Calibrating a Discrete-Event Simulation for Quantification of Sex-Specific Colorectal Neoplasia Development

Carolina Vivas-Valencia
  Purdue University West Lafayette

Nan Kong (✉ nkong@purdue.edu)
  Purdue University West Lafayette

Aditya Sai
  Carbon Health

Thomas F Imperiale
  Indiana University School of Medicine, Indiana University

Abstract

Background: Medical evidence collected from new observational studies can sometimes significantly alter our understanding of disease incidence and progression. This requires efficient and accurate calibration of disease models to help quantify the differences between observed cohorts. However, in model calibration, it is common to encounter overfitting with many model parameters but few observational outcomes. Additionally, the difficulty in evaluating fitting performance is significant due to a large degree of outcome variation and expensive computations for even a single simulation run.

Methods: We developed a two-phase calibration procedure to address the above challenges. As a proof-of-the-concept study, we verified the procedure with a discrete-event-simulation-based study on sex-specific colorectal neoplasia development. For the study, we estimated eight disease model parameters that govern colorectal adenoma incidence risk and growth rates at three distinct states: non-advanced, advanced adenoma, and adenoma becoming cancerous. For the calibration, we defined the likelihood measure by a relative weighted sum-of-squares difference between the three actual prevalence values reported in a recent publication and those predicted by a discrete-event colorectal cancer simulation. In phase I of the calibration procedure, we performed a series of low-dimensional sampling-based grid searches to identify reasonably good candidate parameter designs. In phase II, we performed a local search-based approach to further improve the model fit.

Results: Overall, our two-phase procedure showed better goodness of fit than a straightforward implementation of the Nelder-Mead algorithm, yielding a 10-fold reduction in calibration error (0.0025 vs. 0.0251 for an all-white mixed-family-history male cohort on the likelihood measure defined above). Further, the two-phase procedure was more effective in calibrating a validated simulation model for a female cohort than a male cohort. Finally, in phase II, performing local search on each of the parameters sequentially is more effective than searching the entire parameter space simultaneously.

Conclusions: The proposed two-phase calibration procedure is effective for estimating computationally expensive stochastic dynamic disease models. In addition, initial parameter search range truncation and sensitivity analysis on various parameters can be computationally cost-effective.

Background

Colorectal cancer (CRC) is the third most commonly diagnosed cancer in the United States [1]. In 2017, there were roughly 135,000 new CRC cases, with 45% of men and 39% of women younger than 65 at the age of diagnosis [2]. The disease starts with non-cancerous polyps growing in the walls of the colon and rectum. Then the disease invades other parts of the body through lymph and blood vessels. The five-year survival rate is 90% when cancer is confined to the colon and rectum, whereas the five-year survival rate declines to 12% when CRC has spread to distant locations[3]. Hence, for CRC prevention and treatment, it is important to detect precancerous polyps early before they evolve to cancerous polyps. Colonoscopy is the screening test that can most accurately detects polyps. In addition, during colonoscopy, precancerous

adenomas can be removed, which contributes to CRC incidence reduction. However, colonoscopy is not universally accepted among the screen-eligible population. And while considered cost-effective, it is the most costly screening test too. These drawbacks give rise to the need for alternative screening tests, including stool-based tests (for occult blood with or without DNA mutations), flexible sigmoidoscopy, and virtual colonoscopy, which are less costly and less invasive. However, these methods are much less accurate than colonoscopy. The U.S. Preventive Services Task Force recommend these tests for CRC diagnosis with some distinction; see[4] for more detail.

On the other hand, there is a need to improve the prediction of precancerous polyp progression so that CRC diagnostic screening and surveillance may be better targeted to those at high-risk for rapid progression. In clinical practice, patients are further classified by detection of advanced precancerous polyps, which include adenomas and sessile serrated polyps ≥ 10 mm and adenomas with villous histology or high-grade dysplasia [5], [6]. Individuals with an advanced adenoma are more likely to develop another advanced adenoma and CRC, as are persons with multiple non-advanced adenomas (4–6 mm)[7]. Thus, with improved prediction, we can conduct population-wide surveillance more effectively and cost-effectively, e.g., having differentiated risk of subsequent risk for advanced neoplasia (combination of advanced precancerous polyps and CRC).

It is well-established that sex plays an important role in risk for both advanced precancerous polyps and CRC. More men than women are diagnosed with CRC. While men and women have similar genetic predispositions in terms of adenoma-carcinoma sequence, there are substantial differences in cancer incidence between the two sexes [8], e.g., the American Cancer Society reported a 30% higher annual incidence rate in men than women in the United States [1]. Murphy et al. (2011) [9] found that the rates of finding cancer were higher for men than women at almost all anatomic subsites, based on data from the U.S. National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER). Several studies suggested that female patients diagnosed with CRC have significantly longer survival rates than male patients [10], [11]. The sex disparities have been associated with exposure to behavioral risk factors, including smoking [12], red-meat diet [13], and other lifestyle-related risk factors [14]–[16].

Furthermore, men have a higher prevalence of adenomas than women by a ratio of nearly 2:1. Ferlitsch et al. (2011)[17] reported that adenomas prevalence was higher among men than women by an absolute difference of 10%, through a study of more than 44 thousand participants of a national screening colonoscopy program in Austria. Regula et al. (2006)[18] reported that advanced precancerous polyp was found with a significantly higher percentage in men than women, from a study of more than 50,000 Polish participants. Brenner et al. (2013)[19] reported that adenoma prevalence (both advanced and non-advanced) was substantially higher in men than women for different age groups, from an observational study of more than 3.6 million German participants.

