Discrete-Event Simulation Model 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, Palo Alto, CA.
[2]Hutch Institute for Cancer Outcomes Research, Biostatistics Program, Public Health Sciences Division, Fred Hutch Cancer Center, Seattle, WA.
[3]The Comparative Health Outcomes, Policy, and Economics (CHOICE) Institute, University of Washington, Seattle, WA.
[4]Stanford Center for Biomedical Informatics Research, Stanford University, Palo Alto, CA.
[5]Stanford Clinical Excellence Research Center, Stanford University, Palo Alto, CA.
[6]Division of Hospital Medicine, Stanford University, Palo Alto, CA.
[7]Department of Health Policy, School of Medicine, Stanford University, Stanford, CA.
[8]Center for Health Policy, Freeman-Spogli Institute for International Studies, Stanford University, Stanford, CA.

*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 efficient and 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 determine the optimal strategy to recommend, assess confidence in the recommendation, 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 and other downstream tasks. To fill this gap, we introduce the **DES** Model for **C**ancer **I**nterventions and **P**opulation **H**ealth in **R** (DESCIPHR), an open-source framework and codebase integrating a flexible DES model for the natural history of cancer, Bayesian calibration for parameter estimation, and screening strategy evaluation. We also introduce an automated method to generate data-informed parameter prior distributions and enhance the accuracy and flexibility 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.

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 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. For a disease progression and screening model, calibration involves gathering and combining data on clinical, biological, or epidemiological targets, such as disease incidence and prevalence of subtypes (calibration targets), and estimating parameters such that the model produces matching outputs in a simulated population [8, 9]. Bayesian calibration methods have the ability to output parameter distributions that reflect the uncertainty in calibration targets [10], and though they are computationally demanding, advances in model emulator methods, statistical approximations, improved computing resources, and parallelization have made them faster to implement and deploy [11, 12].

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 [13], and open-source DES tools, such as the `simmer` R package [14] and the `SimPy` [15] and `Ciw` Python libraries [16], 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 [17], but the tutorial does not cover parameter estimation and calibration. Open-source cancer simulation models, such as the Colon Modeling Open Simulation Tool (CMOST) [18] and Microsimulation Lung Cancer (MILC) model [10, 19] 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, 12, 20–22].

To facilitate the development of individual-level decision models for cancer screening evaluation, we introduce the **DES** Model for **C**ancer **I**nterventions and **P**opulation **H**ealth in **R** (DESCIPHR), an open-source framework and codebase for 1) structuring a DES model for cancer natural history and screening; 2) estimating model parameters with state-of-the-art Bayesian calibration methods; and 3) evaluating screening strategies with the calibrated model accounting for parameter uncertainty. We illustrate the framework with an example that is adaptable to many real-world policy decision contexts. The pipeline also augments Bayesian Calibration using Artificial Neural Networks (BayCANN) [11] 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 [13, 23–25]. 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 [25]. To our knowledge, DESCIPHR is the first open-source, end-to-end DES disease

natural history model that demonstrates model parameter estimation using modern Bayesian calibration methods and health policy evaluation using the parameters' joint posterior distribution. We provide the code for the entire pipeline on GitHub at https://github.com/sjpi22/tutorial_cancer_modeling_des.

2 Cancer screening model

We begin by describing the structure, implementation, and outcomes of the cancer screening model. The model is composed of a natural history microsimulation that captures the sequence of landmark events in cancer progression and an intervention component that outlines how screening and treatment may interrupt the disease process [26, 27]. We then discuss the choice of cohort size and the derivation of relevant summary statistics from the simulated individual-level life histories.

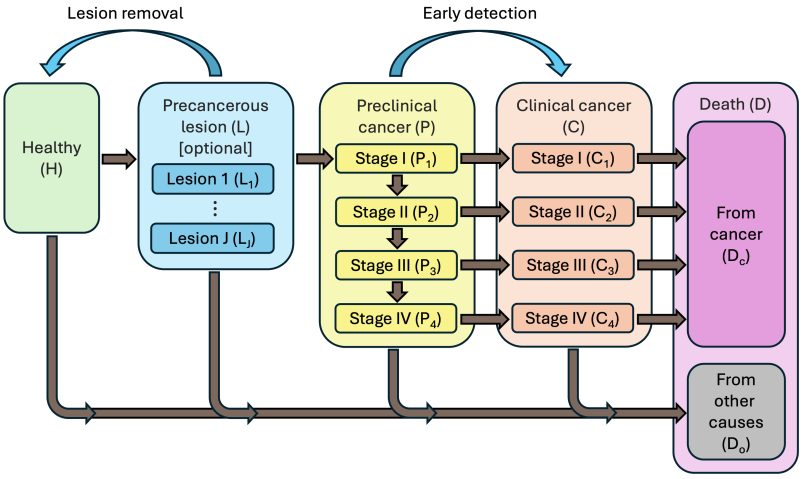


Fig. 1: Cancer natural history model schema with effects of screening in blue arched arrows

2.1 Natural history model

We model the natural history of cancer by splitting the carcinogenic process into different health states based on prior cancer simulation models [18, 28, 29]. Relevant health states and possible transitions between states are illustrated in the model diagram in Figure 1. We assume that transitions between health states i and j occur in continuous time, governed by rates λ_{ij}. Rates can be functions of the time t spent in state i, age a, and time period p, *i.e.*, $\lambda_{ij}(t, a, p)$, to account for age- and period-dependent risk factors and population dynamics, such as environmental influences, treatment effects, population changes, and age-related biological responses.

In the model, patients are assumed to be born into a healthy state (H). If the natural history of the cancer is known to include precursor lesions, such as adenomas for CRC or ductal carcinoma in situ for breast cancer, the onset of a lesion causes patients to transition to the precancerous lesion state (L). For simplicity, the current implementation of the pipeline assumes a single lesion type, no lesion regression, and a constant rate at which additional lesions develop after the first lesion. At the first conversion of any lesion to detectable cancer, patients enter the preclinical, or pre-detection, cancer stage (P). The model can also account for a direct transition from the healthy to 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 (C). Following a competing risks model, we associate each cancer stage with two variables, 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, with the time to progression at the final stage conceptualized as infinite. Transition rates to cancer onset and progressive stages before diagnosis are typically unobserved but can be informed by model calibration, described in Section 3. Individuals who develop clinical cancer face a risk of dying from cancer (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 survival associated with current technology. All individuals are at risk of death from other causes (D_o), following an age-specific mortality rate that is derived from life tables and independent of the cancer pathway. The time of death (D) is determined by the earliest of death from cancer and death from other causes.

