Discrete-Event Simulation Modeling Framework for Cancer Interventions and Population Health in R (DESCIPHR): An Open-Source Pipeline

Selina Pi[1*], Carolyn M. Rutter[2], Carlos Pineda-Antunez[3], Jonathan H. Chen[4,5,6], Jeremy D. Goldhaber-Fiebert[7,8], Fernando Alarid-Escudero[7,8*]

[1]Department of Biomedical Data Science, School of Medicine, Stanford University, 300 Pasteur, Edwards, Floor 3, Palo Alto, CA 94304.
[2]Hutch Institute for Cancer Outcomes Research, Biostatistics Program, Public Health Sciences Division, Fred Hutch Cancer Center, 1100 Fairview Ave, Seattle, WA 98109.
[3]The Comparative Health Outcomes, Policy, and Economics (CHOICE) Institute, University of Washington, 1959 NE Pacific St, Seattle, WA 98195.
[4]Stanford Center for Biomedical Informatics Research, Stanford University, 3180 Porter Dr, Palo Alto, CA 94304.
[5]Stanford Clinical Excellence Research Center, Stanford University, 453 Quarry Rd, Palo Alto, CA 94304.
[6]Division of Hospital Medicine, Stanford University, 453 Quarry Rd, Palo Alto, CA 94304.
[7]Department of Health Policy, School of Medicine, Stanford University, 615 Crothers Way, Stanford, CA 94305.
[8]Center for Health Policy, Freeman-Spogli Institute for International Studies, Stanford University, 615 Crothers Way, Stanford, CA 94305.

*Corresponding author(s). E-mail(s): sjpi@stanford.edu; falarid@stanford.edu;
Contributing authors: crutter@fredhutch.org; cpinedaa@uw.edu; jonc101@stanford.edu; jeremygf@stanford.edu;

Abstract

Simulation models inform health policy decisions by integrating data from multiple sources and forecasting outcomes when there is a lack of comprehensive evidence from empirical studies. Such models have long supported health policy for cancer, the first or second leading cause of death in over 100 countries. Discrete-event simulation (DES) and Bayesian calibration have gained traction in the field of Decision Science because they enable flexible modeling of complex health conditions and produce estimates of model parameters that reflect real-world disease epidemiology and data uncertainty given model constraints. This uncertainty is then propagated to model-generated outputs, enabling decision makers to assess confidence in recommendations and estimate the value of collecting additional information. However, there is limited end-to-end guidance on structuring a DES model for cancer progression, estimating its parameters using Bayesian calibration, and applying the calibration outputs to policy evaluation. To fill this gap, we introduce the **DES** Modeling Framework for **C**ancer **I**nterventions and **P**opulation **H**ealth in **R** (DESCIPHR), an open-source codebase integrating a flexible DES model for the natural history of cancer, Bayesian calibration for parameter estimation, and an example application of screening strategy evaluation. To illustrate the framework, we apply DESCIPHR to calibrate bladder and colorectal cancer models to real-world cancer registry targets. We also introduce an automated method for generating data-informed parameter prior distributions and increase the functionality of a neural network emulator-based Bayesian calibration algorithm. We anticipate that the adaptable DESCIPHR modeling template will facilitate the construction of future decision models evaluating the risks and benefits of health interventions.

Keywords: decision-analytic modeling, discrete-event simulation, microsimulation, cancer, screening, Bayesian calibration

Key points for decision makers

1. For simulation models to be useful for decision-making, they should accurately reproduce real-world outcomes and their uncertainty.
2. The DESCIPHR framework and code repository address a gap in open-source resources to fit an individual-level model for cancer progression to real-world data and forecast the impact of cancer screening interventions while accounting for data uncertainty.
3. The codebase is designed to be highly adaptable for researchers who wish to apply DESCIPHR for economic evaluation or for studying methodological questions.

Running title: Discrete-Event Simulation Model for Cancer Interventions and Population Health in R (DESCIPHR)

1 Introduction

Decision-analytic models inform policy decisions by integrating available evidence in a mathematical framework to forecast outcomes under alternative intervention strategies and potential counterfactual scenarios [1]. In particular, microsimulation models have had concrete effects on cancer screening policy, for instance informing the United States (US) Preventive Services Task Force's (USPSTF) breast cancer screening guidelines starting from 2009 [2] and a reduction in the recommended start age for US colorectal cancer (CRC) screening in 2021 [3, 4]. Compared to discrete-time microsimulation methods that model the state of a system at fixed time steps, discrete-event simulation (DES) models gain efficiency by only focusing on the time points of events at which a system changes [5] and more easily incorporate competing risks and time-varying patient characteristics [6]. These features have made DES a compelling alternative for model-based economic evaluations and operations research in healthcare settings, leading to its application for modeling health systems, disease progression, screening, and patient behavior [7].

Microsimulation models often contain deep, or unobservable, parameters that must be estimated through calibration. Calibration seeks to identify parameters such that the model reproduces real-world clinical, biological, or epidemiological outcomes, such as disease incidence and subtype prevalence [8, 9]. Bayesian calibration methods, in particular, have become widely used in cancer modeling [10–17] and other health applications, such as modeling infectious diseases [18, 19] or smoking behavior [20], to output parameter distributions that capture the measurement uncertainty of the real-world quantities used as calibration targets. Though Bayesian calibration can be computationally demanding, advances in model emulator methods, statistical approximations, and computing resource utilization have made them faster to implement and deploy [13, 21].

Though published methods, codebases, and best practice guidelines provide some instruction for implementing DES and Bayesian calibration separately, there is limited integrated and publicly available guidance with tutorial code on implementing and calibrating DES cancer models. Fewer than 10% of publications with healthcare DES models release open-source versions of their models [22], and open-source DES tools, such as the `simmer` R package [23] and the `SimPy` [24] and `Ciw` Python libraries [25], are more suited for operations questions than for disease natural history modeling. The steep learning curve of DES with existing bespoke open-source models has prompted the release of a tutorial on cost-effectiveness analyses of screening in R using a simplified DES-based cancer model [26], but the tutorial does not cover parameter estimation and calibration. Open-source cancer simulation models, such as the Colon Modeling Open Simulation Tool (CMOST) [27] and Microsimulation Lung Cancer (MILC) model [10, 28] are implemented with discrete-time methods or restrictive distributional assumptions and more limited calibration techniques. Open-source programs for Bayesian calibration of disease models are limited in scope, stopping short of integrating the required data pre-processing to prepare calibration targets for the model and applying the calibration outputs to decision analyses [11, 13–15, 21, 29].

To facilitate the development of individual-level models for cancer policy evaluation, we introduce the **DES** Modeling Framework for **C**ancer **I**nterventions and **P**opulation **H**ealth in **R** (DESCIPHR) with an open-source codebase for 1) structuring a DES model for cancer natural history, 2) estimating model parameters with state-of-the-art Bayesian calibration methods, and 3) applying the calibrated model for policy analyses accounting for parameter uncertainty, illustrated with a screening evaluation example.

The pipeline also augments Bayesian Calibration using Artificial Neural Networks (BayCANN) [21] to include hyperparameter tuning and multiple output types for constrained calibration targets. The model is built in R because of its accessibility and well-maintained and documented packages for statistical analysis and calibration, making it the most commonly used free and open-source programming language for healthcare DES models [22, 30–32]. The structure of the code repository follows the Decision Analysis in R for Technologies in Health (DARTH) framework for decision modeling, designed to facilitate model transparency and adaptability [32]. Section 2.1 introduces the DES cancer natural history model. Sections 3 and 4 explain parameter estimation of the model using modern Bayesian calibration methods and screening policy evaluation using the parameters' joint posterior distribution, respectively. In Section 5, we apply DESCIPHR to bladder and colorectal cancer to demonstrate how it can be adapted to real-world cancer registry targets and used to simulate two different types of cancers: one with precancerous lesions, colorectal, and one without, bladder. We provide the code for the entire pipeline and cancer site-specific analyses on GitHub at https://github.com/sjpi22/tutorial_cancer_modeling_des.

2 Cancer natural history model

Microsimulation modeling for policy analysis requires simulating a disease under natural history, that is, in the absence of intervention, as a baseline against which the impacts of interventions are compared [33]. There are multiple approaches to modeling the natural history of cancer, ranging from highly detailed cellular-level simulations [34, 35] to population-level cohort and microsimulation models [28, 36–38]. In practice, the key is to build the simplest model with just enough structural detail to appropriately evaluate the intervention(s) of interest and maintain face validity [33, 39]. Accordingly, the modeling process should begin by articulating the cancer control interventions to evaluate (e.g., primary prevention, screening, active surveillance, treatment), identifying the aspects of disease natural history that are influenced by the intervention (for example, the transition from undetected to screen-detected disease, removal of precursor lesions, or post-diagnosis survival), and specifying any other features that differentiate the intervention's effect in the model. In this section, we describe the structure, DES implementation, and summary outcomes of a model capturing the sequence of landmark events in cancer natural history needed to evaluate screening. In Appendix A, we provide guidance on implementing the model and extending it to other applications, such as cancer surveillance and treatment evaluation.

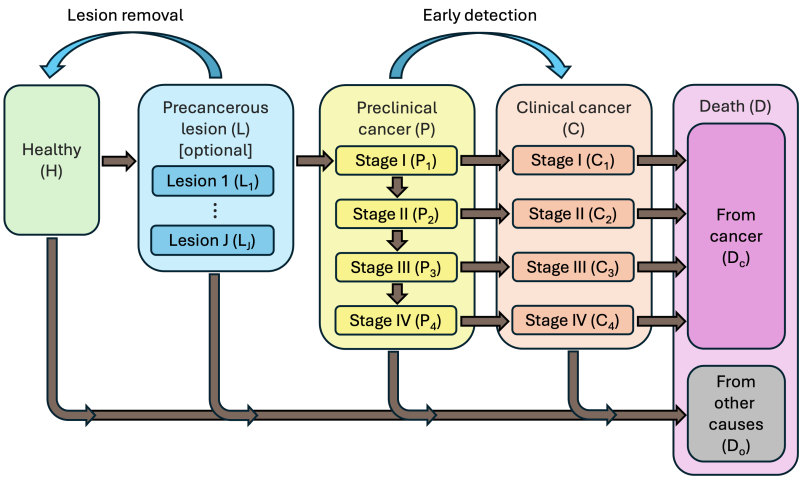


Fig. 1: Cancer natural history model schema with effects of screening in blue arched arrows

2.1 Disease states and transitions

We model the natural history of cancer by splitting the carcinogenic process into health states reflecting disease onset, progression, diagnosis, and death, consistent with the structure used in numerous general and site-specific cancer simulation models [12, 27, 28, 37, 40–45]. Relevant health states and possible transitions between states are illustrated in the model diagram in Figure 1. In the model, patients are assumed to be born into a healthy state, labeled as state H in Figure 1. If the natural history of a cancer

site includes precursor lesions, such as adenomas for CRC [46–48] or cervical intraepithelial neoplasia (CIN) for cervical cancer [49], the onset of a lesion causes patients to transition to the precancerous lesion state, labeled as state L. For simplicity, the current implementation of the model assumes a single lesion type and no lesion regression, although guidance on incorporating additional cancer subtypes and lesion growth patterns is provided in Appendix A.4. Additional lesions may develop over the course of the individual's life. At the first conversion of any lesion to detectable cancer, patients enter the preclinical, or pre-detection, cancer stage (state P). The model can also account for a direct transition from the healthy to the preclinical cancer state in the absence of precancerous lesions.

Within the preclinical cancer state, patients advance through progressive stages of disease until symptom-based diagnosis, upon which they transition to the clinical cancer state (state C). There are multiple approaches to modeling cancer progression, such as simulating the sojourn time and tumor size before assigning the stage at detection based on size [12] or modeling the hazard of stage progression and detection as proportional to tumor volume [50]. In our implementation, for each stage before the final stage, we sample the time to progression to the next stage and the time to symptomatic detection within the stage. The stage and time of diagnosis are determined by the first stage during which the time to detection occurs earlier than the time to progression, or the time to detection within the last stage if detection has not occurred before then. Individuals who develop clinical cancer face a risk of dying from cancer, thus transitioning to the state D_c, based on a stage- and disease-specific mortality rate derived from relative survival data from cancer registries. Cancer treatment effects are not explicitly modeled, but we assume that stage-specific relative survival distributions account for the survival associated with current technology. All individuals are at risk of death from other causes (state D_o), following an age-specific mortality rate derived from life tables and independent of the cancer pathway. The time of death is determined by the earlier of death from cancer and death from other causes, with the death state denoted as D. Guidance on customizing the model structure, for example, by adding an infection state preceding the precursor lesion state, is provided in Appendix A.4.

2.2 DES implementation

We simulate the mathematical model described in Section 2.1 using a DES approach [51] by defining the events as transitions between states. When generating patient trajectories using DES, the time from entering state i to entering a consecutive state j is characterized as a random variable $T_{i,j}$ that follows a time-to-event (TTE) distribution with probability density function $f_{i,j}$, cumulative distribution function (CDF) $F_{i,j}$, and hazard function $\lambda_{i,j} = \frac{f_{i,j}}{1-Fi,j}$. The hazard can be a function of the time t spent in state i, age a, time period p, and other demographic and risk factors x, *i.e.*, $\lambda_{i,j}(t, a, p, x)$, to account for heterogeneity related to population dynamics, genetic and environmental influences, treatment effects, and age-related biological responses.

Individual transition times and state characteristics can be generated using inverse transform sampling given the corresponding CDF, meaning that for each individual n, a value $u^n_{i,j}$ is drawn from a standard uniform distribution and used to calculate the time $t^n_{i,j}$ at which $F_{i,j}$ would equal $u^n_{i,j}$:

$$u^n_{i,j} \sim U(0, 1) \tag{1}$$

$$t^n_{i,j} = F^{-1}_{i,j}(u^n_{i,j}) \tag{2}$$

TTE variables may also be sampled using intensity-based methods [52]. For the results to be reproducible, a fixed random seed should be set before running an analysis based on random sampling.

After the times between consecutive states are sampled, the time from entry into the health state i to entry into a non-consecutive future health state k can be calculated as the sum of times within consecutive states from i to k. The model generates a matrix with a row for each individual in the simulated cohort and a column for each simulated patient characteristic, TTE variable, or health state characteristic. For models that include the precancerous lesion state, an additional lesion-level event matrix is generated with a row for each lesion that develops during a patient's lifetime and columns for the times of lesion onset and conversion to preclinical cancer, as well as other modeled lesion characteristics. More details on the implementation of the probability distributions and sampling procedures are provided in Appendix A.

2.3 Data inputs and parameters

While some distributions characterizing the model transitions and state characteristics may be obtained from published literature or empirically estimated with individual-level data using well-established survival analysis methods, others are unobserved but may be estimated through calibration to real-world data sources, a process described in Section 3. In the former category, the time to death from other causes is often modeled empirically using background mortality rates derived from actuarial life tables, as done in several established cancer models [12, 38, 41], with adjustments for disease-specific rates if they substantially impact all-cause mortality [53, 54]. In addition, the time from cancer diagnosis to death from cancer is commonly sampled from empirical relative survival distributions by stage at diagnosis, which are reported by the US National Cancer Institute's (NCI) Surveillance, Epidemiology, and End Results (SEER) program [55].

Conversely, no data exist to directly inform most pre-diagnosis distributions, such as the time to cancer onset, dwell time within preclinical stages, and rate of precancerous lesion development, which need to be estimated via calibration. These unobserved transitions are modeled by assigning a distributional form and then calibrating the distribution parameters to selected real-world outcomes, known as calibration targets. Epidemiological outcomes commonly used as calibration targets include the prevalence of precancerous lesions and incidentally detected cancer in screening or autopsy studies [11, 13, 15, 56, 57], the incidence and stage distribution of clinical cancer from cancer registry data in the absence of screening [2, 11, 15, 56, 58], the proportions of the population with different subtypes of precancerous lesions or cancer [57], and the multiplicity and size distribution of lesions [12, 15]. For a cancer natural history model, age-specific preclinical cancer prevalence and symptomatic cancer incidence are important for informing the transition rates to cancer onset and symptom-based detection by age, while the stage distribution informs the distribution of disease progression at diagnosis, which predicts survival outcomes. Targets for lesion characteristics inform tumor growth rates and heterogeneity in cancer pathways. Outcomes from screening and surveillance trials [45, 59] have also been used for calibration.