In summary, it is evident that men and women differ in the risk of adenoma progression. However, the current recommendations on diagnostic screening and surveillance have not taken the sex difference into account. Thus, our objective here is to quantify age- and sex-specific colorectal adenoma progression

with a computational model, which will in turn, help to inform the optimal age for each sex to initiate colonoscopy testing. In this paper, we conduct stochastic modeling of age- and sex-specific dwell durations at four states (i.e., adenoma before initiation, non-progressive non-advanced adenoma, progressive non-advanced adenoma, and advanced adenoma). Since clinical data does not provide necessary information about age- and sex-specific adenoma progression, the above unknown quantities cannot be directly modeled from clinical data using conventional statistical methods such as regression. In response, we have adapted a discrete-event simulation model and performed model calibration to estimate progression parameters against sex-specific prevalence data on key disease stages, i.e., non-advanced adenoma, advanced adenoma, and adenoma becoming cancerous.

Only a handful number of papers in the current CRC disease modeling literature have reported their model calibration work in detail. In several studies [20]–[22], the authors conducted model calibration to estimate unobservable disease progression parameters against benchmark statistics (e.g., national CRC incidence). More recently, Erenay et al. (2011) [23] developed an individually-based event-driven state transition simulation that mimics the natural history of metachronous colorectal cancer (MCRC) for a 5-year period following the treatment of primary CRC. The model comprises five states, namely polyp free, polyp, MCRC, metastatic-MCRC, and MCRC-related death. The authors estimated six unknown parameters of the natural history of MCMC through calibrating the simulation mentioned above against two calibration targets, 5-year MCRC incidence and mortality rate, in principle of least sum-of-squared error of the two calibration targets. For the calibration, the authors simply ran the simulation model exhaustively with every possible combination of the unknown parameters and selected those with simulated outputs matching the benchmark statistics of a well-defined patient cohort, derived from the SEER database. Rose et al. (2014) [24] proposed an individually-based state transition model consisting of two interacting submodels: a continuous-time disease-progression submodel and a discrete-time Markov submodel for surveillance and retreatment. The key components on disease progression are recurring transitions to unresectability and to the point of symptom onset, either of which is determined by the transition timing and modeled with an exponential distribution. The author estimated seven unknown parameters of disease progression through calibrating the simulation mentioned above against seven observable outcomes, reported by Pietra et al. (1998) [25]. The authors developed an efficient calibration procedure that consists of several rounds of calibration with increasingly narrowed candidate parameter sets and against a series of specific calibration targets.

In this research, we adapted the Vanderbilt-NC State (V/NCS) simulation model, which is a discrete-event-based stochastic model emulating the age- and sex-dependent growth of each adenoma. To calibrate this computationally expensive simulator, we used a two-phase approach to estimate the values of eight unknown model parameters over an ample design space. The first phase used sampling to identify reasonably good candidate parameter designs; whereas the second phase used local search to further improve the model fitting. We set the sum of squared deviation of the prevalence values as the loss function to minimize and used benchmark statistics of three disease states (e.g., non-advanced adenoma, advanced adenoma, and adenoma becoming cancerous) extracted from a German cohort study by Brenner et al. (2013) [19]. Brenner and colleagues analyzed national screening colonoscopy

registry data from nearly 3.6 million German participants. To facilitate the calibration, we relied on subject matter expertise to select both model parameters and target responses for the calibration procedure. To efficiently adapt the local search idea, we compared two variants of the optimization procedure (more axial based vs. globally). At the end, we quantified the sex-specific adenoma-carcinoma sequence for different age groups.

Our approach may be applied to calibrating "black-box" disease models with many unknown input parameters, wide value ranges, and multiple target outcomes. Our main contributions are 1) development of an efficient calibration procedure for complex discrete-event disease simulation models; and 2) quantification of age-dependent sex differences on the adenoma-carcinoma sequence based on empirical observations.

Methods

Overview of approach

In this study, we used the Vanderbilt-NC State (V/NCS) discrete-event microsimulation model developed by researchers at Vanderbilt and North Carolina State University [20].

The V/NCS model mimics the natural history of colorectal neoplasia for each hypothetic entity in the cohort. In addition, the model can be used to evaluate colorectal neoplasia screening strategies. One can specify the input cohort by sex, birth year, and family history of each simulated entity. Then for each entity, discrete events are simulated along the adenoma-carcinoma sequence. In the V/NCS model, events trigger changes at discrete times to the states of each adenoma along the progression pathways and the hypothetic person as a whole. These events lead to the collection of statistics and the creation of new events. Events that are relevant to our work include *new person creation*, *natural death*, *cancer death*, *non-advanced adenoma incidence*, *advanced adenoma incidence*, *cancer incidence from an adenoma*. Other events in the model include *regional cancer*, *distance cancer*, *cancer symptomatic*, *terminal cancer*, *colonoscopy*, *recover from cancer*, *terminal cancer charge*, and *age-based utility*. For more information, we refer to Roberts et al. (2007) [20] .

