = f(y; \phi) = f(y; \theta, M)$. A schematic diagram of the relationship between input parameters θ and

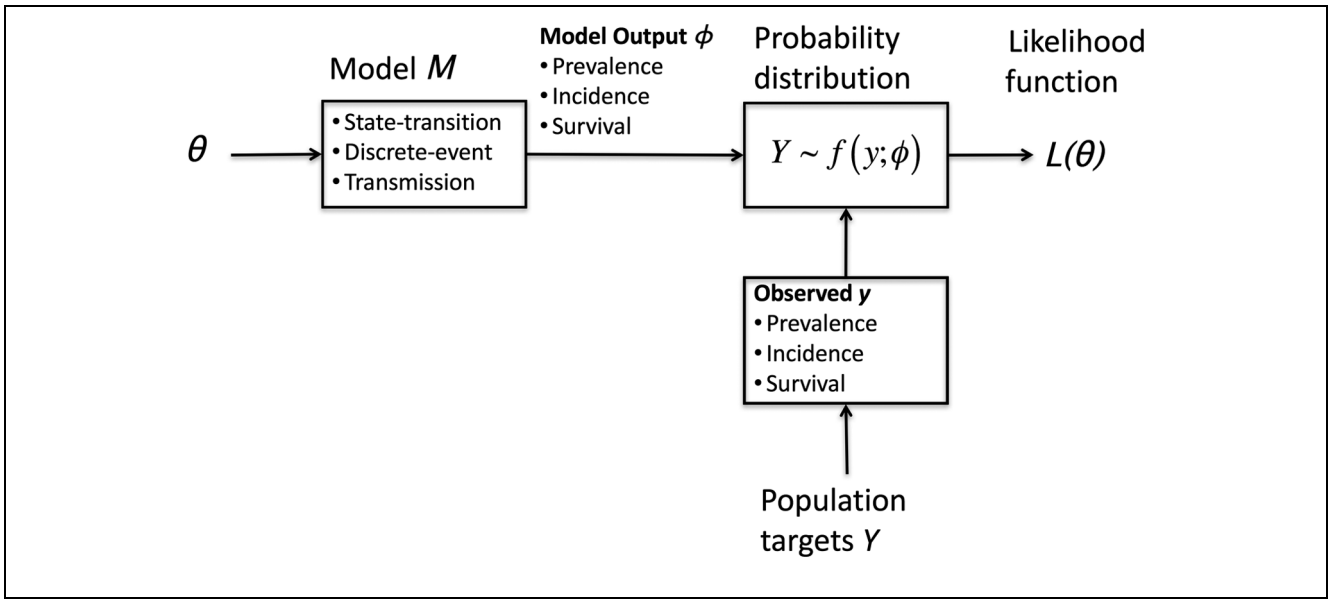

Figure 1 Schematic diagram of the relation between input parameters θ and calibration targets **y** to create a likelihood function L(θ) for model calibration in medical decision making.

calibration targets y to create a likelihood function $L(\theta)$ for model calibration in medical decision making is provided in Figure 1. A more detailed conceptualization of calibration as an estimation process has been proposed previously to account for both model inadequacy and observation error.[1,3] However, modeling these discrepancies falls outside the scope of this article.

Nonidentifiability

Identifiability refers to whether the specified model and the chosen calibration targets are sufficient to yield a unique set of values for the calibrated input parameters.[28–32] From a statistical perspective, a set of parameters θ is said to be identifiable if different values of θ correspond to different probability distributions $f(y;\phi)$: if $\theta \neq \theta'$ then $f(y;\phi) \neq f(y;\phi')$.[33] Therefore, the parameters θ are identifiable if $f(y;\phi)$ has a one-to-one mapping for all $\theta \in \Theta$.[25]

Nonidentifiability can arise at different stages of model calibration represented in Figure 1. The most conventional sense of nonidentifiability would occur when, given different parameter sets $\theta \neq \theta'$, the model M produces the same output values corresponding to the calibration targets, $\phi = \phi'$, which in turn could yield multiple peaks of similar magnitude in the fit function. Furthermore, it might not be possible to estimate parameters of a nonidentifiable model even with an infinite amount of data,[11,31] where data are referred to as the sample size informing the selected calibration targets. Note that the model outputs corresponding to the calibration targets, ϕ, are a subset of the complete model output (e.g., the distribution of the population across all health states at every time point), meaning that nonidentifiability can arise when matching the calibration targets even if the mathematical model itself is well defined. To construct an identifiable calibration problem in this case, one might need to include additional calibration targets (if data became available)[11] or further restrict the search space of possible values for unknown parameters based on biological plausibility or other a priori knowledge of the disease or population of interest.

Even if $M(\theta) = \phi$ has a one-to-one mapping for all $\theta \in \Theta$, eliminating the possibility of conventional nonidentifiability, calibration nonidentifiability can still occur in the mapping of model outputs to a goodness-of-fit value, $L(\phi;y)$. If the goodness-of-fit function maps different model outputs $\phi \neq \phi'$ to the same values $L(\phi;y) = L(\phi';y)$, then there again may exist multiple peaks of similar magnitude in the fit function. A simple example would be the case of calibration to 2 (scalar) calibration targets using a sum of squared difference between the model outputs and target values as the goodness-of-fit measure. If one set of parameter values fits the first target well and the second target poorly, while another set of parameter values fits the second

target well and the first target poorly to the same extent, then both sets of parameters would result in the same value of fit. If these were also the best fits that could be achieved for this case (i.e., there were no other sets of parameter values that fit both targets well), then the fit function would exhibit multiple peaks. In this case, nonidentifiability could potentially be alleviated by reconsidering the measure of goodness of fit. While theoretically the goodness-of-fit function should reflect the true preferences of the analyst in constructing the model, these decisions are often arbitrary or not carefully thought out. Therefore, it is important that the analyst knows what the assumptions are behind using different goodness-of-fit measures. For example, the sum of squared differences assumes equal weight to all targets while the weighted sum of squared differences allows for different weights for different targets, which could be computed with the variances of each target. Alternatively, other approaches that eliminate the need to produce a single summary measure could be employed, such as the Pareto frontier approach.[21] However, in this study, we are interested in nonidentifiability issues at the goodness-of-fit level. If the goodness-of-fit function was carefully constructed and the problem of nonidentifiability remains, the issue could still be addressed by including additional calibration targets and restricting the unknown parameter search space, as in conventional nonidentifiability.

