Bayesian Calibration of Microsimulation Models

Carolyn M. RUTTER, Diana L. MIGLIORETTI, and James E. SAVARINO

Microsimulation models that describe disease processes synthesize information from multiple sources and can be used to estimate the effects of screening and treatment on cancer incidence and mortality at a population level. These models are characterized by simulation of individual event histories for an idealized population of interest. Microsimulation models are complex and invariably include parameters that are not well informed by existing data. Therefore, a key component of model development is the choice of parameter values. Microsimulation model parameter values are selected to reproduce expected or known results though the process of model calibration. Calibration may be done by perturbing model parameters one at a time or by using a search algorithm. As an alternative, we propose a Bayesian method to calibrate microsimulation models that uses Markov chain Monte Carlo. We show that this approach converges to the target distribution and use a simulation study to demonstrate its finite-sample performance. Although computationally intensive, this approach has several advantages over previously proposed methods, including the use of statistical criteria to select parameter values, simultaneous calibration of multiple parameters to multiple data sources, incorporation of information via prior distributions, description of parameter identifiability, and the ability to obtain interval estimates of model parameters. We develop a microsimulation model for colorectal cancer and use our proposed method to calibrate model parameters. The microsimulation model provides a good fit to the calibration data. We find evidence that some parameters are identified primarily through prior distributions. Our results underscore the need to incorporate multiple sources of variability (i.e., due to calibration data, unknown parameters, and estimated parameters and predicted values) when calibrating and applying microsimulation models.

KEY WORDS: Colorectal cancer; Identifiability; Markov chain Monte Carlo.

1. INTRODUCTION

Microsimulation models (MSMs) provide a way to estimate population-level effects of medical interventions on health outcomes by integrating results from randomized controlled trials, observational studies, and expert opinion. MSMs have been developed for a wide range of diseases, including prostate cancer (Etzioni et al. 1999; Feuer et al. 2004), breast cancer (Berry et al. 2005), and colorectal cancer (Loeve et al. 1999). MSMs, which have the potential to influence policy making decisions, have been used to estimate the impact of overdiagnosis due to screening on breast cancer incidence (de Koning et al. 2006), to compare the effect of risk factors and screening behaviors on colorectal cancer (CRC) mortality rates (Vogelaar et al. 2006), to evaluate the cost-effectiveness of cervical cancer (van den Akker-van Marle et al. 2002) and breast cancer screening (Shen and Parmigiani 2005), and to examine mortality and reoperation rates after aortic valve replacement (Puvimanasinghe et al. 2004). The National Cancer Institute recognized the value of microsimulation models when it formed the Cancer Intervention and Surveillance Modeling Network (CISNET; *cisnet.cancer.gov*) (Croyle 2006).

MSMs are characterized by simulation of individual event histories for an idealized population of interest. These individual event histories catalog landmarks in the disease process, such as the development of an incident cancer or onset of myocardial infarction. Simulation of event histories requires mathematical models for key components of the disease process. In general, these processes are not directly observable, although their outcomes may be. For example, the process of developing CRC cannot be observed, but the prevalence of both precancerous lesions (adenomas) and preclinical CRC can be estimated from screening trials.

MSM model calibration involves selecting parameter values that are consistent with observed data and expected findings. Once parameters are selected, MSMs can be used to make predictions about hypothetical interventions and/or future trends in population disease outcomes. The simplest calibration method involves perturbing parameters one at a time and subjectively judging the goodness of fit to calibration data (Ramsey et al. 2000). More recently, chi-squared (Ness et al. 2000) and deviance (Loeve et al. 2004) statistics have been used to evaluate how closely MSM estimates match calibration data at different points in the parameter space. Statistical criteria for MSM goodness of fit also have been combined with search algorithms to determine the best parameters. Salomon et al. (2002) used simulation-based estimates of likelihoods in combination with a sequential search algorithm to determine the best-fitting model parameters. Chia et al. (2004) demonstrated the use of three maximum likelihood methods for estimating parameters associated with a complex two-parameter model describing breast cancer tumor growth, including grid search using the Nelder–Mead algorithm (Nelder and Mead 1965); the Kiefer–Wolfowitz algorithm (Kiefer and Wolfowitz 1952), based on likelihood differences estimated by simulating the likelihood; and the Robbins–Monro algorithm (Robbins and Monro 1951), based on directly simulating the likelihood gradient.

Here we propose a Bayesian method for calibrating MSM parameters that parallels Bayesian sampling-based estimation approaches. We place prior distributions on all parameters and use Markov chain Monte Carlo (MCMC) to estimate parameters using data from multiple sources. The Bayesian approach allows us to describe an appropriately complex model for disease

Carolyn M. Rutter is Senior Investigator, Group Health Center for Health Studies, Seattle, WA 98101 (E-mail: *rutter.c@ghc.org*) and Affiliate Professor, Departments of Biostatistics and Health Services, University of Washington, WA 98195. Diana L. Miglioretti is Senior Investigator, Group Health Center for Health Studies, Seattle, WA 98101 and Affiliate Associate Professor, Department of Biostatistics, University of Washington, WA 98195. James E. Savarino is Programmer, Group Health Center for Health Studies, Seattle, WA 98101. This work was supported by National Cancer Institute grant U01 CA97427 (*http://cisnet.cancer.gov*).

Journal of the American Statistical Association
Accepted for publication
DOI: 10.1198/jasa.2009.ap07466

processes while incorporating information from both prior distributions and observed data. By using informative prior distributions, we can include model parameters that otherwise might not be identifiable based on available data. This approach acknowledges uncertainty in parameters as opposed to fixing selected parameters to obtain the identifiability of a potentially inferior model (Gustafson 2005). We suggest using model diagnostics to examine parameter identifiability. Finally, our approach allows point and interval estimation of both model parameters and functions of these parameters, including posterior predictive estimates that can be used to assess model fit.

In the next section we briefly describe our MSM for the natural history of CRC. In Section 3 we describe our calibration approach, examine its relationship to Bayesian analysis, and present its asymptotic characteristics. In Section 4 we describe a simulation study used to examine the behavior of our MCMC calibration with finite sample sizes. We provide calibration results for our CRC MSM in Section 5 and concluding remarks in Section 6.

2. MICROSIMULATION MODEL FOR THE NATURAL HISTORY OF COLORECTAL CANCER

CRC is the second leading cause of cancer death in the United States (American Cancer Society 2008; Jemal et al. 2008). CRC mortality rates have declined since 1980 (Wingo et al. 2003; Stewart et al. 2004; American Cancer Society 2008); and CRC incidence declined between 1985 and 1995 and then stabilized (Howe et al. 2001, 2006). The reasons behind these trends are unclear. Clinical trials have shown that screening for CRC by fecal occult blood tests can reduce CRC mortality (Hardcastle et al. 1996; Kronborg et al. 1996; Towler et al. 1998); however, a 1999 study found low rates of CRC screening, with fewer than half of Americans over age 50 reporting ever being screened (CDC 2001). The possibility that lifestyle changes have reduced the population's risk for CRC holds promise for primary prevention. The use of MSMs allows estimation of the effects of screening and population risk on changes in CRC incidence and mortality, offering new insights into factors associated with observed trends. In this article, we develop and calibrate an MSM for CRC.