The purpose of this study is to estimate a set of model parameters that specify four quantities of cohort heterogeneity, which are used to govern the progression of each adenoma created and consequently the natural history of colorectal neoplasia. These quantities cannot be directly observable in an observational study and sex-specific estimates of these parameters are not available to simulation models currently available in the literature. Further, it is somewhat unethical to follow the continued progression of an adenoma once it is observed in clinical practice. Instead, polypectomy is recommended, and the natural progression is halted. Therefore, we resorted to calibrating the V/NCS model against prevalence data from a published study of Brenner et al. (2013)[19]. While the V/NCS model offers sufficient fidelity to the adenoma-carcinoma sequence and lends flexibility on candidate parameter design selection with a user-friendly interface, we faced a severe computational burden. It took 40–50 minutes to do one simulation

run with a cohort size of 10,000 on a regular personal computer. In response, we developed an efficient two-phase calibration procedure.

Disease Progression Modeling In The V/ncs Model

CRC begins as a non-visible, benign adenoma. Once such an adenoma appears, it transitions to the next stage, depending on the pathway to cancer it follows. The V/NCS model includes three types of progression (i.e., pathways to cancer): non-progressive (i.e. #1 in Fig. 1), slowly progressive (#2), and immediately progressive (#3). The first type is non-progressive. An adenoma of this type has no chance ever to become cancerous, but can grow as a benign adenoma (i.e., advanced adenoma) to match the data on the portion of adenomas that can be detected. The second type is progressive. A non-advanced adenoma of this type can either become an advanced adenoma defined by its histology or become cancerous directly, with the former being more common. The transition of this type is then modeled as a competing process between the above two possibilities. The third type of progression is immediately progressive, which implies that an adenoma with this type immediately progresses to becoming cancerous upon its initiation. Regardless of the second or third type, as long as an adenoma becomes cancerous, it follows the usual CRC pathway (i.e., from pre-clinical phase to clinical phase and through cancer stages).

In the V/NCS model, there are several key assumptions on adenoma incidence and progression. First, to each individual, the incidence and progression of each adenoma are independent of other adenomas. Second, both adenoma incidence rate and progression time are influenced by an individualized risk (Liebsch 2003) [26]. Obviously, family history (i.e., presence of CRC in first-degree relatives) is an important characteristic affecting the risk. Sex and race, to a lesser degree, also affect the risk. Specific risks to individuals may further include factors like fibrous diet, lack of exercise, familiar adenomatous polyposis, and heredity nonpolyposis CRC, but the exact effect of these factors is not known and therefore not included in the model. The risk is modeled with a JohnsonSB distribution for individuals with or without family histories. In both cases, the JohnsonSB distribution has a minimum of 0.0 and a maximum of 1.0, which implies the absolute individual risk on a scale between 0.0 and 1.0. Further, both distributions are highly positively skewed with a mode close to zero, and a mean of 0.11 for individuals with family history and a mean of 0.056 for individuals without any family history. These additional specifications match with the published research findings through a meta-analysis (Johns and Houlston 2001) [27].

In addition to the risk adjustment, the baseline adenoma incidence follows a non-homogenous Poisson process with the incidence rate modeled by an age-dependent piecewise linear function. The baseline time taken to progress to an advanced adenoma either from a non-progressive non-advanced adenoma (i.e., NP_NON) or from a slowly progressive non-advanced adenoma (i.e., P_NON) follows a JohnsonSB distribution with a minimum of 0.0 and a maximum of 60.0. The baseline time for advanced adenoma becoming cancerous also follows a JohnsonSB distribution with a minimum of 0.0 and a maximum of

60.0. Given that an adenoma does not occur likely before the age of 40, the maximum of 60.0 implies the longest allowable time for making a transition is 60 years and thus gives sufficient time-lapse during an individual's lifetime. For the above three quantities, the actual transition time is further adjusted by the personal risk.

We list several additional assumptions on adenoma progression and cancer incidence in the V/NCS model that are only indirectly relevant to our calibration work as follows. First, the distribution of adenoma to progression type is dependent upon age. As the body repair mechanism deteriorates, the ability of the body to deal with abnormal cells begins to decline. Thus, it is assumed that the percentage of progressive adenomas increases as a person's age increases. Second, for immediately progressive adenomas, they become cancerous as soon as they are initiated. Since these adenomas progress immediately, it is further assumed that they then only take no more than 10 years until cancer becomes symptomatic. The actual duration follows a JohnsonSB distribution as well. Third, the time to cancer incidence follows a JohnsonSB distribution with a mean of 20 years and a mode of 22 years.

V/NCS model assumptions on cancer staging and mortality include (1) regional and distance metastasis rates are independent processes; (2) pre-clinical cancer stage progression and symptom development are lesion-specific and independent of personal risk; (3) times to clinical cancer at the regional and distance stages follow Johnson SB distributions; and (4) rate of progression of death, and potential survival (from CRC) is determined by cancer stage at the time of symptoms.

V/ncs Model Adaptation For The Calibration

With the V/NCS model platform, one can input a realistic population that matches the U.S. census or a hypothetic population of any arbitrary size, and enter a specific simulation start year. The simulation can then trace colorectal neoplasia development of the cohort to a pre-determined end year. One beneficial feature of the V/NCS model platform is that it generates a trace statement that summarizes a sequence of time-stamped events capturing disease development. To process the trace statement, we developed a procedure to extract all the events that take place in the input population (see Fig. 2). For each individual, we created a state transition chart to record the time at which his/her events occur. By following each individual through the simulation duration, one can characterize the health state (i.e., NOV, ADV, and CRC) of each individual at any specific point. The state NOV includes individuals who have had at least one P_NON, NP_NON, or at least one adenoma that immediately progresses to cancer, but no advanced adenoma. The state ADV includes those individuals who have had at least one advanced adenoma but none has become cancerous. The state CRC includes those individuals who have had cancerous adenomas or have developed CRC. For each of the five age groups (54–59, 60–64, 65–69, 70–74 and 75–79), we then counted its population at the end of the simulation horizon (i.e., at a particular year) and calculated the portion of the corresponding population subgroup in each of the three states as the three corresponding prevalence values.

