*The Annals of Applied Statistics*
2019, Vol. 13, No. 4, 2189–2212
https://doi.org/10.1214/19-AOAS1279
© Institute of Mathematical Statistics, 2019

# MICROSIMULATION MODEL CALIBRATION USING INCREMENTAL MIXTURE APPROXIMATE BAYESIAN COMPUTATION

BY CAROLYN M. RUTTER[*,1], JONATHAN OZIK[†,‡,2], MARIA DEYOREO[*,1] AND NICHOLSON COLLIER[†,‡,2]

*RAND Corporation*[*], *University of Chicago*[†] *and Argonne National Laboratory*[‡]

Microsimulation models (MSMs) are used to inform policy by predicting population-level outcomes under different scenarios. MSMs simulate individual-level event histories that mark the disease process (such as the development of cancer) and the effect of policy actions (such as screening) on these events. MSMs often have many unknown parameters; calibration is the process of searching the parameter space to select parameters that result in accurate MSM prediction of a wide range of targets. We develop Incremental Mixture Approximate Bayesian Computation (IMABC) for MSM calibration which results in a simulated sample from the posterior distribution of model parameters given calibration targets. IMABC begins with a rejection-based ABC step, drawing a sample of points from the prior distribution of model parameters and accepting points that result in simulated targets that are near observed targets. Next, the sample is iteratively updated by drawing additional points from a mixture of multivariate normal distributions and accepting points that result in accurate predictions. Posterior estimates are obtained by weighting the final set of accepted points to account for the adaptive sampling scheme. We demonstrate IMABC by calibrating CRC-SPIN 2.0, an updated version of a MSM for colorectal cancer (CRC) that has been used to inform national CRC screening guidelines.

## 1. Introduction.

Microsimulation models (MSMs) are used to inform policy by predicting population-level outcomes under different policy scenarios. MSMs are characterized by simulation of *agents* that represent individual members of an

Received August 2018; revised June 2019.

[1] Supported by a grant from the National Cancer Institute (U01-CA-199335) as part of the Cancer Intervention and Surveillance Modeling Network (CISNET). The contents of this article are the sole responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute.

[2] Supported by NIH Grants 1R01GM115839 and 1S10OD018495, the NCI-DOE Joint Design of Advanced Computing Solutions for Cancer program, and through resources provided by the Computation Institute and the Biological Sciences Division of the University of Chicago, the University of Chicago Research Computing Center, and Argonne National Laboratory. This material is based on work supported by the U.S. Department of Energy, Office of Science, under contract number DE-AC02-06CH11357.

*Key words and phrases.* Adaptive ABC, agent-based models, colorectal cancer.

idealized population of interest. For each agent, the model simulates event histories that catalog landmarks in the disease process. In general, disease processes modeled are not directly observable, though outcomes from these processes may be observed. For example, the process of developing colorectal cancer (CRC) cannot be observed, but the prevalence of both precursor lesions (adenomas) and preclinical (asymptomatic) CRC can be estimated from screening trials, and CRC incidence can be observed from national registry data.

Model calibration involves selecting parameter values that result in model predictions that are consistent with observed data and expected findings. Once parameters are selected, MSMs can be used to make predictions about population trends in disease outcomes, effectiveness of interventions and the comparative effectiveness of interventions, especially those without direct empirical comparisons. For example, models have been used to inform U.S. Preventive Services Task Force screening guidelines for breast (Mandelblatt et al. (2016)), cervical (Kim, Burger, Regan and Sy (2019)), colorectal (Knudsen et al. (2016)) and lung cancer (de Koning et al. (2014)) by comparing the effectiveness of different screening regimens.

MSM calibration involves searching a high-dimensional parameter space to predict many targets. Several approaches have been proposed. The simplest calibration method involves perturbing parameters one at a time and evaluating the goodness of fit to calibration data, but this is only feasible when calibrating a few parameters. Directed searches, such as the Nelder–Mead algorithm (Nelder and Mead (1965)), provide a derivative-free hillclimb to identify a single best value for each parameter. Kong, McMahon and Gazelle (2009) used search algorithms from engineering (simulated annealing and a genetic algorithm) for model calibration. Bayesian calibration methods estimate the joint posterior distribution of MSM parameters which provides information about parameter uncertainty and enables estimation of functions of parameters. Rutter, Miglioretti and Savarino (2009) used Markov chain Monte Carlo (MCMC) to simulate draws from the posterior distribution of MSM parameters given calibration targets. However, MCMC can be difficult and costly to apply to MSM calibration and, because MCMC is based on a process of sequentially updating draws, it is not easy to parallelize the process to take advantage of modern computing resources.

Approximate Bayesian Computation (ABC) offers an alternative approach to MSM calibration. ABC is a likelihood-free technique for simulating draws from the posterior distribution that approximates likelihood-based algorithms by choosing parameters that produce a close match to data rather than calculating the likelihood (Conlan et al. (2012), Marin et al. (2012), Sisson, Fan and Beaumont (2019)). The validity of ABC algorithms, in the sense that they result in samples from the approximate posterior distribution, relies on the validity of the corresponding exact algorithms (Sisson, Fan and Tanaka (2007)). The idea underlying ABC is simple. For a parameter $\theta$ with prior distribution $\pi(\theta)$ and observed data $y$, we can write the posterior probability as $p(\theta|y) = p(y|\theta)\pi(\theta)$ implying that we can approximate $p(\theta|y)$ by sampling $\theta$ from $\pi(\cdot)$ and retaining only points with $p(y|\theta) \approx 1$. However, ABC is inefficient and can fail when the parameter space is high dimensional, when there are many calibration targets, or when the prior distributions are very different from the posterior distributions. McKinley et al. (2018) found that popular ABC variants that improve the algorithm's efficiency were not computationally feasible for calibrating stochastic epidemiological models. We propose an Incremental Mixture ABC (IMABC) approach for MSM model calibration that begins with a basic rejection-sampling ABC step (e.g., Pritchard et al. (1999)) and then incrementally adds points to regions where targets are well predicted.

In the next sections we describe the CRC-SPIN MSM for the natural history of colorectal cancer (CRC) (Section 2), calibration targets used to inform CRC-SPIN model parameters (Section 3), the IMABC calibration approach (Section 4) and results of CRC-SPIN model calibration based on IMABC (Section 5). We conclude with general remarks about the proposed approach and discussion of future work (Section 6).

**2. Microsimulation model for the natural history of colorectal cancer.** The ColoRectal Cancer Simulated Population Incidence and Natural history model (CRC-SPIN) (Rutter, Miglioretti and Savarino (2009), Rutter and Savarino (2010)) describes the natural history of CRC based on the adenoma-carcinoma sequence (Leslie et al. (2002), Muto, Bussey and Morson (1975)). Four model components describe the natural history of CRC: (1) adenoma risk; (2) adenoma growth; (3) transition from adenoma to preclinical cancer; and (4) transition from preclinical to clinical cancer (sojourn time).

CRC-SPIN has been used to provide guidance to the Centers for Medicare and Medicaid Services (CMS) (Zauber et al. (2009)) and to inform U.S. Preventive Services Task Force CRC screening guidelines (Knudsen et al. (2016)). Model validation, based on comparison of model predictions to observed outcomes, revealed that while CRC-SPIN predicted many aspects of CRC well (including clinically detected cancer, cancer mortality and the effectiveness of screening), it predicted detection of too few preclinical cancers at screening, indicating that the simulated times spent in the preclinical cancer phase (sojourn times) were too short (Rutter et al. (2016)). In this paper we present CRC-SPIN 2.0, an update to the original CRC-SPIN 1.0. CRC-SPIN 2.0 contains 22 calibrated parameters (Table 1). Because this is a model recalibration, prior distributions are based on results from the previous calibration of CRC-SPIN 1.0 (Rutter, Miglioretti and Savarino (2009)). In this section we provide an overview of the model. Additional details are provided in Appendix A and online at `cisnet.cancer.gov` (National Cancer Institute (2018)).

2.1. *Adenoma risk model.* The occurrence of adenomas is modeled using a nonhomogeneous Poisson process with a piecewise age effect. We assume zero

TABLE 1

*Summary of CRC microsimulation model components. Calibrated parameters associated with the four components of the natural history model, including parameter notation, associated equations, prior distributions and posterior estimates (mean and 95% credible interval). $\mathrm{TN}_{[a,b]}(\mu,\sigma)$ denotes a truncated normal distribution with mean $\mu$ and standard deviation $\sigma$, restricted to the interval $(a, b)$. $U(a,b)$ denotes a uniform distribution over $(a, b)$. Refer to Section 2 for details of the four model components*