Demographics and other risk factors can be generated for individual patients that result in heterogeneity in the hazard of progression between disease states, as long as the functional relationships with the time-to-event (TTE) distributions are specified. The output of the natural history model is a matrix with a number of rows equal to the user-specified cohort size and each column containing values for a patient characteristic, TTE variable, or health state characteristic. For models including the precancerous lesion state, an additional lesion-level event matrix is generated with a row for each lesion that develops in each patient's lifetime and columns for the times of lesion onset and conversion to preclinical cancer. Guidance on customizing the model structure, for example by adding an infection state preceding the precursor lesion state, is provided in Appendix B.3.

2.2 Screening model

Screening may prevent cancer by removing precancerous lesions, thereby interrupting their progression to cancer, or by modifying the disease process through the early detection of cancer, potentially at a more treatable stage. However, mass screening in individuals without disease is associated with clinical and economic burdens, including the cascade of care triggered by false positive results [30, 31]. To evaluate the effects of screening, a cohort of individual trajectories is first simulated without it (i.e., under natural history), and then screening regimens are applied to generate counterfactual trajectories and the corresponding effects of screening on cancer-related outcomes.

We define a screening regimen by its starting age, stopping age, routine screening interval, test characteristics, and rules for confirmatory testing and surveillance. Test characteristics required for modeling include the sensitivity p, which may vary with cancer stage and lesion characteristics such as size or histological subtype, and specificity q. The sensitivity determines the probability that a test produces a positive result among individuals with cancer or at least one precancerous lesion, whereas the specificity is the probability that a test performed for an individual without disease produces a negative result.

Some non-invasive screening tests, such as mammography for breast cancer or stool testing for CRC, require a follow-on test to confirm diagnosis and remove both precursor lesions and cancer after a positive result. Depending on the screening test modality, the confirmatory test may be applied to all lesions present during screening or only the subset of lesions flagged by the initial test. The confirmatory test is assumed to have perfect specificity because it is associated with biopsy and pathology. It may also have higher sensitivity than the non-invasive test, and the time to the next screening test after a negative confirmatory test may be longer than the routine screening interval. In some cases, such as colonoscopy for CRC, the gold-standard confirmatory test doubles as a screening test, so positive results imply diagnosis of preclinical cancer and removal of detected lesions if applicable.

Individuals suspected of having a higher than average risk of cancer, for example due to the detection of precancerous lesions, may be put on a surveillance regimen. Surveillance regimens are defined by their test modalities, test interval, and stopping criteria. Examples of stopping criteria include returning to routine screening after two negative tests or reaching an age older than the stopping age for routine screening. If any 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.

2.3 DES implementation and data inputs

We simulate the mathematical model described in 2.1 using a DES approach [32]. When generating patient trajectories using DES, the time from state i to a consecutive state j is characterized as a random variable T_{ij} that follows a time-to-event (TTE) distribution with probability density function

f_{ij}, cumulative distribution function (CDF) F_{ij}, and hazard function $\lambda_{ij} = \frac{f_{ij}}{1-Fij}$. 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_{ij}^n is drawn from a standard uniform distribution and used to calculate the time t_{ij}^n at which F_{ij} would equal u_{ij}^n:

$$u_{ij}^n \sim U(0,1) \tag{1}$$

$$t_{ij}^n = F_{ij}^{-1}(u_{ij}^n) \tag{2}$$

TTE variables may also be sampled using intensity-based methods [33]. Given the stochasticity of sampling, a random seed is set before running the model for reproducibility.

The TTE distributions may be either empirically derived model inputs or unobserved with parameters estimated through calibration (see Table 1). For instance, the time to death from other causes is modeled empirically using background mortality rates from actuarial life tables under the assumption that the rate of death from the modeled disease is low enough not to substantially impact all-cause death rates. In addition, the time from cancer diagnosis to death from cancer is sampled from empirical relative survival distributions by stage at diagnosis. Conversely, no data exist to directly inform most pre-diagnosis variables, such as the time to first precancerous lesion. This distribution is modeled by default in DESCIPHR with a Weibull distribution parametrized by shape α and scale σ with CDF $F(t) = 1 - e^{-(t/\sigma)^\alpha}$, where $\alpha > 1$ indicates an increasing hazard rate over time, consistent with the increasing risk of developing cancer with age.

The age at health state j is the sum of times between consecutive states from birth to state j. Table 1 displays the distributions or calculations associated with each TTE variable and state characteristic in the lesion-inclusive model used to illustrate the pipeline, which is formally introduced in Section 2.6. For more detail on adapting the model to include custom data and configurations, as well as a description of how probability distributions are stored and sampled, see Appendix B.

2.4 Choice of cohort size

The choice of the cohort size, or the number of life trajectories to simulate, in individual-level models incurs many of the same challenges as the choice of sample size in prospective trials; as the cohort size increases, the variance of the estimates of the summary statistics decreases [24], but the time required to run the simulation increases [22]. As a rule of thumb, the Monte Carlo error of the generated summary statistics should be low enough to distinguish any quantities of interest. To facilitate convergence during calibration, the Monte Carlo error should not exceed the estimation error of the targets [22]. For strategy evaluation, the Monte Carlo variability should be much smaller than the mean differences in outcomes between strategies [34].

To determine the minimum sample size N_{target} required for the 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 [34]. 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 [22]. 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 [35].

2.5 Outcomes

A range of summary statistics for the outcomes of interest may be calculated from the simulated individual-level data for various purposes, including calibrating the model, evaluating interventions, and understanding the disease process. Generating such model outputs is useful for calibration and validation of the natural history model, as both involve comparing modeled outcomes to relevant epidemiological endpoints estimated from empirical data, such as cancer registries and screening studies. For these comparisons, it is important for the simulated population to be representative of the real-world population from which the target of interest is derived. Regarding epidemiological outcomes used as calibration targets, we calculate the prevalence of precancerous lesions and preclinical cancer, the incidence and stage distribution of clinical cancer, and the distribution of lesion multiplicity according to [36]. Prevalence is

(a) Variables directly informed by external data sources

Variable	Description	Distribution	Source
M	Male sex	Bernoulli	Population sex distribution
T_{HD_o}	Time from birth to death from other causes	Empirical	Age- and sex-specific mortality rates from life tables
T_{CD_c}	Time from clinical cancer to death from cancer	Empirical	Relative survival by stage at cancer diagnosis from cancer registries

(b) Variables with unobserved distributions

Variable	Description	Distribution	Hazard	Ground truth parameters
T_{HL}	Time from birth to onset of first lesion	Weibull	$\frac{\alpha}{\sigma}\left(\frac{t}{\sigma}\right)^{\alpha-1}$	$\alpha = 2$, $\sigma = 100$
N_L	Number of additional lesions before death from other causes	Poisson	$\lambda_L(T_{HD_o} - T_{HL})$ for time to next lesion	$\lambda_L = 0.0625$
T_{L_jP}	Time from onset of lesion j to conversion to preclinical cancer	Weibull	$\frac{\kappa}{\mu}\left(\frac{t}{\mu}\right)^{\kappa-1}$	$\kappa = 3$, $\mu = 50$
$T_{P_iP_{i+1}}$	Time from stage i to $i+1$ of preclinical cancer	Exponential	λ_{P_i}	$\lambda_{P_1} = 0.33$, $\lambda_{P_2} = 0.4$, $\lambda_{P_3} = 0.6$
$T_{P_iC_i}$	Time from stage i preclinical cancer to stage i clinical cancer	Exponential	λ_{C_i}	$\lambda_{C_1} = 0.09$, $\lambda_{C_2} = 0.02$, $\lambda_{C_3} = 0.5$, $\lambda_{C_4} = 0.5$