To adapt the V/NCS model for our calibration study, we made further specifications on the model calibration. First, we set an all-white mixed-family-history cohort in the V/NCS model. Second, we set the base incidence of non-visible adenomas to the same as in the original version of the V/NCS model. Note that there is no evidence showing significant differences in the division of having family history and not in an all-white cohort in the US and the Germany cohort studied by Brenner et al. (2014) [28], which is believed to be predominantly white. Third, we assumed that the time for some non-advanced progressive adenoma (i.e., P_NON) directly becoming cancerous does not change from the original version of V/NCS to our work. As it is quite rare for an adenoma to follow this pathway, we did not expect this fixation affects the sex-based comparative results significantly.

Consequently, our calibration efforts were focused on quantifying the following random variates: (1) individual risk; (2) if an adenoma created is progressive, its baseline dwell duration in the non-advanced adenoma state before transitioning to the advanced precancerous polyps state (termed transition time of P_NON to ADV); (3) if the adenoma created is non-progressive, its baseline dwell duration in the non-advanced adenoma state before transition to the advanced adenoma state (termed transition time of NP_NON to ADV); and (4) if the adenoma is in the advanced adenoma state, its baseline dwell duration before transitioning to the cancerous adenoma state (termed transition time of ADV to CRC). Further, as stated earlier, each of the above four random variables is modeled with a JohnsonSB distribution. A JohnsonSB distribution is a four-parameter distribution family where the shape and location of the distribution are governed by two model parameters δ and γ. The other two parameters, *minimum* and *maximum*, specify the scale. For these four random variates, we assumed the minimum and maximum to be unchanged with the Germany cohort. Hence, for either sex, we had eight calibration variables, i.e., four δ's and γ's, to estimate based on three system responses, namely sex-specific age-group-aggregate NON, ADV, and CRC percentages.

A Two-phase Calibration Procedure

We took a two-phase approach for the model calibration. In phase I (a preliminary phase), we performed a series of low-dimensional searches on subsets of model parameters (i.e., a pair of δ and γ) in an ad-hoc manner. We conducted these searches progressively against varied calibration targets that are aggregate over age groups. The purpose was to identify a set of promising values on each model parameter, which would serve as the multiple starting points (parameter designs) in phase II. The sequence of these searches and design inclusion criteria were developed from discussions with a domain expert. In phase II, we viewed the model calibration task as a nonlinear optimization problem and performed the Nelder-Mead algorithm (simplex search algorithm), one of the best-known algorithms for multidimensional unconstrained optimization without derivatives. Given the high-dimensionality of the "black-box" optimization problem, we explored two variants of the search procedure, namely one-shot globally over the entire model parameter space and sequentially based on interconnections in subsets of model parameters. Design of the loss functions was consulted with the domain expert as well. As for our calibration targets, we used age-specific prevalence data from Brenner et al. [18]. More specifically, we

used the prevalence of both men and women in five age groups (54−59, 60−64, 65−69, 70−74 and 75−79).

Phase I (Preliminary Phase) -- Identify promising initial search points for Phase II.

In our study, it is computational challenging to quantify the joint influence of the eight input model parameters (calibration variables) on the fifteen age-group-specific system responses. We thus performed low-dimensional searches on the four pairs of model parameters over respective pre-specified search ranges. Note that the original version of the V/NCS model was previously calibrated against the SEER (Surveillance, Epidemiology and End Results Program) data for U.S. populations. We consulted the domain expert to finalize each search range, which is centered at the previously calibrated value. Through our preliminary simulation analysis, we observed that in each pair of δ and γ, the prevalence values are a lot more sensitive to changes in δ than in γ. Thus, we set a larger range for each δ than the paired γ. We divided the search subspace of (δ_0, γ_0) with a five-by-five grid and divided each of $\delta_1, \gamma_1, \delta_2, \gamma_2, \delta_3, \gamma_3$ with ten even intervals.

Basically, we followed the adenoma-carcinoma sequence to calibrate the model parameters progressively, i.e., first adenoma progression propensity, then transition from NON to ADV, and finally transition from ADV to CRC. In the first step, we performed a grid search on (δ_0, γ_0) and fixed the other parameter values as it is first to mimic the adenoma progression risk distribution of the simulated cohort. Our calibration targets are all three prevalence values over age groups. At the end of this step, we identified promising (δ_0, γ_0) values such that all three predicted prevalence values are reasonably close to the observations (less than 15% relative error). Next, we fixed (δ_0, γ_0) values to the identified ones and performed orthogonal sampling in the subspace formed by (δ_1, γ_1). The use of a sampling-based search as opposed to a grid search is because multiple promising (δ_0, γ_0) values were identified and thus using all of them for ensuing search would be computationally expensive. Our calibration targets are aggregate prevalence values of NON and ADV over age groups. We then followed the same idea to search in the subspace formed by (δ_2, γ_2) and again used the same calibration targets. We perturbed (δ_1, γ_1) first because there were many more transitions from P_NON to ADV than from NP_NON to ADV. At the end of this step, we identified promising (δ_1, γ_1) and (δ_2, γ_2) designs such that both predicted prevalence values (i.e., at states NON and ADV) were further closer to the observations (less than 10% relative error). Finally, we fixed (δ_0, γ_0), (δ_1, γ_1), (δ_2, γ_2) to be the identified values and performed orthogonal sampling in the subspace formed by (δ_3, γ_3). Our calibration targets are aggregate prevalence values of NON, ADV, and CRC over age groups. At the end of this step, we identified promising (δ_3, γ_3) designs such that all three predicted prevalence values fall in a close range of the target values (less than 10% relative error on NON, less than 10% relative error on ADV, and less than 5% relative error on CRC).