Identifying Nonidentifiability

There are a number of different methods for detecting nonidentifiability.[34–36] One such method, collinearity analysis,[37] involves computing a collinearity index γ_K that reflects the degree of near-linear dependence of summary measures on a subset of parameters while fixing all other parameters to certain values. The summary measures could simply be the model outcomes or, in the context of identifiability analysis, the likelihood, SSD, or WSSD. The collinearity index γ_K of a subset K parameters is defined as

$$\gamma_K = \frac{1}{\sqrt{\min\left(\mathrm{EV}\left[\tilde{S}_K^T \tilde{S}_K\right]\right)}},$$

where EV[·] refers to the operator calculating the set of eigenvalues of the argument, and $\tilde{S}_K$ denotes a subset of the $\tilde{S}$ matrix containing the columns corresponding to the parameters in K. $\tilde{S}$ is a normalized sensitivity matrix whose jth column is defined as $\tilde{S}_j = \frac{S_j}{||S_j||}$, and S_j represents a column of derivatives of the summary measures with respect to the jth parameter. Thus, $\tilde{S}_j$ is a normalized measure of the importance of individual parameters on the summary measures.[38] To compute $\tilde{S}$, it is required to evaluate the model j times, one evaluation for each parameter. The collinearity index takes values from 1 to infinity, where higher values are associated with nonidentifiability problems.[39] For example, a collinearity index of 1 means that the columns of $\tilde{S}$ are orthogonal and the parameter set is identifiable. A high value of γ_K implies that the parameter set K is poorly identifiable, and the higher the value, the less identifiable the parameter set. In practical terms, it is suggested that parameters with collinearity indices lower than 10 are identifiable, between 10 and 15 are poorly identifiable, and greater than 15 are nonidentifiable.[38]

Likelihood profiling is another approach for identifying issues of nonidentifiability. The profile likelihood, pL, is a 1-dimensional representation of the likelihood function indicating the values of a parameter subset controlled by the influence of the rest of the parameters of interest.[40] Profiling the likelihood involves evaluating the log-likelihood as a function of the k values of one parameter θ_i while the rest of the parameters θ_j, $j \neq i$ are reoptimized at each value of k.[41] In general, if θ_I is a set of parameters of interest and θ_C is a set of complementary (i.e., nuisance) parameters, the profile likelihood for θ_I is $pL(\theta_I) = \max_{\theta_C} L(\theta_I, \theta_C)$. In practical terms, profiling the likelihood of p different parameters on k different values each requires p^k optimization routines. Having more than 1 minimum in the negative likelihood profile is an indicator of nonidentifiability.

Simple Model of Cancer Relative Survival

We illustrate the nonidentifiability problem by calibrating the transition probabilities of a simple deterministic Markov model to observed relative survival (RS) as reported by the Surveillance, Epidemiology, and End Results (SEER) Program. RS represents "cancer survival in the absence of other causes of death."[42] Therefore, relative survival at time t can be represented as the probability of not dying of cancer at time t in the absence of being at risk of dying of non-cancer-related causes. Most cancer-related deaths are attributed to metastatic recurrence,[43,44] which is not directly observed in SEER. Because the cause of death information in cancer registries can be misclassified, SEER computes relative survival "as the ratio of the proportion of observed survivors in a cohort of cancer patients to the proportion of expected survivors in a comparable set of cancer-free individuals."[42,45] To model the mechanism of cancer death, Markov models of cancer RS from an early stage

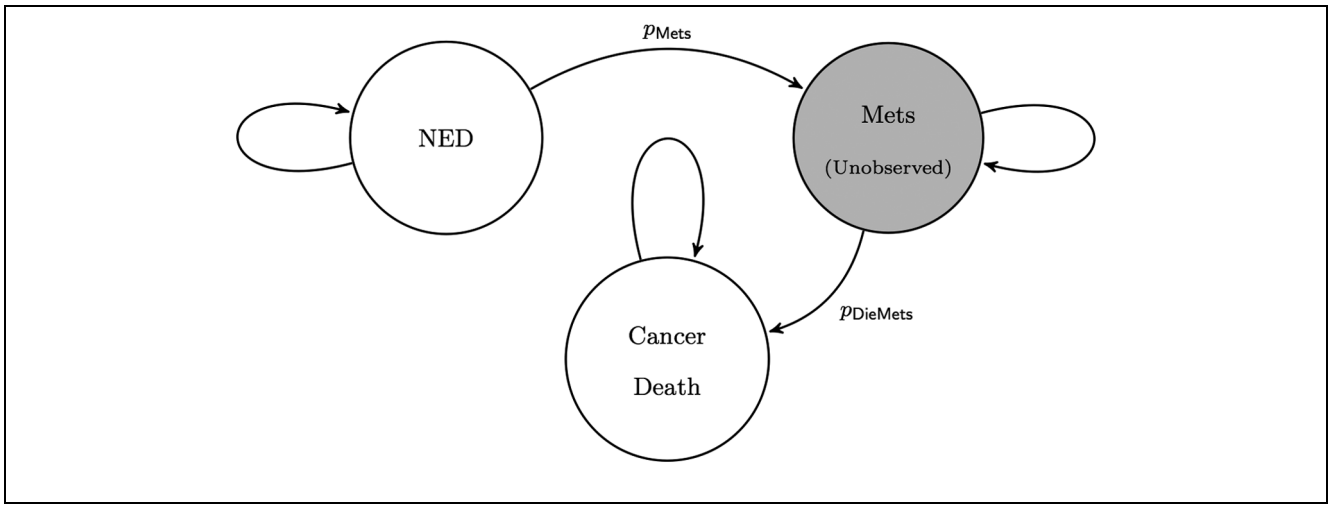

Figure 2 State-transition diagram of a 3-state Markov model of cancer relative survival. Mets, metastasis; NED, no evidence of disease.