(c) Variables with assumed distributions

Variable	Description	Distribution
T_{LL_j}	Time from first lesion to onset of additional lesion j	Uniform$(0, T_{HD_o} - T_{HL})$

(d) Variables calculated using other variables

Variable	Description	Calculation
T_{LP}	Time from onset of first lesion to preclinical cancer	$\min_j T_{L_jP}$
T_{HP}	Time from birth to onset of preclinical cancer	$T_{HL} + T_{LP}$
S	Stage at cancer diagnosis	$\arg\min_i T_{P_iC_i} < T_{P_iP_{i+1}}$
T_{PC}	Time from cancer onset to diagnosis	$\sum_{i=1}^{S-1} T_{P_iP_{i+1}} + T_{P_SC_S}$
T_{HD_c}	Time from birth to death from cancer	$T_{HP} + T_{PC} + T_{CD_c}$
T_{HD}	Time from birth to death from all causes	$\min(T_{HD_o}, T_{HD_c})$

Table 1: Model parameters and default distributions

defined as the proportion with the condition out of the population at risk [37]. Incidence is defined as the number of new cases per person-year of exposure among those at risk [38].

For decision analysis, tradeoffs associated with screening are calculated in terms of life years gained (LYG) and screening test burden, though other metrics, such as quality-adjusted life years, could also be used. Letting $\mathcal{A}$ denote the set of patients alive without diagnosed cancer by age t_1 in the baseline non-intervention scenario (i.e., $\mathcal{A} = \{n | \min(T^n_{HD}, T^n_{HC}) > t_1\}$), the LYG per 1,000 is calculated as

$$LYG(t_1) = \frac{1}{|\mathcal{A}|} \sum_{n \in \mathcal{A}} \left(\tilde{T}^n_{HD} - T^n_{HD} \right) \times 1,000, \tag{3}$$

where T^n_{HD} and $\tilde{T}^n_{HD}$ respectively denote the lifetime of patient n in the baseline and intervention scenarios and $|\mathcal{A}|$ is the number of individuals in $\mathcal{A}$. We also estimate intermediate outcomes of cancer natural history that are difficult to observe in practice, including the mean sojourn time of cancer and the mean dwell time of precancerous lesions, according to the formulas in [36]. We account for model input parameter uncertainty for all outcome measures by randomly sampling from the joint posterior distribution obtained from the Bayesian calibration and running the simulation with each sampled parameter set.

2.6 Example model

To illustrate the modeling and calibration approaches, we follow the process documented in Figure 2 and provide a publicly available GitHub code repository configured to run the entire analysis on an example cancer model with simulated data. The model includes a precancerous lesion state and 4 stages of cancer, assumes a 50% probability of being male, and follows the ground truth variable distributions in Table 1. 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, 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, study ages are uniformly sampled within the age ranges, and individuals censored before the study 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 empirical distributions for cancer-specific mortality by stage at diagnosis.

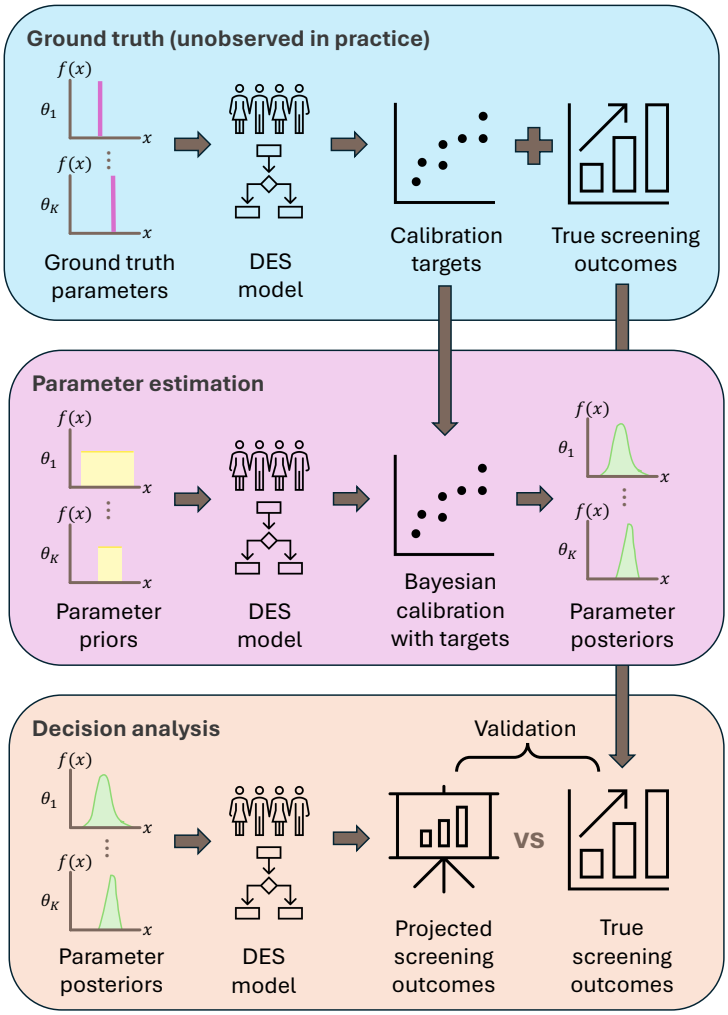
For the calibration and decision analysis, we cloak the ground truth parameter values, resulting in 13 unknown parameters, and replace the distributions for cancer-specific mortality with the generated empirical distributions. We otherwise maintain the same model structure, distributional forms, and background mortality distribution. 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 and incidence 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. The input data and specific model configurations can be customized and the analysis can be easily rerun according to a downstream user's needs.

3 Parameter estimation with Bayesian calibration

Deep or unobservable parameters in disease models are often not estimable using individual-level data. For instance, the time of cancer onset and the duration of each stage of cancer cannot be monitored directly in humans, so the parameters underlying their distributions are unknown; however, they may be estimated by fitting the disease model's outputs to real-world aggregated data in a process known as model calibration [10]. Consistency between disease natural history targets and corresponding outputs of the simulation model is typically assessed using a distance function [39]. Various methods exist to search the parameter space for optimal-fitting parameters, including manual tuning, empirical search, directed search, and Bayesian techniques [27]. We focus on Bayesian calibration methods because they effectively combine prior knowledge about the parameters of interest with the uncertainty associated with the calibration targets.