Our MSM for the natural history of CRC is based on the adenoma–carcinoma sequence (Muto, Bussey, and Morson 1975; Leslie et al. 2002) and assumes that all CRCs arise from an adenoma. Four model components describe the natural history of CRC: adenoma risk, adenoma growth, transition from adenoma to preclinical cancer, and transition from preclinical to clinical cancer. Table 1 summarizes each component's parameters and prior distributions.

2.1 Adenoma Risk Model

We modeled the occurrence of adenomas using a nonhomogeneous Poisson process with a piecewise age effect. The ith individual's instantaneous risk of an adenoma at time t is given by

$$\psi_i(t) = \exp\left(\alpha_{0i} + \alpha_1 sex_i + \sum_{k=1}^{4} \delta(A_k < age_i(t) \le A_{k+1}) \times \left\{ age_i(t)\alpha_{2k} + \sum_{j=2}^{k} A_j(\alpha_{2\,j-1} - \alpha_{2j}) \right\}\right), \quad (2.1)$$

where α_{0i} describes an individual's baseline risk, α_1 describes the difference in risk for men ($sex_i = -1$) versus women ($sex_i = +1$), α_{2k} describes changes in risk with age (in years) in the kth interval, and $\delta(\cdot)$ is an indicator function with $\delta(x) = 1$ when x is true and $\delta(x) = 0$ otherwise. We assumed that before age 20, individuals are not at risk of developing adenomas and allowed for the following age-risk intervals: $A_1 = 20, A_2 = 50, A_3 = 60, A_4 = 70$, and $A_5 = 120$.

The prior distributions for α_0, α_1, α_{2k}, and α_{3k}, $k \in \{1, 2, 3, 4\}$, were based on results from a Bayesian meta-analysis of autopsy data that modeled a log-linear age effect that remained constant across age and sex groups (Rutter, Miglioretti, and Yu 2007). The meta-analysis did not provide information about σ_α, so we used a minimally informative uniform prior distribution for this parameter.

Once adenomas were initiated, we assigned their location (colon vs. rectum). We specified $P(\text{rectum}) = 0.09$ based on analysis of data from 9 autopsy studies (Blatt 1961; Chapman 1963; Stemmermann and Yatani 1973; Eide and Stalsberg 1978; Rickert et al. 1979; Williams, Balasooriya, and Day 1982; Bombi 1988; Johannsen, Momsen, and Jacobsen 1989; Szczepanski, Urban, and Wierzchowski 1992a) and one colonoscopy study (Szczepanski, Urban, and Wierzchowski 1992b).

2.2 Adenoma Growth Model

We modeled adenoma growth using an extension to the Janoschek growth curve model (Janoschek 1957; Gille and Salomon 2000), $d_{ij}(t) = d_\infty - (d_\infty - d_0)e^{-\lambda_{ij}t}$, where $d_{ij}(t)$ is the diameter of the jth adenoma in the ith individual at time t after initiation, d_∞ is the maximum possible adenoma diameter, d_0 is the minimum possible adenoma diameter, and λ_{ij} is the growth rate for the jth adenoma within the ith individual. This model is asymmetric, with early exponential growth that slows to allow an asymptote at d_∞.

To improve our ability to obtain prior information from clinicians, we reparameterized the growth model in terms of the time needed for the adenoma to grow to 10 mm in diameter, $t_{10\text{ mm}} = -\frac{1}{\lambda}\ln(\frac{d_\infty - 10}{d_\infty - d_0})$. We assumed that $t_{10\text{ mm}}$ has a type 2 extreme value distribution with cumulative distribution function given by

$$F(t) = \exp\left[-\left(\frac{t}{\beta_1}\right)^{-\beta_2}\right] \quad (2.2)$$

for $t \ge 0$. Because the type 2 extreme value distribution has a long right tail, with skewness that can persist as the mean moves away from 0, increasing the proportion of slow-growing adenomas does not also require increasing the proportion of fast-growing adenomas. The parameter β_1 roughly determines the location of the distribution, and β_2 corresponds to the scale. We modeled separate growth distributions for colon and rectal adenomas, with parameters (β_{1c}, β_{2c}) and (β_{1r}, β_{2r}), respectively. We chose relatively uninformative prior distributions for the β's. Under these priors, the median time to 10 mm ranged from 1.1 to 144.3 years.

Table 1. Summary of CRC MSM components

Component	Prior distribution	Posterior estimates		Estimated overlap
		Mean	95% CI	
Adenoma risk [eqn. (2.1)]				
Baseline log-risk	$\alpha_0 \sim \text{N}(-6.7, 0.27)$	−6.6	(−7.5, −5.9)	0.79
Main sex effect	$\alpha_1 \sim \text{N}(-0.3, 0.04)$	−0.24	(−0.35, −0.14)	0.53
Standard deviation, baseline log-risk	$\sigma_\alpha, \sim \text{U}(0.10, 3.0)$	1.1	(0.35, 1.6)	0.46
Age effect, age ∈ [20, 50)	$\alpha_{20} \sim \text{TN}_0(0.03, 0.003)$	0.037	(0.024, 0.049)	0.42
Age effect, age ∈ [50, 60)	$\alpha_{21} \sim \text{TN}_0(0.03, 0.003)$	0.031	(0.011, 0.047)	0.51
Age effect, age ∈ [60, 70)	$\alpha_{22} \sim \text{TN}_0(0.03, 0.003)$	0.029	(0.011, 0.047)	0.50
Age effect, age ≥ 70	$\alpha_{23} \sim \text{TN}_0(0.03, 0.003)$	0.030	(0.011, 0.049)	0.51
Time to 10 mm [eqn. (2.2)]				
Colon: Location	$\beta_{1c} \sim \text{U}(1, 100)$	28.6	(24.3, 34.2)	0.14
Scale	$\beta_{2c} \sim \text{U}(1, 4)$	2.7	(1.1, 3.9)	0.89
Rectum: Location	$\beta_{1r} \sim \text{U}(1, 100)$	10.3	(5.9, 13.8)	0.11
Scale	$\beta_{2r} \sim \text{U}(1, 4)$	2.7	(1.4, 3.9)	0.83
Transition to cancer [eqn. (2.3)]				
Men				
Colon, size	$\gamma_{1cm} \sim \text{U}(0.02, 0.05)$	0.045	(0.040, 0.049)	0.37
Colon, age at initiation	$\gamma_{2cm} \sim \text{U}(0.0, 0.02)$	0.008	(0.002, 0.016)	0.64
Rectum, size	$\gamma_{1rm} \sim \text{U}(0.02, 0.05)$	0.035	(0.021, 0.049)	0.90
Rectum, age at initiation	$\gamma_{2rm} \sim \text{U}(0.0, 0.02)$	0.010	(0.001, 0.019)	0.91
Women				
Colon, size	$\gamma_{1cf} \sim \text{U}(0.02, 0.05)$	0.048	(0.044, 0.050)	0.23
Colon, age at initiation	$\gamma_{2cf} \sim \text{U}(0.0, 0.02)$	0.005	(0.000, 0.013)	0.58
Rectum, size	$\gamma_{1rf} \sim \text{U}(0.02, 0.05)$	0.043	(0.030, 0.050)	0.57
Rectum, age at initiation	$\gamma_{2rf} \sim \text{U}(0.0, 0.02)$	0.015	(0.008, 0.019)	0.56
Mean sojourn time [Lognormal($\mu, \tau\mu$)]				
Colon	$\mu_c \sim \text{U}(0.5, 5.0)$	1.9	(1.0, 3.9)	0.55
	$\tau_c \sim \text{U}(0.1, 1.5)$	0.80	(0.15, 1.4)	0.89
Rectum	$\mu_r \sim \text{U}(0.5, 5.0)$	2.7	(1.1, 4.7)	0.80
	$\tau_r \sim \text{U}(0.1, 1.5)$	0.84	(0.15, 1.4)	0.90