diagnosis typically include a distant metastasis (Mets) state, which is not directly observed in cancer registries.[46] Accordingly, we developed a 3-state, 2-parameter Markov model of relative survival. In the Markov model, a simulated cohort of individuals starts in a "no evidence of disease" state (NED) and faces a monthly risk p_{Mets} of being diagnosed with distant metastasis (Mets). Individuals who develop distant metastasis face a monthly risk $p_{DieMets}$ of dying of cancer. To be consistent with the dynamics of RS, we assume that individuals are only allowed to die of cancer if they develop Mets. The state-transition diagram of the relative survival model is shown in Figure 2. To translate this calibration exercise into the terminology defined in the previous sections, y is RS, M represents the Markov model of cancer recurrence and RS, $\theta_u = \{p_{Mets}, p_{DieMets}\}$, and $\theta_k = \{\varnothing\}$ (i.e., there are no fixed input parameters). To compute RS from the Markov model over time, y_t, we sum the proportion of the cohort in the NED and Mets states at each time t.

Calibration of Simple Cancer Relative Survival Model

For illustration purposes and to avoid issues of model misspecification, we generated the target data, y, by running the Markov model as a microsimulation with known parameter values ($p_{Mets} = 0.10$ and $p_{DieMets} = 0.05$, that is $\theta_{true} = (0.10, 0.05)$) and stochastically simulating 200 independent individuals over 60 months.[47] Using the information of these simulated individuals, we estimated a RS curve, y_t, with its corresponding standard error σ_t in monthly intervals. The likelihood function was constructed by assuming that the targets y_t were normal deviations from the model outputs ϕ_t with standard deviation σ_t at each time t. That is,

$$y_t \sim \text{normal}(\phi_t, \sigma_t).$$

Therefore, the likelihood function is given by

$$L_t(\theta|y_t) = f(y_t|\theta),$$

where $f(y_t|\theta)$ denotes the normal density function for the target y_t.

We assumed independence across targets to compute an aggregated likelihood function by multiplying the likelihood value at each time t as follows:

$$L(\theta|Y) = \prod_{t=1}^{T} L_t(\theta|y_t).$$

If there are reasons to believe that targets are not independent, other likelihood functions can be specified such as multivariate normal or multinomial distributions. However, if the choice is a multivariate normal distribution, the analyst is required to define a correlation matrix across the targets. We used the Nelder-Mead (NM) algorithm initialized at 100 random starting values to identify cancer RS model parameter values that minimize the negative log-likelihood.[48]

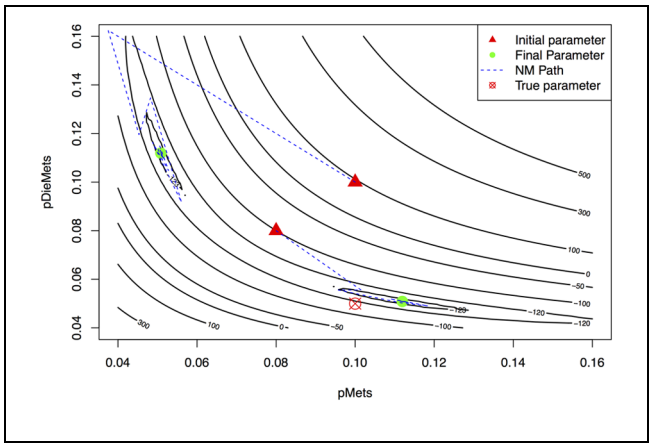

Figure 3 Contour plot of the negative log-likelihood function for the calibration of the cancer relative survival Markov model with 2 different search paths. NM, Nelder-Mead.

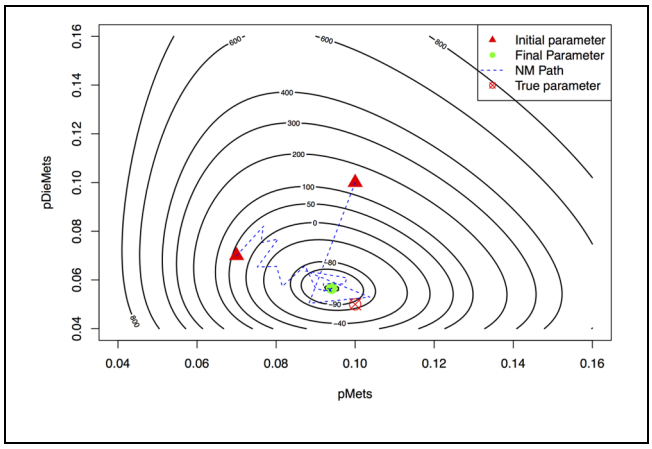

Figure 4 Contour of the negative log-likelihood function for the calibration of the cancer relative survival Markov model with information on the ratio of the simulated cohort between no evidence of disease (NED) and metastasis (Mets) states with 2 different search paths. NM, Nelder-Mead.