Choosing the distributional form for unobserved transitions is a somewhat subjective process but may be guided by the calibration targets. Since most cancer types display an increasing risk with age [60], a parametric distribution with an accelerating hazard, such as a Weibull distribution, can be a good starting point to model the time of precancerous lesion or cancer onset [37, 61], though sigmoidal functions have also been used to reflect tapering incidence at older ages [27]. Due to their simplicity, exponential distributions, which have a constant transition rate, are also commonly used for unobserved distributions [11, 41, 44, 62], although they may not adequately describe transitions with a nonzero mode or long tail [63]. Misspecification of the distributional form may bias estimates of outcomes [37], so conducting model validation and sensitivity analyses is encouraged to examine which distributional assumptions produce a reasonable model fit after calibration (see Section 3.4).

2.4 Summary outcomes

After generating event data using an individual-level simulation model with a given set of parameters, the data must often be aggregated into summary statistics reflecting population-level outcomes for various purposes, including model calibration, validation, and policy evaluation. Both calibration and validation of the natural history model, described in Section 3, involve comparing modeled outcomes to epidemiological endpoints estimated from empirical data, as discussed in Section 2.3. For prevalence, incidence, stage distribution, multiplicity, and survival targets, we simulate a population representative of the real-world sample from which the endpoint is derived and calculate the corresponding epidemiological outcomes using the equations described by [64]. We may also estimate intermediate outcomes of cancer natural history that are often not observed in practice, including the mean sojourn time of cancer and the mean dwell time of precancerous lesions, according to the formulas in [64]. For decision analysis, tradeoffs associated with screening may be calculated in terms of quality-adjusted or absolute life years gained (LYG) and the economic or resource burden from additional testing compared to no screening. Average LYG is calculated by summing total life-years, optionally among individuals meeting specific criteria (e.g., alive and cancer-free at a given age), dividing by the cohort size to obtain per capita estimates, repeating the calculation under screening (see Section 4.2), and computing the difference. Incremental testing burden can be calculated similarly. Because life trajectories in the microsimulation model are generated through random sampling, the simulated outcomes exhibit stochastic variation that depends on the simulated population size. In Appendix A.3, we provide an approach to determine the cohort size accounting for this variation.

3 Parameter estimation with Bayesian calibration

As discussed in Section 2.2, disease models may include deep parameters that are not estimable using individual-level data but may be inferred through calibration, which involves adjusting the parameters so that the model outputs, when summarized according to Section 2.4, are consistent with real-world calibration targets [10]. Various methods exist to search the parameter space for optimal-fitting parameters, including manual tuning, empirical search, directed search, and Bayesian techniques [33]. Whereas directed search algorithms output a single parameter set, we focus on Bayesian calibration methods because they combine prior knowledge about the parameters of interest with the uncertainty associated with the calibration targets to output a distribution of parameters reflecting their joint uncertainty.

In Bayesian calibration, the modeler specifies prior distributions for unknown model parameters and measures the goodness-of-fit for the targets given the parameters using a loss function, such as a likelihood. The output is the joint posterior distribution for the parameters, which can then be sampled to simulate distributions for model outcomes of interest. For each outcome, we can generate base case estimates from the expected value and 95% posterior model-prediction intervals (PI) from the 2.5th and 97.5th percentiles of the simulated outcomes from the posterior distribution. An early limitation of Bayesian calibration was the computational intensity required to simulate large numbers of model samples in sequence to achieve accurate estimates of the joint posteriors [10, 65, 66]. However, advances in computational capacity and statistical approximations have made Bayesian methods increasingly feasible.

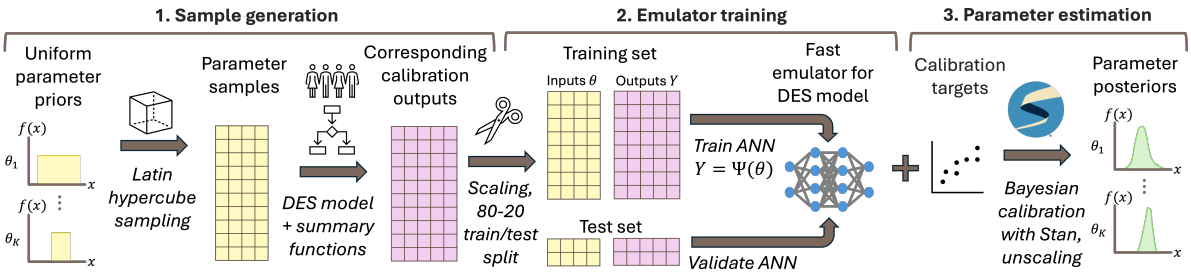
Approximate Bayesian computation (ABC) refers to a family of likelihood-free methods that infer model parameters by simulating outputs and assessing how closely they match observed or target data, thereby approximating the likelihood [67]. Incremental mixture ABC (IMABC) leverages ABC in an initial rejection-sampling step and then performs adaptive sampling of regions consistent with targets; the adaptive sampling is similar to incremental mixture importance sampling (IMIS) and allows IMABC to overcome inefficiencies of ABC in searching high-dimensional parameter spaces and calibrating to multiple targets [13]. Since the most time-consuming aspect of Bayesian calibration is simulating the model with different parameter samples, emulators that quickly and accurately map model inputs to desired outputs, such as the Bayesian Calibration using Artificial Neural Networks (BayCANN) algorithm, could drastically reduce runtime [15, 21]. IMABC and some steps in BayCANN can be parallelized to optimize computational resources and speed up processing.

3.1 Choice of priors

Assigning prior distributions to calibrated model parameters may be guided by meta-analysis of evidence, expert opinion, biological knowledge, and common sense to reflect the state of knowledge before model fitting [65, 66, 68]. When there is little evidence to inform the parameters, one can perform empirical calibration, such as Nelder-Mead optimization followed by a grid search, to set bounds for mutually independent, uniformly distributed diffuse priors [10, 29]. In Appendix B, we also describe a target-informed approach to derive priors for the parameters underlying the time from birth to disease onset. Prior to calibration, to ensure the model can produce outputs in the ranges of the calibration targets, we perform a model coverage analysis by sampling from the joint prior distribution, plotting the model outputs, and demonstrating that the model-predicted outputs have a wider range than the 95% confidence intervals of the calibration targets. For BayCANN, the coverage analysis can be conducted using the parameter and output set generated to train the neural network emulator.

3.2 IMABC

IMABC improves on earlier Bayesian calibration algorithms, such as Markov Chain Monte Carlo (MCMC) and ABC, due to its ability to handle high-dimensional parameter spaces and its compatibility with parallel processing. The initial rejection-based sampling step draws N_o parameter sets from the prior distribution using Latin hypercube sampling (LHS) and accepts sets that produce model outputs within a user-specified tolerance threshold from the targets [13]. Parameter sets are ranked using a distance-based measure [69]. At each iteration of the subsequent updating phase, the algorithm samples from a mixture of multivariate normal distributions centered at the $N^{(c)}$ highest ranking parameter sets and keeps the sets whose outputs fall within incrementally narrowing bounds around the targets. The algorithm stops when either the number of accepted parameter sets within the final target bounds reaches a user-specified value N_{post} or the maximum number of iterations is reached. It is recommended



Abbreviations: ANN = artificial neural network, DES = discrete-event simulation

Fig. 2: Steps of BayCANN

to use a large N_o (for example, 1000 samples per calibrated parameter) and wide initial tolerance intervals to increase the algorithm's likelihood of identifying acceptable parameter sets in the first step [13]. Additional guidance for choosing the values of these arguments is provided in [13] and the IMABC GitHub documentation. We used the `IMABC` R package to implement Bayesian calibration with IMABC. The output of the calibration is a set of accepted parameters with weights that account for the adaptive sampling. A weighted sample of the posterior distribution is used to generate decision-analytic outcomes.

3.3 BayCANN

A metamodel, or emulator, is a function that approximates a more computationally intensive simulation model by predicting model outputs based on the same inputs [70]. BayCANN uses an artificial neural network (ANN) as an emulator to map parameters of the cancer natural history model to summary statistics corresponding to the calibration targets [21] and has been applied to emulate other models, including a seasonal influenza model [71] and the CISNET CRC models [14]. Popular in machine learning due to their accuracy in modeling nonlinear interactions between input variables, ANNs are a type of regression consisting of layers of interconnected nodes that resemble the interconnected neurons of the brain [72]. BayCANN includes sample generation, emulator training, and parameter estimation steps as illustrated in Figure 2. We increase its functionality from the original implementation with two modifications: incorporating hyperparameter tuning during ANN training to increase emulator accuracy and enabling multiple output layers in the ANN, allowing it to handle different activation functions and directly integrate constraints on the outputs.

3.3.1 Sample generation

To create a sample for ANN fitting and validation, we draw parameter sets from the prior distributions using LHS; simulate a cohort of life trajectories for each set; and calculate the simulated outcomes corresponding to the calibration targets for each cohort. The sets of parameters and outcomes are divided for training and testing, often with a ratio of 80-20, and standardized to ensure stable gradients during ANN training.

3.3.2 Emulator training

We fit the ANN by minimizing a loss function, such as the mean squared error or binary cross-entropy, between the ANN's predicted outputs and the cancer model outcomes for the training set. The training set may be further split into training and validation sets for hyperparameter tuning, which is the process of testing various values of ANN hyperparameters to identify the configuration that yields the most accurate ANN. Hyperparameters include the number of layers, number of nodes per layer, type of activation function, inclusion of dropout layers, and dropout rate. We may assign groups of outputs with different output activation functions, losses, evaluation metrics, and contribution weights in the total loss. For the output layer, sigmoid activation is consistent with targets such as prevalence and standardized outputs that range from 0 to 1, softmax activation for groups of variables that must sum to 1, exponential or rectified linear unit (ReLU) activation for nonnegative outputs, and linear activation for outputs with no range constraints. We implement the ANN and hyperparameter tuning using the `keras3` and `tfruns` R packages. The ANN can be validated by calculating the loss between the predicted and true outputs in the held-out test set or plotting the ANN-predicted outputs against the model outputs. Lines of points along the diagonals indicate that the emulator accurately predicts the model outputs.

3.3.3 Parameter estimation

The hyperparameters, weights, and biases of the final `keras3` ANN are extracted and input into an ANN coded in the probabilistic programming language Stan via the `rstan` interface [73]. Using the Hamiltonian Monte Carlo algorithm, Stan generates a collection of parameters that approximates the joint posterior distribution and that can be uniformly sampled to produce new model outputs. To assess the quality of the Stan model fit, we check the mixing of the chains and the R-hat convergence diagnostic. The posteriors are returned to the original scale of the parameters.

3.4 Internal validation

Validating the calibration results establishes trust in the model's predictions and involves evaluating the consistency of the model with the calibration targets (internal validation), other models (comparative model validation), or data held out from the calibration process (external validation) [33, 74, 75]. For internal validation, we visually examine whether the model outputs produced by sampling from the joint parameter posterior distribution (see Section 4.1) replicate the targets and their uncertainty bounds [75]. The IMABC algorithm automatically generates the calibration outputs associated with the posteriors. For BayCANN, the user may choose to sample from the posterior as described in Section 4.1 and simulate the calibration outputs and decision outcomes together for efficiency. For a more detailed discussion of the implementation and reporting of comparative and external validation, which are stronger forms of validation, we refer to [74–76].

3.5 Nonidentifiability

If available calibration targets are insufficient to inform all calibrated parameters, these parameters may be nonidentifiable, meaning that there are multiple parameter sets that fit equally well with the calibration targets but may have different policy implications [77]. For instance, a natural history characterized by a slow onset time but fast sojourn time may fit observed cancer incidence data just as well as one with a fast onset time and slow sojourn time, but screening intervals would need to be shorter in the former case to intercept cancers at the same rate. In such cases, it is important to transparently report how the model transitions and parameter prior distributions are constructed and to simulate outcomes from the entire posterior distribution of the parameters, ensuring that policy analyses reflect their uncertainty. Parameter nonidentifiability can be addressed by using informative priors, but this approach requires justification, as informative priors may bias the posterior [68, 78].

4 Model-based screening policy evaluation

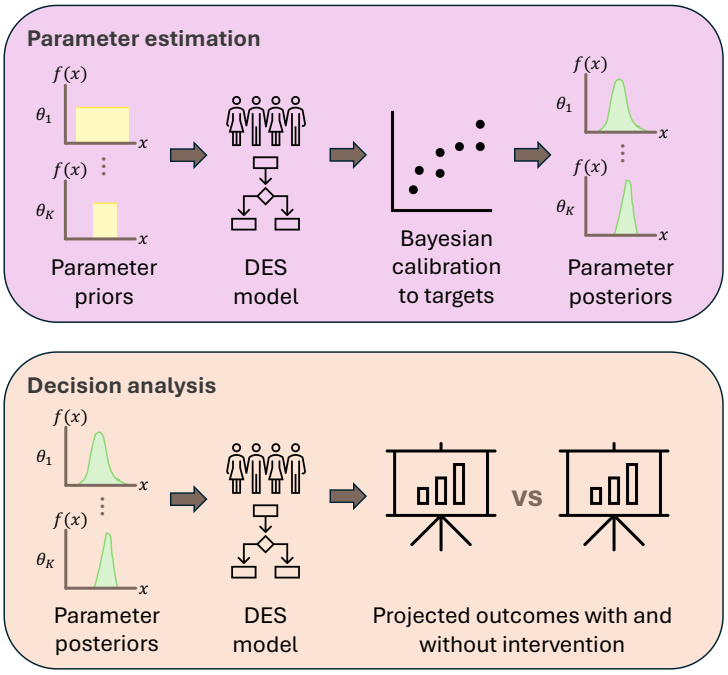
We illustrate an application of the modeling and calibration pipeline by sampling from the parameter posterior distribution and other uncertain model inputs to estimate the effects of screening. In this section, we first provide general guidance on how to apply the calibration results to policy evaluation and then describe how screening interventions interact with the natural history model.

4.1 Sampling from the posterior

The outputs obtained from Bayesian calibration with IMABC and BayCANN are a random sample of parameter sets from the joint posterior distribution of the model parameters. For each parameter set in the sample, we can estimate the impact of a particular intervention by simulating a cohort of individual trajectories with the natural history model, sampling from distributions that capture the uncertainty of quantities related to the intervention (for example, the sensitivity of a screening test or the hazard ratio of a cancer treatment), updating the life trajectories in response to applying the intervention to the cohort, and comparing outcomes of interest between the scenarios with and without the intervention. The distribution of outputs associated with the parameter posterior sample and other uncertain quantities constitutes a probabilistic sensitivity analysis (PSA) of the impact of the intervention. This procedure can be repeated for different interventions to compare outcomes, determine the optimal policy recommendations, and assess the uncertainty in the recommendations and estimates.

4.2 Modeling screening

Screening may intervene in the disease process in two ways: it can prevent cancer by removing precursor lesions, thereby interrupting their progression to cancer, as in the case of CRC or cervical cancer; or it



Abbreviations: DES = discrete-event simulation
Fig. 3: Diagram of DESCIPHR modeling framework

may modify the disease process through the early detection of cancer, potentially at a more treatable stage, as in the case of prostate and breast cancer. However, mass screening in individuals without disease is associated with clinical and economic burdens, including the cascade of care triggered by false positive results [79, 80]. To evaluate the effects of screening, a cohort of individual trajectories is first simulated without it (i.e., under natural history). Next, screening regimens are applied to generate counterfactual trajectories, from which we estimate the effects of screening on cancer-related outcomes. We consider the LYG and test burden compared to the natural history scenario, which are two of many outcomes, including quality of life, adverse events, healthcare costs, and equity, that policymakers may consider when deciding whether to adopt a new screening intervention. Costs can also be assigned to each test to capture the economic burden of screening.

We define a screening regimen by its starting age, stopping age, screening interval, and test characteristics. Individuals who are alive without diagnosed cancer by the screening starting age will undergo screening until the earliest of the screening stopping age, the diagnosis of cancer, or death. Test characteristics required for modeling include the sensitivity p, which is the probability of a positive result among individuals with disease, and specificity q, or the probability of a negative result for an individual without disease. If any precancerous lesions are removed at a particular screening event, the time to preclinical cancer onset is recalculated as the earliest conversion time of any remaining lesions over the individual's lifetime. Before applying the next test, the time to death from cancer is recalculated by adding the original times from preclinical cancer onset to clinical cancer and then to death from cancer, assuming independence between the age of preclinical cancer onset and the time from onset to symptom-based diagnosis. Under the same assumption, the stage of symptom-based diagnosis remains unchanged. Meanwhile, the confirmation of asymptomatic cancer causes the patient to transition immediately to the detected cancer state, and the time to death from cancer is resampled based on the stage at diagnosis. The time to death is then recalculated as the minimum time to death from either cancer or other causes but restricted to occur no earlier than in the no-screening scenario.