| | Prior | Posterior Estimates | |
|---|---|---|---|
| Component | Distribution | Mean | 95% CI |
| Adenoma Risk (eqn (1)) | | | |
| Baseline log-risk | $A \sim \mathrm{TN}_{[-7.5,-5.9]}(-6.6, 0.4)$ | −6.56 | −6.89 to −6.11 |
| Standard deviation, baseline log-risk | $\sigma_\alpha \sim \mathrm{TN}_{[0.25,1.75]}(1.1, 0.3)$ | 1.62 | 1.43 to 1.74 |
| Female | $\alpha_1 \sim \mathrm{TN}_{[-0.5,0.1]}(-0.5, 0.1)$ | −0.62 | −0.70 to −0.49 |
| Age effect, age $\in [20, 50)$ | $\alpha_{20} \sim \mathrm{TN}_{[0.02,0.05]}(0.04, 0.01)$ | 0.044 | 0.034 to 0.050 |
| Age effect, age $\in [50, 60)$ | $\alpha_{50} \sim \mathrm{TN}_{[0.01,0.05]}(0.03, 0.01)$ | 0.038 | 0.021 to 0.049 |
| Age effect, age $\in [60, 70)$ | $\alpha_{60} \sim \mathrm{TN}_{[-0.01,0.05]}(0.03, 0.01)$ | 0.022 | −0.006 to 0.047 |
| Age effect, age $\geq 70$ | $\alpha_{70} \sim \mathrm{TN}_{[-0.02,0.03]}(0.03, 0.03)$ | −0.002 | −0.018 to 0.022 |
| Time to 10mm (eqn (2)) | | | |
| Shape, colon | $\beta_{1C} \sim \mathrm{U}(1.1, 5)$ | 1.43 | 1.16 to 1.70 |
| Shape, rectum | $\beta_{1R} \sim \mathrm{U}(1.1, 5)$ | 3.39 | 1.97 to 4.76 |
| Scale, colon* | $\beta_{2C} \sim \mathrm{U}(10.7, 40)$ | 37.7 | 34.0 to 39.9 |
| Scale, rectum* | $\beta_{2R} \sim \mathrm{U}(10.7, 40)$ | 15.3 | 11.8 to 19.5 |
| Adenoma Growth Curve (eqn (3)) | | | |
| Shape parameter | $p \sim \mathrm{TN}_{[0.5,3.2]}(1.0, 0.5)$ | 0.69 | 0.56 to 0.83 |
| Transition to Preclinical Cancer (eqn (4)) | | | |
| Intercept | $\gamma_0 \sim \mathrm{TN}_{[2.6,3.6]}(3.1, 0.5)$ | 3.16 | 3.03 to 3.35 |
| Female (versus male) | $\gamma_1 \sim \mathrm{TN}_{[-0.5,0.3]}(-0.06, 0.2)$ | −0.13 | −0.19 to −0.07 |
| Rectal (versus colon) | $\gamma_2 \sim \mathrm{TN}_{[-0.5,0.5]}(0.25, 0.25)$ | −0.05 | −0.24 to 0.13 |
| Female & rectal | $\gamma_3 \sim \mathrm{TN}_{[-0.35,0.25]}(-0.14, 0.2)$ | 0.07 | −0.03 to 0.17 |
| Age at initiation | $\gamma_4 \sim \mathrm{TN}_{[-0.24,0.02]}(-0.08, 0.04)$ | −0.11 | −0.14 to −0.08 |
| Squared age at initiation | $\gamma_5 \sim \mathrm{U}(-1.5, 1.5)$ | 0.01 | 0.00 to 0.02 |
| Standard deviation | $\sigma_\gamma \sim \mathrm{U}(0.5, 1.0)$ | 0.56 | 0.50 to 0.65 |
| Sojourn Time (eqn (5)) | | | |
| Scale | $\lambda_1 \sim \mathrm{U}(2.25, 4.25)$ | 2.57 | 2.27 to 3.06 |
| Shape | $\lambda_2 \sim \mathrm{U}(2.0, 5.0)$ | 3.72 | 2.20 to 4.92 |
| log-hazard ratio, rectal cancer | $\lambda_3 \sim \mathrm{U}(-1.0, 1.0)$ | −0.35 | −0.96 to 0.67 |

*Scale parameters, $\beta_2$, were also restricted to range from $10(-\ln(0.25))^{1/\beta_1}$ to $(-\ln(0.0001))^{1/\beta_1}$, corresponding to the probability of an adenoma reaching 10mm within 10 years ranging from 0.0001 to 0.25.

risk before age 20. We focus on CRC in adults because CRC is very rare before age 20 with incidence of about one in 10 million (Koh et al. (2015)). The $i$th agent's baseline instantaneous risk of an adenoma at age $a = 20$ years is given by $\psi_i(20) = \exp(\alpha_{0i} + \alpha_1 \mathrm{female}_i)$ where $\alpha_{0i} \sim N(A, \sigma_\alpha)$ and $\alpha_1$ captures the difference in risk for women ($\mathrm{female}_i = 1$ indicates agent $i$ is female). Adenoma risk changes over time, generally increasing with age, a process we model using a linear change-point for log-risk with knots at ages 50, 60 and 70.

$$
\begin{aligned}
\ln(\psi_i(a)) &= \alpha_{0i} + \alpha_1 \mathrm{sex}_i + \delta(a \geq 20)\min(a-20, 30)\alpha_{20} \\
&\quad + \delta(a \geq 50)\min((a-50), 10)\alpha_{50} \\
&\quad + \delta(a \geq 60)\min((a-60), 10)\alpha_{60} \\
&\quad + \delta(a \geq 70)(a-70)\alpha_{70}.
\end{aligned} \tag{1}
$$

2.2. *Adenoma growth model.* For each adenoma we simulate a hypothetical time to reach 10mm, $t_{10\mathrm{mm}}$ which may exceed the agent's lifespan. We assume that $t_{10\mathrm{mm}}$ has a Frèchet distribution with shape parameter $\beta_1$, scale parameter $\beta_2$, and cumulative distribution function given by

$$
F(t) = \exp\left[-\left(\frac{t}{\beta_2}\right)^{-\beta_1}\right] \tag{2}
$$

for $t \geq 0$. Prior distributions for adenoma growth parameters specify that most adenomas grow very slowly. We allow different scale and shape parameters for adenomas in the colon and rectum.

Adenoma size at any point in time is simulated using the Richard's growth model which incorporates a wide range of sigmoidal growth patterns (Tjørve and Tjørve (2010)). The diameter of the $j$th adenoma in the $i$th agent at time $t$ after initiation is given by

$$
d_{ij}(t) = d_\infty\left[1 + \left(\left(\frac{d_0}{d_\infty}\right)^{1/p} - 1\right)\exp(-\lambda_{ij}t)\right]^p, \tag{3}
$$

where $d_0 = 1$mm is the minimum adenoma diameter in millimeters (mm) and $d_\infty = 50$ is the maximum adenoma diameter. The calibrated parameter $p$ determines the shape of the growth curve. The growth rate for the $j$th adenoma within the $i$th agent, $\lambda_{ij}$, is calculated by setting $t = t_{10\mathrm{mm}}$ and $d = 10$ in equation (3).

2.3. *Model for transition from adenoma to preclinical invasive cancer.* For the $j$th adenoma in the $i$th agent, the size at transition to preclinical cancer (in mm) is simulated using a log-normal distribution; the underlying (exponentiated) normal distribution is assumed to have standard deviation $\sigma_\gamma$ and mean

$$
\begin{aligned}
\mu_{ij} &= \gamma_0 + \gamma_1 \mathrm{female}_i + \gamma_2 \mathrm{rectum}_{ij} + \gamma_3 \mathrm{female}_i \mathrm{rectum}_{ij} \\
&\quad + \gamma_4 \mathrm{age}_{ij} + \gamma_5 \mathrm{age}_{ij}^2,
\end{aligned} \tag{4}
$$

where $\mathrm{rectum}_{ij}$ is an indicator of rectal versus colon location and $\mathrm{age}_{ij}$ is the age at adenoma initiation in decades, centered at 50 years. Based on this model, the probability that an adenoma transitions to preclinical cancer increases with increasing size. Most adenomas do not reach transition size, and small adenomas are unlikely to transition to cancer.

2.4. *Model for sojourn time.* Sojourn time is the time between the transition to preclinical (asymptomatic) CRC and the transition to clinical (symptomatic and detected) cancer. We simulate sojourn time using a Weibull distribution with survival function

$$S(t) = \exp\left(-\left(\frac{t}{\lambda_1}\right)^{\lambda_2}\right), \tag{5}$$

for preclinical cancer in the colon, and assume a proportional hazards model, with hazard ratio $\exp(\lambda_3 \text{rectum}_{ij})$, to allow sojourn time to systematically differ for preclinical cancers in the colon and rectum.

2.5. *Simulation of lifespan and colorectal cancer survival.* Once a cancer becomes clinically detectable, we simulate stage and tumor size at clinical detection based on SEER data from 1975 to 1979, prior to diffusion of CRC screening (National Cancer Institute (2004)). Survival time after CRC diagnosis is simulated using relative survival estimates from analysis of SEER data from individuals diagnosed with CRC from 1975 through 2003 (Rutter et al. (2013)). CRC survival is based on the first diagnosed CRC and depends on sex, age at diagnosis, cancer location (colon or rectum) and stage at diagnosis. We assume proportional hazards of CRC and other-cause mortality within sex and birth-year cohorts. Other-cause mortality is modeled using survival probabilities based on product-limit estimates for age and birth-year cohorts from the National Center for Health Statistics Databases (National Center for Health Statistics (2000)).

**3. Calibration data.** Calibration data are derived from published studies and typically take the form of summary statistics with known distributions, such as binomial, multinomial and Poisson. For example, when we calibrate to incidence rates from a population of a given size, we assume, given the population size, the number of incident cancers follows a binomial distribution. This is a unique feature of our calibration problem, in that calibration targets are summary statistics drawn from published studies rather than multiple user-defined summaries of a single dataset. We calibrate to 40 targets from six sources, SEER registry data (National Cancer Institute (2004), 20 targets, Section 3.1) and five published studies (20 targets, Section 3.2). We also bounded adenoma growth parameters, based on information from a recent study of repeated screening colonoscopies (Ponugoti and Rex (2017)), so that the probability of an adenoma reaching 10mm within 10 years ranged from 0.0001 to 0.25, by requiring $10(-\ln(0.25))^{1/\beta_1} \leq \beta_2 \leq 10(-\ln(0.0001))^{1/\beta_1}$.

Calibration targets are based on individual-level data that is reported in aggregate. Calibration requires simulating targets by simulating a set of agents with risk that is similar to the study population based on age, gender, prior screening patterns and the time period of the study which may affect both overall and cancer-specific mortality.