Results

Calibration

The NM algorithm converged to 2 parameter sets with similar log-likelihood values: $\hat{\theta}_1^{NI} = (0.05, 0.11)$ and $\hat{\theta}_2^{NI} = (0.11, 0.05)$. Figure 3 shows the contour of the likelihood and the 2 regions with similar values. For illustration purposes, in Figure 3, we plot only 2 different optimization paths; however, these are representative of the 100 different runs of the NM algorithm that we performed. The RS curve as the only calibration target was not sufficient to ensure identifiability.

For illustration purposes, we also simulated the ratio between NED and Mets over time and added it as a second calibration target to demonstrate that parameters can become identifiable with an additional type of target that provides information on the specific dynamics of the unobserved state Mets. This information is not provided by SEER but could be obtained from alternative sources, such as a clinical trial. To compute the likelihood of this additional target, we assume that the logarithm of the ratio between NED and Mets follows a normal distribution. We assume independence between types of targets to compute an aggregated measure of the overall likelihood.

When the ratio between NED and Mets over time is incorporated as an additional target, the parameter space gets constrained in terms of the most likely region making the likelihood function unimodal, exhibiting a unique solution (Figure 4), suggesting that the model is now identifiable. In this case, regardless of the starting value, NM consistently recovers the same set of parameter estimates, $\hat{\theta}_3^{ID} = (0.09, 0.05)$.

Policy Implications

We estimated the benefit of a hypothetical treatment that reduces p_{Mets} by 30% implemented as a relative risk (RR) of 0.7 using the 3-state Markov model. We ran the model with the 2 different sets of calibrated parameter values and the true values. Benefits were quantified in terms of life expectancy (LE) gains under the intervention, calculated as a percentage of the LE gains predicted by the model using the true parameters (see Table 1). In the absence of treatment, both parameter sets from the nonidentifiable problem, $\hat{\theta}_1^{NI}$ and $\hat{\theta}_2^{NI}$, result in the same LE. With the intervention, the benefit of treatment is overestimated by 88.6% when using $\hat{\theta}_1^{NI}$ and underestimated by 10.4% when using $\hat{\theta}_2^{NI}$. However, when using the parameter set from the identifiable problem, $\hat{\theta}_3^{ID}$, the benefit of treatment is only slightly overestimated (by 6.6%), which is much closer to the truth.

Identifying Nonidentifiability

We computed the collinearity index for the parameters of the simple model using the R package FME.[49] When the RS curve is used as the only target to calibrate the model, the collinearity index on both parameters tends to infinity, indicating nonidentifiability (Table 2). Similarly, when only the ratio between NED and Mets is

Table 1 LE, LE Gains, and Bias in LE Gains from a Hypothetical Treatment Intervention That Reduces the Risk of Developing Mets by 70% (i.e., Relative Risk = 0.7) using the True Parameters θ_{true}, 2 Sets of Nonidentifiable Parameter Estimates $\hat{\theta}_1^{NI}$ and $\hat{\theta}_2^{NI}$, and the Identifiable Set of Parameter Estimates $\hat{\theta}_3^{ID}$

Scenario	Parameters		LE without Treatment, years	LE with Treatment, years	LE Gains	% Difference
	p_{Mets}	$p_{DieMets}$				
θ_{true}	0.10	0.05	2.33	2.68	0.35	0.0
θ_1^{NI}	0.05	0.11	2.21	2.88	0.67	88.6
θ_2^{NI}	0.11	0.05	2.21	2.52	0.31	–10.4
θ_3^{ID}	0.09	0.06	2.19	2.56	0.37	6.6

LE, life expectancy; Mets, distant metastasis; $p_{DieMets}$, monthly risk of dying from Mets; p_{Mets}, monthly risk of developing Mets.

Table 2 Collinearity Indices for Different Combinations of Calibration Targets and Subsets of Parameters to Be Estimated

Subset of Parameters	Calibration Targets	Collinearity Index
p_{Mets} and $p_{DieMets}$	Survival	$\rightarrow \infty$
	Log-ratio NED/Mets	10.0
	Survival and log-ratio NED/Mets	3.5
p_{Mets} or $p_{DieMets}$	Survival	1.0
	Log-ratio NED/Mets	1.0
	Survival and log-ratio NED/Mets	1.0

Mets, metastasis; NED, no evidence of disease' $p_{DieMets}$, monthly risk of dying from Mets; pMets, monthly risk of developing Mets.

included, the calibration problem is again poorly identifiable with a collinearity index of 10. Once both targets are considered for calibration, the parameters become identifiable with a collinearity index of 3.5, suggesting that calibration of both parameters should yield a unique solution, consistent with our findings (Figure 4). If instead it were possible to reduce the calibration to only a single unknown parameter (the other being fixed to some known value), the calibration becomes identifiable with a collinearity index of 1 for any subset of targets (Table 2). This might occur if some new information becomes available that allows one of the unknown parameters to be directly estimated (provided that this information pertains to the population of interest and is unbiased), obviating the need to estimate it through calibration.

The likelihood profile of the parameters of our simple model using only the survival curve as target shows 2 local minima in the negative log-likelihood (see Figure 5), which correspond to the 2 parameter sets identified by NM in Figure 3. Having more than 1 minimum in the negative likelihood profile is an indicator of nonidentifiability.

To illustrate the potential issue of nonidentifiability and the application of collinearity analysis to a more realistic modeling setup, we developed a state-transition model (STM) of the natural history of colorectal cancer (CRC) implemented in discrete annual cycles based on a model structure originally proposed by Wu et al.[50] Briefly, this model has 9 different health states that include absence of the disease, precancerous lesions (i.e., adenomatous polyps), and preclinical and clinical cancer states by stage. The model has 11 parameters in total, from which we assume 9 are unknown and need to be calibrated, which reflects the typical unknown parameters in this type of models. Similar to the simple model of cancer relative survival, we generated 4 different age-specific targets by running independent individuals through the model as a microsimulation.[47] The targets are prevalence of adenomas, proportion of small adenomas, and CRC incidence for early and late cancer stages. A more detailed description of the model and the generation of the calibration targets is presented in the online Supplemental Appendix.