5 Application to bladder and colorectal cancer

To illustrate the modeling and calibration approaches, we apply DESCIPHR to two cancer types that differ in their natural history structure: bladder cancer, which has no precursor pathway, and CRC, which does.

5.1 Bladder cancer

Bladder cancer is associated with a lifetime risk of 3.3% in men and 1.0% in women [55]. It is typically diagnosed after blood is detected in a routine urine test, leading to a cystoscopy to visualize the bladder wall [81]. Currently, there is no population-based bladder cancer screening program in the US, but there is interest in researching and modeling the potential benefits of early detection and active surveillance [82, 83].

5.1.1 Data

For the background mortality distribution informing the time to death from non-cancer causes, we use the 1933 life table for the overall US population from the Human Mortality Database [84]. For the time from diagnosis to death from cancer, we use SEER relative survival rates from 2000 to 2021 [55]. Since relative survival is only reported up to 10 years from diagnosis, we extrapolate beyond 10 years by fitting a concave B-spline to the cumulative hazard of cancer-specific mortality. For calibration targets, we use SEER data on age-specific incidence in 5-year intervals from 0 to 85 and aggregated for the population 85 and older, and stage distribution (localized, regional, and distant) of bladder cancer from 1989 to 1993, the earliest period reported in the SEER archives [85]. We exclude unstaged cancer and increase the percentages of the other categories proportionally so that they sum to 1.

5.1.2 Natural history model

We model the natural history of bladder cancer using the model described in Section 2.1 without the precursor lesion state between the healthy and preclinical cancer states. That is, healthy individuals transition directly to preclinical cancer. The transitions between states are parameterized by the distributions listed in Table 1. Although the natural history of bladder cancer includes carcinoma in situ as a stage preceding the localized stage, we only model progression starting from the localized stage due to data limitations. The natural history model ultimately contains 8 parameters that must be calibrated to the incidence and stage distribution targets.

5.1.3 Parameter estimation with IMABC

To determine the prior bounds for the calibrated parameters, we apply the procedure in Appendix B to estimate a range of parameters for the time to preclinical cancer onset and use a wide range of plausible priors for the other parameters. Since data on the age-specific prevalence of preclinical bladder cancer was unavailable, we use the ratio of incidental to known bladder cancer diagnoses in autopsy data from [86] and SEER cancer incidence to estimate age-specific prevalence for calculating the prior bounds. We use the $(1 - 0.0001) \times 100\%$ confidence interval of the calibration targets as the final bounds for the simulated outputs, as the sample size of the SEER registry leads to narrow bounds. Following the procedure in Appendix A.3 and multiplying the target cohort size by 2, we calculate that a cohort size of 1,500,000 would achieve a Monte Carlo error within these bounds.

For the initial bounds of the calibration targets, we expand the final bounds by three times their difference from the target value. We run IMABC with an initial sample of 4000 parameter sets (500 samples per calibrated parameter) drawn from the prior distributions using LHS to identify sets associated with model outputs within the initial bounds. After the initial step, IMABC continues to run by sampling 50 parameter sets per center from a multivariate normal mixture distribution with 5 centers for up to 1200 iterations or until the posterior distribution achieves an effective sample size of 1000. The algorithm ran for 12 hours on 16 cores on a high-performance computing cluster, ending with 395 in-range parameter sets. For internal validation, the 95% model-predicted posterior ranges of outputs are shown in gray against the final bounds of the calibration targets in red in Figure 4, showing good overlap.

5.1.4 Screening policy evaluation

For illustrative purposes, we evaluate the LYG and test burden associated with implementing population-wide screening using a hypothetical test at 5-year intervals from age 55 to 75. The test has a beta-distributed sensitivity with mean 0.9 and sample size 100 and a beta-distributed specificity with mean 0.95 and sample size 100. We sample from the parameter posterior distribution as well as the beta distribution for sensitivity to run the model and calculate a distribution of outcomes.

(a) Variables directly informed by external data sources

Variable	Description	Distribution	Value	Source
M	Male sex	Bernoulli	0.5	Assumption
T_{H,D_o}	Time from birth to death from other causes	Empirical		Human Mortality Database 1933 US overall life table [84]
T_{C,D_c}	Time from clinical cancer to death from cancer	Empirical		SEER relative survival by stage at bladder cancer diagnosis [55]

(b) Variables with unobserved distributions (assumed or informed by calibration)

Variable	Description	Distribution	Parameter priors
$T_{H,P}$	Time from birth to onset of preclinical cancer	Weibull cure $\left(\text{CDF: } F_{H,P}(t) = \kappa\left(1 - e^{-(t/\sigma)^{\alpha}}\right)\right)$	$\kappa \sim \text{Uniform}(0.04,\ 0.1)$ $\alpha \sim \text{Uniform}(3.3,\ 11.9)$ $\sigma \sim \text{Uniform}(40.5,\ 133.1)$
T_{P_1,P_2}	Time from localized onset to regional onset of preclinical cancer	Exponential $\left(\text{CDF: } F_{P_1,P_2}(t) = \left(1 - e^{-\lambda_1 t}\right)\right)$	$\lambda_1 \sim \text{Uniform}(0.05,\ 2)$
T_{P_1,C_1}	Time from localized onset of preclinical cancer to diagnosis within localized stage	Exponential $\left(\text{CDF: } F_{P_1,C_1}(t) = 1 - e^{-\frac{(d_1+\delta_1)\lambda_1 t}{1-d_1-\delta_1}}\right)$ where d_1 is the target for proportion diagnosed at the localized stage	$\delta_1 \sim \text{Uniform}(-0.15,\ 0.03)$
T_{P_2,P_3}	Time from regional onset to distant onset of preclinical cancer	Exponential $\left(\text{CDF: } F_{P_2,P_3}(t) = 1 - e^{-\rho_1\lambda_1 t}\right)$	$\rho_1 \sim \text{Uniform}(0.7,\ 3)$
T_{P_2,C_2}	Time from regional onset of preclinical cancer to diagnosis within regional stage	Exponential $\left(\text{CDF: } F_{P_2,C_2}(t) = 1 - e^{-\frac{(d_2+\delta_2)\rho_1\lambda_1 t}{1-\sum_{i=1}^{2}(d_i-\delta_i)}}\right)$ where d_1 and d_2 are the targets for proportion diagnosed at the localized and regional stages, respectively	$\delta_2 \sim \text{Uniform}(-0.05,\ 0.05)$
T_{P_3,C_3}	Time from distant onset of preclinical cancer to diagnosis	Exponential $\left(\text{CDF: } F_{P_3,C_3}(t) = 1 - e^{-\rho_1\rho_2\lambda_1 t}\right)$	$\rho_2 \sim \text{Uniform}(0.7, 3)$

Table 1: Bladder cancer model parameters and distributions

5.2 Colorectal cancer

We use DESCIPHR to simulate the natural history of CRC for a hypothetical cohort of women in the U.S. The natural history of CRC is characterized by the emergence of benign precursor lesions, or polyps, that can progress to cancer. After the development of the flexible colonoscope in the 1960s to visualize the interior of the colon and the recognition that polyp removal could interrupt the development of CRC in the 1970s [87], medical institutions began to recognize the lifesaving potential of CRC screening in the US. USPSTF guidelines recommend colonoscopy every 10 years or stool testing every 1-3 years for average-risk adults starting from age 45 as of 2021 [4]. For illustrative purposes, we assume that all cases of CRC arise from adenomatous polyps.

5.2.1 Data

For background mortality, we use the annual probabilities of mortality from the 1980 female life table from the Human Mortality Database [84]. For calibration targets, we use the age-specific prevalence of adenomas from a meta-analysis of screening and autopsy studies [88], as well as the stage distribution and age-specific incidence of CRC from 1975-1979, before widespread screening for CRC in the US [85]. To simulate the time to death from cancer by stage at diagnosis, we use SEER relative survival rates from 2000 to 2021 [55]. Since stages reported for relative survival are aggregated as localized, regional, and distant, we use the relative survival of the localized stage to simulate mortality from a diagnosis at stage I, the regional stage for stages II and III, and the distant stage for stage IV.

5.2.2 Natural history model

We implement a CRC natural history model with all states described in Section 2.1, including an intermediary precursor lesion state. The transitions between states are parameterized by the distributions

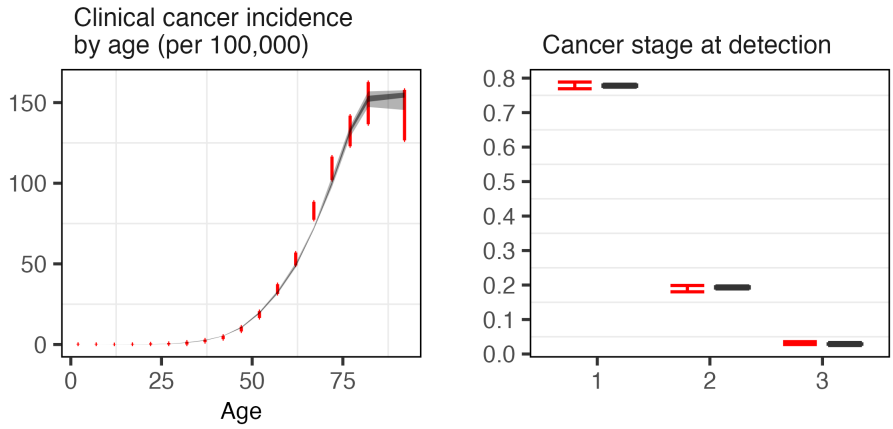


Fig. 4: Internal validation of calibrated bladder cancer model with model outputs in gray and $(1 - 0.0001) \times 100\%$ CIs of calibration targets in red

listed in Table 2. The natural history model ultimately contains 13 parameters that must be calibrated to adenoma prevalence and colorectal cancer incidence and stage distribution targets.

5.2.3 Parameter estimation with BayCANN

Prior distributions for the time to lesion onset are determined using the procedure in Appendix B. For other parameters, we use a wide range of reasonable priors. Following the procedure in Appendix A.3 and multiplying the target cohort size by 1.5, we proceed with the calibration using a cohort size of 3,100,000. To generate the training sample for BayCANN, we draw 500 samples per calibrated parameter from the prior distributions using LHS for a total of 6500 parameter sets and simulate the corresponding calibration outcomes. The parameter inputs and calibration outputs other than stage distribution are scaled from 0 to 1, with 0 associated with the minimum value of each data type and 1 the maximum. 20% of the sample is held out to validate the ANN, and an 80-20 split is used for hyperparameter tuning. For training the ANN, we use a batch size of 128, a maximum number of training epochs of 5000, and a patience of 30. We perform hyperparameter tuning by randomly sampling 30% of permutations of the following hyperparameters: 1 to 4 hidden layers; 32, 64, or 128 hidden nodes per layer; including or excluding dropout layers with a dropout rate of 0.25; and reLU, tanh, or sigmoid as activation functions. The ANN architecture that minimized the loss included 2 hidden layers with the tanh activation function, 32 nodes per hidden layer, and no dropout. We run Stan with 4 chains for a maximum of 300,000 iterations and a thinning parameter of 100 to address autocorrelation, resulting in a posterior sample of 6000 parameter sets. The sample generation ran for 4.1 hours on 8 cores on a high-performance computing cluster. Hyperparameter tuning, ANN training, and Stan calibration were run on 8 cores on an Apple MacBook Pro (M1 chip, 16 GB RAM, macOS Sequoia v15.7.2), with run times of 25, 1, and 10 minutes, respectively. The calibration produced two sets of divergent chains, indicating two separate parameter regions that fit with the targets (Figure 5).

5.2.4 Screening policy evaluation

To demonstrate an application of the calibrated model, we evaluate the LYG and test burden associated with colonoscopy every 10 years from age 45 to 75. We assume a beta-distributed test sensitivity with mean 0.95 and sample size for cancer and mean 0.7 and sample size 100 for precancerous lesions. The specificity is assumed to be 100%. We sample from each region of the parameter posterior distribution as well as the beta distribution for sensitivity to run the model and calculate the distribution of outcomes associated with each region.

6 Discussion

We introduce the DESCIPHR modeling framework to fill a gap in open-source, flexible, end-to-end resources for implementing and deploying health decision models. The framework spans considerations for building a DES model for the natural history of cancer, estimating unobservable model parameters using

(a) Variables directly informed by external data sources

Variable	Description	Distribution	Value	Source
T_{H,D_o}	Time from birth to death from other causes	Empirical		Human Mortality Database 1980 US female life table [84]
T_{C,D_c}	Time from clinical cancer to death from cancer	Empirical		SEER relative survival by stage at CRC diagnosis [85]

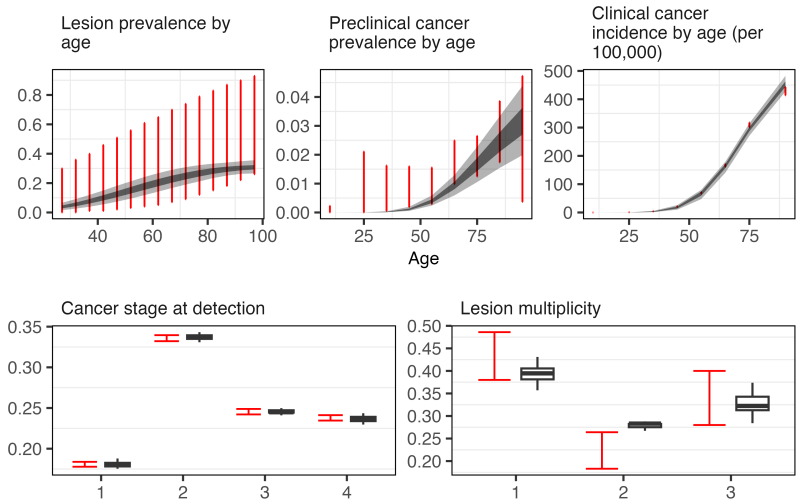
(b) Variables with unobserved distributions (assumed or informed by calibration)

Variable	Description	Distribution	Parameter priors
$T_{H,L}$	Time from birth to onset of first lesion	Weibull cure $\left(\text{CDF: } F_{H,P}(t) = \kappa\left(1 - e^{-(t/\sigma)^\alpha}\right)\right)$	$\kappa \sim \text{Uniform}(0.5,\ 1)$ $\alpha \sim \text{Uniform}(1.1,\ 6.2)$ $\sigma \sim \text{Uniform}(46.1,\ 115.2)$
N_L	Number of additional lesions before death from other causes	Poisson with rate ν	$\nu \sim \text{Uniform}(0.02, 0.1)$
T_{L,L_j}	Time from first lesion to onset of additional lesion j	Uniform from 0 to $T_{H,D_o} - T_{H,L}$	N/A
$T_{L_j,P}$	Time from onset of lesion j to conversion to preclinical cancer	Weibull $\left(\text{CDF: } F_{H,P}(t) = 1 - e^{-(t/\sigma)^\alpha}\right)$	$\alpha \sim \text{Uniform}(1.5,\ 7)$ $\sigma \sim \text{Uniform}(40,\ 150)$
T_{P_1,P_2}	Time from onset of stage I to stage II preclinical cancer	Exponential $\left(\text{CDF: } F_{P_1,P_2}(t) = \left(1 - e^{-\lambda_1 t}\right)\right)$	$\lambda_1 \sim \text{Uniform}(0.05,\ 0.5)$
T_{P_1,C_1}	Time from onset of stage I preclinical cancer to diagnosis within stage I	Exponential $\left(\text{CDF: } F_{P_1,C_1}(t) = 1 - e^{-\frac{(d_1+\delta_1)\lambda_1 t}{1-d_1-\delta_1}}\right)$ where d_1 is the target for proportion diagnosed at stage I	$\delta_1 \sim \text{Uniform}(-0.1,\ 0.03)$
T_{P_2,P_3}	Time from onset of stage II to stage III preclinical cancer	Exponential $\left(\text{CDF: } F_{P_2,P_3}(t) = \left(1 - e^{-\rho_1\lambda_1 t}\right)\right)$	$\rho_1 \sim \text{Uniform}(0.7,\ 3)$
T_{P_2,C_2}	Time from onset of stage II preclinical cancer to diagnosis within stage II	Exponential $\left(\text{CDF: } F_{P_2,C_2}(t) = 1 - e^{-\frac{(d_2+\delta_2)\rho_1\lambda_1 t}{1-\sum_{i=1}^{2}(d_i-\delta_i)}}\right)$ where d_i is the target for proportion diagnosed at stage i	$\delta_2 \sim \text{Uniform}(-0.1,\ 0.03)$
T_{P_3,P_4}	Time from onset of stage III to stage IV preclinical cancer	Exponential $\left(\text{CDF: } F_{P_3,P_4}(t) = \left(1 - e^{-\rho_2\rho_1\lambda_1 t}\right)\right)$	$\rho_2 \sim \text{Uniform}(0.7,\ 3)$
T_{P_3,C_3}	Time from onset of stage III preclinical cancer to diagnosis within stage III	Exponential $\left(\text{CDF: } F_{P_3,C_3}(t) = 1 - e^{-\frac{(d_3+\delta_3)\rho_1\rho_2\lambda_1 t}{1-\sum_{i=1}^{3}(d_i+\delta_i)}}\right)$ where d_i is the target for proportion diagnosed at stage i	$\delta_3 \sim \text{Uniform}(-0.03, 0.03)$
T_{P_4,C_4}	Time from onset of stage IV preclinical cancer to diagnosis	Exponential $\left(\text{CDF: } F_{P_4,C_4}(t) = \left(1 - e^{-\rho_1\rho_2\rho_3\lambda_1 t}\right)\right)$	$\rho_3 \sim \text{Uniform}(0.7, 3)$