For any identified parameter values, we used a built-in interactive visual tool to graph the corresponding Johnson SB distributions (mainly their shapes and locations) and checked them with our domain expert. We then discarded some of the parameter designs according to the domain expert's suggestions.

Phase 2. Local-Search based Nonlinear Optimization

In this phase, we employed the Nelder-Mead algorithm for gradient-free nonlinear optimization to further improve the model fitting. We set the parameter designs identified on Phase 1 as the starting points for the Nelder-Mead algorithm. We used the weighted sum squared of the relative errors on the three aggregate prevalence values as a similarity measure and the objective function of the unconstrainted nonlinear optimization problem (i.e., loss function of the calibration variables). Through consulting with our domain expert, we assigned a larger weight to the set of CRC similarities than to the set of ADV similarities and the set of NON similarities.

Considering the resultant optimization problem is computationally expensive due to the fact that it takes a long time to evaluate just one parameter design, we designed two search paths that differ by the search space chosen along the solution process. In the base case algorithm, we considered an 8-dimensional search space at any moment of the solution (i.e., all eight model parameters are possible to be varied). We termed this strategy the "full-space local search strategy." As an alternative option, we considered four 2-dimensional subspaces progressively (i.e., (δ_0, γ_0) first, then (δ_1, γ_1), then (δ_2, γ_2), then (δ_3, γ_3)). We termed this strategy the "sequential local search strategy." When we performed Nelder-Mead in one subspace, other were fixed at the initial values. The above order used in this alternative algorithm followed the same idea as in phase I. That is, the system responses are more sensitive to (δ_0, γ_0), than (δ_1, γ_1), than (δ_2, γ_2), and finally (δ_3, γ_3).

Results

We set the simulated cohort in the V/NCS model to be either a population of 1000 white males or females with no family history and born in 1949. In this way, we were able to utilize additional parameters previously made available in the V/NCS model, e.g., mortality rate. We first examined the efficiency of our two-phase approach in comparison with the straightforward execution of the Nelder-Mead algorithm. For any implementation of the Nelder-Mead algorithm, we simply called the corresponding Matlab function and ran it to natural termination with default algorithm parameter settings. To make fair comparisons, we also set the same two search strategies, i.e., full-space and sequential. We report the comparative results in Table 1.

Overall, our two-phase approach showed better goodness of fit than the straightforward Nelder-Mead implementation. For example, both with the sequential search strategy for Nelder-Mead, the two-phase approach yielded a loss function value of 0.0025 whereas the straightforward calibration with Nelder-Mead yielded a loss function value of 0.0251 (ten-fold reduction). Furthermore, when comparing the two local search strategies, the full-space search strategy yielded a lower loss function value than the

sequential search strategy (male, 0.0025 vs. 0.0056; female, 0.0005 vs. 0.0008). Finally, these preliminary results suggested that the two-phase procedure was more effective in calibrating the V/NCS model for a female cohort than a male cohort.

Table 1
Calibration Strategy Comparison

	Male		Female	
	Sequential	Full-space	Sequential	Full-space
Direct local search	0.0251	0.0465	0.0230	0.2969
Two-phase approach	0.0025	0.0056	0.0005	0.0008

With the above result, we conclude that our two-phase procedure with the sequential local search strategy is more effective. We next specified the number of simulation replications to be 10 to ensure the statistical confidence on stochastic dominance for each comparison. We collected prevalence statistics for five different age groups in the range of 55 to 79. Table 2 shows the percentage of people with advanced adenoma for men and women within each of the five age groups. Our results show that the model with calibrated variables underestimated male advanced adenoma prevalence and overestimated female advanced adenoma prevalence for younger age groups, whereas it overestimated male advanced adenoma prevalence and underestimated female advanced adenoma prevalence. On the other hand, the comparative results on the prevalence of adenomas having become cancerous were just the opposite except for the age group 55 to 59 years. Overall, our results supported our calibration of the V/NCS simulation for sex-specific colorectal neoplasia development modeling.