NOTE: Shown are calibrated parameters associated with the 4 components of the natural history model: including parameter notation, associated equations, prior distributions and posterior estimates (mean and 95% credible interval). $\text{N}(\mu, \sigma)$ denotes a Normal distribution with mean μ and standard deviation σ. $\text{TN}_0(\mu, sigma)$ denotes a truncated Normal distribution with mean μ and standard deviation *sigma*, based on a Normal distribution restricted to range over $(0, \infty)$. $\text{U}(a, b)$ denotes a Uniform distribution over (a, b). The estimated overlap statistic is based on $\int \min(g(\theta), h(\theta|y))\, d\theta$ (Garret and Zeger 2000).

2.3 Model for Transition From Adenoma to Preclinical Invasive Cancer

Our adenoma transition model and priors for associated parameters are informed by an analysis of colorectal polyp registry data reported by Nusko et al. (1997) and a recent study of follow-up colonoscopy that provides evidence suggesting that the probability of transition depends on an individual's age at the time of adenoma onset (Yamaji et al. 2006). We use a lognormal cumulative distribution function to describe the cumulative transition probability as a function of sex, size, and age at adenoma onset. For an adenoma initiated at age a in the colon of a man, the probability of transition to preclinical cancer at or before size s is given by the lognormal cumulative distribution,

$$\xi_c(s, a) = \Phi\big(\{\ln(\gamma_{1cm} s) + \gamma_{2cm}(a - 50)\}/\gamma_3\big), \qquad (2.3)$$

where $\Phi(\cdot)$ is the standard Normal cumulative distribution function.

Cumulative transition probabilities for adenomas in the male rectum have the same form and are parameterized by γ_{1rm}, γ_{2rm}, and γ_3. Similarly, the probability of adenoma transition in women is based on eq. (2.3), parameterized by γ_{1cf}, γ_{2cf}, and γ_3 for adenomas in the colon and γ_{1r}, γ_{2rf}, and γ_3 for adenomas in the rectum. The γ_1 parameters rescale adenoma size and are associated with the magnitude of the survival function. The γ_2 parameters allow the probability of transition to increase with age at adenoma onset. The γ_3 parameter is associated with the standard deviation of the lognormal distribution and determines the shape of the cumulative probability curve (i.e., how rapidly the probability of transition increases with size) and is closely associated with the probability of transition in small (≤10 mm) adenomas.

Because we lacked sufficient calibration data to estimate γ_3, we set $\gamma_3 = 0.5$ to restrict the transition probability for small (≤10 mm) adenomas, based on expert opinion that such small adenomas are unlikely to undergo malignant transition. Priors on the remaining transition parameters impose clinically reasonable restrictions on the range of cumulative transition probabilities; for example, cumulative transition probabilities of a 10-mm adenoma can range from 0.0006 to 0.08 for a 50-year-old and up to 0.28 for a 70-year-old with the maximum age ef-

fect. Cumulative transition probabilities for a 20-mm adenoma can range from 0.03 to 0.50 at age 50 up to 0.79 at age 70.

2.4 Model for Time From Preclinical to Clinical Cancer (Sojourn Time)

We define sojourn time as the time from the onset of preclinical cancer to clinical detection. For individuals with multiple cancers, each sojourn time is assumed to be independent, and cancer survival is dependent on the earliest clinically detected cancer. Sojourn time is modeled with a lognormal distribution that depends on whether an adenoma is located in the colon or in the rectum. Let t represent sojourn time. For adenomas in the colon, $\log(t) \sim \text{Normal}(\xi_c, \nu_c)$, where ξ_c and ν_c denote the mean and standard deviation of $\log(t)$, and $t \sim$ lognormal with mean μ_c and standard deviation $\tau_c\mu_c$, where $\mu_c = \exp(\xi_c + \frac{1}{2}\nu_c^2)$ and $\tau_c = \sqrt{\exp(\nu_c^2) - 1}$. Sojourn time for adenomas in the rectum is modeled by a lognormal distribution with mean μ_r and standard deviation $\tau_r\mu_r$. Prior distributions for mean sojourn time are based on data from the TAMACS study Chen et al. (1999), which reported an estimated mean sojourn time of 2.85 years with a 95% confidence interval (CI) of 2.15–4.30. Under our prior distributions for sojourn time parameters (μ_c, τ_c, μ_r, τ_r), sojourn time has a median ranging from 0.3 to 5.0 years, a 5th percentile ranging from 0 to 4.2 years, and a 95th percentile ranging from 0.6 to 16.5 years.

Once a cancer becomes clinically detectable, we simulate size and stage at clinical detection and survival. We specify an overall (unconditional) distribution for tumor size at clinical detection and a conditional distribution of stage given size, using observed SEER data describing size at detection in 1975–1979 (Surveillance, Epidemiology, and End Results (SEER) Program (*www.seer.cancer.gov*) 2004). Little or no CRC screening was done during this time period. We simulate survival time based on a Cox proportional hazards model for relative survival. Using SEER data on cancer survival for cases diagnosed in 1975–1979, we estimate multiple proportional hazard models using the CANSURV program (*http://srab.cancer.gov/cansurv/*). These proportional hazards models are stratified by location (colon or rectum) and AJCC stage, with age and sex included as covariates. Finally, we model other-cause mortality using survival probabilities based on product-limit estimates for age and birth year cohorts from the National Center for Health Statistics Databases (US Life Tables 2000).

2.5 Calibration Data

Table 2 presents calibration data from 4 studies based on the evaluation of minimally screened asymptomatic individuals. These data include the prevalence of adenomas and preclinical cancers in individuals undergoing screening colonoscopy (Strul et al. 2006), the size of the largest adenoma found in veterans participating in a study of screening colonoscopy (Lieberman et al. 2000), the number and size of adenomas found in individuals participating in a study comparing virtual and optical colonoscopy (Pickhardt et al. 2003), and the prevalence of adenomas found in individuals participating in a workplace study of screening colonoscopy (Imperiale et al. 2000).