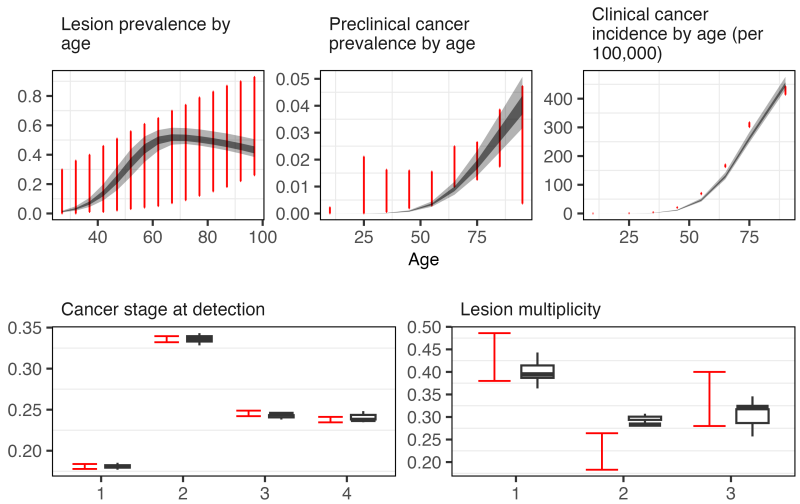
Table 2: Colorectal cancer model parameters and distributions

Bayesian calibration, and applying the resulting posterior distributions for screening policy analysis. The DES implementation provides flexibility in modeling individual-level transitions between states compared to simpler Markov or cohort-based approaches and increases computational efficiency compared to discrete-time implementations. Improving on prior DES and calibration tutorials, the DESCIPHR codebase also demonstrates the model pre-processing and setup required to run two Bayesian calibration methods, IMABC and BayCANN, and apply the resulting joint parameter posterior distribution to make forecasts that account for the uncertainty in model parameters. We use IMABC and BayCANN because they increase efficiency relative to older Bayesian calibration techniques, such as MCMC and ABC, by leveraging adaptive sampling and metamodeling, respectively [13, 21].

In addition to linking existing modeling concepts and methods in an open-source pipeline, we make several novel methodological contributions to facilitate decision modeling with DESCIPHR. First, our target-informed approach to estimating prior distributions for the parameters governing the time to disease onset and cancer progression (described in Appendix B) circumvents the need to guess a reasonable range for a grid search manually. Despite advances in Bayesian calibration methods, assigning parameter



(a) Outputs associated with chains 1, 2, and 4



(b) Outputs associated with chain 3

Fig. 5: Internal validation of calibrated CRC model with model outputs in gray and 95% CIs of calibration targets in red

prior distributions is still considered more of an art than a science, requiring extensive trial and error. Even though guidance suggests performing an initial grid search to find a reasonable range of parameters that could be used as priors, the grid search step still requires setting initial bounds within which to search. Initial bounds that are too narrow risk missing the target parameter space entirely. However, excessively wide bounds are prone to the curse of dimensionality; as the number of parameters and the width of the bounds increase, more samples are needed to capture a region in the parameter space close enough to the targets for calibration to be successful.

Second, our augmentations to BayCANN allow for automated hyperparameter optimization, which is a critical part of empirical machine learning because the prediction accuracy of machine learning models is sensitive to the hyperparameter configuration [89, 90]. We also add capabilities for users to distinguish output types, such as outputs that must sum to 1 versus merely range from 0 to 1, and automatically propagate their characteristics to the emulator model. These additions minimize the setup required for a modeler to train an accurate model and ensure that output constraints are satisfied. Future work could

integrate more advanced and automated approaches for parameter estimation, such as reinforcement learning-based calibration [91].

To balance the learnability and adaptability of DESCIPHR, the implemented cancer natural history and screening model relies on some simplifying assumptions, only modeling a single tumor type, screening sensitivity independent of lesion size and cancer stage, and perfect adherence to screening. However, the model can be easily extended to incorporate other natural history assumptions or evaluate additional interventions based on the guidance provided in Appendix A.4. The modularity of the DESCIPHR codebase also allows researchers to easily substitute variable distributions, input data, calibration methods, and outcomes according to their needs, making the framework adaptable to specific cancer sites, as well as other diseases characterized by progressive stages that have preclinical and clinical states.

Furthermore, without targets on the stage-specific prevalence of preclinical cancer, the parameters for distributions related to cancer stage progression may not be identifiable. However, as long as the decision outcomes associated with the posteriors accurately account for the uncertainty due to nonidentifiability, a modeler may still determine the optimal decision alternative across the range of plausible parameter sets. Frameworks for decision-making under deep uncertainty can be further explored in the robust decision-making literature [92–94].

Another limitation is that comparative and external validation are not explicitly included in the DESCIPHR codebase, although the natural history and screening models can be adapted to perform them. Independently developed models have been used for comparative model validation of policy recommendations [2, 3, 95]. Unlike internal and comparative validation, external validation tests the accuracy of the model's structure and predictions by comparing simulated outputs to empirical data not used for model construction or calibration [74–76], such as cancer mortality and incidence outcomes from screening trials [96, 97] or observational data [76, 98].

Models act as approximations of reality, and their predictive ability for decision-making carries inherent uncertainties, especially when there is insufficient data for precise calibration [8]. However, this does not belie their usefulness for predicting information that would not otherwise be measurable, especially given their history of successful inference of health trends [38, 96, 97]. For decades, simulation modeling has successfully informed cancer screening policy, though typically with bespoke, closed-source implementations [4, 95, 99]. Though this convention has the advantage of validation by consensus when diverse, independently developed models produce the same conclusion, these highly heterogeneous implementations come with a steep learning curve, a high barrier to entry, and a lack of transparency. We believe that releasing an adaptable open-source model will accelerate the progress of health decision modeling applications and methods development. As new cancer screening modalities and treatments emerge, there will be no shortage of policy decisions to make before the arrival of long-term comparative evidence. DESCIPHR represents a significant step towards streamlining model development and increasing transparency to optimally inform these decisions.

7 Conclusion

The DESCIPHR framework integrates the entire process of implementing a DES model for cancer natural history, calibrating the model parameters to real-world data using Bayesian methods, and applying the results to evaluate the impact of cancer screening strategies. The modular organization of the DES model and pipeline, the model's flexibility in accepting parametric and non-parametric inputs, and the demonstration of the pipeline with a ground truth simulation facilitate downstream users' ability to understand the pipeline and customize it to their modeling specifications. We provide and illustrate all steps of DESCIPHR in an open-source code repository as a foundation for developing, calibrating, testing, and deploying future decision models for cancer and other progressive diseases.

8 Declarations

8.1 Funding

SP was supported by the US National Institutes of Health (NIH Grant T15LM007033) and the National Science Foundation (NSF) Graduate Research Fellowship Program (Grant DGE-2146755). CMR was supported by grants U01-CA253913 from the NCI as part of the Cancer Intervention and Surveillance Modeling Network (CISNET). FAE was supported by grants U01-CA253913 and U01-CA265750 from the NCI as part of CISNET. Any opinions, findings, and conclusions or recommendations expressed in

this material are those of the authors and do not necessarily reflect the views of the NIH, NSF, NCI, or CISNET.

8.2 Conflicts of interest

Authors have no conflicts of interest to declare.

8.3 Availability of data and material

All data generated for this manuscript are provided in a GitHub repository accessible at https://github.com/sjpi22/tutorial_cancer_modeling_des.

8.4 Ethics approval

Not applicable.

8.5 Consent to participate

Not applicable.

8.6 Code availability

All code used for this manuscript is provided in a GitHub repository accessible at https://github.com/sjpi22/tutorial_cancer_modeling_des.

8.7 Author contributions

Conceptualization, methodology: FAE, CMR, SP, JGF. Data curation, validation, writing — original draft: SP. Formal analysis: SP, FAE, CMR, CPA. Project administration: FAE, SP, CMR. Resources: JGF, FAE, JHC. Software: SP, CPA, FAE, CMR. Supervision: FAE, CMR, JHC, JGF. Visualization: SP, CPA, FAE. Writing — review and editing: all authors.

9 Acknowledgments

We are grateful to Michael C. Higgins, PhD, for feedback on the natural history model. Additionally, some of the computing for this project was performed on the Sherlock cluster. We would like to thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to these research results.

References

[1] Siebert U. When should decision-analytic modeling be used in the economic evaluation of health care? The European Journal of Health Economics, formerly: HEPAC. 2003 Sep;4(3):143–150. https://doi.org/10.1007/s10198-003-0205-2.

[2] Alagoz O, Berry DA, de Koning HJ, Feuer EJ, Lee SJ, Plevritis SK, et al. Introduction to the Cancer Intervention and Surveillance Modeling Network (CISNET) Breast Cancer Models. Medical Decision Making. 2018 Apr;38(1_suppl):3S–8S. Publisher: SAGE Publications Inc STM. https://doi.org/10.1177/0272989X17737507.

[3] Knudsen AB, Rutter CM, Peterse EFP, Lietz AP, Seguin CL, Meester RGS, et al. Colorectal Cancer Screening: An Updated Modeling Study for the US Preventive Services Task Force. JAMA. 2021 May;325(19):1998–2011. https://doi.org/10.1001/jama.2021.5746.

[4] Ng K, May FP, Schrag D. US Preventive Services Task Force Recommendations for Colorectal Cancer Screening: Forty-Five Is the New Fifty. JAMA. 2021 May;325(19):1943–1945. https://doi.org/10.1001/jama.2021.4133.

[5] Banks J, Carson JS. Introduction to discrete-event simulation. In: Proceedings of the 18th conference on Winter simulation - WSC '86. Washington, D.C., United States: ACM Press; 1986. p. 17–23. Available from: http://portal.acm.org/citation.cfm?doid=318242.318253.

[6] Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, Möller J. Modeling Using Discrete Event Simulation: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force–4. Medical Decision Making. 2012 Sep;32(5):701–711. Publisher: SAGE Publications Inc STM. https://doi.org/10.1177/0272989X12455462.

[7] Zhang X. Application of discrete event simulation in health care: a systematic review. BMC Health Services Research. 2018 Sep;18(1):687. https://doi.org/10.1186/s12913-018-3456-4.

[8] Hofmann M. On the Complexity of Parameter Calibration in Simulation Models. The Journal of Defense Modeling and Simulation. 2005 Oct;2(4):217–226. Publisher: SAGE Publications. https://doi.org/10.1177/154851290500200405.

[9] Kennedy MC, O'Hagan A. Bayesian Calibration of Computer Models. Journal of the Royal Statistical Society Series B: Statistical Methodology. 2001 Sep;63(3):425–464. https://doi.org/10.1111/1467-9868.00294.

[10] Chrysanthopoulou SA, Rutter CM, Gatsonis CA. Bayesian versus Empirical Calibration of Microsimulation Models: A Comparative Analysis. Medical Decision Making. 2021 Aug;41(6):714–726. Publisher: SAGE Publications Inc STM. https://doi.org/10.1177/0272989X211009161.

[11] Whyte S, Walsh C, Chilcott J. Bayesian Calibration of a Natural History Model with Application to a Population Model for Colorectal Cancer. Medical Decision Making. 2011 Jul;31(4):625–641. https://doi.org/10.1177/0272989X10384738.

[12] Rutter CM, Miglioretti DL, Savarino JE. Bayesian Calibration of Microsimulation Models. Journal of the American Statistical Association. 2009 Dec;104(488):1338–1350. Publisher: ASA Website _eprint: https://doi.org/10.1198/jasa.2009.ap07466. https://doi.org/10.1198/jasa.2009.ap07466.

[13] Rutter CM, Ozik J, DeYoreo M, Collier N. Microsimulation model calibration using Incremental Mixture Approximate Bayesian Computation. The annals of applied statistics. 2019 Dec;13(4):2189–2212. https://doi.org/10.1214/19-aoas1279.

[14] Pineda-Antunez C, Seguin C, van Duuren LA, Knudsen AB, Davidi B, Nascimento de Lima P, et al. Emulator-Based Bayesian Calibration of the CISNET Colorectal Cancer Models. Medical Decision Making. 2024 Jul;44(5):543–553. Publisher: SAGE Publications Inc STM. https://doi.org/10.1177/0272989X241255618.

[15] Vahdat V, Alagoz O, Chen JV, Saoud L, Borah BJ, Limburg PJ. Calibration and Validation of the Colorectal Cancer and Adenoma Incidence and Mortality (CRC-AIM) Microsimulation Model Using Deep Neural Networks. Medical Decision Making. 2023 Aug;43(6):719–736. Publisher: SAGE Publications Inc STM. https://doi.org/10.1177/0272989X231184175.

[16] DeYoreo M, Rutter CM, Ozik J, Collier N. Sequentially calibrating a Bayesian microsimulation model to incorporate new information and assumptions. BMC Medical Informatics and Decision Making. 2022 Jan;22(1):12. https://doi.org/10.1186/s12911-021-01726-0.

[17] Stout NK, Knudsen AB, Kong CY, McMahon PM, Gazelle GS. Calibration Methods Used in Cancer Simulation Models and Suggested Reporting Guidelines:. PharmacoEconomics. 2009 Jul;27(7):533–545. https://doi.org/10.2165/11314830-000000000-00000.

[18] Semochkina D, Walsh CD. Incorporating Additional Evidence as Prior Information to Resolve Non-Identifiability in Bayesian Disease Model Calibration: A Tutorial. Statistics in Medicine. 2025 Mar;44(6):e70039. https://doi.org/10.1002/sim.70039.

[19] Menzies NA, Soeteman DI, Pandya A, Kim JJ. Bayesian Methods for Calibrating Health Policy Models: A Tutorial. PharmacoEconomics. 2017 Jun;35(6):613–624. https://doi.org/10.1007/s40273-017-0494-4.

[20] Wade S, Sarich P, Vaneckova P, Behar-Harpaz S, Ngo PJ, Grogan PB, et al. Using Bayesian evidence synthesis to quantify uncertainty in population trends in smoking behaviour.

Statistical Methods in Medical Research. 2025 Mar;34(3):545–560. https://doi.org/10.1177/09622802241310326.

[21] Jalal H, Trikalinos TA, Alarid-Escudero F. BayCANN: Streamlining Bayesian calibration with artificial neural network metamodeling. Frontiers in Physiology. 2021 May;12:662314. https://doi.org/10.3389/fphys.2021.662314.

[22] Monks T, Harper A. Computer model and code sharing practices in healthcare discrete-event simulation: a systematic scoping review. Journal of Simulation. 2023 Sep;0(0):1–16. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/17477778.2023.2260772. https://doi.org/10.1080/17477778.2023.2260772.

[23] Ucar I, Smeets B, Azcorra A. **simmer** : Discrete-Event Simulation for *R*. Journal of Statistical Software. 2019;90(2). https://doi.org/10.18637/jss.v090.i02.

[24] Matloff N.: Introduction to Discrete-Event Simulation and the SimPy Language.

[25] Palmer GI, Knight VA, Harper PR, Hawa AL. Ciw: An open-source discrete event simulation library. Journal of Simulation. 2019 Jan;13(1):68–82. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/17477778.2018.1473909. https://doi.org/10.1080/17477778.2018.1473909.

[26] Lin YS, O'Mahony JF, van Rosmalen J. A Simple Cost-Effectiveness Model of Screening: An Open-Source Teaching and Research Tool Coded in R. PharmacoEconomics Open. 2023 Jun;7(4):507–523. https://doi.org/10.1007/s41669-023-00414-1.