	Percentage (%) people with advanced adenoma				Percentage (%) people with cancerous neoplasia			
	Male	Target	Female	Target	Male	Target	Female	Target
55–59	6.25	6.6	3.85	3.50	0.73	0.60	0.27	0.30
60–64	7.63	8.2	4.98	4.50	0.82	1.00	0.56	0.50
65–69	10.41	9.2	5.38	5.30	0.85	1.30	1.03	0.70
70–74	11.31	9.9	5.76	6.40	2.20	1.90	0.92	1.10
75–79	13.17	10.4	6.71	6.80	2.62	2.50	1.44	1.60

Discussion And Conclusions

In this paper, we introduce an efficient and effective two-phase calibration procedure approach to estimate the natural history of unobservable CRC parameters. This work showcases an essential step in assessing population-level cost and effectiveness of CRC screening strategies for a population whose prevalence data were recently acquired. We took into consideration the adenoma-carcinoma sequence to calibrate the model parameters progressively, i.e., first adenoma progression propensity, then transition from NON to ADV, and finally transition from ADV to CRC. Moreover, we quantified the sex-specific adenoma-carcinoma sequence among different age groups based on observations from a large cohort study. Our results demonstrate that the combined first phase involving a direct local search and nonlinear optimization methods in the second phase increases the accuracy of model-based parameter estimation. In addition, the proposed model-based parameter estimation approach is a general procedure that could be extended to other disease model calibration problems that involve the adaptation of some robust simulation platform (e.g., the V/NCS).

While we developed an efficient model-based parameter estimation procedure, our study has several inherent limitations. First, the cost of the simulation is very expensive; therefore, we used a smaller cohort size, which is not comparable to the cohort studied by Brenner et al. (2014) [28]. Second, it is possible that when using Nelder-Mead in phase II, the algorithm converged on a local minimum and not a global minimum. Third, the use of calibration targets estimated from the German cohort may not reflect CRC natural history characteristics in the U.S., which can be crucial to further cost-effectiveness analysis on screening strategies for U.S. populations. Fourth, underweighted endpoints implemented in the objective function may result in biased calibration outputs, thus undermining the validity of our parameter estimation. Fifth, while the progressive calibration procedure could bring valuable insights into efficient

calibration for tailored natural history models, due to the limited access to the V/NCS model, some stages of the calibration procedure were based on empirical evidence and a few hunches.

Abbreviations

CRC
Colorectal Cancer; V/NCS:Vanderbilt-NC State simulation model

Declarations

Authors' contributions

C.V.V conducted the experiments and prepared the tables, C.V.V and N.K wrote the manuscript and prepared the figures, C.V.V, N.K, and T.F.E did the research conceptualization, A.S, and T.F.I reviewed the manuscript

References

1. American Cancer Society, "Cancer Facts & Figs. 2017," *American Cancer Society*, 2017.

2. R. L. Siegel *et al.*, "Colorectal Cancer Statistics, 2017," *CA Cancer J Clin*, vol. 67, no. 3, pp. 177–193, 2017, doi: 10.3322/caac.21395.

3. M. Fleming, S. Ravula, S. F. Tatishchev, and H. L. Wang, "Colorectal carcinoma: Pathologic aspects," *Journal of Gastrointestinal Oncology*, vol. 3, no. 3. pp. 153–173, 2012. doi: 10.3978/j.issn.2078-6891.2012.030.

4. K. Bibbins-Domingo *et al.*, "Screening for colorectal cancer: US preventive services task force recommendation statement," *JAMA - Journal of the American Medical Association*, vol. 315, no. 23, pp. 2564–2575, 2016, doi: 10.1001/jama.2016.5989.

5. W. Atkin *et al.*, "Adenoma surveillance and colorectal cancer incidence: a retrospective, multicentre, cohort study," *The Lancet Oncology*, vol. 18, no. 6, pp. 823–834, 2017, doi: 10.1016/S1470-2045(17)30187-0.

6. D. H. Kim, P. J. Pickhardt, and A. J. Taylor, "Characteristics of advanced adenomas detected at CT colonographic screening: Implications for appropriate polyp size thresholds for polypectomy versus surveillance," *American Journal of Roentgenology*, vol. 188, no. 4, pp. 940–944, 2007, doi: 10.2214/AJR.06.0764.

7. H. Brenner, M. Hoffmeister, C. Stegmaier, G. Brenner, L. Altenhofen, and U. Haug, "Risk of progression of advanced adenomas to colorectal cancer by age and sex: estimates based on 840,149 screening colonoscopies.," *Gut*, vol. 56, no. 11, pp. 1585–9, Nov. 2007, doi: 10.1136/gut.2007.122739.

8. A. White, L. Ironmonger, R. J. C. Steele, N. Ormiston-Smith, C. Crawford, and A. Seims, "A review of sex-related differences in colorectal cancer incidence, screening uptake, routes to diagnosis, cancer stage and survival in the UK," *BMC Cancer*, vol. 18, no. 1, 2018, doi: 10.1186/s12885-018-4786-7.

9. G. Murphy, S. S. Devesa, A. J. Cross, P. D. Inskip, K. A. Mcglynn, and M. B. Cook, "Sex Disparities in Colorectal Cancer Incidence by Anatomic Subsite, Race and Age," *Int J Cancer Int J Cancer April*, vol. 1, no. 1287, pp. 1668–1675, 2011, doi: 10.1002/ijc.25481.

10. K. E. Storli *et al.*, "Overall survival after resection for colon cancer in a national cohort study was adversely affected by TNM stage, lymph node ratio, gender, and old age," *International Journal of Colorectal Disease*, vol. 26, no. 10, pp. 1299–1307, 2011, doi: 10.1007/s00384-011-1244-2.

11. C. S. McArdle, D. C. McMillan, and D. J. Hole, "Male gender adversely affects survival following surgery for colorectal cancer," *British Journal of Surgery*, vol. 90, no. 6, pp. 711–715, 2003, doi: 10.1002/bjs.4098.