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 likelihood function or likelihood approximation. 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, 39, 40]. However, advances in computational capacity and statistical approximations have made Bayesian methods increasingly feasible.

Approximate Bayesian computing (ABC) is useful for calibrating complex microsimulation models when the likelihood function for outputs given parameters would be intensive or infeasible to compute [22]. 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 [12]. 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 [11, 21]. IMABC and some steps in BayCANN can be parallelized to optimize computational resources



Abbreviations: DES = discrete-event simulation
Fig. 2: Diagram of DESCIPHR pipeline

and speed up processing. In DESCIPHR, we demonstrate model calibration using IMABC and BayCANN with the targets listed in Section 2.6.

3.1 Choice of priors

Assigning prior distributions to unknown parameters may be guided by meta-analysis of evidence, expert opinion, or a non-informative prior with wide variance [39, 40]. Bayesian calibration outputs could be sensitive to the values of prior distributions, but weakly informative priors induce less bias in the final parameter estimate than misspecified priors with low variance [40, 41]. 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 initial bounds for mutually independent, uniformly distributed diffuse priors [10, 22].

We also demonstrate a target-informed approach to derive parameter priors for probability distributions. Using targets characterizing each health state, such as the incidence of clinical cancer, prevalence of preclinical cancer, and prevalence of precancerous lesions by age, we derive estimates of the CDFs of the time to onset of each state (see Appendix C). The 95% CIs of the targets are used to estimate lower and upper bounds for the CDFs. Parametric distributions are then fit to the bounds for the CDFs to calculate ranges for the prior distributions. In our modeling pipeline, we use this approach to define the prior distributions for the parameters underlying the time from birth to initial disease onset (precancerous lesion onset for the lesion model and preclinical cancer onset for the de novo cancer model).

We make two alterations to the model parametrization for calibration to take advantage of intuition about cancer biology and model structure. First, instead of setting independent priors for the rates λ_{P_i} of the preclinical cancer stage progression TTE variables $T_{P_iP_{i+1}}$, we merely 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 $\lambda_{P_{i+1}}$. The HR priors are set to range from 0.7 to 2, indicating 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},

$$\mathbf{P}\left(T_{P_iC_i} < T_{P_iP_{i+1}}\right) = \frac{\lambda_{C_i}}{\lambda_{P_i} + \lambda_{C_i}}, \tag{4}$$

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,

$$\mathbf{P}\left(T_{P_iC_i} < T_{P_iP_{i+1}}\right) = D_i^* | S \geq i = \frac{D_i^*}{\sum_{j \geq i} D_j^*}. \tag{5}$$

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 $d_i = D_i^* - D_i$ to correct for this difference, solving for λ_{C_i} in Equation 4 by substituting the left-hand side using Equation 5 and D_i^* for $D_i + d_i$. The constraint that the stage distribution must sum to one removes a degree of freedom, so d_i is fixed at 0 for the final stage. Otherwise, prior distributions for d_i are set at small values between -0.05 and 0.05, with earlier stages weighted towards more negative values.

For the demonstration, prior distributions for other parameters are created by adding random noise around the ground truth values. In total, the model has 13 parameters for 11 probability distributions that need to be estimated via calibration. Prior to calibration, to ensure the model can produce outputs

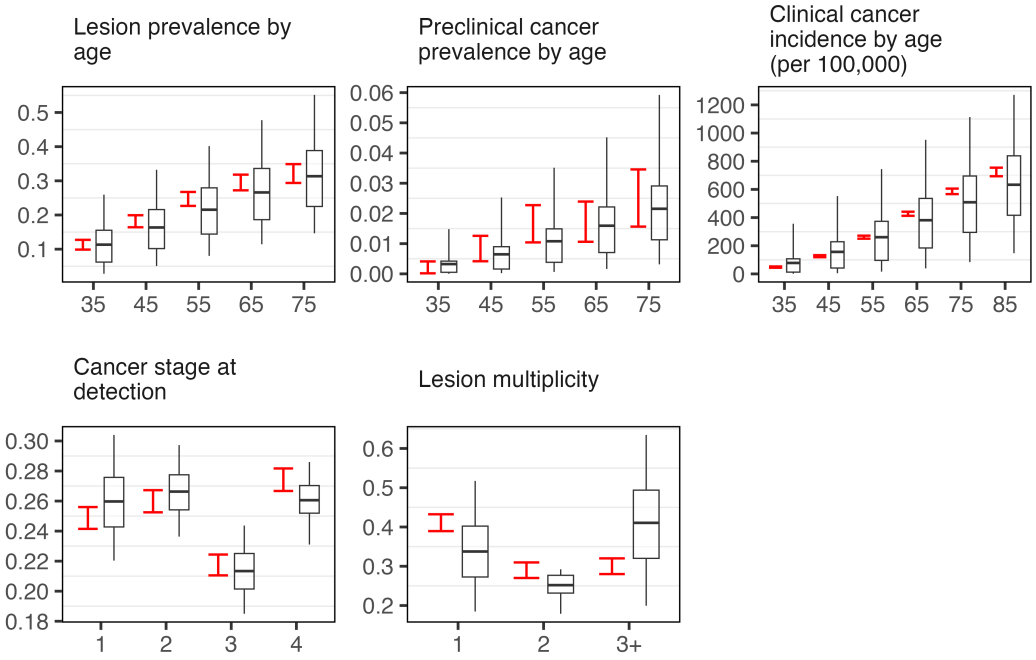


Fig. 3: Target coverage of prior distributions, with box plots corresponding to simulated outputs and 95% CIs of targets in red

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. In Figure 3, the box plots of simulated outputs cover most of the corresponding target 95% CIs in red. 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 IMABC algorithm starts with a rejection-based sampling step that draws 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 [12]. Parameter sets are ranked by their p-value proximity to the calibration targets as measured by the minimum probability across targets of observing a value at least as extreme as the target for a normal distribution centered at the analogous model output with standard deviation equal to the standard error of the target. Ties in the ranking are broken by choosing the parameter set with the lower total squared percentage difference from the targets among outputs outside the final tolerance bounds. New parameter sets are sampled from a mixture of multivariate normal distributions centered at the highest probability parameter sets. With the updated sample, the process of accepting parameter sets compatible with the tolerance threshold and adaptively resampling repeats. The tolerance thresholds are incrementally reduced as the number of accepted sets and the median p-value proximity of the resampled outputs increase until there is a sufficient number of parameter sets with model outputs in the final user-specified bounds.