TABLE 2
*Observed and predicted annual incidence of clinically detected cancers in* 1975–1979, *per* 100,000 *individuals*

| | | | | | Posterior Predicted | |
|---|---|---|---|---|---|---|
| Location | Gender | Age | Observed Mean | Tolerance Interval | Mean | 95% CI |
| Colon | Female | 20–49 | 4.9 | 3.1 to 6.7 | 3.9 | 3.1 to 5.2 |
| | | 50–59 | 43.3 | 32.9 to 53.7 | 42.7 | 34.9 to 51.2 |
| | | 60–69 | 101.5 | 83.2 to 119.9 | 105.6 | 89.3 to 118.7 |
| | | 70–84 | 221.5 | 191.8 to 251.2 | 217.7 | 194.5 to 244.8 |
| | | 85+ | 308.4 | 225.4 to 391.5 | 343.0 | 281.0 to 388.0 |
| Colon | Male | 20–49 | 4.7 | 2.9 to 6.4 | 4.1 | 3.0 to 5.4 |
| | | 50–59 | 46.0 | 34.9 to 57.1 | 48.4 | 39.3 to 56.1 |
| | | 60–69 | 122.4 | 100.7 to 144.2 | 124.0 | 105.9 to 141.2 |
| | | 70–84 | 274.5 | 233.5 to 315.5 | 261.0 | 235.5 to 294.9 |
| | | 85+ | 399.1 | 259.3 to 538.9 | 427.1 | 350.1 to 507.2 |
| Rectal | Female | 20–49 | 1.9 | 0.8 to 3.0 | 2.0 | 1.2 to 2.9 |
| | | 50–59 | 20.5 | 13.3 to 27.6 | 19.3 | 14.3 to 25.6 |
| | | 60–69 | 42.9 | 31.0 to 54.9 | 42.5 | 32.7 to 53.2 |
| | | 70–84 | 75.2 | 57.8 to 92.5 | 75.0 | 60.5 to 89.9 |
| | | 85+ | 105.1 | 56.6 to 153.7 | 102.8 | 69.7 to 139.1 |
| Rectal | Male | 20–49 | 2.3 | 1.1 to 3.6 | 2.9 | 2.0 to 3.5 |
| | | 50–59 | 29.9 | 20.9 to 38.9 | 29.9 | 23.4 to 36.8 |
| | | 60–69 | 71.6 | 55.0 to 88.3 | 66.1 | 56.1 to 79.7 |
| | | 70–84 | 129.8 | 101.6 to 158.0 | 117.3 | 102.8 to 138.2 |
| | | 85+ | 164.9 | 74.9 to 254.9 | 157.6 | 116.4 to 203.2 |

3.1. *SEER registry data.* SEER colon and rectal cancer incidence rates in 1975–1979 are a key calibration target (Table 2). Incidence rates reported are per 100,000 individuals. These rates are based on the first observed invasive colon or rectal cancer during the years 1975–1979, the most recent period prior to dissemination of CRC screening tests. We assume that given the SEER population size, the number of incident CRC cases in any year follows a binomial distribution.

To simulate SEER incidence rates, we generate a population of individuals from 20 to 100, with an age and sex distribution that matches the SEER 1978 population (to capture risk levels within each age category), who are free from clinically detected CRC. Model-predicted CRC incidence is based on the number of people who develop CRC in the next year.

3.2. *Other published targets.* Table 3 summarizes calibration targets from five studies. To simulate these targets, we generated separate populations for each target that match the age and gender distribution of study participants during the time-period of the study. One study (Church (2004)) describing the pathology of lesions (i.e., adenomas and preclinical cancers) did not provide information about

TABLE 3
*Observed and predicted calibration targets from published studies*

| | | | Posterior Predicted | |
|---|---|---|---|---|
| Target | Mean | Tolerance Interval | Mean | 95% CI |
| Corley et al. (2013) | | | | |
| Adenoma Prevalence, Women 50–54 | 15 | 12.9 to 20.8 | 18.9 | 17.0 to 20.7 |
| Adenoma Prevalence, Women 55–59 | 18 | 15.5 to 25.0 | 22.5 | 20.3 to 24.5 |
| Adenoma Prevalence, Women 60–64 | 22 | 19.4 to 30.1 | 26.1 | 23.7 to 28.2 |
| Adenoma Prevalence, Women 65–69 | 24 | 20.6 to 33.4 | 29.5 | 26.9 to 31.6 |
| Adenoma Prevalence, Women 70–74 | 26 | 21.5 to 37.0 | 32.5 | 29.9 to 34.7 |
| Adenoma Prevalence, Women $\geq$75 | 26 | 20.8 to 37.7 | 35.6 | 32.6 to 37.8 |
| Adenoma Prevalence, Men 50–54 | 25 | 22.1 to 34.2 | 27.7 | 25.0 to 30.5 |
| Adenoma Prevalence, Men 55–59 | 29 | 25.6 to 39.7 | 32.3 | 29.2 to 35.2 |
| Adenoma Prevalence, Men 60–64 | 31 | 27.5 to 42.3 | 36.6 | 33.4 to 39.7 |
| Adenoma Prevalence, Men 65–69 | 34 | 29.6 to 46.9 | 40.6 | 37.1 to 43.8 |
| Adenoma Prevalence, Men 70–74 | 39 | 33.2 to 54.6 | 44.1 | 40.3 to 47.5 |
| Adenoma Prevalence, Men $\geq$75 | 38 | 31.6 to 53.9 | 47.5 | 43.3 to 51.1 |
| Pickhardt et al. (2003)* | | | | |
| Percent of Detected Adenomas $\leq$ 5mm | 62.0 | 55.3 to 68.8 | 63.2 | 59.0 to 66.2 |
| Percent of Detected Adenomas 6–9mm | 28.7 | 22.4 to 35.0 | 24.7 | 22.5 to 28.6 |
| Percent of Detected Adenomas $\geq$ 10mm | 9.2 | 5.2 to 13.2 | 12.0 | 10.7 to 13.1 |
| Imperiale et al. (2000) | | | | |
| Detected Preclinical Cancers per 1000 People | 6.0 | 0.3 to 117.1 | 2.6 | 2.1 to 3.3 |
| Lieberman et al. (2008)* | | | | |
| Preclinical CRCs per 1000 Lesions 6–9mm | 2.5 | 0.0 to 8.4 | 7.5 | 6.2 to 8.3 |
| Preclinical CRCs per 1000 Lesions $\geq$ 10mm | 32.8 | 11.6 to 54.0 | 27.9 | 23.1 to 34.7 |
| Church (2004) | | | | |
| Preclinical CRCs per 1000 Lesions [6, 10)mm | 2.4 | 0.0 to 10.3 | 7.1 | 5.9 to 8.4 |
| Preclinical CRCs per 1000 Lesions $\geq$ 10mm | 42.3 | 12.6 to 72.1 | 16.4 | 12.8 to 22.0 |

*Size was reported categorically as $\leq$ 5mm, 6 to 9mm, and $\geq$ 10mm. We operationalized these categories as: $[1, 5.5)$mm, $[5.5, 9.5)$mm and $\geq$ 9.5mm.

the age or sex of patients, and so we simulated a population that was 50% male with an average age of 65 (standard deviation of five) and an age range of 20 to 90 years.

Simulation of targets in Table 3 also requires simulating the detection of lesions (adenomas and preclinical cancers). Sensitivity is a function of lesions size and is informed by back-to-back colonoscopy studies (Hixson et al. (1990), Rex et al. (1997); additional details provided in Appendix A). We assume that study participants are free from symptomatic (clinically detectable) CRC and have not been screened for CRC prior to the study. This is a reasonable assumption because studies used for model calibration were conducted prior to widespread screening or were based on minimally screened samples. CRC screening guidelines have been in place since the late 1990s (Winawer et al. (1997)), and screening rates have since risen steadily (Centers for Disease Control (2011), Meissner et al. (2006)).

## 4. Posterior inference via incremental mixture approximate Bayesian computation (IMABC).

The basic rejection-based ABC algorithm (Pritchard et al. (1999), Tavare et al. (1997)) generates model parameter vectors $\theta$ from the prior distribution, $\pi(\theta)$, and then uses the model to simulate data, $y^*$. Draws that result in simulated data that are similar to observed data, $y$, are accepted. Similarity between $y^*$ and $y$ is based on user-defined summary statistics, a distance metric and a tolerance level that defines the distance of acceptable points.

In practice, simulating $\theta$ from the prior distribution can be very inefficient because the prior and posterior distributions are often poorly aligned. Many versions of ABC have been developed to address inefficiencies. Two popular variants are ABC-MCMC (Marjoram et al. (2003)) and sequential Monte Carlo ABC (ABC-SMC, Sisson, Fan and Tanaka (2007), Toni et al. (2009)). ABC-MCMC involves proposing a new value of $\theta$ by sampling $u$ from a user-specified jumping distribution, $q(\theta \mid \theta^{(t)})$ that is centered at zero with $\theta^{(t+1)} = \theta^{(t)} + u$. If simulated data based on $\theta^{(t+1)}$ are within tolerance levels for observed data, then, similarly to MCMC, $\theta^{(t+1)}$ is accepted with a probability equal to the minimum of 1 and

$$\frac{q(\theta^{(t+1)} \mid \theta^{(t)})\pi(\theta^{(t+1)})}{q(\theta^{(t)} \mid \theta^{(t+1)})\pi(\theta^{(t)})}.$$

Drawbacks of ABC-MCMC include the usual problems with MCMC, such as correlated samples, low acceptance rates, the possibility of getting stuck in low posterior probability regions and slow mixing requiring simulation of very long chains. ABC-SMC is based on importance sampling with the prior used as the proposal distribution. ABC-SMC starts by simulating a set of draws from the prior distribution. Each subsequent set of draws is simulated by drawing an (importance) weighted sample from the previous set of draws and for each sampled point adding a random deviate $u$ that is drawn from a user-specified jumping distribution. For each sampled point this process is repeated until the perturbed point is accepted (i.e., falls within the tolerance interval). When using the ABC-SMC approach, users specify the total number of iterations, $T$, and a sequence of $T$ increasingly stringent tolerance intervals, which require accepted points to be nearer to targets as the algorithm proceeds. After $T$ iterations, draws from the posterior distribution are simulated by drawing a weighted sample of $\theta$'s using final importance weights that are based on the sequence of jumping distributions. The population Monte Carlo ABC algorithm (ABC-PMC) is closely related to ABC-SMC and also draws on importance sampling (Beaumont et al. (2009), Marin et al. (2012)). ABC-PMC uses a multivariate normal jumping distribution with covariance matrix that is based on prior draws.