[27] Prakash MK, Lang B, Heinrich H, Valli PV, Bauerfeind P, Sonnenberg A, et al. CMOST: an open-source framework for the microsimulation of colorectal cancer screening strategies. BMC medical informatics and decision making. 2017 Jun;17(1):80. https://doi.org/10.1186/s12911-017-0458-9.

[28] Chrysanthopoulou SA. MILC: A Microsimulation Model of the Natural History of Lung Cancer. International Journal of Microsimulation. 2017;10(3):5–26. Publisher: International Microsimulation Association.

[29] Shewmaker P, Chrysanthopoulou SA, Iskandar R, Lake D, Jutkowitz E. Microsimulation model calibration with approximate Bayesian computation in R: A tutorial. Medical decision making : an international journal of the Society for Medical Decision Making. 2022 Jul;42(5):557–570. https://doi.org/10.1177/0272989X221085569.

[30] Jalal H, Pechlivanoglou P, Krijkamp E, Alarid-Escudero F, Enns E, Hunink MGM. An Overview of R in Health Decision Sciences. Medical Decision Making. 2017 Oct;37(7):735–746. Publisher: SAGE Publications Inc STM. https://doi.org/10.1177/0272989X16686559.

[31] Krijkamp EM, Alarid-Escudero F, Enns EA, Jalal HJ, Hunink MM, Pechlivanoglou P. Microsimulation modeling for health decision sciences using R: a tutorial. Medical decision making : an international journal of the Society for Medical Decision Making. 2018 Apr;38(3):400. https://doi.org/10.1177/0272989X18754513.

[32] Alarid-Escudero F, Krijkamp EM, Pechlivanoglou P, Jalal H, Kao SYZ, Yang A, et al. A need for change! A coding framework for improving transparency in decision modeling. PharmacoEconomics. 2019;37(11). https://doi.org/10.1007/s40273-019-00837-x.

[33] Rutter CM, Zaslavsky AM, Feuer EJ. Dynamic Microsimulation Models for Health Outcomes: A Review. Medical Decision Making. 2011 Jan;31(1):10–18. Publisher: SAGE Publications Inc STM. https://doi.org/10.1177/0272989X10369005.

[34] Beckman RA, Kareva I, Adler FR. How Should Cancer Models Be Constructed? Cancer Control. 2020 Oct;27(1):1073274820962008. https://doi.org/10.1177/1073274820962008.

[35] Moolgavkar S, Luebeck G. Multistage Carcinogenesis: A Unified Framework for Cancer Data Analysis. In: Almudevar A, Oakes D, Hall J, editors. Statistical Modeling for Biological Systems: In

Memory of Andrei Yakovlev. Cham: Springer International Publishing; 2020. p. 117–136. Available from: https://doi.org/10.1007/978-3-030-34675-1_7.

[36] Putter H, Van Der Hage J, De Bock GH, Elgalta R, Van De Velde CJH. Estimation and Prediction in a Multi-State Model for Breast Cancer. Biometrical Journal. 2006 Jun;48(3):366–380. https://doi.org/10.1002/bimj.200510218.

[37] Cheung LC, Albert PS, Das S, Cook RJ. Multistate models for the natural history of cancer progression. British Journal of Cancer. 2022 Oct;127(7):1279–1288. https://doi.org/10.1038/s41416-022-01904-5.

[38] Alagoz O, Caswell-Jin JL, de Koning HJ, Huang H, Huang X, Lee SJ, et al. Mathematical Modeling to Address Questions in Breast Cancer Screening: An Overview of the Breast Cancer Models of the Cancer Intervention and Surveillance Modeling Network. Journal of Breast Imaging. 2025 Feb;p. wbaf003. https://doi.org/10.1093/jbi/wbaf003.

[39] Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling Good Research Practices—Overview: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. Value in Health. 2012 Sep;15(6):796–803. https://doi.org/10.1016/j.jval.2012.06.012.

[40] Lange JM, Gogebakan KC, Gulati R, Etzioni R. Projecting the Impact of Multi-Cancer Early Detection on Late-Stage Incidence Using Multi-State Disease Modeling. Cancer Epidemiology, Biomarkers & Prevention. 2024 Jun;33(6):830–837. https://doi.org/10.1158/1055-9965.EPI-23-1470.

[41] Loeve F, Boer R, van Oortmarssen GJ, van Ballegooijen M, Habbema JDF. The MISCAN-COLON Simulation Model for the Evaluation of Colorectal Cancer Screening. Computers and Biomedical Research. 1999 Feb;32(1):13–33. https://doi.org/10.1006/cbmr.1998.1498.

[42] Shwartz M. A Mathematical Model Used to Analyze Breast Cancer Screening Strategies. Operations Research. 1978;26(6):937–955. Publisher: INFORMS.

[43] Day NE, Walter SD. Simplified Models of Screening for Chronic Disease: Estimation Procedures from Mass Screening Programmes. Biometrics. 1984;40(1):1–13. Publisher: [Wiley, International Biometric Society]. https://doi.org/10.2307/2530739.

[44] Hubbell E, Clarke CA, Aravanis AM, Berg CD. Modeled Reductions in Late-stage Cancer with a Multi-Cancer Early Detection Test. Cancer Epidemiology, Biomarkers & Prevention. 2021 Mar;30(3):460–468. https://doi.org/10.1158/1055-9965.EPI-20-1134.

[45] Meza R, Ten Haaf K, Kong CY, Erdogan A, Black WC, Tammemagi MC, et al. Comparative analysis of 5 lung cancer natural history and screening models that reproduce outcomes of the NLST and PLCO trials. Cancer. 2014 Jun;120(11):1713–1724. https://doi.org/10.1002/cncr.28623.

[46] Zauber AG, Winawer SJ, O'Brien MJ, Lansdorp-Vogelaar I, Van Ballegooijen M, Hankey BF, et al. Colonoscopic Polypectomy and Long-Term Prevention of Colorectal-Cancer Deaths. New England Journal of Medicine. 2012 Feb;366(8):687–696. https://doi.org/10.1056/NEJMoa1100370.

[47] Risio M. The natural history of adenomas. Best Practice & Research Clinical Gastroenterology. 2010 Jun;24(3):271–280. https://doi.org/10.1016/j.bpg.2010.04.005.

[48] Morson BC. Genesis of Colorectal Cancer. Clinics in Gastroenterology. 1976 Sep;5(3):505–525. https://doi.org/10.1016/S0300-5089(21)00305-9.

[49] Park TW, Fujiwara H, Wright TC. Molecular biology of cervical cancer and its precursors. Cancer. 1995 Nov;76(S10):1902–1913. https://doi.org/10.1002/1097-0142(19951115)76:10+⟨1902::AID-CNCR2820761306⟩3.0.CO;2-0.

[50] Plevritis SK, Salzman P, Sigal BM, Glynn PW. A natural history model of stage progression applied to breast cancer. Statistics in Medicine. 2007 Feb;26(3):581–595. https://doi.org/10.1002/sim.2550.

[51] Caro JJ, Möller J, Karnon J, Stahl J, Ishak J. Discrete Event Simulation for Health Technology Assessment. New York: Chapman and Hall/CRC; 2015.

[52] Trikalinos TA, Sereda Y. The nhppp package for simulating non-homogeneous Poisson point processes in R. PLOS ONE. 2024 Nov;19(11):e0311311. Publisher: Public Library of Science. https://doi.org/10.1371/journal.pone.0311311.

[53] Gangnon RE, Stout NK, Alagoz O, Hampton JM, Sprague BL, Trentham-Dietz A. Contribution of Breast Cancer to Overall Mortality for US Women. Medical Decision Making. 2018 Apr;38(1_suppl):24S–31S. https://doi.org/10.1177/0272989X17717981.

[54] Rosenberg MA. Chapter 3: Competing Risks to Breast Cancer Mortality. JNCI Monographs. 2006 Oct;2006(36):15–19. https://doi.org/10.1093/jncimonographs/lgj004.

[55] Surveillance Research Program, National Cancer Institute.: SEER*Explorer: An interactive website for SEER cancer statistics. Available from: https://seer.cancer.gov/statistics-network/explorer/.

[56] Campos NG, Demarco M, Bruni L, Desai KT, Gage JC, Adebamowo SN, et al. A proposed new generation of evidence-based microsimulation models to inform global control of cervical cancer. Preventive Medicine. 2021 Mar;144:106438. https://doi.org/10.1016/j.ypmed.2021.106438.

[57] Burger EA, De Kok IMCM, Groene E, Killen J, Canfell K, Kulasingam S, et al. Estimating the Natural History of Cervical Carcinogenesis Using Simulation Models: A CISNET Comparative Analysis. JNCI: Journal of the National Cancer Institute. 2020 Sep;112(9):955–963. https://doi.org/10.1093/jnci/djz227.

[58] Mandelblatt JS, Near AM, Miglioretti DL, Munoz D, Sprague BL, Trentham-Dietz A, et al. Common Model Inputs Used in CISNET Collaborative Breast Cancer Modeling. Medical Decision Making. 2018 Apr;38(1_suppl):9S–23S. Publisher: SAGE Publications Inc STM. https://doi.org/10.1177/0272989X17700624.

[59] Rose J, Augestad KM, Kong CY, Meropol NJ, Kattan MW, Hong Q, et al. A simulation model of colorectal cancer surveillance and recurrence. BMC Medical Informatics and Decision Making. 2014 Apr;14(1):29. https://doi.org/10.1186/1472-6947-14-29.

[60] Armitage P, Doll R. The Age Distribution of Cancer and a Multi-stage Theory of Carcinogenesis. British Journal of Cancer. 1954 Mar;8(1):1–12.

[61] Pinsky PF. Estimation and Prediction for Cancer Screening Models Using Deconvolution and Smoothing. Biometrics. 2001;57(2):389–395. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.0006-341X.2001.00389.x. https://doi.org/10.1111/j.0006-341X.2001.00389.x.

[62] Chen CD, Yen MF, Wang WM, Wong JM, Chen TH. A case–cohort study for the disease natural history of adenoma–carcinoma and de novo carcinoma and surveillance of colon and rectum after polypectomy: implication for efficacy of colonoscopy. British Journal of Cancer. 2003 Jun;88(12):1866–1873. Number: 12 Publisher: Nature Publishing Group. https://doi.org/10.1038/sj.bjc.6601007.

[63] Madadi M, Heydari M, Zhang S, Pohl E, Rainwater C, Williams DL. Analyzing overdiagnosis risk in cancer screening: A case of screening mammography for breast cancer. IISE Transactions on Healthcare Systems Engineering. 2018 Jan;8(1):2–20. https://doi.org/10.1080/24725579.2017.1396512.

[64] Pi S, Goldhaber-Fiebert JD, Alarid-Escudero F.: Calculating epidemiological outcomes from simulated longitudinal data. Available from: https://www.medrxiv.org/content/10.1101/2025.04.30.25326766v1.full.

[65] Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. Statistical Methods in Medical Research. 2001 Aug;10(4):277–303. Publisher: SAGE Publications Ltd STM.

https://doi.org/10.1177/096228020101000404.

[66] Vanni T, Karnon J, Madan J, White RG, Edmunds WJ, Foss AM, et al. Calibrating Models in Economic Evaluation. PharmacoEconomics. 2011 Jan;29(1):35–49. https://doi.org/10.2165/11584600-000000000-00000.

[67] Marin JM, Pudlo P, Robert CP, Ryder RJ. Approximate Bayesian computational methods. Statistics and Computing. 2012 Nov;22(6):1167–1180. https://doi.org/10.1007/s11222-011-9288-2.

[68] Van Dongen S. Prior specification in Bayesian statistics: Three cautionary tales. Journal of Theoretical Biology. 2006 Sep;242(1):90–100. https://doi.org/10.1016/j.jtbi.2006.02.002.

[69] Maerzluft CE, Rutter CM, Ozik J, Collier N.: imabc: Incremental Mixture Approximate Bayesian Computation (IMABC). Available from: https://CRAN.R-project.org/package=imabc.

[70] Degeling K, IJzerman MJ, Lavieri MS, Strong M, Koffijberg H. Introduction to Metamodeling for Reducing Computational Burden of Advanced Analyses with Health Economic Models: A Structured Overview of Metamodeling Methods in a 6-Step Application Process. Medical Decision Making. 2020 Apr;40(3):348–363. https://doi.org/10.1177/0272989X20912233.

[71] Lee K, Jalal H, Raviotta JM, Krauland MG, Zimmerman RK, Burke DS, et al. Estimating the Impact of Low Influenza Activity in 2020 on Population Immunity and Future Influenza Seasons in the United States. Open Forum Infectious Diseases. 2022 Jan;9(1):ofab607. https://doi.org/10.1093/ofid/ofab607.

[72] Zou J, Han Y, So SS. Overview of Artificial Neural Networks. In: Livingstone DJ, editor. Artificial Neural Networks: Methods and Applications. Totowa, NJ: Humana Press; 2009. p. 14–22. Available from: https://doi.org/10.1007/978-1-60327-101-1_2.

[73] Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: A Probabilistic Programming Language. Journal of statistical software. 2017;76:1. https://doi.org/10.18637/jss.v076.i01.

[74] 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. Medical Decision Making. 2012 Sep;32(5):733–743. https://doi.org/10.1177/0272989X12454579.

[75] Alarid-Escudero F, Gulati R, Rutter CM. Validation of Microsimulation Models Used for Population Health Policy. In: Apostolopoulos Y, Lemke MK, Hassmiller Lich K, editors. Complex Systems and Population Health. Oxford University Press; 2020. p. 0. Available from: https://doi.org/10.1093/oso/9780190880743.003.0016.

[76] Goldhaber-Fiebert JD, Stout NK, Goldie SJ. Empirically Evaluating Decision-Analytic Models. Value in Health. 2010 Jul;13(5):667–674. https://doi.org/10.1111/j.1524-4733.2010.00698.x.

[77] Alarid-Escudero F, MacLehose RF, Peralta Y, Kuntz KM, Enns EA. Nonidentifiability in model calibration and implications for medical decision making. Medical decision making : an international journal of the Society for Medical Decision Making. 2018 Oct;38(7):810–821. https://doi.org/10.1177/0272989X18792283.

[78] Van De Schoot R, Depaoli S. Bayesian analyses: Where to start and what to report. European Health Psychologist. 2014;16(2):75–84.

[79] Brawley OW, Kramer BS. Cancer screening in theory and in practice. Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology. 2005 Jan;23(2):293–300. https://doi.org/10.1200/JCO.2005.06.107.

[80] Russell LB. Educated Guesses: Making Policy about Medical Screening Tests. vol. 16. University of California Press; 1994. Publisher: SAGE Publications Inc. Available from: https://doi.org/10.1177/027046769601600190.

[81] Lopez-Beltran A, Cookson MS, Guercio BJ, Cheng L. Advances in diagnosis and treatment of bladder cancer. BMJ. 2024 Feb;p. e076743. https://doi.org/10.1136/bmj-2023-076743.

[82] Fradet Y. Screening for bladder cancer: the best opportunity to reduce mortality. Canadian Urological Association Journal. 2013 May;3(6-S4):180. https://doi.org/10.5489/cuaj.1192.

[83] Sereda Y, Alarid-Escudero F, Bickell NA, Chang SH, Colditz GA, Hur C, et al. Approaches to developing de novo cancer population models to examine questions about cancer and race in bladder, gastric, and endometrial cancer and multiple myeloma: the Cancer Intervention and Surveillance Modeling Network incubator program. Journal of the National Cancer Institute Monographs. 2023 Nov;2023(62):219–230. https://doi.org/10.1093/jncimonographs/lgad021.

[84] French Institute for Demographic Studies (France), Max Planck Institute for Demographic Research (Germany), the University of California, Berkeley (USA).: HMD. Human Mortality Database. Available from: https://www.mortality.org/Country/Country?cntr=USA.

[85] Ries LAG, Kosary CL, Hankey BF, Miller BA, Edwards BK. SEER Cancer Statistics Review, 1973-1995. Bethesda, MD: National Cancer Institute; 1998. Available from: https://seer.cancer.gov/archive/csr/1973_1995/citation.pdf.

[86] Suen KC, Lau LL, Yermakov V. Cancer and old age.An autopsy study of 3,535 patients over 65 years old. Cancer. 1974 Apr;33(4):1164–1168. https://doi.org/10.1002/1097-0142(197404)33:4⟨1164::AID-CNCR2820330440⟩3.0.CO;2-O.

[87] Winawer SJ. The History of Colorectal Cancer Screening: A Personal Perspective. Digestive Diseases and Sciences. 2015 Mar;60(3):596–608. https://doi.org/10.1007/s10620-014-3466-y.