Data from 2 clinical series inform our adenoma transition model (Table 3). These studies focused on adenomas and did not include individual characteristics, such as age and sex. The first series, reported by Church (2004), describes the pathology of 5,722 adenomas removed between January 1995 and September 2002 in a single endoscopist's practice. The second series, reported by Odom et al. (2005), describes the pathology of 3,225 adenomas removed between January 1999 and December 2003, excluding those obtained from bowel resection and CRC not associated with a polyp. We calibrate to preclinical cancer rates in adenomas ≥ 5 mm, because the rates of preclinical cancer are near 0 in smaller adenomas (0.05% in the Church data and 0.03% in the Odom data), which causes problems with embedded simulations (as described in the next section).

Table 2. Calibration data from studies reporting individual-level outcomes

Prevalence of adenomas and preclinical cancers (Strul et al. 2006) ($m = 50{,}000$)					
Age range	Mean age (SD)	% male	Sample size		Prevalence
40–49	49.2 (3.0)	49.2%	183		0.10
50–75	61.2 (7.6)	47.5%	917		0.22
76–80	77.7 (1.4)	37.7%	77		0.29
Size of largest adenoma given at least 1 adenoma (Lieberman et al. 2000) ($m = 10{,}000$)					Proportion of 1,411
Age range	Mean age (SD)	% male	Sample size		adenomas ≥ 10 mm
50–75	61.2 (7.3)	96.8%	3121		0.23
Overall number and size distribution of adenomas (Pickhardt et al. 2003) ($m = 10{,}000$)					
Age range	Mean age (SD)	% male	Sample size		Size distribution of 554 adenomas
40–79	57.8	59%	1233	< 6 mm:	0.620
				[6, 10) mm:	0.287
				≥ 10 mm:	0.092
Number of preclinical cancers (Imperiale et al. 2000) ($m = 50{,}000$)					
Age range	Mean age (SD)	% male	Sample size		Prevalence
50+	59.8 (8.3)	59%	1994		0.0035

NOTE: Age is given in years. m indicates the number of embedded simulations used to estimate the data parameters for each Metropolis–Hastings step.

Table 3. Calibration studies reporting adenoma transition to preclinical disease

Source	Sample size	Preclinical cancer rate by adenoma size		
Church (2004)		6–10 mm	>10 mm	
	1,341	1/666 (0.15%)	21/675 (3.1%)	
Odom et al. (2005)		6–10 mm	11–20 mm	>20 mm
	374	1/152 (0.65%)	3/191 (1.6%)	6/31 (19.4%)

NOTE: Data parameters associated with multinomial distributions are based on $m = 200{,}000$ embedded simulations.

SEER colon and rectal cancer incidence rates in 1975–1979 are a key calibration component (Table 7, Section 5.2). The incidence rates reported are per 100,000 individuals. These rates are based on the first observed invasive colon or rectal cancer during 1975–1979, the last period before the dissemination of CRC screening tests.

3. BAYESIAN MICROSIMULATION MODEL CALIBRATION

We propose a MCMC calibration (MCMCC) approach that is similar to MCMC estimation methods for Bayesian analysis (Gelman et al. 1995). Parametric data analysis has two basic components: data, y, and a data distribution that is conditional on some unknown parameter vector, $f(Y|\theta)$. Bayesian models add a third component, the prior distribution of the unknown parameters, $\theta \sim \pi(\theta)$.

In the context of MSM calibration, θ represents MSM parameters, and prior distributions are specified for these θ. But the distribution of the calibration data, $y \sim f(Y|g(\theta))$, is parameterized by an unknown function of MSM parameters, $g(\theta)$.

3.1 Markov Chain Monte Carlo Calibration

Let $M(\theta)$ denote a MSM parameterized by the vector θ. We assume that the MSM is composed of K separate components and that the corresponding parameter vectors are distinct and a priori mutually independent. Let $M_i(\theta_i)$ denote the ith model component and let $\pi_i(\theta_i)$ denote the prior distribution associated with θ_i, where $\theta' = (\theta_1', \theta_2', \ldots, \theta_K')$.

Let $\mathbf{Y} = \{y_1, \ldots, y_N\}$ denote the full set of calibration data from N independent sources. We assume that each y_j has a known distributional form, $y_j|\theta \sim f_j(g_j(\theta))$. In general, calibration data are statistics from published studies that can be described using distributions for count and categorical data, such as the binomial, multinomial, and Poisson distributions.

We calibrate θ by simulating draws from the posterior distribution of θ given Y, $h(\theta|Y)$, using MCMC with Metropolis–Hastings (MH) steps within Gibbs iterations. Gibbs iterations are based on the distinct parameter sets associated with each of the K model components. One complete iteration of the algorithm involves sequentially drawing candidate values (θ_i^*) and accepting or rejecting these candidates for each of the K components (Smith and Roberts 1993). We use a random-permutation sweep approach to improve the mixing of the sampler (Roberts and Sahu 1997). Within Gibbs steps, we use a random-walk MH algorithm to approximately sample from the full conditional distributions, $h(\theta_i|\theta_{(-i)}, Y)$ (Tierney 1994), where $\theta_{(-i)}$ denotes the parameter vector excluding the ith component.

For a symmetric jumping distribution, such as the Normal centered at the current value, the MH algorithm accepts θ_i^* based on the transition probability $\alpha(\theta_i, \theta_i^*)$, where

$$\alpha(\theta_i, \theta_i^*) = \begin{cases} \min(r_i(\theta_i, \theta_i^*), 1) & \text{if } \pi_i(\theta_i)\prod_{j=1}^{N} f_j(y_j|\theta) > 0 \\ 1 & \text{if } \pi_i(\theta_i)\prod_{j=1}^{N} f_j(y_j|\theta) = 0 \end{cases}$$

and

$$r_i(\theta_i, \theta_i^*) = \frac{h(\theta_i^*, \theta_{(-i)}|Y)}{h(\theta|Y)} = \frac{\pi_i(\theta_i^*)\prod_{j=1}^{N} f_j(y_j|\theta_i^*, \theta_{(-i)})}{\pi_i(\theta_i)\prod_{j=1}^{N} f_j(y_j|\theta_i)}.$$

In most data analyses, calculating $r_i(\theta_i, \theta_i^*)$ is straightforward, because both the priors, $\pi(\theta)$, and the data distributions, $f(y|\theta)$, are known and have closed form. In the case of MSM calibration, $r_i(\theta_i, \theta_i^*)$ cannot be calculated directly, because $g_j(\theta)$ are unknown functions of θ. For example, when calibrating to observed cancer incidence, the associated probability of incident clinical cancer contains information about the full disease process, described by multiple MSM parameters, from the risk of an adenoma through the transition to clinically detected cancer, and the exact functional form of this relationship is unknown.