We computed the collinearity index, γ_K, on all possible combinations of the calibrated parameters for different combinations of targets. Computing these collinearity indexes required evaluating the model 9 times in total. The subset of calibration targets included in the calibration influenced the number of model parameters that could be identified (having $\gamma_K \leq 15$),[38,39] as summarized in the Supplemental Figures S4 to S6. When all

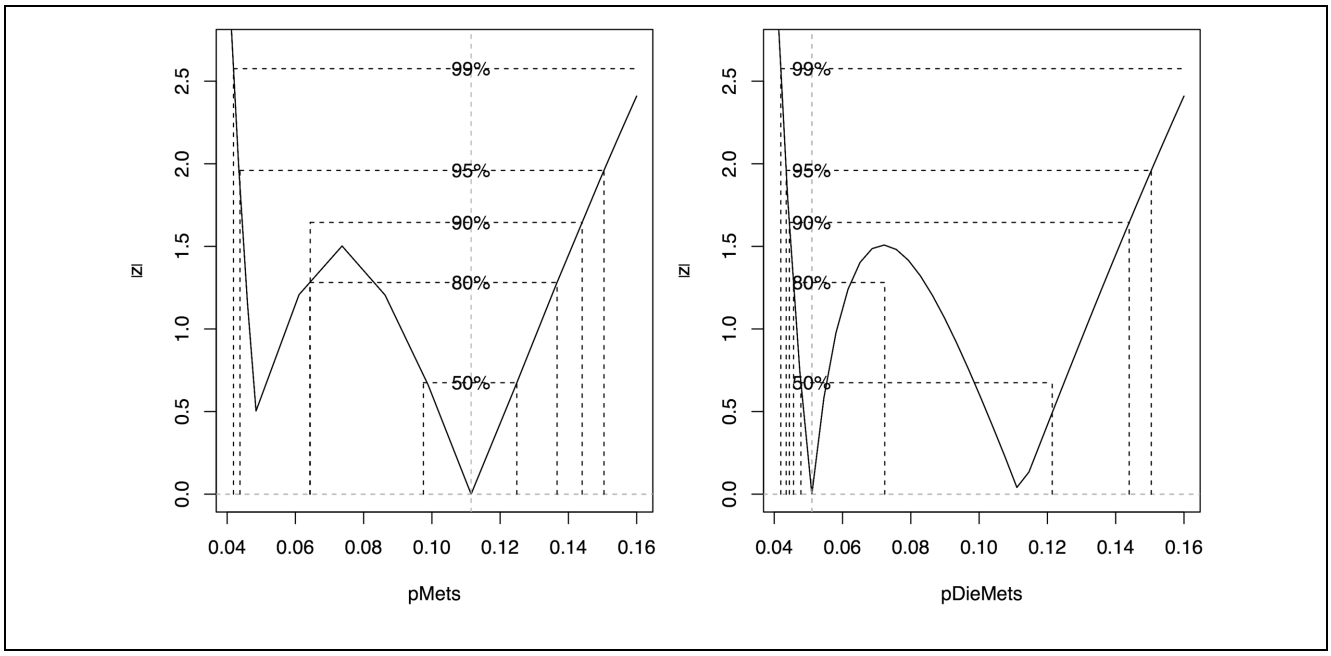

Figure 5 Negative-likelihood profile (black solid line), confidence levels (black dashed line), and minimum value (gray dashed-line) of parameters pMets and pDieMets. Mets, distant metastasis; $p_{DieMets}$, monthly risk of dying from Mets; p_{Mets}, monthly risk of developing Mets.

Table 3 Collinearity Index (γ_k) for Different Combinations and Number of Parameters (N) of the Natural History Model of Colorectal Cancer Using All 4 Targets[a]

l	g	λ_2	λ_3	λ_4	λ_5	λ_6	$p_{adenoma}$	p_{small}	N	γ
1	1	1	1	1	1	1	1	1	9	110
1	0	1	1	1	1	1	1	1	8	14
0	1	1	1	1	1	1	1	1	8	14
1	1	1	1	1	1	1	1	0	8	92
1	1	1	1	1	1	1	0	1	8	81
1	1	1	1	1	1	0	1	1	8	109
1	1	1	1	1	0	1	1	1	8	108
1	1	1	1	0	1	1	1	1	8	108
1	1	1	0	1	1	1	1	1	8	109
1	1	0	1	1	1	1	1	1	8	102

[a]If parameter is present, the cell equals 1 and 0 otherwise. The parameters are described in detail in the online Supplemental Appendix.

4 targets were included and only 2 parameters were unknown, the calibration was almost always identifiable (35 of 36 possible combinations; Figure 6). However, if 8 parameters were to be estimated through calibration, only 2 combinations were identifiable using all 4 calibration targets. Even with all the targets, it is not possible to define an identifiable calibration problem for all the model parameters, which have a combined γ_K of 110, shown in the first row of Table 3 and in Figure 6.