[88] Rutter CM, Yu O, Miglioretti DL. A hierarchical non-homogenous Poisson model for meta-analysis of adenoma counts. Statistics in Medicine. 2007 Jan;26(1):98–109. https://doi.org/10.1002/sim.2460.

[89] Bergstra J, Bengio Y. Random search for hyper-parameter optimization. J Mach Learn Res. 2012 Feb;13(null):281–305.

[90] Kohavi R, John GH. Automatic Parameter Selection by Minimizing Estimated Error. In: Prieditis A, Russell S, editors. Machine Learning Proceedings 1995. San Francisco (CA): Morgan Kaufmann; 1995. p. 304–312. Available from: https://www.sciencedirect.com/science/article/pii/B9781558603776500451.

[91] Tian Y, Chao MA, Kulkarni C, Goebel K, Fink O. Real-time model calibration with deep reinforcement learning. Mechanical Systems and Signal Processing. 2022 Feb;165:108284. https://doi.org/10.1016/j.ymssp.2021.108284.

[92] Nascimento de Lima P. Robust Decision Making in Health Policy: Applications to COVID-19 and Colorectal Cancer. Pardee RAND Graduate School; 2022. Available from: https://www.rand.org/pubs/rgs_dissertations/RGSDA2531-1.html.

[93] Hadka D, Herman J, Reed P, Keller K. An open source framework for many-objective robust decision making. Environmental Modelling & Software. 2015 Dec;74:114–129. https://doi.org/10.1016/j.envsoft.2015.07.014.

[94] Lempert RJ. Robust Decision Making (RDM). In: Marchau VAWJ, Walker WE, Bloemen PJTM, Popper SW, editors. Decision Making under Deep Uncertainty: From Theory to Practice. Cham: Springer International Publishing; 2019. p. 23–51. Available from: https://doi.org/10.1007/978-3-030-05252-2_2.

[95] Meza R, Jeon J, Toumazis I, ten Haaf K, Cao P, Bastani M, et al. Evaluation of the Benefits and Harms of Lung Cancer Screening With Low-Dose Computed Tomography: Modeling Study for the US Preventive Services Task Force. JAMA. 2021 Mar;325(10):988–997. https://doi.org/10.1001/jama.2021.1077.

[96] Rutter CM, Knudsen AB, Marsh TL, Doria-Rose VP, Johnson E, Pabiniak C, et al. Validation of Models Used to Inform Colorectal Cancer Screening Guidelines: Accuracy and Implications. Medical Decision Making. 2016 Jul;36(5):604–614. Publisher: SAGE Publications Inc STM. https://doi.org/10.1177/0272989X15622642.

[97] Berg DMNvd, Lima PNd, Knudsen AB, Rutter CM, Weinberg D, Lansdorp-Vogelaar I, et al. NordICC Trial Results in Line With Expected Colorectal Cancer Mortality Reduction After Colonoscopy: A Modeling Study. Gastroenterology. 2023 Oct;165(4):1077–1079.e2. Publisher: Elsevier. https://doi.org/10.1053/j.gastro.2023.06.035.

[98] Caswell-Jin JL, Sun LP, Munoz D, Lu Y, Li Y, Huang H, et al. Analysis of Breast Cancer Mortality in the US—1975 to 2019. JAMA. 2024 Jan;331(3):233–241. https://doi.org/10.1001/jama.2023.25881.

[99] Trentham-Dietz A, Alagoz O, Chapman C, Huang X, Jayasekera J, Ravesteyn NTv, et al. Reflecting on 20 years of breast cancer modeling in CISNET: Recommendations for future cancer systems modeling efforts. PLOS Computational Biology. 2021 Jun;17(6):e1009020. Publisher: Public Library of Science. https://doi.org/10.1371/journal.pcbi.1009020.

[100] Kuntz KM, Weinstein MC. Modelling in economic evaluation. In: Drummond M, McGuire A, editors. Economic Evaluation in Health Care: Merging theory with practice. Oxford University Press; 2001. p. 0. Available from: https://doi.org/10.1093/oso/9780192631770.003.0007.

[101] Hoveling LA, Lepe A, Boissonneault M, de Beer JAA, Smidt N, de Kroon MLA, et al. Educational inequalities in metabolic syndrome prevalence, timing, and duration amongst adults over the life course: a microsimulation analysis based on the lifelines cohort study. The International Journal of Behavioral Nutrition and Physical Activity. 2023 Sep;20:104. https://doi.org/10.1186/s12966-023-01495-1.

[102] El-Serag HB, Mason AC. Risk Factors for the Rising Rates of Primary Liver Cancer in the United States. Archives of Internal Medicine. 2000 Nov;160(21):3227. https://doi.org/10.1001/archinte.160.21.3227.

[103] Wroblewski LE, Peek RM, Wilson KT. *Helicobacter pylori* and Gastric Cancer: Factors That Modulate Disease Risk. Clinical Microbiology Reviews. 2010 Oct;23(4):713–739. https://doi.org/10.1128/CMR.00011-10.

[104] Jeon J, Meza R, Krapcho M, Clarke LD, Byrne J, Levy DT. *Chapter 5* : Actual and Counterfactual Smoking Prevalence Rates in the U.S. Population via Microsimulation. Risk Analysis. 2012 Aug;32(s1). https://doi.org/10.1111/j.1539-6924.2011.01775.x.

[105] Meester RGS, Ladabaum U. Impact of the serrated pathway on the simulated comparative effectiveness of colorectal cancer screening tests. JNCI Cancer Spectrum. 2024 Oct;8(5):pkae077. https://doi.org/10.1093/jncics/pkae077.

[106] Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, et al. Multitarget Stool DNA Testing for Colorectal-Cancer Screening. New England Journal of Medicine. 2014 Apr;370(14):1287–1297. Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMoa1311194. https://doi.org/10.1056/NEJMoa1311194.

[107] Kim JJ, Kuntz KM, Stout NK, Mahmud S, Villa LL, Franco EL, et al. Multiparameter Calibration of a Natural History Model of Cervical Cancer. American Journal of Epidemiology. 2007 Jun;166(2):137–150. https://doi.org/10.1093/aje/kwm086.

[108] Campos NG, Burger EA, Sy S, Sharma M, Schiffman M, Rodriguez AC, et al. An Updated Natural History Model of Cervical Cancer: Derivation of Model Parameters. American Journal of Epidemiology. 2014 Sep;180(5):545–555. https://doi.org/10.1093/aje/kwu159.

[109] Cyhaniuk A, Coombes ME. Longitudinal adherence to colorectal cancer screening guidelines. The American Journal of Managed Care. 2016 Feb;22(2):105–111.

[110] Mohl JT, Ciemins EL, Miller-Wilson LA, Gillen A, Luo R, Colangelo F. Rates of Follow-up Colonoscopy After a Positive Stool-Based Screening Test Result for Colorectal Cancer Among Health Care Organizations in the US, 2017-2020. JAMA Network Open. 2023 Jan;6(1):e2251384. https://doi.org/10.1001/jamanetworkopen.2022.51384.

[111] Bretthauer M, Løberg M, Wieszczy P, Kalager M, Emilsson L, Garborg K, et al. Effect of Colonoscopy Screening on Risks of Colorectal Cancer and Related Death. New England Journal of Medicine. 2022 Oct;387(17):1547–1556. Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMoa2208375. https://doi.org/10.1056/NEJMoa2208375.

[112] Sabatino SA, Thompson TD, White MC, Villarroel MA, Shapiro JA, Croswell JM, et al. Up-to-Date Breast, Cervical, and Colorectal Cancer Screening Test Use in the United States, 2021. Preventing Chronic Disease. 2023 Oct;20:E94. https://doi.org/10.5888/pcd20.230071.

[113] Zhu X, Parks PD, Weiser E, Fischer K, Griffin JM, Limburg PJ, et al. National Survey of Patient Factors Associated with Colorectal Cancer Screening Preferences. Cancer Prevention Research. 2021 May;14(5):603–614. https://doi.org/10.1158/1940-6207.CAPR-20-0524.

[114] Ayer T, Alagoz O, Stout NK, Burnside ES. Heterogeneity in Women's Adherence and Its Role in Optimal Breast Cancer Screening Policies. Management Science. 2016 May;62(5):1339–1362. https://doi.org/10.1287/mnsc.2015.2180.

[115] D'Andrea E, Ahnen DJ, Sussman DA, Najafzadeh M. Quantifying the impact of adherence to screening strategies on colorectal cancer incidence and mortality. Cancer Medicine. 2019 Nov;9(2):824–836. https://doi.org/10.1002/cam4.2735.

[116] Cronin KA, Yu B, Krapcho M, Miglioretti DL, Fay MP, Izmirlian G, et al. Modeling the dissemination of mammography in the United States. Cancer Causes & Control. 2005 Aug;16(6):701–712. https://doi.org/10.1007/s10552-005-0693-8.

[117] Surveillance, Epidemiology, and End Results (SEER) Program, National Cancer Institute.: Rate Algorithms. Available from: https://seer.cancer.gov/help/seerstat/equations-and-algorithms/rate-algorithms.

[118] Hernán MA. The Hazards of Hazard Ratios. Epidemiology (Cambridge, Mass). 2010 Jan;21(1):13–15. https://doi.org/10.1097/EDE.0b013e3181c1ea43.

[119] Juckett DA, Rosenberg B. Comparison of the Gompertz and Weibull functions as descriptors for human mortality distributions and their intersections. Mechanisms of Ageing and Development. 1993 Jun;69(1):1–31. https://doi.org/10.1016/0047-6374(93)90068-3.

A Adapting the model

A.1 Model configurations

The YAML files in the `configs` folder provide a central location for model structure specifications. Under `params_model`, parameters that serve as inputs to the `load_model_params()` function in `R/01_model_inputs_functions.R` can be defined, such as the indicator for whether to include a precancerous lesion state (`lesion_state`) and the list of cancer stages (`v_cancer`). Similarly, customized parameters for IMABC, BayCANN, Monte Carlo error analyses, target coverage analyses, and screening strategy analyses can be set respectively in `params_imabc`, `params_baycann`, `params_montecarlo`, `params_coverage`, and `params_screening`.

Each random variable in the simulation model may be customized by providing the distribution name, parameters, and source in a list. For example, we may create a list object named `d_time_H_L` for the distribution of the time from the healthy state to the first precancerous lesion as follows, using the base R accelerated failure time parametrization of the Weibull distribution:

```
d_time_H_L <- list(distr = 'weibull',
                   params = list(shape = 2, scale = 75),
```

```
                    src = 'unknown')
```

With this framework, the user may substitute different parametric or empirical distributions for each of the TTE variables. We may then draw 10 samples from the distribution using the `query_distr` function as follows:

```
time_H_L <- query_distr(target = 'r',
                        x = 10,
                        distr = d_time_H_L$distr,
                        params = d_time_H_L$params)
```

The `query_distr()` function appends the `target` letter to the distribution name and calls the resulting function with `x` as the first input and `params` as the following inputs, following R's syntax for probability distributions, in which `d` indicates the density, `p` the probability, `q` the quantile, and `r` random generation. As a result, the value of `distr` should be a recognized function when appended to one of the four values of `target`; acceptable values corresponding to R base distributions include `exp`, `gamma`, and `norm`, while some functions may be loaded from other packages (for instance, functions for the Gompertz distribution are available from the `VGAM` or `flexsurv` packages).

We have also provided functions to sample from empirical distributions in the file `R/utils/distr_empirical.R`. The required input data include an ordered discrete variable `xs`, such as an age interval, and the corresponding probability mass `probs`. The function `dempirical()` retrieves the value of the probability mass function for the interval containing the queried value. The function `pempirical()` calculates the cumulative probability by summing the values of `probs` up to the queried value of `xs`. Its inverse is `qempirical()`, which outputs the corresponding quantile of a probability input. The random generator function `rempirical()` samples from the categorical variables weighted by the probability mass. For continuous variables such as age, the function includes an option to apply a uniform correction to achieve values in between the discrete values of `xs` and `probs`.

To modify the time-to-event distributions used in the simulation model, the user may do the following:

1. Specify distributions in a CSV file such as `data/priors.csv`, which updates the distributions of the natural history model accordingly when used as an input to `load_calib_params()` in `03_calibration_general_functions.R` as `file_priors` or to `load_calib_params()` in `R/01_model_inputs_functions.R` as `file.distr`
2. Manually change the distributions after loading the model parameters.
3. Change the default distributions set in `01_model_inputs_functions.R/load_model_params()`

A.2 Data

To run an analysis with custom input data and calibration targets, replace the default datasets in the `data` folder with custom data and make the following modifications according to the type of input.

1. Model inputs (background mortality, relative survival)
   (a) Update the data file paths for background mortality (`file.mort`) and relative survival (`file.mort`) in one of the YAML files in the `configs` folder under `params_model`.
   (b) If necessary, modify `load_lifetables()` for the background mortality data and `load_surv_data()` in `R/utils/data_processing.R` for the relative survival data so that the data inputs can respectively be converted to probability distribution objects using `set_mort_distr()` and `set_surv_distr()`, also defined in `R/utils/data_processing.R`. These functions are all called in `load_model_params()`, which is defined in `R/01_model_inputs_functions.R` and loads the cancer natural history model parameters.
2. Calibration targets
   (a) Replace or add parameters for calibration targets under `params_calib:l_params_outcome` for non-lesion calibration targets, such as prevalence of preclinical cancer, incidence, and stage distribution, and `params_calib:lesion_state_true:l_params_outcome` for precancerous lesion-specific calibration targets in the the applicable YAML file in the `configs` folder. Each calibration target should be associated with a uniquely labeled list of parameters, including
      (i) `file_path`: path to target data

(ii) `outcome_type`: label for target type, which indicates the functions to load the data in `R/utils/data_processing.R` (`load_[outcome_type]()`) and calculate the analogous outcome from simulated data in `R/utils/epi_functions.R` (`calc_[outcome_type]()`)
(iii) `categorical`: indicator for whether the target is a categorical distribution (i.e., percentages that must sum to 1)
(iv) `get_params`: named list of inputs to `calc_[outcome_type]()` that must be retrieved using `get()`
(v) `lit_params`: named list of literal inputs to `calc_[outcome_type]()`

(b) If necessary, modify or create functions of the form `load_[outcome_type]()` in `R/utils/data_processing.R` to load and process each type of target. The output should include the following columns: "target_names" with unique labels for each target, "target_groups" to label the entire category of outcomes, "target_index" for the indices associated with each value of the target, "targets" for the target value at each index, and "se" for the standard error
(c) For any new types of calibration targets, create functions of the form `calc_[outcome_type]()` in `R/utils/epi_functions.R` to calculate the corresponding targets from a matrix of simulated patient trajectories

A.3 Choice of cohort size

When implementing an individual-level simulation model, a key question is how many life trajectories to simulate. Selecting the cohort size involves similar considerations as choosing a sample size in prospective trials; as the cohort size increases, the Monte Carlo error of the summary outcome estimates decreases [31], but the time required to run the simulation increases [29]. The acceptable Monte Carlo error of the summary statistics generated depends on the task at hand (e.g., calibration, validation, policy evaluation). To facilitate convergence during calibration, the Monte Carlo error of the simulated outcomes should not exceed the estimation error of the corresponding calibration targets [29]. For strategy evaluation, the Monte Carlo variability should be much smaller than the mean differences in outcomes between strategies [100].

To determine the minimum sample size N_{target} required for Monte Carlo variability to be lower than each target's standard error SE_{target}, a modeler could first calculate the minimum standard error SE_n of the outcomes of interest with a small preliminary sample n, such as $n = 1,000$, then solve $N_{target} = n(SE_n/SE_{target})^2$, as the standard error and the square root of the sample size are inversely proportional [100]. The modeler may then analyze the computation time for varying sample sizes to determine a cohort size reasonably larger than N_{target} that falls within their computational limits [29]. One study alternatively determined its microsimulation cohort size by performing 20 simulations each for sample sizes starting at 100,000 and incrementing by 100,000 until the standard deviation of the parameter estimates no longer decreased meaningfully, thus achieving near-minimum Monte Carlo uncertainty [101].

A.4 Further augmentations

This section describes ways to adapt the current model implementation to accommodate additional cancer natural history assumptions and interventions. For an overview of where to add functions and parameters and how to specify model configurations to support these adaptations, refer to Appendix A.2.

A.4.1 Additional states

Modeling certain cancer types or evaluating certain non-screening interventions may require additional states from those that are currently implemented in the model. For example, infection may precede precancerous dysplasia or cancer onset in the case of human papillomavirus for cervical cancer [56], hepatitis B or C for liver cancer [102], and *H. pylori* for gastric cancer [103]. Thus, modeling interventions to prevent or treat such infections may require adding a state to reflect reduced transition rates to cancer through that pathway. Furthermore, to evaluate surveillance strategies for cancer recurrence or therapies based on the line of treatment, states for recurrence and detection of recurrence must be added between initial diagnosis and death from cancer.