In general, ABC and its variants can be impractical or can fail when the parameter space is high dimensional, or there are many summary statistics that the simulated data must approximate (Blum and François (2010)). McKinley et al. (2018) encountered this issue and proposed using a history matching algorithm, which is similar to ABC, to identify regions of the parameter space that produce acceptable matches to data. Implausibility measures are used to sequentially rule out regions of the input space. Here, we propose a new ABC approach that we call incremental mixture approximate Bayesian computation (IMABC), which is well suited to MSM calibration, that involves both high-dimensional parameter spaces and many calibration targets. IMABC is an approximate Bayesian version of adaptive importance sampling, similar to IMIS (Raftery and Bao (2010), Steele, Raftery and Emond (2006)), with samples drawn from the parameter space using a proposal distribution that is a mixture of normal distributions. Posterior estimates are based on accepted draws that are weighted to account for differences between the prior and proposal distributions. IMABC is most similar to the ABC-PMC approach (Beaumont et al. (2009)). IMABC adds new points in regions near a subset of points that produce simulated targets closest to observed targets, whereas ABC-PMC samples points based on an approximation to the joint distribution using importance weights.

4.1. *The IMABC algorithm.* The IMABC algorithm begins with a rejection-sampling ABC step, and updates this initial sample by adding points near a set of "best" points that result in simulated targets that are closest to corresponding observed targets.

Let $O_1, \dots, O_J$ denote the $J$ calibration targets, which we assume are summary statistics. We specify tolerance bounds around targets based on $(1 - \alpha_j) \times 100\%$ confidence intervals, for $j = 1, \dots, J$. Let $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_J)$. The IMABC algorithm updates tolerance intervals so they become more stringent in later iterations. Let $\alpha^{(0)}$ be the alpha levels used for tolerance intervals for the initial ABC step, $\alpha^{(t)}$ are alpha levels for the $t$th iteration and $\alpha^*$ are the final (user-specified) alpha-levels, corresponding to convergence of the IMABC algorithm. When searching a high-dimensional parameter space, it is practical to begin with very wide tolerance intervals, corresponding to small values of $\alpha$. If tolerance intervals are too narrow, there may be very few points that lie within tolerance regions, which is inefficient for exploring the parameter space to identify the regions with high posterior probability. Final alpha levels used to calculate tolerance intervals may vary across targets, depending on the quality of and confidence in calibration targets.

Let $S_{ij}$ denote the $j$th simulated target (corresponding to $O_j$) for the $i$th sampled point, and let $\delta_j(\theta_i, \alpha_j) = 1$ if $S_{ij}$ falls within the $(1 - \alpha_j) \times 100\%$ confidence interval for target $O_j$. We use an intersection criterion for acceptance (Conlan et al. (2012), Ratmann et al. (2014)), where $\theta_i$ is accepted when all $S_{ij}$ lie within ABC tolerance bounds, so that $\delta(\theta_i, \alpha) = \prod_{j=1}^{J} \delta_j(\theta_i, \alpha_j)$.

At the first IMABC step, a sample of $N_0$ points is drawn from the prior distribution of model parameters, $\pi(\theta)$. The algorithm then enters an updating phase. Iteration $(t + 1)$ in the IMABC algorithm proceeds as outlined below:

*Step 1: Identify the best points and sample new points nearby.*

1A. Calculate $p$-values, $\rho_{ij}$, for each accepted $\theta_i$, based on two-sided tests of $\mathrm{H}_0$: $S_{ij} = O_j$ vs. $\mathrm{H}_A$: $S_{ij} \neq O_j$ for $j = 1, \dots, J$, treating $S_{ij}$ as fixed and $O_j$ as estimated with error. Often, as in our application, $O_j$ is a summary statistic and is approximately normally distributed. The calculated $p$-values are comparable across targets as they are on the same scale regardless of the sample size that generated the targets or the distribution of the test statistic, so we summarize model fit across multiple targets with $\rho_{i\cdot} = \min_i(\rho_{ij})$ the worst fit across the $J$ targets.

1B. Select the $N^{(c)}$ points with the largest $\rho_{i\cdot}$. When there are ties, calculate the distance between the simulated and observed targets, $d_{i\cdot} = \sum_j d_{ij}$ where $d_{ij} = (S_{ij} - O_j)^2/O_j^2$, and select points with the largest $\rho_{i\cdot}$ and smallest $d_{i\cdot}$ which are the best fitting points based on the $p$-value criteria and in terms of distance from observed targets.

1C. Simulate $B$ new draws around each of the $\theta_{(k)}^{(t+1)}$, $k = 1, \dots, N^{(c)}$ best points by sampling from a normal distribution with mean $\theta_{(k)}^{(t+1)}$ and covariance $\Sigma_{(k)}^{(t+1)}$.

Let $p$ be the dimension of $\theta$ (i.e., the number of calibrated parameters). If there are fewer than $25p$ accepted points, then $\Sigma_{(k)}^{(t+1)}$ is set to a diagonal covariance matrix with standard deviation set to half the prior distribution standard deviation for each parameter. If there are at least than $25p$ accepted points, $\Sigma_{(k)}^{(t+1)}$ is calculated using the $25p$ accepted points nearest to $\theta_{(k)}^{(t+1)}$. This means that until the algorithm accepts $25p$ points, the same covariance matrix is used for all normal mixtures. This approach allows wider exploration of the parameter space during initial iterations and, after a moderate number of points have been accepted, more directed sampling of new points based on a mixture distribution with a correlation structure computed from a neighborhood of nearby points that are within tolerance intervals.

1D. Simulate calibration targets, $S_{ij}$, for each new draw, and resimulate targets at center points, $\theta_{(k)}^{(t+1)}$. Accept or reject new draws and previously sampled center points based on $\delta(\theta_i, \alpha^{(t)})$. Resimulation of targets at center points enables the algorithm to move away from center points with $S_{ij}$ that are, by chance, similar to $O_i$.

*Step 2: Update tolerance intervals.*

If any $\alpha_j^{(t)} < \alpha_j^*$ and there are $50p$ or more accepted points, check to see if the tolerance can be updated. Identify $i'$ associated with the median $\rho_{i\cdot}$ with $d_{i\cdot}$ as a tie breaker. For each potentially updated tolerance level, set $\alpha_j^{(t+1)} = \min(\rho_{i'j}, \alpha_j^*)$, then update the accepted $\theta$'s, so that they are based on $\delta(\theta_i, \alpha^{(t+1)})$, removing up to half of the previously accepted points that are furthest from the targets.

*Step 3: Evaluate stopping criteria.*

If $\alpha^{(t+1)} = \alpha^*$, calculate sampling weights and the corresponding effective sample size (ESS). Sampling weights account for sampling of points from the normal mixture rather than the prior distribution, $w_i = \pi(\theta_i)/q_t(\theta_i)$. The mixture sampling distribution, $q_t$, is given by $q_t = \frac{N_0}{N_t}\pi + \frac{B}{N_t}\sum_{s=1}^{t}\sum_{k=1}^{N^{(c)}} H_k^{(s)}$ where $H_k^{(s)}$ is the $k$th normal distribution at iteration $s$, given by $N(\theta_{(k)}^{(s)}, \Sigma_{(k)}^{(s)})$, and $N_t = N_0 + N^{(c)}Bt$, the total number of draws through iteration $t$.

ESS provides the expected number of unique points obtained from a weighted random sample of size equal to the number of in range points. The closer ESS is to the number of in-range points, the better our representation of the posterior distribution. ESS for the $N_{(t+1)}$ draws is equal to $(\sum_{i=1}^{N_{(t+1)}} w_i^2)^{-1}$, where $w_i = 0$ if $\delta(\theta_i, \alpha^{(t+1)}) = 0$ (Kish (1965), Liu (2001)). The algorithm stops when ESS $\geq N_{\text{post}}$, having obtained the desired number of draws from the posterior distribution. If $\alpha^{(t+1)} = \alpha^*$ and ESS $< N_{\text{post}}$ the algorithm continues to iterate but without further updates to tolerance intervals.

Once the IMABC algorithm is complete, independent draws from the posterior distribution are simulated by taking a weighted sample from accepted points with replacement, using the $w_i$. Alternatively, posterior means and 95% credible intervals (CIs) can be estimated using weighted means and percentiles based on all accepted draws.