We propose an approximate MH algorithm that includes an embedded simulation to estimate $g(\theta)$. We assume that the MSM, $M(\theta)$, and data distributions, $f(y|g(\theta))$, are specified correctly and use $M(\theta)$ to simulate m draws from $f_j(y_j|g_j(\theta))$: $\{\tilde{y}_{j1}, \ldots, \tilde{y}_{jm}\}$. Using the simulated sample, we calculate the maximum likelihood estimate of $g_j(\theta)$, denoted by $\hat{g}_j(\theta)$; for example, for Poisson and binomial distributions, we estimate $\hat{g}_j(\theta) = \frac{1}{m}\sum \tilde{y}_{ji}$. To simulate draws from $h(\theta|\mathbf{Y})$, we use the Metropolis-within-Gibbs steps described earlier, substituting $f_j(y_j|\hat{g}_j(\theta))$ for $f_j(y_j|g_j(\theta))$. The resulting transition probability function, $\hat{\alpha}(\theta, \theta^*)$, is based on $\hat{r}(\theta_i, \theta_i^*)$, where

$$\hat{r}_i(\theta_i, \theta_i^*) = \frac{\pi_i(\theta_i^*)\prod_{j=1}^{N} f_j(y_j|\hat{g}_j(\theta_i^*, \theta_{(-i)}))}{\pi_i(\theta_i)\prod_{j=1}^{N} f_j(y_j|\hat{g}_j(\theta_i))}.$$

To maintain ergodicity of the chain, we obtain independent estimates of $\hat{g}(\theta)$, with a new embedded simulation for both candidate and current parameter values at each iteration. Heuristically, this repeated simulation is required because $f(y|\hat{g}(\theta))$ may be large by chance. The MH algorithm tends to accept candidate values associated with large estimated likelihoods, including both those with large true likelihoods and those with small true likelihoods that by chance are estimated

to be large. Without repeated simulation of $g(\theta)$ at both candidate and current values of θ, the sampler can get stuck where the true likelihood is small but an estimated likelihood is large. In the Appendix we show that this approach, based on an approximate jumping rule, converges to the target posterior distribution as m increases.

4. SIMULATION STUDY

We performed a simulation study to explore the characteristics of MSM parameter estimates based on MCMCC with finite m and fixed chain length, focusing on the effects of the number of embedded simulations (m) and the number of estimated MSM parameters on bias and credible interval length and coverage. Our simulation study was based on a relatively simple hypothetical MSM describing transitions among 4 disease states with 6 potentially unknown parameters.

4.1 Hypothetical Microstimulation Model

Under the hypothetical MSM, individuals begin in a disease-free state (S_0) and may progressively transition through 3 disease-related states (S_1, S_2, and S_3), with transition to a death state (S_d) possible from any of these 4 states. We select true parameter values so that the hypothetical model produces prevalence patterns that are similar to those seen for cancers, with a precursor state (S_1) that is relatively common and preclinical and clinical cancer outcomes (S_2 and S_3) that are rare. Let t_k indicate the transition time from S_{k-1} to S_k, and let t_d indicate the transition time from S_0 to S_d. We assume that $t_1 \sim$ exponential(λ_1), where $\log(\lambda_1) \sim \mathrm{N}(\Lambda_1, \sigma_1)$. Given a transition to S_1, individuals have the potential to transition to S_2 with probability p_2. For individuals with this potential, we assume $t_2 \sim$ exponential(λ_2) and $t_3 \sim$ lognormal with mean μ_3 and standard deviation $\tau_3\mu_3$. Finally, $t_d \sim$ Gumbel(λ_d, σ_d). Although the time scale is arbitrary, we refer to units of time as years and to a simulated individual's time since initiation as his or her age. True values for model parameters are set to $\Lambda_1 = 4$, $\sigma_1 = 0.5$, $p_2 = 0.05$, $\lambda_2 = 10$, $\mu_3 = 3$, $\sigma_3 = 0.5$, $\lambda_d = 65$, and $\sigma_d = 15$. Associated population characteristics based on this hypothetical MSM, calculated for 10 million simulated individuals, are shown in Table 4. Under the hypothetical true MSM, membership in S_1 is relatively common and prevalence increases with age, but membership in S_2 and transition into S_3 are rare. In this sense, the hypothetical MSM is similar to CRC, in which adenomas are relatively common and prevalence increases with age, but preclinical cancer and transition to clinically detectable cancer are rare.

Table 4. Estimated population characteristics for the hypothetical 3-state MSM, based on 10,000,000 simulated individuals

Time, T	$P(\in S_1)$	$P(\in S_2)$	$P(\sum_{i=1}^{3} t_i \in [T, T+10))$	$P(t_D > T)$
10	0.18	0.0014	0.059	1.00
20	0.31	0.0018	0.064	1.00
30	0.42	0.0016	0.057	0.99
40	0.51	0.0014	0.049	0.99
50	0.57	0.0011	0.043	0.93
60	0.63	0.0010	0.038	0.75
70	0.68	0.0008	0.033	0.51
80	0.71	0.0006	0.029	0.31

NOTE: For $i \in (1, 2, 3)$, t_i is time from S_{i-1} to S_i, and t_D is the time to S_D, the death state.

4.2 Simulated Calibration Data

For each run of our simulation study, we used the hypothetical MSM model with the true MSM parameter values to generate 3 calibration data sets. Calibration set 1 describes the number of individuals in S_1 at age 10, 20, 30, 40, and 50 years, with 250 individuals at each age. Calibration set 2 describes the number of individuals in S_2 at age 10, 20, and 50 years, based on 5,000 individuals at each age. Calibration set 3 describes the proportion of individuals (per life year) who transition to S_3 within 20-year age intervals, $[0, 20)$, $[20, 40)$, $[40, 60)$, and $[60, 80)$, based on a cohort of 10,000 individuals.

A subset of simulations examines the effect of additional data on parameter estimates by adding a fourth calibration data set. Calibration set 4 provides the categorical distributions of t_2 and t_3 across 5 intervals (t_2: [0, 2), [2, 5), [5, 10), [10, 15), and $\geq$15; t_3: [0, 1), [1, 2), [2, 3), [3, 4), and $\geq$4) for 500 individuals transitioning from S_1 to S_2 and from S_2 to S_3, respectively.

4.3 Simulation Study Design

We examined 4 estimation scenarios that vary in terms of the number of MSM parameters estimated and the calibration sets used for estimation. Scenarios 1, 2, and 3 all use calibration sets 1, 2, and 3, differing only in terms of the number of MSM parameters estimated. Scenario 1 examines estimation of Λ_1 and σ_1. Scenario 2 examines estimation of Λ_1, σ_1, p_2, and σ_2. Scenario 3 examines estimation of all 6 model parameters. Scenario 4 is identical to scenario 3, but with the addition of calibration data set 4.

We used different embedded simulation sizes to estimate the data parameters associated with each calibration data set. For calibration set 1, we simulated m_1 individuals who are distributed equally across the 5 corresponding calibration ages. For calibration set 2, we simulated m_2 individuals who are distributed equally across the 3 corresponding calibration ages. For calibration set 3, we simulated a cohort of m_3 individuals. For calibration set 4, we fixed the embedded simulation size, simulating 5,000 individuals undergoing each transition (S_1 to S_2 and S_2 to S_3). We examined 3 combinations of embedded simulation sizes: small ($m_1 = 6{,}250$, $m_2 = 75{,}000$, $m_3 = 50{,}000$), medium ($m_1 = 12{,}500$, $m_2 = 150{,}000$, $m_3 = 100{,}000$), and large ($m_1 = 25{,}000$, $m_2 = 300{,}000$, $m_3 = 200{,}000$).