Additional states in the natural history process can be integrated into the code by performing the following steps:

1. Updates to `01_load_model_inputs.R`
   (a) Assign a label for the new state as a letter different from existing state labels (H, L, P, C, and D) and assign its order relative to other states.

(b) Add a binary indicator to the arguments of the `load_model_params()` function for the inclusion of the state in the decision model, analogous to the indicator for the precancerous lesion state (L).
(c) Update the vector of states `v_states` to include the new state if the indicator argument for the state is `TRUE`.
(d) Create placeholder lists for relevant probability distributions for time-to-event variables leading to and from the new state and state. Wrap the lists in an if-statement that checks whether the state label is in `v_states`. Adjust the code to default to the correct placeholders if the state is not included.

2. Updates to `02_decision_model_functions.R`
   (a) Create one or more subfunctions to generate the events and characteristics related to the new state, including variables that lead into the new state and the following state (for example, one could create a `simulate_infection_onset()` function to simulate the time of a precipitating infection). The function(s) should take the patient matrix `m_times` and model parameters `l_params_all` as inputs. If the new state progresses into the lesion or preclinical cancer state, update `simulate_disease_onset()` to integrate the time of disease onset with the state.
   (b) Call the subfunctions in the main function `run_base_model()` in the desired order of the states. Wrap the subfunctions in an if statement calling them if the new state is in `v_states`.
3. Sanity checks
   (a) Load model parameters and check that the distributions for time-to-event variables connecting to and from the new state exist.
   (b) Run the model with the loaded parameters and check that the patient-level matrix outputs contain the time-to-event variables. Check that they sum up as you would expect.

A.4.2 Risk factors

Risk factors that are treated as time-invariant in the model, such as insurance status, family history, or geographic region at the time of screening initiation, may be added in one line to the `simulate_baseline_data()` module. Their population distributions, as well as their interactions with other distributions' parameters (for example, as covariates in the hazard of cancer onset) may be specified in `load_model_params()`.

For time-varying risk factors, such as smoking behaviors [104], a module can be added using the steps in A.4.1 to simulate the trajectory of the risk factor over each individual's lifespan. A separate module can be added to specify how disease risk or mortality changes with the risk factor, as implemented by the CISNET Lung Group models [95].

A.4.3 Additional subtypes

Heterogeneity in precancerous lesion pathways, as in colorectal cancer [105], or cancer subtypes, such as for breast cancer [38], may be important to model if associated with differential outcomes. To simulate multiple precancerous lesion types, one may sample an onset time and lifetime lesion count for each type in the `simulate_disease_onset()` subfunction, potentially with an individual lesion risk index to induce correlation. Then, in the `simulate_additional_lesions()` subfunction, one may loop over the lesion types to simulate the time to cancer progression using type-specific progression rates. In the screening module, the screening sensitivity inputs and code for sampling positive test results can be modified to use different detection probabilities by lesion type.

Assuming that individuals only develop one of several mutually exclusive cancer subtypes in the model, one may sample from the observed distribution of subtypes from cancer incidence data, then sample an age of onset by subtype. Differential cancer progression rates, symptomatic detection rates, survival, and intervention efficacy may also be specified based on subtype. If multiple subtypes of cancer per individual are modeled, one may generate event times starting from cancer onset at the tumor level, as is done for precancerous lesions.

A.4.4 Lesion size, stages, and regression

Precancerous lesion or tumor size may be important to model if screening test sensitivity varies with lesion size, as in the case of stool tests for colorectal cancer [106]. The trajectory of lesion growth can be modeled using discrete size categories with specified transition distributions between categories. Similar methods can be used to model precancerous lesion stages, such as cervical intraepithelial neoplasia (CIN) 1, 2, and 3 [107]. Tumor growth has also been modeled in continuous time using exponential, Gompertz,

and Janoschek curves [12, 28, 50]. Parameters for transition distributions and growth curves can be calibrated to real-world targets on the distribution of lesion size. In the screening module, a function for the test sensitivity based on lesion size would be added as an input so that the probability of detection based on the lesion size at the time of the screening test can be set.

In addition, some cancer types, such as cervical cancer [107, 108], show evidence of precancerous lesion regression. To model regression, after instantiating lesions, one can sample a binary variable indicating whether a lesion will regress or progress. If modeling lesion size categories, one may sample the times to the maximum lesion size and transitions to smaller categories until the lesions that will regress no longer exist. Continuous trajectories for lesion shrinkage over time can also be used. In the screening module, one can either assign a test sensitivity of zero for lesions that are too small to be detected or exclude lesions that have completely regressed from screening.

A.4.5 Incomplete resection

To model incomplete resection of precancerous lesions during an invasive screening or diagnostic test, the probability of incomplete resection, the distribution of the size of the remaining lesion, the growth trajectory after resection, and assumptions for cancer progression after resection would need to be added as inputs or otherwise specified. Using these parameters, for lesions detected at each screening, one could sample whether the lesion is fully resected, flag completely resected lesions as removed, reset the size of incompletely resected lesions, sample a new time to preclinical cancer for incompletely resected lesions, and calculate the lesion size at the next screening test.

A.4.6 Adherence

Given the imperfect real-world adherence to U.S. cancer screening guidelines [109–112] and differential preferences in screening modalities [113, 114], there is increasing interest in adherence as a parameter in models evaluating screening strategies, such as in [115]. In the simplest case, adherence may be modeled as a single probability that is applied to each individual and independent between tests, though this may not be realistic. To implement this form of adherence, at each time that individuals are due for screening, before sampling test results, one may sample whether individuals participate and proceed with calculating test outcomes (e.g., false positives, detection of lesions or cancer) only among adherent individuals. Alternatively, at the time that individuals are due for screening, one may sample from a time-to-event (TTE) distribution for the delay in testing to simulate realistic uptake patterns. To incorporate heterogeneity in screening adherence, one can assign individuals to categories of adherence and either set a fixed screening interval or sample from separate TTE distributions for each category. For example, in [116], individuals are categorized as annual screeners, biennial screeners, and irregular screeners, each with separate survival curves for the time to next breast cancer exam.

A.4.7 Additional decision outcomes

Though screening test costs are incorporated in the pipeline already, more outcomes, such as quality-adjusted life years (QALYs) and cancer treatment costs, may be needed for a comprehensive cost-effectiveness analysis, weighing the financial burden of screening programs against the life extension and cost savings of screening-mediated cancer prevention or early detection. Since the DES model already simulates the time in each health state, one can apply state-specific utility weights to calculate QALYs as well as annual costs of living in a particular health state. At the time of each screening test, one can also sample the occurrence of screening-related adverse events and their associated costs.

B Target-informed priors

For an acyclic disease model with progressive states, we show how to quantitatively derive prior distributions for the parameters governing the time to disease onset given epidemiological data for each state. In the cancer setting, we take data on the incidence of clinical cancer $I_C(t)$ and prevalence of preclinical cancer $P_P(t)$ across multiple age ranges. The midpoint of the i-th age range is denoted by t_i. First, we derive the CDF of clinical cancer $F_C(t)$ from the incidence rates. Next, we use the CDF of clinical cancer and the prevalence of preclinical cancer to estimate the CDF for the time to preclinical cancer onset $F_P(t)$. The same methods can be used to derive the CDF for precancerous lesion onset $F_L(t)$ given data on the prevalence $P_L(t)$ of precancerous lesions by age. In this appendix, we also discuss how cancer stage distribution targets can inform the parameter priors for cancer stage progression.

B.1 Deriving the CDF of clinical cancer from the incidence rate

The National Cancer Institute (NCI)'s Surveillance, Epidemiology, and End Results (SEER) Program calculates cancer incidence as the number of new cases in a year divided by the total population [117]. Using the `cobs` R package, we fit a constrained B-spline $\hat{I}_C(t)$ to the incidence values $I_C(t_i)$ at each midpoint of the age range t_i, with knots at 0, t_2, every 3rd midpoint after t_2, and the maximum upper bound of the age ranges. The knots are sparse to prevent overfitting. For improved extrapolation beyond the data ranges, constraints can be set so that the fitted spline is negative at 0 (to prevent high incidence rates at 0), increasing, or nonnegative. Since incidence is nonnegative, any negative values of the fitted spline are assumed to be set to 0 in $\hat{I}_C(t)$.

To convert the SEER incidence to a hazard rate or probability density that can be integrated to calculate the CDF, we must respectively subtract the living clinical cancer cases from the denominator or add the deceased clinical cancer cases to the denominator. Both methods would require calculating death rates from cancer, so we proceed with the latter, which is more direct. Letting $F_{CD_c}(t|T_C)$ denote the CDF for the time from cancer diagnosis to death from cancer conditional on the age at diagnosis, we estimate the probability $F_{D_c}(t)$ of dying from cancer by age t by integrating over possible diagnosis and death times using

$$F_{D_c}(t) = \int_0^t \hat{I}_C(u) F_{CD_c}(t-u|T_C = u)du. \tag{3}$$

We then calculate the probability density of clinical cancer $f_C(t)$ by scaling the clinical cancer incidence rates by the proportion of people who have not died from clinical cancer using

$$f_C(t) = \hat{I}_C(t)\left(1 - F_{D_c}(t)\right). \tag{4}$$

Finally, the CDF is calculated by integrating over the PDF using

$$F_C(t) = \int_0^t f_C(u)du. \tag{5}$$

The risk of death from other causes is not considered because we assume that it is independent of the cancer disease process.

We can derive a crude estimate of $F_{CD_c}(t|T_C)$ for deriving priors assuming that relative survival rates are independent of age at diagnosis and are exponentially distributed. To calculate the rate of cancer death, we start with SEER data on relative survival over 10 years by stage s at diagnosis, denoted as $S_{C_sD_c}(t) = 1 - F_{C_sD_c}(t)$. We extrapolate relative survival beyond 10 years by fitting constrained splines or parametric distributions to produce estimates $\hat{S}_{C_sD_c}(t)$. The average relative survival for each stage, $\bar{T}_{C_sD_c}$, is estimated using a Monte Carlo simulation from $\hat{S}_{C_sD_c}(t)$, and the overall average time to death from cancer $\bar{T}_{CD_c}$ weights the stage-specific averages by the proportion of the population diagnosed at each stage. $1/\bar{T}_{CD_c}$ serves as the exponential rate for $F_{CD_c}(t)$.

Alternatively, if the incidence rate excludes prior clinical cancer cases from the denominator, it is equivalent to the hazard rate of clinical cancer [118]. The fitted spline for the incidence rate can then be integrated directly to calculate the cumulative hazard $H_C(t)$ and the CDF as follows:

$$H_C(t) = \int_0^t \hat{I}_C(u)du \tag{6}$$

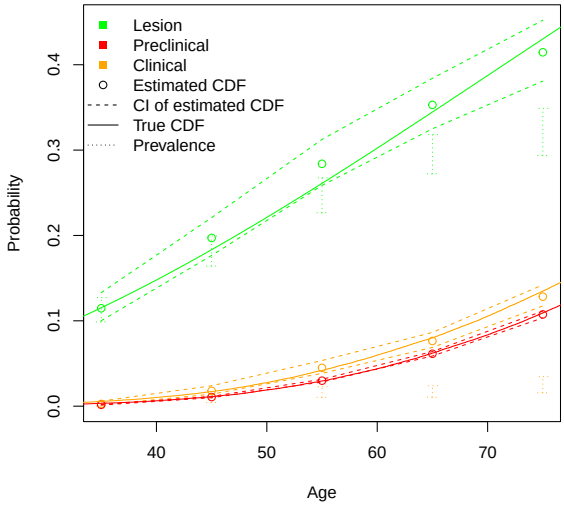
$$F_C(t) = 1 - e^{-H_C(t)} \tag{7}$$

B.2 Deriving the CDF of disease onset

Prevalence data are assumed to come from screening studies that exclude patients already diagnosed with cancer. To calculate the cumulative probability of preclinical cancer, we must therefore rescale preclinical cancer prevalence as a proportion of the whole population and add the proportion of people who have passed the preclinical cancer stage as follows:

$$F_P(t) = P_P(t)(1 - F_C(t)) + F_C(t). \tag{8}$$

Using the 95% CIs of incidence and prevalence, we calculate lower and upper bounds for the CDF of the time to cancer onset.



Abbreviations: CDF = cumulative distribution function, CI = confidence interval
Fig. 6: Coverage of ground truth simulated CDFs by the CDF estimates

If precancerous lesion prevalence is similarly derived from screening studies that include preclinical cancer but exclude clinical cancer cases, we can likewise calculate the CDF for precancerous lesion onset using:

$$F_L(t) = P_L(t)(1 - F_C(t)) + F_P(t) \tag{9}$$

Figure 6 demonstrates the results of this algorithm, showing that the ground truth simulated CDFs (solid lines) for precancerous lesion (green), preclinical cancer (orange), and clinical cancer (red) onset are within the 95% CIs (dashed lines) of the respective CDFs (points) estimated from the prevalence and incidence targets.

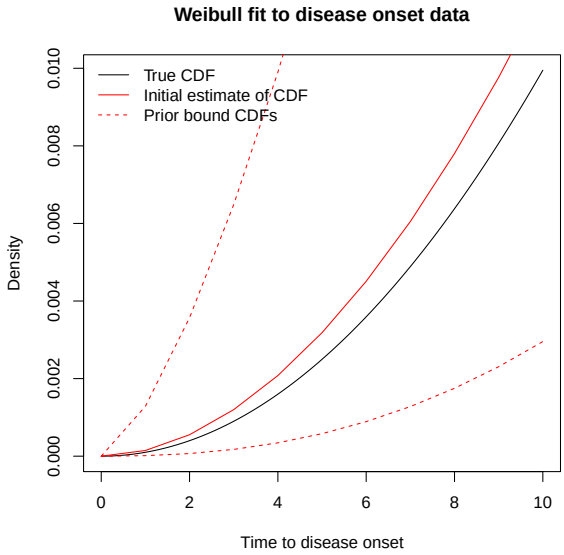
Denoting $F_X(t)$ as the CDF estimate for the time from birth to disease onset, whether it be precancerous lesion or preclinical cancer onset, we let $F_X^{LB}(t)$ and $F_X^{UB}(t)$ be the lower and upper bounds, respectively. Using these CDFs, we may fit parametric distributions and derive bounds for a plausible range of parameters. In our example, we assume that $F_X(t)$ follows a Weibull distribution. Since the CDF of a Weibull distribution equals $1 - e^{(x/\sigma)^\alpha}$, where α and σ are the shape and scale parameters, respectively, we can perform a weighted least squares regression by transforming the CDF equation into

$$y_i = ax_i + b, \tag{10}$$

where $y_i = \ln(-\ln(1 - F_X(t_i)))$, $a = \alpha$, $x_i = \ln t_i$, and $b = -\alpha \ln \sigma$ [119]. Each point is weighted by $1/\left(y_i^{UB} - y_i^{LB}\right)^2$, where y_i^{UB} is the transformation y_i applied to $F_X^{UB}(t)$ and likewise for y_i^{LB}. This weighting scheme downweights values of the CDF for which there is less certainty as reflected by wider CIs. With the coefficient a and intercept b from the regression, we solve for $\alpha = a$ and $\sigma = e^{-b/a}$. Heteroskedasticity-robust standard errors of the regression parameters are used to derive 95% CIs of the parameters, which are converted to bounds for α and σ. The final bounds of the priors for α and σ are calculated by expanding the CI bounds by 20% to ensure that the targets would be covered. As shown in Figure 7, the range of priors for the parameters governing the time to disease onset contains the ground truth parameters.

B.3 Deriving priors for cancer progression

Rather than setting independent priors for the rates λ_{P_i} of the preclinical cancer stage progression TTE variables $T_{P_iP_{i+1}}$, we make two alterations to the model parametrization for calibration to take advantage of intuition about cancer biology and model structure. First, we set a prior for λ_{P_1} and define new parameters h_i as the unknown hazard ratios (HRs) applied to λ_{P_i} to calculate $\lambda_{P_{i+1}}$. For example, HR



Abbreviations: CDF = cumulative distribution function
Fig. 7: Coverage of ground truth Weibull distribution by priors

priors that are set to range from 0.7 to 2 indicate a belief that the average progression time for each stage is no more than twice as fast as that of the previous stage and no less than 30% slower.