When implementing the IMABC algorithm, we recommend using a large initial sample size, $N_0$, to ensure exploration of the parameter space and because few initially sampled points may lie in high posterior probability regions. The number of normal mixture components used to draw new points at each step, $N^{(c)}$, can be selected to optimize use of computing resources, as new points from each center (mixture component) can be drawn in parallel. The number of centers should be chosen to balance computational constraints and gains, arising from the number of available processors, with the total number of points to be drawn at each iteration, $BN^{(c)}$. The effective sample size of the final set of accepted points, $N_{\text{post}}$, will depend on the planned uses of the calibrated targets. For example, 2000 is a good choice when the goal is to provide interval estimates of model predictions based on percentile intervals, but larger samples may be desired when estimating functions of parameters.

Using IMABC to calibrate an MSM requires multiple model evaluations at each parameter draw, and the user needs to specify $m_j$, the number of simulated agents used to obtain $S_{ij}$. The number of simulated agents may be smaller for common outcomes (such as adenoma prevalence) and larger for rare outcomes (such as cancer incidence). Setting $m_j$ too low will result in too much stochastic variation in

$S_{ij}$ and inaccurate identification of acceptable $\theta_i$. Setting $m_j$ too high will unnecessarily slow the algorithm.

## 5. CRC-SPIN 2.0 calibration results.

5.1. *IMABC implementation.* To calibrate CRC-SPIN 2.0, we used $N_0 = 22{,}000$ with Latin hypercube sampling from the prior distribution to ensure coverage of the parameter space at the initial draw. With the exception of the SEER target, we began with $\alpha^{(0)} = 0.0001$ and worked toward $\alpha^* = 0.001$. For SEER targets we began with $\alpha^{(0)} = 0$, accepting all points regardless of nearness to SEER targets, and worked toward $\alpha^* = 1 \times 10^{-7}$ which results in narrow bands around these registry-based incidence rates. Tolerance intervals are wider for study-derived targets because of the smaller sample sizes. These wider tolerance intervals also reflect the greater uncertainty in these targets due to a range of factors related to their simulation, including uncertainty about population characteristics, sensitivity of lesion detection and lesion size measurement and categorization.

We specified asymmetric tolerance limits for the Corley et al. (2013) target, extending the upper tolerance range by adding $0.25O_j$ to the upper tolerance limit because of uncertainty about prior screening of the study sample. The Corley et al. (2013) study is based on insured patients who underwent colonoscopies from 1/1/2006 to 12/31/2008. Although Corley et al. (2013) excluded exams from individuals with evidence of prior screening, the study occurred during the CRC screening era and did not restrict their sample to continuously enrolled individuals and so could not completely identify individuals who had a previous colonoscopy. Therefore, the adenoma prevalence estimates from Corley et al. (2013) may be lower than expected in an unscreened population.

To take advantage of high performance computing and parallel processing (Appendix B), we used $N^{(c)} = 10$, drawing $B = 1000$ points from each normal mixture so that 10,000 new points were evaluated at each updating iteration. We assumed a normal distribution for sample statistics when estimating $(1 - \alpha) \times 100\%$ confidence intervals and $p$-values. We set the final effective sample size, $N_{\text{post}}$, to 5000.

When simulating target data, we used $m_j$ equal to $5 \times 10^4$ for Pickhardt et al. (2003); $2 \times 10^5$ for Corley et al. (2013) and Imperiale et al. (2000); $3 \times 10^5$ for Church (2004), $5 \times 10^5$ for Lieberman et al. (2008) and $5 \times 10^6$ for the SEER registry data. To improve efficiency of the IMABC algorithm, we sequentially calculated $S_{ij}$ for each new $\theta_i$ in Step 1 of the algorithm, working from targets that are least to most computationally intensive. After calculating each target, we evaluated $\delta_j(\theta_i, \alpha_j)$ and once $\delta_j(\theta_i, \alpha_j) = 0$; the point is rejected without simulating the remaining, more computationally intensive, targets.

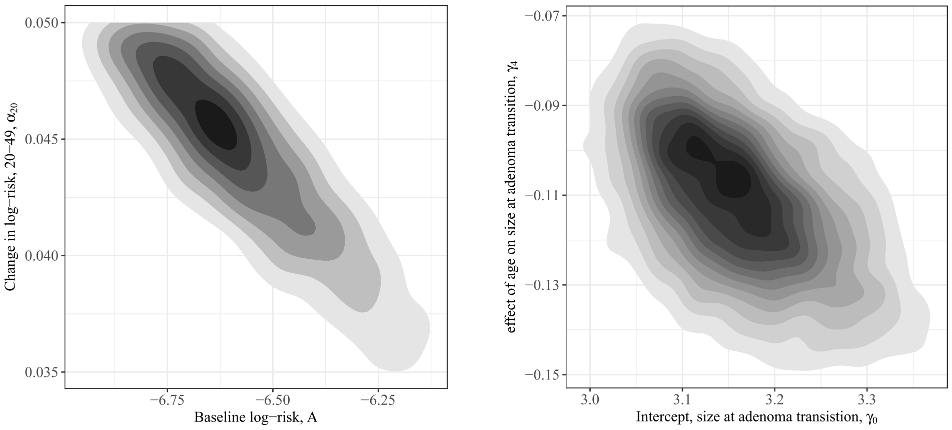
Both the IMABC algorithm and the CRC-SPIN 2.0 model were implemented in the R programming language (R Core Team (2014)). They were coupled to produce an integrated, dynamic, high-performance computing workflow with the use of the Extreme-scale Model Exploration with Swift (EMEWS) framework (Ozik et al. (2016)). Further details about the computing environment are provided in Appendix B.

5.2. *Posterior estimates.* The IMABC algorithm completed 16 iterations, obtaining 5815 parameter draws within tolerance limits with an effective sample size of 5582 draws from the joint posterior distribution. Sampling weights ranged from $9.9 \times 10^{-5}$ to $5.7 \times 10^{-4}$ with a mean and median of $1.7 \times 10^{-4}$.

Posterior estimated means and 95% CIs of model parameters were based on weighted means and percentiles of accepted draws from the joint posterior distribution (shown in Table 1). We estimated that adenoma risk is higher for men than women, increases with age and increases more rapidly at younger (than older) ages. Parameters that govern the time for an adenoma to reach 10mm were tightly estimated with the exception of $\beta_{1R}$. The shape of the CRC-SPIN 2.0 adenoma growth curve is determined by the parameter $p$. Because the posterior mean of $p$ is less than 1.0, this means that CRC-SPIN 2.0 simulates adenomas that initially grow slowly, and this allows the model to simulate a relatively large fraction of small adenomas. Consistent with prior ranges placed on growth parameters, the model predicted that 0.2% of adenomas in the colon reach 10mm within 10 years (95% CI 0.1% to 0.8%) and 4.0% of adenomas in the rectum reach 10mm within 10 years (95% CI 0.2% to 19.7%). The predicted percent of adenomas reaching 10mm within 20 years rises to 8.6% (95% CI 5.9% to 12.1%) of adenomas in the colon and 66.4% (95% CI 38.1% to 86.9%) of adenomas in the rectum.

We estimated that adenomas transition to preclinical cancer at smaller sizes for women, for adenomas in the rectum and for adenomas initiated at older ages. The quadratic age effect slows the decrease in size at transition at older ages. The gender effect on size at transition was stronger for adenomas in the colon than for adenomas in the rectum. There was considerable variability in posterior estimates of sojourn time parameters, indicating little information in targets about these parameters. Estimated mean sojourn times (in years), which are functions of sojourn time parameters, were 2.32 with 95% CI 2.05 to 2.74 for preclinical cancers in the colon and 2.12 with 95% CI 1.62 to 2.95 for preclinical cancers in the rectum.

Sojourn time estimates largely reflect prior distributions. Sojourn time is informed by screening studies, and our targets include a single imprecise screening study (Imperiale et al. (2000)). We explored the impact of reducing the width of the tolerance interval around this target by increasing the alpha-level to 0.05. The resulting tolerance interval was 3.2 to 8.9 preclinical cancers detected per 1000 screened. We found that, although 2486 parameter draws met this criteria (with an ESS equal to 2395), there was little change in posterior mean parameter estimates (data not shown).



(a) Baseline log-risk ($A$) and change in risk ages 20-59 ($\alpha_{20}$)

(b) Transition parameters, intercept ($\gamma_0$) and age effect, $\gamma_4$)

FIG. 1. *Joint posterior distribution of model parameters associated with adenoma risk, and the growth and sojourn time in the colon.*

By simulating draws from the posterior distribution, we were able to examine correlations and relationships among model parameters. For example, Figure 1 displays the bivariate posterior distributions of baseline log-adenoma risk ($A$) and the annual increase in risk between the ages of 20 and 50 years ($\alpha_{20}$). When baseline adenoma risk is lower, risk increases more rapidly from 20 to 50 years to accurately predict observed adenoma prevalence which largely is based on prevalence after age 50 when guidelines recommend initiation of CRC screening (correlation is $-0.87$). Similarly, the intercept term for size at adenoma transition to preclinical cancer, $\gamma_0$, is negatively correlated (correlation is $-0.52$) with the effect of age on size at adenoma transition $\gamma_4$. Larger values of $\gamma_0$ imply that, on average, adenomas transition at larger sizes, and the model compensates for this with a larger decrease in the size at transition with increasing age.

The posterior predicted means of SEER targets were near point estimates of observed rates and posterior 95% CIs include SEER targets (Table 2). Posterior 95% CIs do not always include point estimates of targets (Table 3). For example, the model predicted higher adenoma prevalence than observed by Corley et al. (2013), especially at older ages, acknowledging the possibility that prior screening had occurred in the Corley et al. (2013) population which could explain why our model does not predict these targets accurately. The model also predicted a larger number of adenomas $\geq$10mm than observed by Pickhardt et al. (2003). The probability of detecting preclinical cancer came from three studies (Church (2004), Imperiale et al. (2000), Lieberman et al. (2008)), and the accuracy of model predictions demonstrates how the IMABC calibration approach combines information across targets.