We simulated MCMCC estimation for scenario 1 (estimation of S_1 parameters) using all 3 embedded simulation sizes, for scenarios 2 and 3 using medium and large embedded simulation sizes, and for scenario 4 using medium embedded simulation sizes.

4.4 Implementation of Markov Chain Monte Carlo Calibration

In this simulation study, the MSM was correctly specified, and we set unestimated MSM parameters to their true values. We specified uniform priors for all estimated parameters. We chose the prior mean to be twice the true value, so that (prior mean $-$ true mean)/true mean equals 1 for all parameters and 100% bias corresponds to an average estimate of posterior mean that is equal to the prior mean. We selected starting values to be 3 times the true value, so that true value $<$ prior mean $<$ starting value for all parameters. Table 5 gives the true values

Table 5. Hypothetical MSM parameters: true values, starting values used for MCMC calibration, and the mean and range of associated prior distributions

Parameter	True value	Prior mean	Starting value	Prior range
Λ_1	4	8	12	[0.1, 16.1]
σ_1	0.5	1	1.5	[0.0001, 2.0001]
p_2	0.05	0.1	0.15	[0.0001, 0.2001]
λ_2	10	20	30	[0.1, 40.1]
μ_3	3	6	9	[1, 13]
τ_3	0.5	1	1.5	[0.001, 2.001]

of hypothetical MSM parameters, their prior ranges and means, and starting values used for MCMC calibration.

The parameter estimates were based on a single chain with 32,500 iterations: 5,000 iterations for burn-in, the next 2,500 iterations to estimate the between-parameter covariance (based on every fifth iteration), and every fifth iteration from the last 25,000 to estimate parameters. We took the same approach when calibrating our MSM for CRC, using initial Gibbs steps to estimate the correlation matrix of estimated parameters and subsequent steps based on sampling the full parameter vector using this correlation matrix to direct the random-walk steps. The parameters were estimated by the mean across simulated posterior draws, and the 95% CIs were estimated based on the upper and lower 2.5th percentiles across draws.

4.5 Simulation Results

For each combination of scenario and embedded sample size, we simulated 1,000 runs. For each run, we simulated calibration data and used MCMCC to estimate the parameters. Using these 1,000 estimates, we calculated the percent bias (i.e., bias as a percentage of the true parameter value), the length of the 95% CI, and the coverage of the 95% CI. We assessed the identifiability of model parameters by comparing prior and posterior distributions using an estimated overlap measure proposed by Garrett and Zeger (2000), $\int \min(\pi(\theta), h(\theta|y))\, d\theta$. This measure of overlap ranges from 0 to 1, with values closer to 1 corresponding to greater similarity between the specified prior and estimated posterior distributions.

Table 6 gives results from the simulation study. Increasing embedded simulation sizes (m) reduced the percent bias and decreased the length of the 95% CIs, although the changes tended to be small. Under scenario 1, decreases in percent bias were greatest for the initial increase from small to medium m; another doubling of m (from medium to large) resulted in only minor changes in percent bias. An estimated increase in absolute percent bias of $\hat{\sigma}_1$ was small (0.04 times the standard deviation of the difference, SD_{diff}), reflecting the variability of $\hat{\sigma}_1$, which did not change with increasing m. Under scenarios 2 and 3, increasing m from medium to large resulted in slight decreases in percent bias, ranging from 0.02 times SD_{diff} for $\hat{\mu}_3$ estimated under scenario 3 to 0.12 times SD_{diff} for $\hat{p}_2$ estimated under either scenario 2 or scenario 3.

Across scenarios 1, 2, and 3, increasing m was associated with modest reductions in 95% CI length. Under scenario 1, increasing m from small to medium also decreased the variability of the 95% CI length. In most cases (30/32), the 95% CIs included the true values for at least 94% of the runs.

The overlap statistics of Garret and Zeger (GZ) were generally largest for parameters with largest percent bias and smallest for those with small bias. For example, under scenario 3, $\hat{\tau}_3$ had a percent bias $> 130\%$ and a GZ statistic > 0.80, indicating little difference between the prior and posterior distributions. When additional data were used for estimation (scenario 4), the percent bias of $\hat{\tau}_3$ dropped to 1.9%, and the mean GZ statistic was 0.11, indicating identifiability of τ_3 under this scenario. Similar patterns were seen for μ_3, λ_2, and, to a lesser extent, p_2. In all cases, Λ_1 appeared to be well informed by calibration data, whereas σ_1 appeared to be less well informed by the data and hence more reliant on the prior distribution. Simulated calibration data describes summaries across individuals and thus contain little information about the between-individual variability that σ_1 models. GZ statistics were most sensitive to the data used for calibration and were insensitive to the embedded simulation size, m. In this simulation study, the GZ statistics were not affected by the number of parameters estimated.

These simulations indicate that our proposed MCMCC approach can provide unbiased estimates of informed parameters with a finite number of draws and embedded simulations. They also demonstrate that GZ statistics are a useful measure of parameter identifiability.

5. APPLICATION OF MARKOV CHAIN MONTE CARLO CALIBRATION TO OUR COLORECTAL CANCER MICROSIMULATION MODEL

Parameters associated with our MSM for CRC are grouped into 4 categories corresponding to our 4 model components: adenoma risk, adenoma growth, cancer transition, and cancer detection (Table 1). Initially, each Gibbs iteration updated the 4 parameter vectors in random order, updating the adenoma risk and transition parameter vectors twice in each pass. This unequal sampling approach improved the mixing of the sampler, because the relatively high dimension of the adenoma risk parameter vector resulted in lower acceptance rates of the associated MH steps. After several thousand iterations, we switched to a single draw of all parameters using a covariance matrix for the random walk based on earlier Gibbs steps. This approach saved a significant number of computational steps required to estimate the data parameters, $g(\theta)$.

We ran 2 MCMC chains, beginning at different locations in the parameter space. We evaluated convergence using trace plots and the corrected potential scale-reduction factor based on the ratio of within- and between-chain variability (Brooks and Gelman 1998). Once estimation was complete, we assessed the identifiability of model parameters by comparing prior and posterior distributions visually and using the Garrett and Zeger (2000) overlap statistic, described in Section 4.5.