Next, we leverage that for independent exponential variables $T_{P_iP_{i+1}}$ and $T_{P_iC_i}$ with respective rates λ_{P_i} and λ_{C_i},

$$\mathbf{P}\left(T_{P_iC_i} < T_{P_iP_{i+1}}\right) = \frac{\lambda_{C_i}}{\lambda_{P_i} + \lambda_{C_i}}, \tag{11}$$

where $\mathbf{P}\left(T_{P_iC_i} < T_{P_iP_{i+1}}\right)$ is equivalent to the proportion of cancers detected at stage i conditional on having reached preclinical stage i. Given a true stage distribution d_i^* that would be observed in the absence of death from other causes,

$$\mathbf{P}\left(T_{P_iC_i} < T_{P_iP_{i+1}}\right) = \frac{d_i^*}{\sum_{j \geq i} d_j^*}. \tag{12}$$

However, we expect the observed stage distribution d_i to differ from d_i^* because later stages are more likely to occur at older ages and are more likely to be censored, resulting in a slight overrepresentation of earlier stages in the observed distribution. We thus define unknown parameters $\delta_i = d_i^* - d_i$ to correct for this difference, solving for λ_{C_i} in Equation 11 by substituting the left-hand side using Equation 12 and d_i^* for $d_i + \delta_i$. The constraint that the stage distribution must sum to one removes a degree of freedom, so δ_i is fixed at 0 for the final stage. Otherwise, prior distributions for δ_i are set at small values, such as between -0.05 and 0.05, with earlier stages weighted towards more negative values.

C Testbed model

To illustrate the effectiveness of the calibration approaches, we simulate calibration targets using a ground truth cancer model and assess whether the two Bayesian calibration methods recover the ground truth parameters when calibrating to the simulated targets.

C.1 Data

We simulate data using a ground truth cancer natural history model that includes a precancerous lesion state and 4 stages of cancer, assumes a 50% probability of being male, and follows the variable distributions in Table 3. We first run the ground truth model to generate targets for evaluating the calibration process, including precancerous lesion prevalence, preclinical cancer prevalence, clinical cancer incidence,

(a) Variables directly informed by data sources

Variable	Description	Distribution	Value
M	Male sex	Bernoulli	0.5
T_{H,D_o}	Time from birth to death from other causes	Gompertz $\left(\text{CDF}: 1 - e^{-a/b\left(e^{bt-1}\right)}\right)$	$a_{male} = 0.000078$, $b_{male} = 0.08$, $a_{female} = 0.000076$, $b_{female} = 0.078$
T_{C_i,D_c}	Time from stage i clinical cancer to death from cancer	Exponential $\left(\text{CDF}: 1 - e^{-r_i t}\right)$	$r_1 = 0.05,\ r_2 = 0.08,\ r_3 = 0.16,\ r_4 = 0.5$

(b) Variables with unobserved distributions (assumed or informed by calibration)

Variable	Description	Distribution	Ground truth value
$T_{H,L}$	Time from birth to onset of first lesion	Weibull $\left(\text{CDF: } 1 - e^{-(t/\sigma)^{\alpha}}\right)$	$\alpha = 2$, $\sigma = 100$
N_L	Number of additional lesions before death from other causes	Poisson with rate ν for time to next lesion	$\nu = 0.0625$
T_{L,L_j}	Time from first lesion to onset of additional lesion j	Uniform from 0 to $T_{H,D_o} - T_{H,L}$	N/A
$T_{L_j,P}$	Time from onset of lesion j to conversion to preclinical cancer	Weibull $\left(\text{CDF: } 1 - e^{-(t/\mu)^{\kappa}}\right)$	$\kappa = 3$, $\mu = 50$
$T_{P_i,P_{i+1}}$	Time from stage i to $i+1$ of preclinical cancer	Exponential with rate λ_i	$\lambda_1 = 0.33$, $\lambda_2 = 0.4$, $\lambda_3 = 0.6$
T_{P_i,C_i}	Time from stage i preclinical cancer to stage i clinical cancer	Exponential with rate γ_i	$\gamma_1 = 0.09$, $\gamma_2 = 0.02$, $\gamma_3 = 0.5$, $\gamma_4 = 0.5$

Table 3: Testbed ground truth model parameters

and distributions of lesion multiplicity and cancer stage at diagnosis. The prevalence targets are summarized for 10-year intervals from age 30 to 80, lesion multiplicity from age 50 to 80, and incidence for 10-year intervals from age 30 to 90. All individuals diagnosed with cancer in their lifetimes inform the cancer stage distribution. For the prevalence and multiplicity targets, we sample an observation date at which the presence of disease or number of lesions is assessed for each individual uniformly within the age ranges, and individuals censored due to cancer diagnosis or death before the observation age are excluded. For the multiplicity targets, we consider the percentage of screened individuals with 1, 2, or 3+ lesions out of those with any lesions. In total, there are 23 targets. The ground truth simulation also outputs relative cancer survival for the first ten years after diagnosis stratified by the stage at diagnosis.

C.2 Natural history model

For calibration and decision analysis, we use a natural history model with the same model structure, distributional forms, and background mortality distribution as the ground truth model. The model includes 13 parameters for 11 probability distributions that must be estimated via calibration. To determine the cohort size for calibration, we use an initial sample of 50,000 and conduct 50 simulations to estimate the Monte Carlo error of the prevalence, incidence, and stage distribution outputs. The required sample size for the standard deviation of the outputs to be less than the standard errors of all targets is doubled and rounded up, resulting in a cohort size of 300,000.

C.3 Parameter estimation with IMABC and BayCANN

Uniform prior distributions for the calibrated parameters are created by randomly sampling a lower bound between 16% and 80% of the ground truth value and an upper bound between 120% and 216% of the ground truth value. This procedure generates wide bounds representative of what might realistically be used for Bayesian calibration. The priors for the parameters for time to lesion onset are refined using the procedure described in Appendix B.

We calibrate the model separately with IMABC and BayCANN. For the IMABC parameters, we use the $(1 - 10^{-15}) \times 100\%$ CI of the calibration targets for the starting bounds, the 95% CI for the stopping bounds, 1000 initial samples per unknown parameter, 10 normal mixtures, 50 points per normal mixture, and a target posterior ESS of 1000. For our example, the posterior distribution ultimately contains a total of 1067 parameter sets. For BayCANN, we use the same ANN training scheme and hyperparameters described in the colorectal cancer model example in Section 5.2.3. In Figure 8, the box plots of simulated

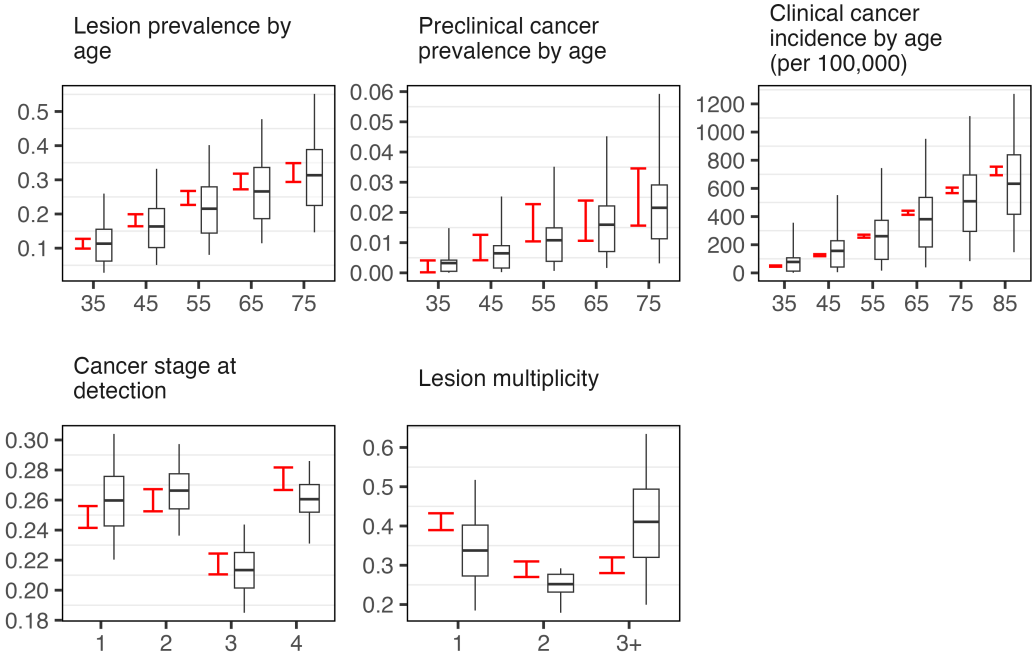
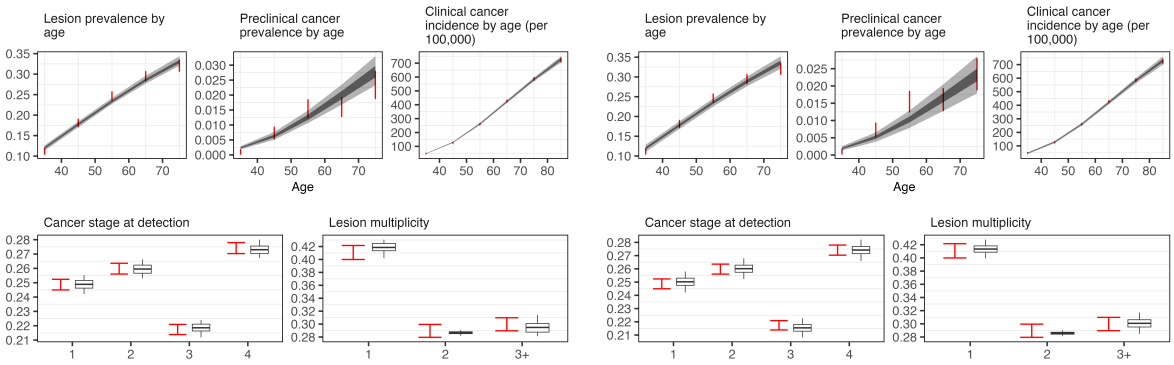


Fig. 8: Target coverage of prior distributions, with box plots corresponding to simulated outputs and 95% CIs of targets in red

outputs associated with a LHS of the parameter prior distributions cover most of the corresponding target 95% CIs in red.

Figure 9 demonstrates the ability of the IMABC and BayCANN posterior distributions to produce model outputs in gray consistent with the calibration targets in red. Although ground truth parameter values are unavailable in practice, we further demonstrate the accuracy of the calibration algorithms for this example by showing the overlap of the posterior distribution histograms and the true parameters, represented by the red vertical lines, in Figure 10.

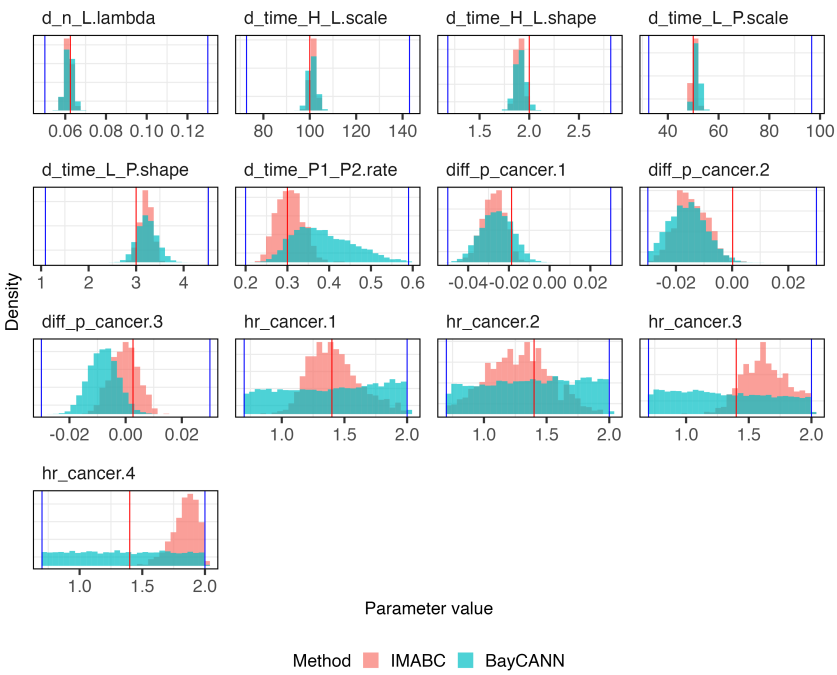
The unimodal likelihood profiles for parameters such as the shape and scale of the Weibull distribution for the time to lesion onset indicate that the calibration algorithms converged to a unique solution [77]. However, the wide range and high correlation (see Figures 11 and 12) of other parameters, such as the hazard ratios for the rates of preclinical cancer progression, indicate that not all parameters are identifiable given the available targets. This is because the preclinical cancer prevalence and clinical cancer incidence targets inform the time from cancer onset to detection but are not fine-grained enough to inform the time between stages of preclinical cancer. These targets are consistent with both faster progression rates in earlier stages combined with slower progression in later stages and vice versa, resulting in non-identifiability.



(a) IMABC (b) BayCANN

Abbreviations: BayCANN = Bayesian Calibration using Artificial Neural Networks, IMABC = Incremental Mixture Approximate Bayesian Computing

Fig. 9: Internal validation of posteriors with model outputs in gray and 95% CIs of calibration targets in red



Abbreviations: BayCANN = Bayesian Calibration using Artificial Neural Networks, IMABC = Incremental Mixture Approximate Bayesian Computing

Fig. 10: Posteriors (blue and red histograms) versus prior bounds (blue vertical lines) and true parameters (red vertical line)

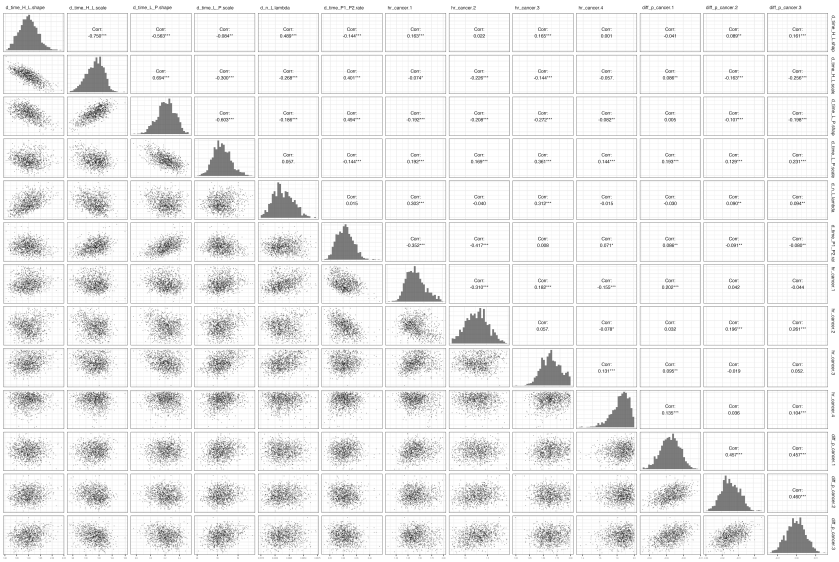


Fig. 11: IMABC correlation plot

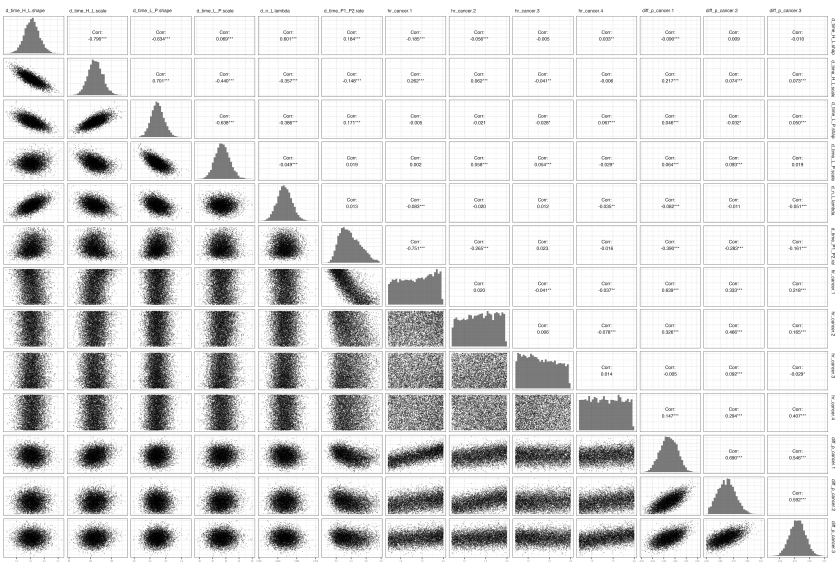


Fig. 12: BayCANN correlation plot