**6. Discussion.** We addressed the problem of calibrating microsimulation models by developing IMABC, an ABC algorithm based on the ideas of incremental mixture importance sampling (IMIS) (Raftery and Bao (2010), Steele, Raftery and Emond (2006)), an adaptive Sampling Importance Resampling algorithm (SIR; Rubin (1987)). We illustrate our approach by calibrating CRC-SPIN 2.0, an MSM for colorectal cancer, a problem that involves a relatively high dimensional parameter space and multiple targets.

Like IMIS, the IMABC algorithm iteratively updates the proposal distribution at each iteration to obtain samples from regions of the parameter space that are consistent with calibration targets. The resulting mixture of normal distributions with locally adaptive covariance matrices is a very flexible distribution, and the algorithm can sample from a distribution that is multimodal to better approximate the posterior distribution. We used a new approach to select tolerance levels, based on $(1-\alpha)\%$ confidence intervals, with $\alpha$-levels updated using $p$-values associated with a test for equality of the simulated and observed targets. This approach implicitly incorporates the precision of calibration targets. IMABC also provides a straightforward approach to tuning these tolerance intervals, requiring users to specify only the initial and final alpha values, whereas ABC-SMC requires prespecification of the sequence of tolerance intervals. The IMABC approach could also be implemented by specifying initial and final tolerance interval widths (distances) for each target, with an updating scheme based on scaled distances (as opposed to $p$-values) between simulated and observed targets.

Other advantages of IMABC include clear stopping rules based on the effective sample size, the ability to specify which targets are most important through final tolerance intervals and the ability to take advantage of parallelized code. The IMABC approach was able to provide posterior samples of parameters that include strong correlation among parameters. This is an important feature of the algorithm because models for the natural history of disease are often complex, and it can be difficult to justify simpler models unless they are scientifically plausible.

A limitation of the IMABC algorithm, especially as applied to MSM calibration, is that IMABC can be computationally demanding. Evaluation of a very large number of points may be necessary, and calibration targets must be simulated for each point. The computational expense can be reduced through the ordering of target evaluations and ceasing evaluation of a point when the first set of targets fails to fall within tolerance bounds. We implemented IMABC as a dynamic high-performance computing (HPC) workflow via the EMEWS framework (Ozik et al. (2016)). While the HPC environment was advantageous for development of the IMABC approach, we found that it was not ultimately necessary for its application.

The application of ABC to MSM calibration differs in important ways from typical ABC applications used to analyze a single dataset. The data used to calibrate an MSM (calibration targets) are summary statistics drawn from published studies.

A goal of MSM calibration is to combine information across these studies. To calibrate the MSM, we require matching across multiple targets, but inconsistencies among targets are possible. In addition, while the behavior of the ABC estimator in a typical application depends on user-defined data summaries, the behavior of the estimator used for MSM calibration will also depend on how well the targets represent and inform the simulated disease process. For example, in our application we found that sojourn time parameters were not well informed by our calibration targets.

The IMABC approach worked well for calibration of CRC-SPIN 2.0, finding points within a high dimensional parameter space that result in a good match to multiple calibration targets. In typical applications the ABC estimator of the posterior mean has an asymptotic normal distribution (Li and Fearnhead (2018)) though coverage of CIs depends on both the large sample behavior of the summary statistics used by the algorithm and tolerance levels (Frazier et al. (2018)); smaller tolerance levels (i.e., those that produce model predictions nearer to summary statistics) are needed to accurately estimate credible intervals at nominal confidence levels than are needed to obtain accurate point estimates for model parameters. This suggests that IMABC can be used to estimate the posterior means of model parameters. However, further research is needed to determine CI interval coverage from MSM model calibration. Future work will use smaller scale models to carry out such exploration. We also plan to release publicly available code to allow others to work with IMABC. In addition, because calibration requires simulation of a large number of model evaluations, each with a large number of agents, we plan to explore ways to improve the efficiency of IMABC model calibration. Finally, we plan to examine efficient approaches to parameter updating when new targets become available (such as screening data to inform CRC-SPIN sojourn time parameters), and sequential calibration approaches that can be used to efficiently build from simpler to more complex models.

## APPENDIX A: CRC-SPIN 2.0: ADDITIONAL MODEL INFORMATION

This appendix provides information about the CRC-SPIN 2.0 model that that may be useful for understanding the model but is not essential to understanding the calibration approach. Complete model description can be found on the `cancer.cisnet.gov` (National Cancer Institute (2018)) in the section describing model profiles.

CRC-SPIN 2.0 is an update to the validated CRC-SPIN 1.0 model (Rutter and Savarino (2010), Rutter et al. (2016)). Therefore, we were able to base prior distributions for many model parameters on the results from CRC-SPIN 1.0 calibration (Rutter, Miglioretti and Savarino (2009)).

**A.1. Adenoma risk model.** Once adenomas are initiated, they are assigned a location. The distribution of adenomas throughout the large intestine follows a multinomial distribution based on data from nine autopsy studies (Blatt (1961), Bombi (1988), Chapman (1963), Eide and Stalsberg (1978), Johannsen, Momsen and Jacobsen (1989), Rickert et al. (1979), Stemmermann and Yatani (1973), Szczepanski, Urban and Wierzchowski (1992), Williams, Balasooriya and Day (1982)). The probabilities associated with six sites in the large intestine (from distal to proximal) are: $\mathsf{Pr}(\text{rectum}) = 0.09$; $\mathsf{Pr}(\text{sigmoid colon}) = 0.24$; $\mathsf{Pr}(\text{descending colon}) = 0.12$; $\mathsf{Pr}(\text{transverse colon}) = 0.24$; $\mathsf{Pr}(\text{ascending colon}) = 0.23$; and $\mathsf{Pr}(\text{cecum}) = 0.08$. For many purposes it is important to distinguish between colon and rectal locations; more detailed location information is sometimes used for determining screening test accuracy.

**A.2. Adenoma growth model.** We parameterized the growth model in terms of the time it takes for the adenoma diameter to reach 10mm ($t_{10\text{mm}}$) to improve our ability to relate adenoma growth to observable data and clinical knowledge. The mean and median time to reach 10mm are $\beta_2\Gamma(1 - 1/\beta_1)$ and $\beta_2 \ln(2)^{-1/\beta_1}$, respectively, where $\Gamma(\cdot)$ is the gamma function.

CRC-SPIN uses a function describing the adenoma growth trajectory to determine adenoma size at any point in time which is needed to determine the time at transition to preclinical cancer and size-dependent sensitivity of screening tests. CRC-SPIN 2.0 generalizes the growth model used by CRC-SPIN 1.0 by calibrating the shape parameter, $p$, in equation (3). CRC-SPIN 1.0 specified $p \equiv 1$, corresponding to the negative exponential model, but this resulted in relatively fast early adenoma growth and resulted in prediction of relatively few small adenomas. The prior distribution for $p$ is centered at 1, with a prior range that is limited to clinically plausible values.

**A.3. Model for transition from adenoma to preclinical invasive cancer.** The CRC-SPIN 2.0 model for adenoma transition is a reparameterized and simplified version of the CRC-SPIN 1.0 model for adenoma transition, restated as a regression model to better evaluate differences based on agent and adenoma characteristics. The mean and median size at adenoma transition to preclinical cancer are $\exp(\mu_\gamma + 0.5\sigma_\gamma^2)$ and $\exp(\mu_\gamma)$, respectively, where $\mu$ is given by equation (4). The variance in the size at transition is $(\exp(\sigma_\gamma^2) - 1)) \exp(2\mu_\gamma + \sigma_\gamma^2)$.

Prior distributions of adenoma transition parameters are based on estimated posterior distributions from CRC-SPIN 1.0 calibration when parameters are functionally related. For other parameters (i.e., $\gamma_5$ and $\sigma_\gamma$) we used uniform prior distributions with wide, but plausible, ranges. The minimum prior value for $\sigma_\gamma$ was set to 0.5 to ensure a minimum amount of between-adenoma variability in the size at adenoma transition, which is a characteristic that may not be well informed by calibration targets, yet can impact the effectiveness of CRC screening.

**A.4. Model for sojourn time.** CRC-SPIN 2.0 uses a Weibull distribution to model sojourn times for preclinical cancers in the colon with scale parameter $\lambda_1$ and shape parameter $\lambda_2$, as shown in equation (5). Under the proportional hazards model sojourn time for preclinical cancers in the rectum also have a Weibull distribution, with scale parameter $\lambda_{1,\text{rectum}} = \lambda_1 \exp(\lambda_3)^{-1/\lambda_2}$ and shape parameter $\lambda_2$. Under this model mean sojourn time is $\lambda_1 \Gamma(1 + 1/\lambda_2)$ for cancers in the colon and $\lambda_{1,\text{rectum}} \Gamma(1 + 1/\lambda_2)$ for cancers in the rectum, where $\Gamma(\cdot)$ is and the gamma function. Prior distributions allow sojourn time to range from 2.0 to 3.9 years for cancers in the colon and from 1.2 to 6.2 years for cancers in the rectum. This range for mean sojourn time is consistent with published estimates (e.g., the TAMACS study (Chen et al. (1999)) reported an estimated mean sojourn time of 2.85 years with a 95% confidence interval 2.15 to 4.30 years) and findings from a validation study (Rutter et al. (2016)) which suggested that mean sojourn times are in the range of two to four years.