12. L. C. Chang, M. S. Wu, C. H. Tu, Y. C. Lee, C. T. Shun, and H. M. Chiu, "Metabolic syndrome and smoking may justify earlier colorectal cancer screening in men," *Gastrointestinal Endoscopy*, 2014, doi: 10.1016/j.gie.2013.11.035.

13. B. Bates, A. Lennox, C. Bates, and G. Swan, "A survey carried out on behalf of the Headline results from Years 1 and 2 (combined) of the," 2008.

14. D. J. Harriss *et al.*, "Lifestyle factors and colorectal cancer risk (2): A systematic review and meta-analysis of associations with leisure-time physical activity," *Colorectal Disease*, vol. 11, no. 7. John Wiley & Sons, Ltd, pp. 689–701, Sep. 01, 2009. doi: 10.1111/j.1463-1318.2009.01767.x.

15. T. Boyle, L. Fritschi, C. Platell, and J. Heyworth, "Lifestyle factors associated with survival after colorectal cancer diagnosis," *British Journal of Cancer*, vol. 109, no. 3, pp. 814–822, 2013, doi: 10.1038/bjc.2013.310.

16. R. R. Huxley, A. Ansary-Moghaddam, P. Clifton, S. Czernichow, C. L. Parr, and M. Woodward, "The impact of dietary and lifestyle risk factors on risk of colorectal cancer: A quantitative overview of the epidemiological evidence," *International Journal of Cancer*, vol. 125, no. 1, pp. 171–180, Jul. 2009, doi: 10.1002/ijc.24343.

17. M. Ferlitsch *et al.*, "Sex-specific prevalence of adenomas, advanced adenomas, and colorectal cancer in individuals undergoing screening colonoscopy.," *JAMA: the journal of the American Medical Association*, vol. 306, no. 12, pp. 1352–8, Sep. 2011, doi: 10.1001/jama.2011.1362.

18. J. Regula *et al.*, "Colonoscopy in colorectal-cancer screening for detection of advanced neoplasia.," *The New England journal of medicine*, vol. 355, no. 18, pp. 1863–72, Nov. 2006, doi: 10.1056/NEJMoa054967.

19. H. Brenner, "Natural History of Colorectal Adenomas: Birth Cohort Analysis Among 3.6 Million Participants of Screening Colonoscopy," 2013. http://cebp.aacrjournals.org/content/22/6/1043.full.pdf (accessed Jan. 11, 2015).

20. S. Roberts, L. Wang, R. Klein, R. Ness, and R. Dittus, "Development of a Simulation Model of Colorectal Cancer," *ACM Transactions onModeling and Computer Simulation*, vol. Vol. 18, 2007, Accessed: Jan. 11, 2015. [Online]. Available: http://delivery.acm.org/10.1145/1320000/1315579/a4-roberts.pdf?ip=128.46.90.19&id=1315579&acc=ACTIVE SERVICE&key = A79D83B43E50B5B8.2BA8E8EA4DBC4DB7.4D4702B0C3E38B35.4D4702B0C3E38B35&CFID = 617753096&CFTOKEN = 49416681&__acm__=1421096368_ea55 c572cc8b5d924

21. C. Hassan *et al.*, "Computed Tomographic Colonography to Screen for Colorectal Cancer, Extracolonic Cancer, and Aortic Aneurysm Model Simulation With Cost-effectiveness Analysis."

22. a L. Frazier, G. a Colditz, C. S. Fuchs, and K. M. Kuntz, "Cost-effectiveness of screening for colorectal cancer in the general population.," *JAMA: the journal of the American Medical Association*, vol. 284, no. 15, pp. 1954–61, Oct. 2000.

23. F. S. Erenay, O. Alagoz, R. Banerjee, and R. R. Cima, "Estimating the unknown parameters of the natural history of metachronous colorectal cancer using discrete-event simulation," *Medical Decision Making*, vol. 31, no. 4, pp. 611–624, 2011, doi: 10.1177/0272989X10391809.

24. J. Rose *et al.*, "A simulation model of colorectal cancer surveillance and recurrence," *BMC Medical Informatics and Decision Making*, vol. 14, no. 1, p. 29, 2014, doi: 10.1186/1472-6947-14-29.

25. N. Pietra, L. Sarli, R. Costi, C. Ouchemi, M. Grattarola, and A. Peracchia, "Role of Follow-Up in Management of Local Recurrences of Colorectal Cancer A Prospective, Randomized Study."

26. C. Liebsch, "Simulation Input Modeling in the Absence of Data.," 2003.

27. L. E. Johns and R. S. Houlston, "A systematic review and meta-analysis of familial colorectal cancer risk.," *The American journal of gastroenterology*, vol. 96, no. 10, pp. 2992–3003, 2001, doi: 10.1111/j.1572-0241.2001.04677.x.