Discussion

In the context of model calibration, much has been written on issues of model specification, structural uncertainty, and search algorithms for performing the estimation.[14,51–55] However, the issue of nonidentifiability is rarely discussed. Identifiability is an important element in model calibration. If a model is nonidentifiable, multiple parameter sets could produce exact or similar best-fitting model outputs. This is of particular relevance in

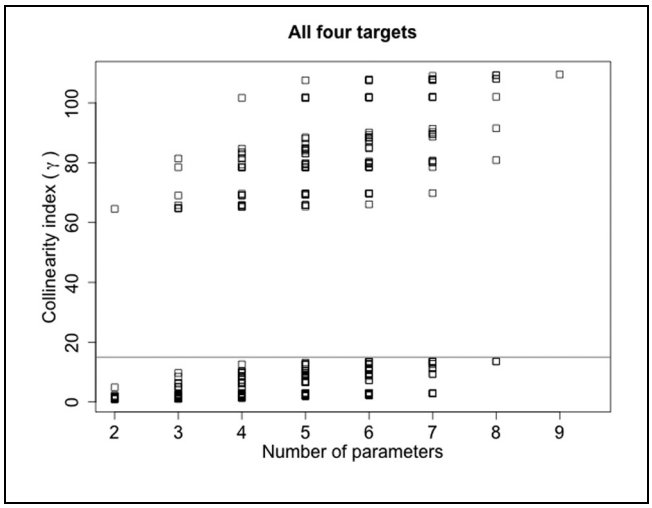

Figure 6 Collinearity index (γ) for all possible combinations of parameters of the natural history model of colorectal cancer using all 4 targets (gray solid vertical line indicates a collinearity index of 15).

the context of medical decision making when using mathematical models that need to be calibrated. Different but equally good-fitting sets of parameters might produce different estimates of the effectiveness of interventions that could potentially influence the optimal decision.[21,56]

The calibration of the simple model using only the RS curve is nonidentifiable because the 2 transition probabilities, p_{Mets} and $p_{DieMets}$, compensate for each other to yield the same RS. In one instance, survival includes greater progression to Mets first with a slower mortality, while in the other, there is slower progression to Mets but higher mortality. While different sets of values of p_{Mets} and $p_{DieMets}$ produce the same RS, the prevalence of metastases in the population is different over time. Specifically, it is much higher when the progression to metastasis, p_{Mets}, is higher with a lower mortality from metastases, $p_{DieMets}$. Adding the ratio between the proportion of the population in NED and Mets states as a calibration target penalizes sets of parameter values that do not match the true prevalence of metastases in the population over time. This allowed the likelihood to have a single peak for parameter values where both RS and the ratio of NED/Mets are well matched over time (Figure 4). To compute the overall likelihood on the simple model of RS, we assumed independence across targets. It is common to assume that targets are independent in calculating an overall goodness-of-fit measure. In practice, this might not be true and might be a nonverifiable assumption; however, while this is a concern in estimating the uncertainty in calibrated parameters, in the case of a nonidentifiable calibrated model, that problem will persist regardless of the independence assumption. The main implication of erroneously assuming independence is an underestimate of the uncertainty of the calibrated parameters.

Scholars have argued that the best approach when nonidentifiability is known or suspected is to acknowledge the problem and, if possible, gather more information to estimate model parameters.[33] Although identifiability might not always be guaranteed in model calibration, incorporating more information or constraining the parameter space based on prior knowledge on the disease of interest could help make the model identifiable.[57] In our example, using an additional target for calibration solved the problem of nonidentifiability, but it was not obvious that this would resolve the issue ex ante. In general, having targets that inform different aspects of the model will help improve and potentially achieve identifiability. If there are targets that are currently not available but could potentially be available at some cost, the analyst could do an exploratory analysis using collinearity analysis to determine if having these additional targets would improve the identifiability of the calibration problem. As a word of caution, seeking out additional data sources to resolve nonidentifiability might introduce bias in the calibrated parameters if the additional calibration targets are derived from a population different from the one being modeled (often called population bias[58]). Therefore, it is important that the target does represent the population of interest, or if this is not feasible, such target should be modified to represent this population. If there is information on how these 2 populations differ, the target could be modified accordingly through different techniques, such as meta-analysis or bias analysis, with the latter being especially designed to account for transferability.[59]

Alternatively, other solutions could be constraining the parameter space using expert information or fixing a subset of the parameters at meaningful values to reduce the number of parameters estimated through calibration. Constraining the parameter space could be done through either adding optimization constrains to the parameter values on a maximum likelihood estimation setup or specifying bounds on the parameters through informative constrained priors (e.g., using uniform priors with user-defined lower and upper bounds) on a Bayesian setup.[37] To verify if either incorporating additional targets or constraining the parameter space helps to improve identifiability, the analyst could compute the collinearity index with this updated information.

To numerically compute the collinearity index or profile the likelihood, the model needs to be reevaluated multiple times at different parameter values. For problems with a relatively large number of parameters, the computational burden to conduct these analyses will increase, but this might not be a problem of concern. For example, the computational burden to conduct collinearity analysis on a realistic calibration setup, such as the calibration of the natural history model of CRC described above, is negligible compared to the number of evaluations needed for calibration. However, as the number of parameters increases, profiling the likelihood quickly becomes computationally intensive,[60] and inference stemming from the likelihood profile might be misleading.[61] In such cases, more efficient methods for exploring the parameter space can be employed. For example, one can use a direct search method, making sure to initialize it at different starting values to look for different converging points or employ a Bayesian approach that is able to recover the whole posterior distribution (e.g., Markov chain Monte Carlo methods).[5] This might not be feasible for simulation models that are computationally time-consuming, such as microsimulation or discrete-event simulation models. However, it is possible to construct statistical emulators (often called metamodels) of the original simulation model that in turn can be evaluated at a fraction of the time.[62] Emulators can and have been used to calibrate the parameters of the original simulation model.[1,63] Thus, different methods to check for nonidentifiability could be applied to the emulator.

Checking for the existence of nonidentifiability should be an important step in model calibration. In the presence of nonidentifiability, it is important to try to make the parameters identifiable by constraining the parameter space either through imposing optimization constraints or constrained priors in a Bayesian setup[33] or by incorporating additional calibration targets.[11] If these approaches are not possible or not sufficient to eliminate nonidentifiability, it is important to report the ranges of the equally good-fitting parameters sets[64] and an estimate of the uncertainty of the calibrated parameters (e.g., variance), which will tend to be high in nonidentifiable parameters.[33] Furthermore, a sensitivity analysis should be conducted on the policy implications of evaluating the different best-fitting solutions, such as we did in Table 1.