**A.5. Simulation of lifespan and colorectal cancer survival.** The CRC-SPIN 2.0 model first simulates the stage at clinical detection given sex and age at detection, and then simulates size at detection conditional on stage. (In contrast, the CRC-SPIN 1.0 model simulated size, and then stage conditional on size.) The CRC-specific probability of survival after diagnosis is calculated given agent sex, cancer location (colon versus rectum), age at diagnosis and year of diagnosis using an approach described by Hakulinen (1977).

**A.6. Simulated screening.** Colonoscopy sensitivity for adenoma and preclinical CRC detection is based on a quadratic function of lesion size ($s$) that was successfully used in the CRC-SPIN 1.0 model. For adenomas we assume $P(\text{miss}|\text{size} = s \le 15\text{mm}) = 0.34 - 0.0349s + 0.0009s^2$, $P(\text{miss}|\text{size} = 15 < s \le 30\text{mm}) = 0.01$, $P(\text{miss}|\text{size} = 30 < s \le 40\text{mm}) = 0.005$ and $P(\text{miss}|\text{size} = s \ge 40\text{mm}) = 0.001$. This function results in sensitivity that is consistent with a observed findings from the 1990s (Hixson et al. (1990), Rex et al. (1997)). Sensitivity is 0.76 for a 3mm adenoma, 0.87 for a 7.5mm adenoma and 0.95 for a 12mm adenoma. For preclinical cancers we assume sensitivity that is the maximum of 0.95 and sensitivity based on adenoma size, so that colonoscopy sensitivity is 0.95 for preclinical cancers 12mm or smaller, and sensitivity is greater than 0.95 for preclinical cancers larger than 12mm.

Participants in the Pickhardt et al. (2003) study underwent both computed tomography colonography (CTC) and colonoscopy for the purposes of evaluating the accuracy of CTC, primarily for adenomas 6mm and larger. To simulate receipt of both CTC and colonoscopy, we assume that the probability of missing an adenoma on CTC, given that it was missed on colonoscopy, is equal to $P(\text{miss}|\text{size} = s)^{0.25}$ to build in correlation in the two tests due to adenoma characteristics.

## APPENDIX B: PROGRAMMING AND COMPUTING ENVIRONMENT

We utilized the EMEWS framework (Ozik et al. (2016)) to implement a dynamic HPC workflow controlled by the IMABC algorithm. EMEWS is free and open source code that is available at https://emews.github.io. EMEWS, built on the general-purpose parallel scripting language Swift/T (Wozniak et al. (2013)), allows for the direct integration of multilanguage software components (in this case IMABC and CRC-SPIN 2.0) and can be used on computing resources ranging from desktops and campus clusters to supercomputers. The resulting IMABC EMEWS workflow is driven directly by the IMABC R source code, obviating the need for porting the code to alternate programming languages or platforms for the sole purpose of running large-scale computational experiments. An IMABC R-package is under development. Collaborators interested in working with a preliminary version of the code should contact corresponding author.

The experiments were performed on the Cray XE6 Beagle at the University of Chicago, hosted at Argonne National Laboratory. Beagle has 728 nodes, each with two AMD Operton 6300 processors, each having 16 cores, for a total of 32 cores per node; the system thus has 23,296 cores in all. Each node has 64 GB of RAM. Experiments were also run on the Midway2 cluster at the University of Chicago Research Computing Center. Midway2 is a hybrid cluster and includes both CPU and GPU resources. For this work the CPU resources were used, consisting of 370 nodes of Intel E5-2680v4 processors, each with 28 cores and 64 GB of RAM. Swift/T, the underlying EMEWS workflow engine, allows for the abstraction of resource specific settings (e.g., scheduler type and compute layouts) for a variety of target computing resources. Thus, once the IMABC EMEWS workflow was developed, it could be run on both the Beagle and Midway2 clusters with only minimal configuration modifications.

The experiment reported here used 80 nodes on Beagle with four worker processes per node (to account for the memory footprint of CRC-SPIN 2.0) for a total of 320 worker processes, each of which could concurrently execute an individual model run. The total compute time was 29.4 hours or 2352 node-hours.

## REFERENCES

BEAUMONT, M. A., CORNUET, J.-M., MARIN, J.-M. and ROBERT, C. P. (2009). Adaptive approximate Bayesian computation. *Biometrika* **96** 983–990. MR2767283

BLATT, L. J. (1961). Polyps of the colon and rectum: Incidence and distribution. *Dis. Colon Rectum* **4** 277–282.

BLUM, M. G. B. and FRANÇOIS, O. (2010). Non-linear regression models for approximate Bayesian computation. *Stat. Comput.* **20** 63–73. MR2578077

BOMBI, J. A. (1988). Polyps of the colon in Barcelona, Spain. *Cancer* **61** 1472–1476.

CHAPMAN, I. (1963). Adenomatous polypi of large intestine: Incidence and distribution. *Ann. Surg.* **157** 223–226.

CHEN, T. H. H., YEN, M. F., LAI, M. S., KOONG, S. L., WANG, C. Y., WONG, J. M., PREVOST, T. C. and DUFFY, S. W. (1999). Evaluation of a selective screening for colorectal carcinoma: The Taiwan multicenter cancer screening (TAMACS) project. *Cancer* **86** 1116–1128.

CHURCH, J. M. (2004). Clinical significance of small colorectal polyps. *Dis. Colon Rectum* **47** 481–485.

CONLAN, A. J. K., MCKINLEY, T. J., KAROLEMEAS, K., POLLOCK, E. B., GOODCHILD, A. V., MITCHELL, A. P., BIRCH, C. P. D., CLIFTON-HADLEY, R. S. and WOOD, J. L. N. (2012). Estimating the hidden burden of bovine tuberculosis in Great Britain. *PLoS Comput. Biol.* **8** e1002730. MR3005930

CORLEY, D. A., JENSEN, C. D., MARKS, A. R., ZHAO, W. K., DE BOER, J., LEVIN, T. R., DOUBENI, C., FIREMAN, B. H. and QUESENBERRY, C. P. (2013). Variation of adenoma prevalence by age, sex, race, and colon location in a large population: Implications for screening and quality programs. *Clin. Gastroenterol. Hepatol.* **11** 172–180.

DE KONING, H. J., MEZA, R., PLEVRITIS, S. K., TEN HAAF, K., MUNSHI, V. N., JEON, J., ERDOGAN, S. A., KONG, C. Y., HAN, S. S. et al. (2014). Benefits and harms of computed tomography lung cancer screening strategies: A comparative modeling study for the US Preventive Services Task Force. *Ann. Intern. Med.* **160** 311–320.

EIDE, T. J. and STALSBERG, H. (1978). Polyps of the large intestine in Northern Norway. *Cancer* **42** 2839–2848.

CENTERS FOR DISEASE CONTROL (2011). Vital signs: Colorectal cancer screening, incidence, and mortality—United States, 2002–2010. *Morb. Mortal. Wkly. Rep.* **60** 884–889.

FRAZIER, D. T., MARTIN, G. M., ROBERT, C. P. and ROUSSEAU, J. (2018). Asymptotic properties of approximate Bayesian computation. *Biometrika* **105** 593–607. MR3842887

HAKULINEN, T. (1977). On long-term relative survival rates. *J. Clin. Epidemiol.* **30** 431–443.

HIXSON, L., FENNERTY, M., SAMPLINER, R., MCGEE, D. and GAREWAL, H. (1990). Prospective study of the frequency and size distribution of polyps missed by colonoscopy. *J. Natl. Cancer Inst.* **82** 1769–1772.

IMPERIALE, T. F., WAGNER, D. R., LIN, C. Y., LARKIN, G. N., ROGGE, J. D. and RANSOHOFF, D. F. (2000). Risk of advanced proximal neoplasms in asymptomatic adults according to the distal colorectal findings. *N. Engl. J. Med.* **343** 169–174.

JOHANNSEN, L. G. K., MOMSEN, O. and JACOBSEN, N. O. (1989). Polyps of the large intestine in Aarhus, Demark. An autopsy study. *Cancer* **24** 799–806.

KIM, J. J., BURGER, E. A., REGAN, C. and SY, S. (2017). Screening for cervical cancer in primary care: a decision analysis for the U.S. Preventive Services Task Force. Technical Report. Agency for Healthcare Research and Quality. Available at https://www.uspreventiveservicestaskforce.org/Home/GetFileByID/3287, Contract No. HHSA-290-2012-00015-I. Last accessed on February 1, 2019.

KISH, L. (1965). *Survey Sampling*. Wiley, New York, USA.

KNUDSEN, A. B., ZAUBER, A. G., RUTTER, C. M., NABER, S. K., DORIA-ROSE, V. P., PABINIAK, C., JOHANSON, C., FISCHER, S. E., LANSDORP-VOGELAAR, I. et al. (2016). Estimation of benefits, burden, and harms of colorectal cancer screening strategies: Modeling study for the US Preventive Services Task Force. *J. Am. Med. Dir. Assoc.* **315** 2595–2609.

KOH, K.-J., LIN, L.-H., HUANG, S.-H. and WONG, J.-U. (2015). CARE—pediatric colon adenocarcinoma: A case report and literature review comparing differences in clinical features between children and adult patients. *Medicine* (*Baltim. Md.*) **94** e503.

KONG, C. Y., MCMAHON, P. M. and GAZELLE, G. S. (2009). Calibration of disease simulation model using an engineering approach. *Value Health* **12** 521–529.

LESLIE, A., CAREY, F. A., PRATT, N. R. and STEELE, R. J. C. (2002). The colorectal adenoma–carcinoma sequence. *Br. J. Surg.* **89** 845–860.