5.1 Data Likelihoods

We assumed that data from different sources and different individuals within sources are independent. We modeled prevalence data using a binomial distribution (Imperiale et al. 2000, 2002; Lieberman et al. 2000; Surveillance, Epidemiology, and End Results (SEER) Program (*www.seer.cancer.gov*) 2004; Strul et al. 2006). For the SEER data, we assumed that the number of clinical CRC cases in each age–sex–location group follow independent binomial distributions. For the Pickhardt

Table 6. Percent bias, length, and coverage of 95% CIs and Garrett and Zeger (2000) "overlap" statistics associated with MSM parameter estimates derived using MCMCC

	Embedded simulation size	Percent bias		95% CI Length		95% CI Coverage	Overlap statistic	
Scenario 1: 2 parameters estimated								
Λ_1	Small	2.2	(17.7)	0.41	(0.97)	0.99	0.04	(0.08)
	Medium	0.1	(1.2)	0.26	(0.02)	0.99	0.03	(0.003)
	Large	0.1	(1.1)	0.23	(0.02)	0.99	0.03	(0.003)
σ_1	Small	3.7	(29.8)	0.93	(0.16)	0.99	0.51	(0.07)
	Medium	−1.1	(28.2)	0.84	(0.10)	0.99	0.47	(0.05)
	Large	−2.8	(28.1)	0.80	(0.09)	0.99	0.45	(0.05)
Scenario 2: 4 parameters estimated								
Λ_1	Medium	1.2	(1.8)	0.37	(0.06)	0.97	0.04	(0.01)
	Large	1.1	(1.6)	0.33	(0.05)	0.98	0.04	(0.01)
σ_1	Medium	45.2	(47.9)	1.3	(0.23)	0.97	0.65	(0.09)
	Large	38.2	(48.3)	1.2	(0.22)	0.96	0.62	(0.09)
p_2	Medium	7.9	(7.6)	0.02	(0.005)	0.98	0.16	(0.03)
	Large	6.7	(7.3)	0.02	(0.004)	0.98	0.14	(0.02)
λ_2	Medium	33.2	(33.5)	16.67	(4.62)	0.94	0.45	(0.09)
	Large	28.6	(33.9)	14.85	(4.17)	0.93	0.42	(0.09)
Scenario 3: 6 parameters estimated								
Λ_1	Medium	1.1	(1.8)	0.36	(0.06)	0.97	0.04	(0.01)
	Large	1.0	(1.7)	0.32	(0.06)	0.96	0.04	(0.01)
σ_1	Medium	44.9	(49.1)	1.30	(0.24)	0.97	0.65	(0.10)
	Large	36.8	(51.0)	1.19	(0.23)	0.95	0.61	(0.10)
p_2	Medium	7.5	(7.7)	0.02	(0.005)	0.97	0.16	(0.03)
	Large	6.2	(7.4)	0.02	(0.004)	0.97	0.14	(0.02)
λ_2	Medium	27.8	(34.1)	16.0	(4.44)	0.95	0.45	(0.09)
	Large	23.7	(34.0)	14.2	(4.09)	0.94	0.41	(0.09)
μ_3	Medium	65.9	(30.4)	7.2	(0.97)	0.94	0.62	(0.07)
	Large	65.1	(32.8)	6.8	(1.02)	0.91	0.60	(0.08)
τ_3	Medium	134.9	(24.1)	1.8	(0.09)	0.99	0.84	(0.06)
	Large	133.7	(26.9)	1.8	(0.09)	0.99	0.83	(0.06)
Scenario 4: 6 parameters estimated with additional data								
Λ_1	Medium	0.8	(5.7)	0.33	(0.19)	0.98	0.04	(0.01)
σ_1	Medium	10.7	(32.2)	0.95	(0.14)	0.99	0.51	(0.06)
p_2	Medium	2.6	(8.6)	0.02	(0.004)	0.99	0.11	(0.02)
λ_2	Medium	3.2	(10.2)	3.7	(0.94)	0.99	0.13	(0.03)
μ_3	Medium	1.6	(6.5)	0.61	(0.27)	0.99	0.08	(0.03)
τ_3	Medium	1.9	(5.8)	0.15	(0.03)	0.99	0.11	(0.01)

NOTE: Results are based on 1,000 runs per combination of scenario and embedded simulation sizes. Means are shown with standard deviations in parentheses. The true values used to simulate data are $\Lambda_1 = 12$, $\sigma_1 = 1.5$, $p_2 = 0.15$, $\lambda_2 = 30$, $\mu_3 = 9$, and $\tau_3 = 1.5$. Further details are provided in Section 4.

et al. (2003) data, we modeled the total number of adenomas across all screened individuals using a Poisson distribution, and modeled the distribution of adenoma size given the total number of adenomas using a multinomial distribution. We modeled the number of preclinical cancers among adenomas grouped by size (Church 2004; Odom et al. 2005) using a multinomial distribution. Tables 2 and 3 note the number of embedded simulations, m, used to estimate data-likelihood parameters. We simulated 2,000,000 individuals to estimate SEER rates. Each embedded simulation corresponds to 1 simulated life history (i.e., 1 pseudoindividual). We chose the number of embedded simulations based on 3 criteria: observed rates, with larger m for rare events; our desire to obtain both point and interval estimates; and limitations on computation time.

When calibrating to colonoscopy data (Imperiale et al. 2000, 2002; Lieberman et al. 2000; Strul et al. 2006), we incorporated the accuracy of colonoscopy into the embedded simulations. We assumed a simple quadratic model for miss rates: $P(\text{miss}|\text{size } = s < 20 \text{ mm}) = 0.34 - 0.035s + 0.0009s^2$ and $P(\text{miss}|\text{size } = s \geq 20 \text{ mm}) = 0$, producing miss rates consistent with observed findings (Hixson et al. 1990; Rex et al. 1997). For the study comparing virtual and optical colonoscopy (Pickhardt et al. 2003), we assumed that all adenomas were found, because individuals underwent both procedures and calibration data include adenomas found by either modality.

The parameters associated with data likelihoods, $\pi(\theta)$, depend on both MSM parameters and the age- and sex-distribution of the data modeled. For SEER data [Surveillance, Epidemiology, and End Results (SEER) Program (*www.seer.cancer.gov*) 2004], embedded simulations used the U.S. age and sex distribution in 1978. For other published data, we made assumptions about the age and sex distribution of study subjects. In general, studies described their samples by reporting the percentage of men and women, their average age, and the standard deviation of age. Unless information by sex was provided, we assumed the same age distribution for men and women. We modeled age distributions (used for embedded simulations) using a truncated Normal distribution based on the reported mean, standard deviation, and age range. We used a grid search to select a truncated Normal distribution with a mean and standard deviation closest to the observed values, based on a simple distance measure. For the adenoma case series data (Church 2004; Odom et al. 2005), no information on either age or sex was provided; for these 2 studies, we assumed an equal probability of being male and female, restricted individuals to be under age 90 years, and specified a truncated Normal distribution on age with a mean of 65 and standard deviation of 5.

5.2 Calibration Results

We began our MCMCC algorithm with 2 chains initiated at different locations in the parameter space. We ran these for 14,500 iterations with block sampling of parameters, followed by 75,000 additional iterations based on draws of the full parameter vector. Drawing the full parameter vector reduces computation due to the embedded simulation, but it requires a reasonable estimate of the posterior covariance matrix for MSM parameters, because this guides the direction of multidimensional random-walk steps. We evaluated convergence using visual assessment of trace plots and Gelman and Rubin statistics for multiple chains (Gelman and Rubin 1992). The chain was slowly mixing but appeared to converge, with Gelman and Rubin statistics ranging from 1.00 to 1.05 for all parameters. We based the estimation on 6,000 iterations resulting from systematically selecting every 25th iteration from the last 75,000 iterations from each chain. Because calculation of prediction intervals is computationally intensive, we based the predictive values on 3,000 iterations resulting from systematically selecting every 50th iteration.