In this article, we showed that nonidentifiability not only is a potential problem in complex simulation models but can also occur in simple models. Checking for the existence of nonidentifiability should be an important step of model calibration. If nonidentifiability is present, its effects on the recommendations driven from the mathematical model should be assessed.

Acknowledgments

We thank Hawre Jalal, Eric F. Lock, and Bryan Dowd for their discussion on this topic with the authors.

ORCID iD

Fernando Alarid-Escudero https://orcid.org/0000-0001-5076-1172

Supplementary Material

Supplementary material for this article is available on the *Medical Decision Making* Web site at http://journals.sagepub.com/home/mdm.

References

1. Kennedy MC, O'Hagan A. Bayesian calibration of computer models. *J R Stat Soc Ser B*. 2001;63(3):425–64.
2. AHRQ. Decision and simulation modeling: review of existing guidance, future research needs, and validity assessment. Available from: http://effectivehealthcare.ahrq.gov/ehc/products/598/1965/modeling-review-draft-140912.pdf
3. Higdon D, Kennedy M, Cavendish JC, Cafeo JA, Ryne RD. Combining field data and computer simulations for calibration and prediction. *SIAM J Sci Comput*. 2004;26(2):448–66.
4. Campbell K. Statistical calibration of computer simulations. *Reliab Eng Syst Saf*. 2006;91(10):1358–63.
5. Gustafson P. *Bayesian Inference for Partially Identified Models: Exploring the Limits of Limited Data*. Boca Raton, FL: CRC Press; 2015.
6. Wang W. Identifiability of linear mixed effects models. *Electron J Stat*. 2013;7(1):244–63.
7. Arendt PD, Apley DW, Chen W. Quantification of model uncertainty: calibration, model discrepancy, and identifiability. *J Mech Des*. 2012;134(10):100908.
8. Arendt PD, Apley DW, Chen W, Lamb D, Gorsich D. Improving identifiability in model calibration using multiple responses. *J Mech Des*. 2012;134(10):100909.
9. Basu S, Galvani AP. Re: "Multiparameter calibration of a natural history model of cervical cancer." *Am J Epidemiol*. 2007;166(8):983.
10. Rutter CM, Miglioretti DL, Savarino JE. Bayesian calibration of microsimulation models. *J Am Stat Assoc*. 2009;104(488):1338–50.
11. Rutter CM, Zaslavsky AM, Feuer EJ. Dynamic microsimulation models for health outcomes: a review. *Med Decis Making*. 2011;31(1):10–8.
12. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Econ*. 2006;15(12):1295–310.

13. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling good research practices—overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. *Med Decis Making*. 2012;32(5):667–77.
14. Briggs AH, Weinstein MC, Fenwick EAL, Karnon J, Sculpher MJ, Paltiel AD. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. *Med Decis Making*. 2012;32(5):722–32.
15. Garnett GP, Kim JJ, French K, Goldie SJ. Modelling the impact of HPV vaccines on cervical cancer and screening programmes. *Vaccine*. 2006;24(Suppl. 3):178–86.
16. Erenay FS, Alagoz O, Banerjee R, Cima RR. Estimating the unknown parameters of the natural history of metachronous colorectal cancer using discrete-event simulation. *Med Decis Making*. 2011;31(4):611–24.
17. Russell LB. Exploring the unknown and the unknowable with simulation models. *Med Decis Making*. 2011;31(4): 521–3.
18. Bilcke J, Chapman R, Atchison C, et al. Quantifying parameter and structural uncertainty of dynamic disease transmission models using MCMC: an application to rotavirus vaccination in England and Wales. *Med Decis Making*. 2015;35(5):633–47.
19. Vanni T, Karnon J, Madan J, et al. Calibrating models in economic evaluation: a seven-step approach. *Pharmacoeconomics*. 2011;29(1):35–49.
20. Karnon J, Vanni T. Calibrating models in economic evaluation: a comparison of alternative measures of goodness of fit, parameter search strategies and convergence criteria. *Pharmacoeconomics*. 2011;29(1):51–62.
21. Enns EA, Cipriano LE, Simons CT, Kong CY. Identifying best-fitting inputs in health-economic model calibration: a Pareto frontier approach. *Med Decis Making*. 2015;35(2): 170–82.
22. Welton NJ, Ades AE. Estimation of Markov chain transition probabilities and rates from fully and partially observed data: uncertainty propagation, evidence synthesis, and model calibration. *Med Decis Making*. 2005;25(6): 633–45.
23. Karnon J, Goyder E, Tappenden P, et al. A review and critique of modelling in prioritising and designing screening programmes. *Health Technol Assess*. 2007;11(52):iii–iv, ix–xi, 1–145.
24. Enns EA, Brandeau ML, Igeme TK, Bendavid E. Assessing effectiveness and cost-effectiveness of concurrency reduction for HIV prevention. *Int J STD AIDS*. 2011;22(10): 558–67.
25. Lehmann EL, Casella G. *Theory of Point Estimation*. 2nd ed. New York: Springer; 1998.
26. van der Steen A, van Rosmalen J, Kroep S, et al. Calibrating parameters for microsimulation disease models: a review and comparison of different goodness-of-fit criteria. *Med Decis Making*. 2016;36(5):652–65.
27. Stout NK, Knudsen AB, Kong CY (Joey), Mcmahon PM, Gazelle GS. Calibration methods used in cancer simulation models and suggested reporting guidelines. *Pharmacoeconomics*. 2009;27(7):533–45.
28. Jacquez JA. The inverse problem for compartmental systems. *Math Comput Simul*. 1982;24(6):452–9.
29. Jacquez JA, Greif P. Numerical parameter identifiability and estimability: integrating identifiability, estimability, and optimal sampling design. *Math Biosci*. 1985;77(1–2):201–27.
30. Bellman R, Åström KJ. On structural identifiability. *Math Biosci*. 1970;7(3–4):329–39.
31. Bickel PJ, Doksum KA. *Mathematical Statistics: Basic Ideas and Selected Topics*. 2nd ed. Upper Saddle River, NJ: Prentice-Hall; 2001.
32. Casella G, Berger R. *Statistical Inference*. 2nd ed. Pacific Grove, CA: Duxbury; 2002.
33. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*. 3rd ed. Boca Raton, FL: CRC Press; 2014.
34. Raue A, Kreutz C, Maiwald T, et al. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*. 2009;25(15):1923–9.
35. Raue A, Kreutz C, Theis FJ, Timmer J. Joining forces of Bayesian and frequentist methodology: A study for inference in the presence of non-identifiability. *Philos Trans R Soc A Math Phys Eng Sci*. 2012;371(1984):20110544.
36. Fröhlich F, Theis FJ, Hasenauer J. Uncertainty analysis for non-identifiable dynamical systems: profile likelihoods, bootstrapping and more. Available from: http://link.springer.com/chapter/10.1007/978-3-319-12982-2_5
37. Belsley DA. *Conditioning Diagnostics: Collinearity and Weak Data in Regression*. New York: John Wiley; 1991.
38. Brun R, Reichert P, Ku HR. Practical identifiability analysis of large environmental simulation models. *Water Resour Res*. 2001;37(4):1015–30.
39. Omlin M, Brun R, Reichert P. Biogeochemical model of Lake Zurich: sensitivity, identifiability and uncertainty analysis. *Ecol Modell*. 2001;141(1–3):105–23.
40. Berger JO, Liseo B, Wolpert RL. Integrated likelihood methods for eliminating nuisance parameters. *Stat Sci*. 1999;14(1):1–28.
41. Kreutz C, Raue A, Kaschek D, Timmer J. Profile likelihood in systems biology. *FEBS J*. 2013;280(11):2564–71.
42. SEER. Relative survival. Available from: https://seer.cancer.gov/seerstat/WebHelp/Relative_Survival.htm
43. Martin TA, Ye L, Sanders AJ, Lane J, Jiang WG. *Cancer Invasion and Metastasis: Molecular and Cellular Perspective*. Austin, TX: Landes Bioscence; 2013. Available from: https://www.ncbi.nlm.nih.gov/books/NBK164700/
44. Chaffer CL, Weinberg RA. A perspective on cancer cell metastasis. *Science*. 2011;331(6024):1559–64.
45. Ederer F, Axtell LM, Cutler SJ. The relative survival rate: a statistical methodology. *Natl Cancer Inst Monogr*. 1961;6:101–21.
46. Huszti E, Abrahamowicz M, Alioum A, Binquet C, Quantin C. Relative survival multistate Markov model. *Stat Med*. 2012;31(3):269–86.