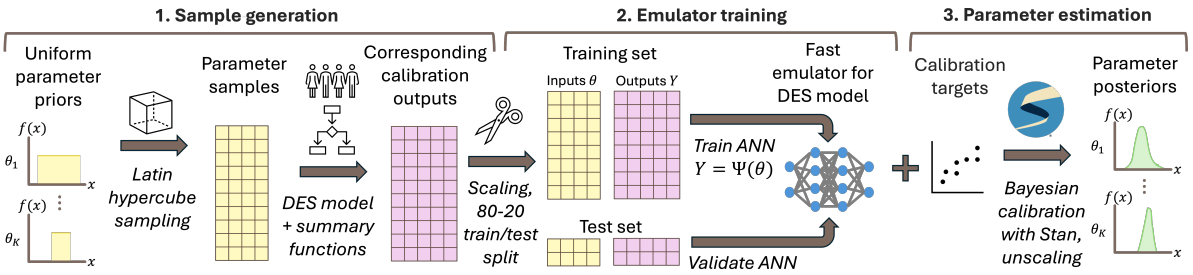
To implement IMABC, the user creates objects for the model parameters with their prior distributions, targets with initial and stopping tolerance intervals, and a function that runs the simulation model with parameter inputs to produce a vector of outputs corresponding to the targets. The user also sets the number of parameter samples in the first iteration N_o, the number of normal mixtures $N^{(c)}$ sampled from at each iteration, the number of points B to draw from each normal mixture, and the effective sample size (ESS) of the final set of accepted posterior samples N_{post}. Guidance for choosing the values of these arguments is provided in [12] and the IMABC GitHub documentation. It is recommended to use a large initial sample size N_o and wide initial tolerance intervals to increase the algorithm's likelihood of identifying acceptable parameter sets in the first step [12]. 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 adaptive sampling. A weighted sample of the posterior distribution is used to generate decision-analytic outcomes.

In our example, 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.

3.3 BayCANN

BayCANN uses an artificial neural network (ANN) as an emulator to map parameter inputs of the cancer natural history model to summary statistics corresponding to the calibration targets [11]. 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 [42]. BayCANN includes sample generation, emulator training, and parameter estimation steps as illustrated in Figure 4. 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.

For ANN fitting and validation, we draw a quasi-random sample of parameter sets from the prior distributions using LHS; simulate a cohort of life trajectories for each set; and calculate the cancer prevalence, incidence, and other target outcomes for each cohort. The sets of parameters and outcomes are divided with an 80-20 train-test split and standardized to ensure stable gradients during ANN training. We fit the ANN by minimizing the loss, 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



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

Fig. 4: Steps of BayCANN

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, with the training parameters specified in Table 2. The ANN is validated by plotting the ANN-predicted outputs against the model outputs, as shown in Figure A.2 for our example. The lines of points along the diagonals indicate that the emulator accurately predicts the model outputs.

ANN training parameters	
Train-test and train-validation split	80-20
Batch size	128
Epochs	1,000
Patience	100
Hidden layers	1, 2, 3, 4
Nodes per hidden layer	32, 64, 128
Inclusion of dropout	true, false
Dropout rate	0.25
Hidden layer activation function	ReLU, tanh, sigmoid
% of hyperparameter combinations to sample	30%
Stan parameters	
Iterations	300,000
Initial periods for thinning	1
Final periods for thinning	100
Chains	4

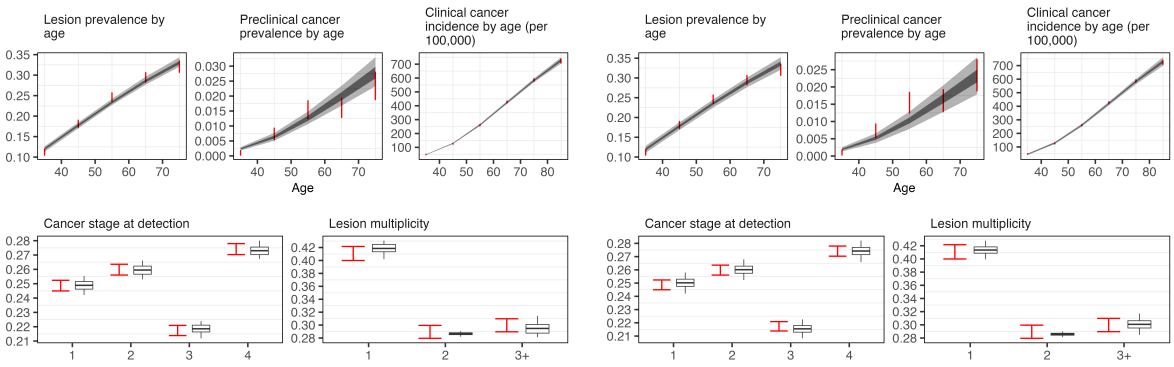
Abbreviations: ANN = artificial neural network, ReLU = rectified linear unit, tanh = hyperbolic tangent

Table 2: BayCANN parameters

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 [43]. 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. We use the Stan parameters in Table 2, starting with a thinning period of 1. After examining the number of periods at which autocorrelation of outputs plateaus, we change the number of periods for thinning to 100 and regenerate the posterior sample. To assess the quality of the Stan model fit, we check the mixing of the chains (Figure A.3) and the R-hat convergence diagnostic. The BayCANN posterior outputs include 6000 parameter sets.

3.4 Validation

To internally validate the calibration, we visually examine whether the model outputs produced by the joint posterior distribution replicate the targets and their uncertainty bounds. The IMABC algorithm automatically generates the calibration outputs associated with the posteriors. For BayCANN, the user may choose to simulate the calibration outputs and decision outcomes together for efficiency to avoid repeatedly generating life trajectories with the same posterior samples. Figure 5 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



(a) IMABC (b) BayCANN

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

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

Modality characteristic	Distribution	Mean (95% CI)
Non-invasive screening test		
Sensitivity		
Lesion	Beta(80, 120)	0.40 (0.33, 0.47)
Cancer	Beta(45, 5)	0.90 (0.80, 0.97)
Specificity	Beta(840, 160)	0.84 (0.82, 0.86)
Confirmatory test		
Sensitivity		
Lesion	Beta(150, 50)	0.75 (0.69, 0.81)
Cancer	Beta(99, 1)	0.99 (0.96, 1.00)
Specificity	Beta(100, 0)	1.00 (1.00, 1.00)