28. H. Brenner, L. Altenhofen, C. Stock, and M. Hoffmeister, "Incidence of colorectal adenomas: Birth cohort analysis among 4.3 million participants of screening colonoscopy," *Cancer Epidemiology Biomarkers & Prevention*, vol. 23, no. September, pp. 1920–1928, 2014, doi: 10.1158/1055-9965.EPI-14-0367.
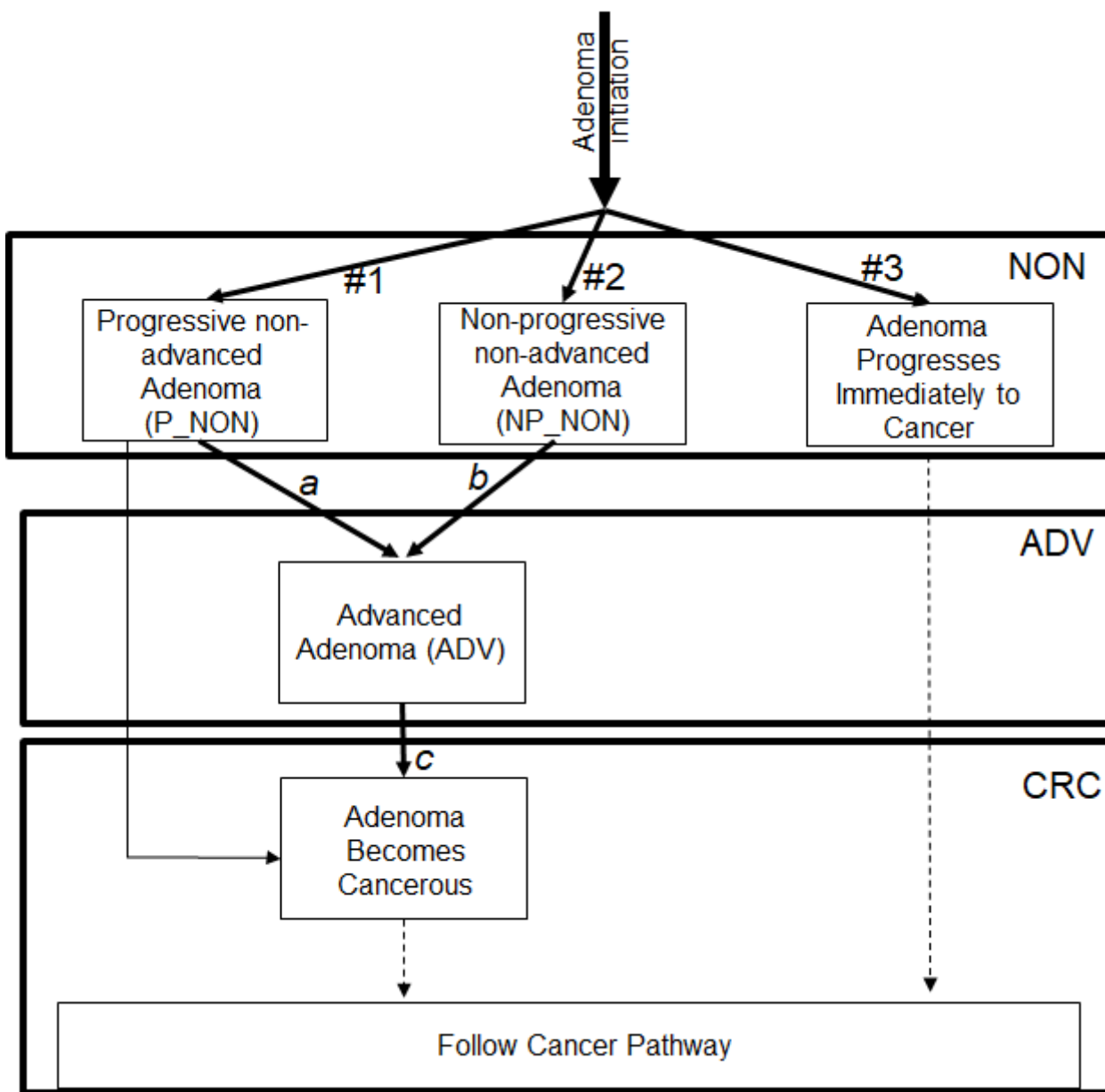
Figures

Figure 1

CRC Preclinical Pathway. (The three progression types are labeled with #1, #2, and #3 with the three thicker solid lines on top. The type distribution is dependent upon a personalized risk influenced by JohnsonSB($\delta 1$, $\gamma 1$, 0, 1). The three dwell durations are labeled with a, b, and c. The three corresponding random variables are also modeled with JohnsonSB distributions.)
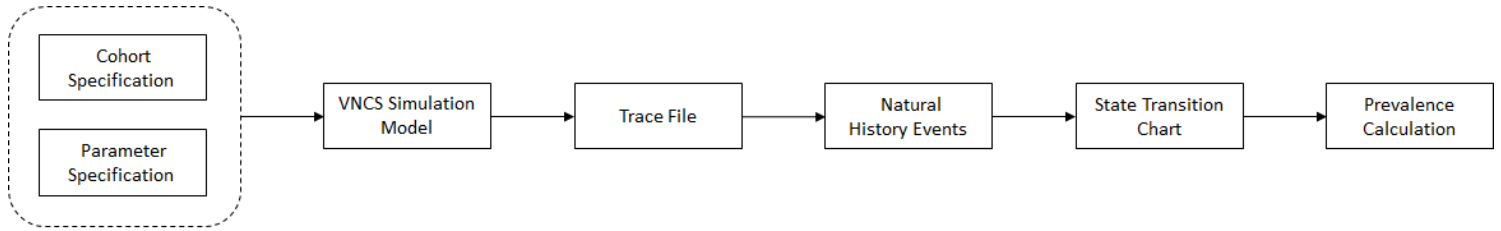

Figure 2

Predicted Outcomes Extraction