LI, W. and FEARNHEAD, P. (2018). On the asymptotic efficiency of approximate Bayesian computation estimators. *Biometrika* **105** 285–299. MR3804403

LIEBERMAN, D., MORAVEC, M., HOLUB, J., MICHAELS, L. and EISEN, G. (2008). Polyp size and advanced histology in patients undergoing colonoscopy screening: Implications for CT colonography. *Gastroenterology* **135** 1100–1105.

LIU, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. *Springer Series in Statistics*. Springer, New York. MR1842342

MANDELBLATT, J. S., STOUT, N. K., SCHECHTER, C. B., VAN DEN BROEK, J. J., MIGLIORETTI, D. L., KRAPCHO, M., TRENTHAM-DIETZ, A., MUNOZ, D., LEE, S. J. et al. (2016). Collaborative modeling of the benefits and harms associated with different US breast cancer screening strategies. *Ann*. *Intern*. *Med*. **164** 215–225.

MARIN, J.-M., PUDLO, P., ROBERT, C. P. and RYDER, R. J. (2012). Approximate Bayesian computational methods. *Stat*. *Comput*. **22** 1167–1180. MR2992292

MARJORAM, P., MOLITOR, J., PLAGNOL, V. and SIMON, T. (2003). Markov chain Monte Carlo without likelihoods. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **100** 15324–15328.

MCKINLEY, T. J., VERNON, I., ANDRIANAKIS, I., MCCREESH, N., OAKLEY, J. E., NSUBUGA, R. N., GOLDSTEIN, M. and WHITE, R. G. (2018). Approximate Bayesian computation and simulation-based inference for complex stochastic epidemic models. *Statist*. *Sci*. **33** 4–18. MR3757500

MEISSNER, H. I., BREEN, N., KLABUNDE, C. N. and VERNON, S. W. (2006). Patterns of colorectal cancer screening uptake among men and women in the United States. *Cancer Epidemiol*. *Biomark*. *Prev*. **15** 389–394.

MUTO, T., BUSSEY, H. J. R. and MORSON, B. C. (1975). The evolution of cancer in the colon and rectum. *Cancer* **36** 2251–2270.

NATIONAL CANCER INSTITUTE (2004). Surveillance, epidemiology, and end results (SEER) program. Available at http://www.seer.cancer.gov, SEER Stat Database: Incidence—SEER 9 Regs Public-Use, Nov 2003 Sub (1973-2001), released April 2004, based on the November 2003 submission.

NATIONAL CANCER INSTITUTE (2018). Cancer INtervention and Surveillance Modeling Network (CISNET). Available at https://cisnet.cancer.gov. Last accessed on February 1, 2019.

NATIONAL CENTER FOR HEALTH STATISTICS (2000). US Life Tables.

NELDER, J. A. and MEAD, R. (1965). A simplex method for function minimization. *Comput*. *J*. **7** 308–313. MR3363409

OZIK, J., COLLIER, N. T., WOZNIAK, J. M. and SPAGNUOLO, C. (2016). From desktop to large-scale model exploration with Swift/T. In *Proceedings of the* 2016 *Winter Simulation Conference* (*WSC*) 206–220. IEEE Press.

PICKHARDT, P. J., CHOI, R., HWANG, I., BUTLER, J. A., PUCKETT, M. L., HILDEBRANDT, H. A., WONG, R. K., NUGENT, P. A., MYSLIWIEC, P. A. et al. (2003). Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. *N*. *Engl*. *J*. *Med*. **349** 2191–2200.

PONUGOTI, P. L. and REX, D. K. (2017). Yield of a second screening colonoscopy 10 years after an initial negative examination in average-risk individuals. *Gastroint*. *Endosc*. **85** 221–224.

PRITCHARD, J. K., SEIELSTAD, M. T., PEREZ-LEZAUN, A. and FELDMAN, M. W. (1999). Population growth of human Y chromosomes: A study of Y chromosome microsatellites. *Mol*. *Biol*. *Evol*. **16** 1791–1798.

R CORE TEAM (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at http://www.R-project.org/. Last accessed on February 1, 2019.

RAFTERY, A. E. and BAO, L. (2010). Estimating and projecting trends in HIV/AIDS generalized epidemics using incremental mixture importance sampling. *Biometrics* **66** 1162–1173. MR2758504

RATMANN, O., CAMACHO, A., MEIJER, A. and DONKER, G. (2014). Statistical modelling of summary values leads to accurate approximate Bayesian computations. Available at arXiv:1305.4283.

REX, D., CUTLER, C., LEMMEL, G., RAHMANI, E., CLARK, D., HELPER, D., LEHMAN, G. and MARK, D. (1997). Colonoscopic miss rates of adenomas determined by back-to-back colonoscopies. *Gastroenterology* **112** 24–28.

RICKERT, R. R., AUERBACH, O., GARFINKEL, L., HAMMOND, E. C. and FRASCA, J. M. (1979). Adenomatous lesions of the large bowel. An autopsy survey. *Cancer* **43** 1847–1857.

RUBIN, D. (1987). The calculation of posterior distributions by data augmentation: Comment: A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The SIR algorithm. *J. Amer. Statist. Assoc.* **82** 543–546.

RUTTER, C. M., MIGLIORETTI, D. L. and SAVARINO, J. E. (2009). Bayesian calibration of microsimulation models. *J. Amer. Statist. Assoc.* **104** 1338–1350. MR2750568

RUTTER, C. M. and SAVARINO, J. E. (2010). An evidence-based microsimulation model for colorectal cancer: Validation and application. *Cancer Epidemiol. Biomark. Prev.* 1055–9965.

RUTTER, C. M., JOHNSON, E. A., FEUER, E. J., KNUDSEN, A. B., KUNTZ, K. M. and SCHRAG, D. (2013). Secular trends in colon and rectal cancer relative survival. *J. Natl. Cancer Inst.* **105** 1806–1813.

RUTTER, C. M., KNUDSEN, A. B., MARSH, T. L., DORIA-ROSE, V. P., JOHNSON, E., PABINIAK, C., KUNTZ, K. M., VAN BALLEGOOIJEN, M., ZAUBER, A. G. et al. (2016). Validation of models used to inform colorectal cancer screening guidelines: Accuracy and implications. *Med. Decis. Mak.* **36** 604–614.

SISSON, S. A., FAN, Y. and BEAUMONT, M. A. (2019). Overview of ABC. In *Handbook of Approximate Bayesian Computation*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 3–54. CRC Press, Boca Raton, FL. MR3889278

SISSON, S. A., FAN, Y. and TANAKA, M. M. (2007). Sequential Monte Carlo without likelihoods. *Proc. Natl. Acad. Sci. USA* **104** 1760–1765. MR2301870

STEELE, R. J., RAFTERY, A. E. and EMOND, M. J. (2006). Computing normalizing constants for finite mixture models via incremental mixture importance sampling (IMIS). *J. Comput. Graph. Statist.* **15** 712–734. MR2291269

STEMMERMANN, G. N. and YATANI, R. (1973). Diverticulosis and polyps of the large intestine. A necropsy study of Hawaii Japanese. *Cancer* **31** 1260–1270.

SZCZEPANSKI, W., URBAN, A. and WIERZCHOWSKI, W. (1992). Colorectal polyps in autopsy material. Part I. Adenomatous polyps. *Patol. Pol.* **43** 79–85.

TAVARE, S., BALDING, D., GRIFFITHS, R. and DONNELLY, P. (1997). Inferring coalescence times from DNA sequence data. *Genetics* **145** 505–518.

TJØRVE, E. and TJØRVE, K. M. (2010). A unified approach to the Richards-model family for use in growth analyses: Why we need only two model forms. *J. Theoret. Biol.* **267** 417–425.

TONI, T., WELCH, D., STRELKOWA, N. and STUMPF, M. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface* **6** 187–202.

WILLIAMS, A. R., BALASOORIYA, B. A. W. and DAY, D. W. (1982). Polyps and cancer of the large bowel: A necropsy study in Liverpool. *Gut* **23** 835–842.

WINAWER, S. J., FLETCHER, R. H., MILLER, L., GODLEE, F., STOLAR, M., MULROW, C., WOOLF, S., GLICK, S., GANIATS, T. et al. (1997). Colorectal cancer screening: Clinical guidelines and rationale. *Gastroenterology* **112** 594–642.

WOZNIAK, J. M., ARMSTRONG, T. G., WILDE, M., KATZ, D. S., LUSK, E. and FOSTER, I. T. (2013). Swift/T: Large-scale application composition via distributed-memory dataflow processing. In 2013 13*th IEEE/ACM International Symposium on Cluster*, *Cloud*, *and Grid Computing* 95–102. IEEE.

ZAUBER, A. G., KNUDSEN, A. B., RUTTER, C. M., LANSDORP-VOGELAAR, I., SAVARINO, J. E., VAN BALLEGOOIJEN, M. and KUNTZ, K. M. (2009). Cost-Effectiveness of CT colonography to screen for colorectal cancer: Report to the Agency for Healthcare Research and Quality from the Cancer Intervention and Surveillance Modeling Network (CISNET) for MISCAN, SimCRC, and CRC-SPIN Models. Technical Report. Available at https://www.cms.gov/

medicare-coverage-database/details/technology-assessments-details.aspx?TAId=58, Project ID: CTCC0608, Last accessed on February 1, 2019.

C. M. RUTTER
M. DEYOREO
RAND CORPORATION
1776 MAIN STREET
SANTA MONICA, CALIFORNIA 90401
USA
E-MAIL: crutter@rand.org
mdeyoreo@rand.org

J. OZIK
N. COLLIER
ARGONNE NATIONAL LABORATORY
BUILDING 221
9700 SOUTH CASS AVENUE
ARGONNE, ILLINOIS 60439
USA
E-MAIL: jozik@anl.gov
ncollier@anl.gov