Abbreviations: CI = confidence interval

Table 3: Test performance

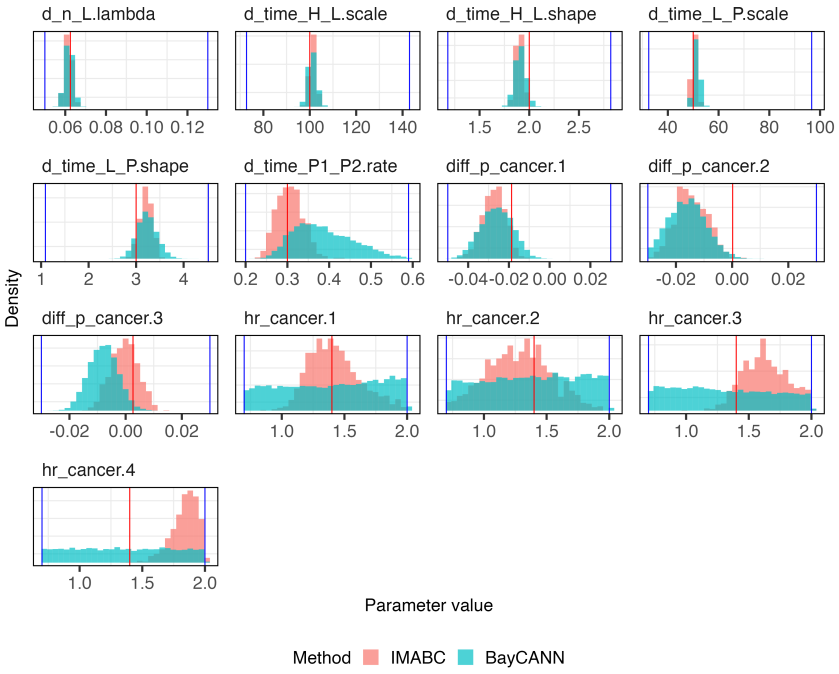
further demonstrate the accuracy of the calibration algorithms for this example by showing the overlap of the posterior distribution histograms and the true parameters, represented by the red vertical lines, in Figure 6.

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 [44]. However, the wide range and high correlation (see Figures A.1 and A.5) 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. However, as long as the decision outcomes associated with the posteriors accurately account for the uncertainty due to non-identifiability, 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 [45–47].

The ultimate external validity test would be for a calibrated model to replicate held-out real-world data, such as the mortality results of a screening trial. For a more detailed discussion of types of microsimulation model validation and choices for validation targets, we refer to [48, 49].

4 Model-based policy evaluation

We illustrate an application of the modeling and calibration pipeline by sampling from the parameter posterior distribution to simulate the LYG and change in testing burden associated with screening among 40-year-olds without a cancer diagnosis. These are two of many outcomes, including quality of life,



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

Fig. 6: Posteriors (blue and red histograms) versus prior bounds (blue vertical lines) and true parameters (red vertical line)

adverse events, healthcare costs, and equity, that policymakers may consider when deciding whether to adopt a new screening intervention.

4.1 Screening strategies

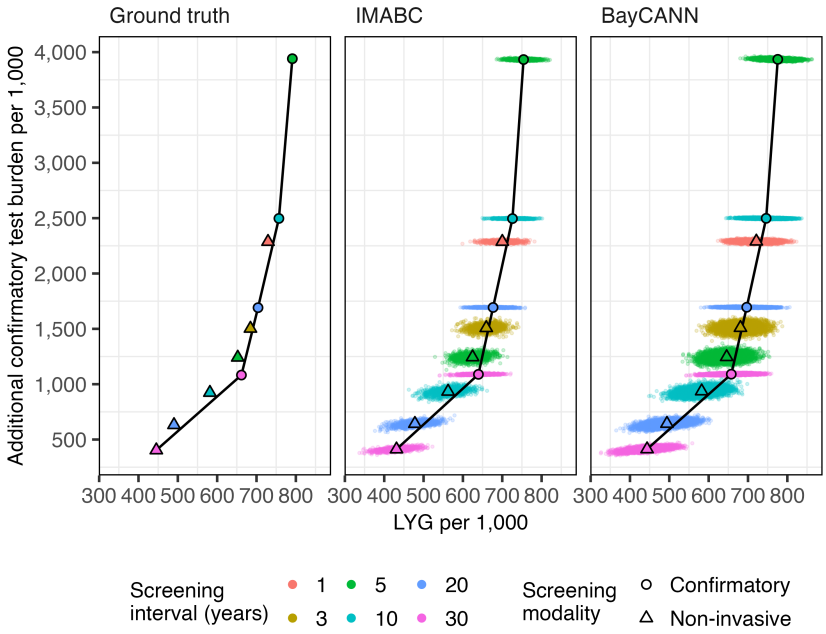
We consider screening strategies applied to individuals aged 50 to 70 using two types of tests, whose performance characteristics are shown in Table 3: an invasive test that doubles as a confirmatory test and a non-invasive test that is one-tenth as costly. For strategies using the confirmatory test as a screening test, we examine routine screening intervals of 10, 20, and 30 years, with a 30-year interval equivalent to one-time screening at age 50. For the strategies with the non-invasive test as the primary screening test, we consider routine screening intervals of 1, 3, 5, 10, 20, and 30 years. Positive results with the non-invasive test are followed immediately by the confirmatory test.

For tests conducted during the precancerous lesion state, we apply the sensitivity of both tests on a per-lesion basis. Only lesions that test positive with the confirmatory test are removed. However, if at least one lesion tests positive with the non-invasive test, the follow-on confirmatory test can detect all lesions present at the time. Individuals who undergo a positive non-invasive screening followed by a negative confirmatory test receive their next screening test after a maximum of 10 years and a routine screening interval.

A confirmatory test that detects lesions puts individuals on a surveillance regimen with follow-up times depending on the number detected. If 1 or 2 lesions are detected, the patient receives another confirmatory test after 7 years. If 3 or more lesions are detected, the patient receives a confirmatory test in another 3 years. After a negative confirmatory test, the individual returns to routine screening with the original test modality. The analysis assumes 100% adherence and that the confirmatory test is used to diagnose all symptom-detected cancer cases.

4.2 Strategy evaluation

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 simulate a cohort of individual trajectories with the natural history model according to Section 2.1, randomly sample test characteristics from the distributions in Table 3, apply each screening



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

Fig. 7: Test burden versus LYG

strategy to update the life trajectories according to Section 2.2, and calculate LYG, test burden, mean dwell time, and mean sojourn time. We retain the cohort size from calibration, though a user may need to adjust the cohort size per Section 2.4. This sampling methodology constitutes a probabilistic analysis that produces a distribution of outcomes accounting for model input parameter uncertainty. To obtain ground truth outcomes for comparison, we take the means across 100 realizations from the ground truth parameters with the same cohort size from calibration.

To visualize the tradeoffs of each strategy, we plot the mean additional confirmatory test burden versus the mean LYG associated with each strategy compared to no screening (Figure 7). Uncertainty around the mean is reflected in the cloud of points generated by sampling from the posterior. The additional confirmatory test burden is calculated as the incremental number of confirmatory tests in the screening versus no-screening scenario plus the number of confirmatory tests equivalent to the value of the total noninvasive tests. The cost-effectiveness frontier, represented by the black line, is consistent between the ground truth and the posteriors generated by both calibration methods. From least to most expensive and beneficial, non-dominated strategies include one-time screening with the non-invasive and confirmatory tests and screening every 20, 10, and 5 years with the confirmatory test. A shorter screening interval and the use of the confirmatory test are associated with a higher LYG and increased testing burden. In addition, we find that the 95% PIs for precancerous lesion dwell time and sojourn time for both calibration algorithms contain the mean estimates sampled from 100 simulations of the ground truth parameters (Table 4).