47. Krijkamp E, Alarid-Escudero F, Enns EA, Jalal H, Hunink MGM, Pechlivanoglou P. Microsimulation modeling for health decision sciences using R: a tutorial. *Med Decis Making*. 2018;38(3):400–22.
48. Nelder JA, Mead R. A simplex method for function minimization. *Comput J*. 1965;7(4):308–13.
49. Soetaert K, Petzoldt T. Inverse modelling, sensitivity and Monte Carlo analysis in R using package FME. *J Stat Softw*. 2010;33(3). Available from: http://www.jstatsoft.org/article/view/v033i03
50. Wu GH-M, Wang Y-M, Yen AM-F, et al. Cost-effectiveness analysis of colorectal cancer screening with stool DNA testing in intermediate-incidence countries. *BMC Cancer*. 2006;6:136.
51. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. *Med Decis Making*. 2012;32(15):678–89.
52. Siebert U, Alagoz O, Bayoumi AM, et al. State-transition modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-3. *Med Decis Making*. 2012;32(5):690–700.
53. Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, Möller J. Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-4. *Med Decis Making*. 2012;32(15):821–7.
54. Pitman R, Fisman D, Zaric GS, et al. Dynamic transmission modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-5. *Med Decis Making*. 2012;32(5):712–21.
55. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. *Med Decis Making*. 2012;32(15):733–43.
56. Taylor DC, Pawar V, Kruzikas DT, Gilmore KE, Sanon M, Weinstein MC. Incorporating calibrated model parameters into sensitivity analyses: deterministic and probabilistic approaches. *Pharmacoeconomics*. 2012;30(2):119–26.
57. Greene WH. *Econometric Analysis*. 7th ed. Upper Saddle River, NJ: Pearson; 2012.
58. Turner RM, Spiegelhalter DJ, Smith GCS, Thompson SG. Bias modelling in evidence synthesis. *J R Stat Soc*. 2009;172(1):21–47.
59. Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. *Int J Epidemiol*. 2014;43(6):1969–85.
60. Raue A, Kreutz C, Maiwald T, et al. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Suppl Bioinform*. 2009;25(15):1923–9.
61. McCullagh P, Nelder JA. *Generalized Linear Models*. 2nd ed. Boca Raton, FL: Chapman and Hall/CRC; 1989.
62. Kleijnen JPC. Design and analysis of simulation experiments. 2nd ed. New York, NY: Springer International Publishing; 2015.
63. Farah M, Birrell P, Conti S, de Angelis D. Bayesian emulation and calibration of a dynamic epidemic model for A/H1N1 influenza. *J Am Stat Assoc*. 2014;109(508):1398–411.
64. Ballnus B, Hug S, Hatz K, Görlitz L, Hasenauer J, Theis FJ. Comprehensive benchmarking of Markov chain Monte Carlo methods for dynamical systems. *BMC Syst Biol*. 2017;1163(11):1–18.