	Mean dwell time (95% PI)	**Mean sojourn time (95% PI)**
Ground truth	32.11	4.06
IMABC	33.22 (31.94, 34.43)	3.98 (3.50, 4.54)
BayCANN	33.65 (31.96, 35.34)	3.53 (2.78, 4.25)

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

Table 4: Natural history outcomes in years

5 Discussion

We introduce the DESCIPHR pipeline to fill a gap in open-source, flexible, end-to-end resources for implementing and deploying health decision models. For decades, simulation modeling has successfully informed cancer screening policy, though typically with bespoke, closed-source implementations [4, 50, 51]. 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, high barrier to entry, and lack of transparency. We believe that releasing an adaptable open-source model will accelerate the progress of health decision modeling applications and methods development. The modularity of DESCIPHR's codebase in keeping with the DARTH framework [25] allows users to easily substitute model structures, variable distributions, input data, calibration methods, parameters, and outcomes. As a general cancer model, the DESCIPHR natural history model can be adapted to specific cancer sites, as well as other diseases characterized by progressive stages.

Improving on prior DES and calibration tutorials, DESCIPHR also demonstrates the model preprocessing and setup required to run two effective Bayesian calibration methods, IMABC and BayCANN, and then apply the calibration outputs to make forecasts that account for the uncertainty of model parameters. Both methods provide viable approximations of parameter posteriors. IMABC leverages an adaptive sampling scheme and multivariate normal approximations to hone in on parameter sets with good fit to real-world targets, thus increasing efficiency relative to the MCMC and ABC Bayesian calibration techniques [12]. The thorough documentation in the IMABC publication, R package, and GitHub repository facilitates the setup and execution of the method. In DESCIPHR, we enhance the implementation by demonstrating the structure of the calibration inputs in the context of our cancer model and outcome calculations, alongside the extraction and application of the calibration outputs, which are a weighted sample of parameter posteriors. BayCANN trains an ANN on samples from the parameter space to create a high-fidelity representation of the relationship between the model and the calibration outputs [11]. Though the algorithm contains a multitude of steps and parameters, the pipeline is set up so a user can run the algorithm out of the box after updating model specifications and input data as needed.

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. Our target-informed approach to estimating prior distributions for the parameters governing the time to disease onset and cancer progression circumvents the need to guess a reasonable range for a grid search manually. Despite advances in Bayesian calibration methods, assigning parameter 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.

The prediction accuracy of machine learning models is sensitive to the hyperparameter configuration, making hyperparameter optimization a critical part of empirical machine learning [52, 53]. Our augmentations to BayCANN allow for automated hyperparameter optimization after specifying the search parameters. 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 [54].

To balance the learnability and adaptability of DESCIPHR, the implemented cancer model relies on some simplifying assumptions, the validity of which should be explored in future research and iterations of the model for specific types of cancer. A limitation of using an exponential distribution to model the stage-specific progression and detection of cancer is the distribution's zero mode, which conflicts with biological intuition [55]. However, exponential distributions and their associated constant hazard rates are commonly used to approximate event times due to the simplicity of calibrating a single-parameter distribution, and it is unclear whether the simplification produces meaningful deviations from reality. The model does not explicitly model cancer recurrence; instead, we assume the survival distributions from cancer diagnosis implicitly account for variation in survival given standards of care during the time range of data collection. This may be acceptable for evaluating a pre-diagnosis screening intervention

but insufficient for evaluating post-diagnosis surveillance and second- or later-line therapies, as has been done for bladder [56], colorectal [57], and prostate cancer [58, 59]. Future work to extend the pipeline to model cancer recurrence and life-extending effects of novel treatments may be worthwhile in assessing their population impacts.

Given the inclusion of parameters for cancer stage progression but a lack of targets to inform the dwell time within each stage, the model also encounters some degree of parameter nonidentifiability. Nonidentifiability is a concern when different parameter sets that can explain the calibration targets equally well produce different outcomes and conclusions about optimal decisions [44]. In our case, despite the interactions between varying progression rates and the likelihood of detection at an earlier cancer stage during screening, the strategies on the cost-effectiveness frontier for both calibration methods are consistent with the ground truth (Figure 7). The effects of nonidentifiability are mitigated by restricting the hazard ratios between progression rates in consecutive cancer stages and using the stage distribution at diagnosis to inform detection rates. We leave an in-depth exploration of the impacts of parameter nonidentifiability and cancer screening decisions to future work.

For the demonstrative example of the pipeline, we assume that sensitivity and specificity apply independently to each test, although, in reality, repeated tests for the same individual may have correlated performance. We also model perfect adherence to screening and complete resection of precancerous lesions to understand best-case outcomes. However, this may overestimate the magnitude of the incremental benefits and risks compared to real-world screening adherence and resection rates. For simplicity, the current implementation assumes a constant test sensitivity for lesion detection. While test sensitivity often increases with lesion size [60], modeling the growth trajectories of lesions introduces significant complexity; the modeler would need to specify distributions for the shape of the growth curve and the size at which the lesion converts to cancer, but in many cases, there are limited longitudinal data on undisturbed lesions to inform these distributions, as clinicians often intervene upon detection in the interest of patients' health. Given the modularity of the pipeline, however, it would not be difficult to add distributions to characterize lesion growth and sample characteristics at the point of screening if desired.

Furthermore, the modeler may need to adjust the target lesion size and multiplicity distributions to account for differential false negative rates based on lesion size; if published studies report these statistics from screening studies using an imperfect test, the targets may underestimate the proportion of smaller lesions and overall lesion counts, as smaller lesions are more likely to have been missed. Test sensitivity is also assumed to be constant across preclinical cancer stages, though this may be a reasonable assumption for cancers without a noticeable upward trend in sensitivity with stage, such as CRC [60]. Future extensions of the model can incorporate more realistic test correlations, adherence patterns, and lesion growth patterns. While the decision outputs from the pipeline reflect input data uncertainty, assuming no bias or uncertainty in model structure, there is an opportunity to extend the pipeline by incorporating systematic evaluations of potential biases and variations in model structure.

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 [61–63]. Future work evaluating the performance of the DESCIPHR modeling pipeline on predicting screening trial outcomes across different cancer types will be necessary for external validation. 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.

6 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.

7 Declarations

7.1 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 National Cancer Institute (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.

7.2 Conflicts of interest

Authors have no conflicts of interest to declare.

7.3 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.

7.4 Ethics approval

Not applicable.

7.5 Consent to participate

Not applicable.

7.6 Code availability

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

7.7 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.

8 Acknowledgments

We are grateful to Michael C. Higgins, PhD, for feedback on the natural history model.

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] 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.

[12] 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.

[13] 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.

[14] 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.

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

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

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

[33] 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.

[34] 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.

[35] 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.

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

[37] Tenny S, Hoffman MR. Prevalence. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2025. Available from: http://www.ncbi.nlm.nih.gov/books/NBK430867/.

[38] Tenny S, Boktor SW. Incidence. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2024. Available from: http://www.ncbi.nlm.nih.gov/books/NBK430746/.

[39] 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.

[40] 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.

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

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] 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.

[47] 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.

[48] 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.

[49] 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.

[50] 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.

[51] 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.

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

[53] 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.

[54] 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.

[55] Parmigiani G, Skates S. Estimating distribution of age of the onset of detectable asymptomatic cancer. Mathematical and Computer Modelling. 2001 Jun;33(12):1347–1360. https://doi.org/10.1016/S0895-7177(00)00320-4.

[56] Heijnsdijk EAM, Nieboer D, Garg T, Lansdorp-Vogelaar I, de Koning HJ, Nielsen ME. Cost-effectiveness of surveillance schedules in older adults with non-muscle-invasive bladder cancer. BJU International. 2019;123(2):307–312. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/bju.14502. https://doi.org/10.1111/bju.14502.

[57] 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.

[58] de Carvalho TM, Heijnsdijk EAM, de Koning HJ. Estimating the risks and benefits of active surveillance protocols for prostate cancer: a microsimulation study. BJU International. 2017;119(4):560–566. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/bju.13542. https://doi.org/10.1111/bju.13542.

[59] Magnani CJ, Hernandez-Boussard T, Baker LC, Goldhaber-Fiebert JD, Brooks JD. Cost-Effectiveness Analysis and Microsimulation of Serial Multiparametric Magnetic Resonance Imaging in Active Surveillance of Localized Prostate Cancer. The Journal of Urology. 2022 Jul;Publisher: Wolters KluwerPhiladelphia, PA. https://doi.org/10.1097/JU.0000000000002490.

[60] 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.

[61] 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.

[62] 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.

[63] 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.

[64] 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.

[65] 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.

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

[67] 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 Additional results figures

See supplemental file for Appendix A.

B Adapting the model

B.1 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 `configs/configs.yaml` 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 `configs/configs.yaml`. 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

B.2 Model configurations

The file `configs/configs.yaml` provides 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 pre-cancerous 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()`

B.3 Further augmentations

Though the pipeline does not include a module that generates time-dependent risk factors or initiating events, such as infection with human papillomavirus in the natural history of cervical cancer [64], additional steps 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.
   (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 are summing up as you would expect.

Risk factors, such as insurance status, family history, or geographic region, may be added in one line to the `simulate_baseline_data()` module, and their population distributions, as well as their interactions with other distributions' parameters, may be specified in `load_model_params()`.

C 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.

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

The National Cancer Institute'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 [65]. 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 age range midpoint 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 may 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

$$F_{D_c}(t) = \int_0^t \hat{I}_C(u) F_{CD_c}(t-u|T_C = u)du. \tag{6}$$

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

$$f_C(t) = \hat{I}_C(t)\left(1 - F_{D_c}(t)\right). \tag{7}$$

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

$$F_C(t) = \int_0^t f_C(u)du. \tag{8}$$

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 by assuming that relative survival rates are independent of the age at diagnosis and exponentially distributed. To calculate the rate of death from cancer, we start with SEER data on the 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 the 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 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 [66]. 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:

$$H_C(t) = \int_0^t \hat{I}_C(u)du \tag{9}$$

$$F_C(t) = 1 - e^{-H_C(t)} \tag{10}$$

C.2 Deriving the CDF of disease onset

Prevalence data is 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:

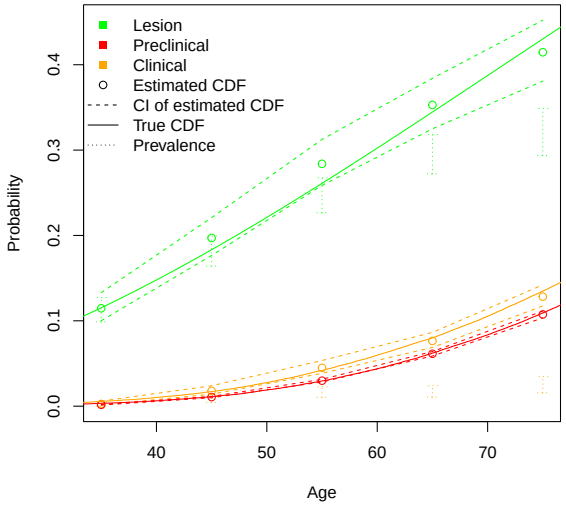
$$F_P(t) = P_P(t)(1 - F_C(t)) + F_C(t). \tag{11}$$

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

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:

$$F_L(t) = P_L(t)(1 - F_C(t)) + F_P(t) \tag{12}$$

Figure 8 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



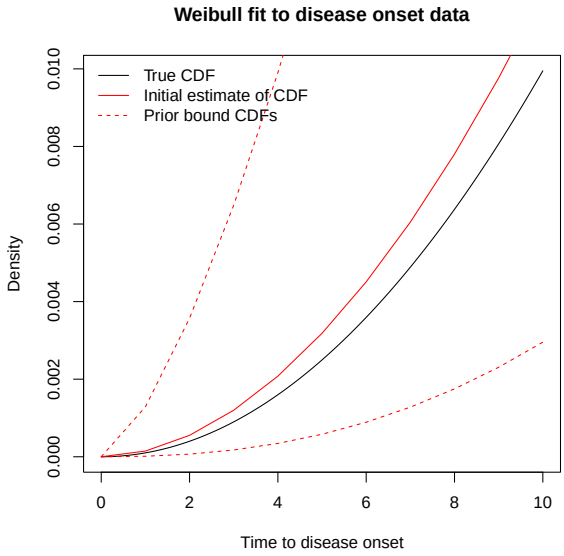
Abbreviations: CDF = cumulative distribution function, CI = confidence interval
Fig. 8: Coverage of ground truth simulated CDFs by the CDF estimates

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 pre-cancerous 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

$$y_i = ax_i + b, \tag{13}$$

where $y_i = \ln(-\ln(1 - F_X(t_i)))$, $a = \alpha$, $x_i = \ln t_i$, and $b = -\alpha \ln \sigma$ [67]. 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 9, the range of priors for the parameters governing the time to disease onset contains the ground truth parameters.



Abbreviations: CDF = cumulative distribution function
Fig. 9: Coverage of ground truth Weibull distribution by priors