Table 1 gives estimated posterior means, 95% CIs, and overlap statistics for all MSM parameters. The estimated posterior distribution of baseline adenoma risk was similar to our prior distribution. Posterior distributions for all other adenoma risk parameters were shifted from prior distributions. Adenoma growth means differed from the wide priors that we specifiedl however, posterior distributions for scale parameters were similar to the uniform distributions specified, indicating that the data provide little information about variation in adenoma growth. Posterior distributions for adenoma transition parameters were shifted from specified priors, except for posterior distributions associated with the transition of rectal adenomas in men. This highly parameterized transition component, with different parameters by sex and location, was required to obtain adequate fit to SEER cancer incidence data. Overlap statistics indicate that information about sojourn time came largely from the prior distributions, with the exception of mean sojourn time associated with colon cancers.

Table 7 gives estimated posterior means together with 95% predictive intervals (PIs) for the MSM model parameters. In general, the MSM demonstrated good prediction of calibration data, particularly for SEER cancer incidence rates. For the SEER cancer rates, estimates are similar to observed data, and PIs cover the observed values, except for rectal cancer rates in men age 20–49 (observed, 2.28 per 100,000; estimated, 3.23 per 100,000; 95% PI = 2.33–4.37). The predicted adenoma prevalence was somewhat high compared with that found by Strul et al. (2006), although predicted adenoma counts across individuals were similar to those reported by Pickhardt et al. (2003), suggesting that model fit might be improved by incorporating additional data describing the number of adenomas within individuals. Predicted adenoma sizes were close to observed data, but with more large adenomas than found by either Lieberman et al. (2000) or Pickhardt et al. (2003). Wide PIs demonstrate uncertainty in predicted transition probabilities. The model accurately predicted preclinical cancer rates in individuals, but predicted too many cancers in small adenomas and too few in large adenomas.

6. DISCUSSION

We have proposed an MCMCC method based on a random-walk MH algorithm that uses an estimated transition rule. Our proposed approach is similar to an approximate MCMC method proposed by Christen and Fox (2005), who based their transition rule on an approximation to the data likelihood. In the scenario that those authors examined, computation of the likelihood was possible but expensive. When calibrating a complex MSM, the likelihood could not be calculated exactly, because the associated data parameters were unknown functions of MSM parameters. Our MCMCC method uses simulation to approximate these data parameters. Although MCMCC is computationally intensive, the price of computation is offset the advantages of MCMCC over previously proposed MSM calibration methods.

MCMCC provides 2 distinct avenues for incorporating information into the MSM: via prior distributions for MSM parameters and via calibration data. Specification of prior distributions formalizes the process of incorporating expert information into MSMs. Uniform priors can be used when little prior knowledge is available. Priors are updated using calibration data to obtain posterior estimates. Care must be taken when choosing priors for parameters that are not informed by calibration data, because there is little or no updating of the prior information. Overlap statistics, used to compare prior and posterior distributions, provide information on how resulting posterior distributions differ from prior beliefs. These overlap statistics also provide insight into parameter identifiability. This information is invaluable, because it can serve to temper findings from MSM studies, indicating areas in which variability in model parameters is especially important to consider. Results describing parameter identifiability also point out areas requiring further research to inform the disease processes.

Simultaneous calibration to multiple data sources via MCMCC is straightforward, and the algorithm implicitly

Table 7. Model fit to calibration data

	Observed	Estimated	95% predictive interval
Adenoma prevalence			
Among individuals 40–49 years old[1]	0.10	0.14	(0.09, 0.21)
Among individuals 50–75 years old[1]	0.22	0.27	(0.21, 0.33)
Among individuals 76–80 years old[1]	0.29	0.38	(0.25, 0.52)
Number of adenomas found in 1,233 individuals[2]	554	575	(450, 747)
Adenoma size			
Percent of individuals with an adenoma $\geq$10 mm, given $\geq$1 adenoma[3]	0.23	0.28	(0.22, 0.35)
Percent of adenomas $<$6 mm[2]	0.62	0.61	(0.53, 0.68)
Percent of adenomas [6, 10) mm[2]	0.29	0.22	(0.16, 0.28)
Percent of adenomas $\geq$10 mm[2]	0.09	0.17	(0.12, 0.22)
Preclinical cancers			
Among employed individuals $\geq$50[4]	0.004	0.003	(0.001, 0.006)
Among adenomas 6–10 mm[5]	0.002	0.008	(0.000, 0.028)
Among adenomas $>$10 mm[5]	0.030	0.048	(0.000, 0.109)
Among adenomas 6–10 mm[6]	0.0065	0.008	(0.000, 0.049)
Among adenomas 11–20 mm[6]	0.0160	0.036	(0.000, 0.133)
Among adenomas $>$20 mm[6]	0.19	0.14	(0.00, 0.667)
Clinical cancers in 1975–1979, per 100,000[7]			
Colon cancer in women			
20–49 y.o.	4.8	4.4	(3.1, 5.8)
50–59 y.o.	43.3	45.5	(37.9, 53.1)
60–69 y.o.	100.7	99.3	(88.1, 111.9)
70–84 y.o.	216.7	207.3	(181.5, 240.7)
Rectal cancer in women			
20–49 y.o.	1.87	2.0	(1.4, 2.8)
50–59 y.o.	20.4	18.6	(14.6, 22.6)
60–69 y.o.	42.5	39.7	(32.8, 46.7)
70–84 y.o.	73.9	82.1	(66.8, 97.2)
Colon cancer in men			
20–49 y.o.	4.51	4.2	(2.9, 5.6)
50–59 y.o.	45.9	50.6	(41.4, 60.1)
60–69 y.o.	121.4	120.0	(106.2, 136.8)
70–84 y.o.	268.4	263.4	(224.9, 306.4)
Rectal cancer in men			
20–49 y.o.	2.3	3.2	(2.3, 4.4)
50–59 y.o.	30.0	29.8	(24.1, 35.5)
60–69 y.o.	71.4	63.2	(53.5, 72.4)
70–84 y.o.	128.0	123.8	(101.5, 145.9)

NOTE: 1. Strul et al. (2006); 2. Pickhardt et al. (2003); 3. Lieberman et al. (2000); 4. Imperiale et al. (2000); 5. Church (2004); 6. Odom et al. (2005); 7. Surveillance, Epidemiology, and End Results (SEER) Program (*www.seer.cancer.gov*) (2004).

weights calibration data based on sampling variability. MCMCC provides both point and interval estimates of model parameters, as well as functions of parameters, offering flexibility in the presentation of results. Calibration methods based on a grid search to obtain the best fit to data do not provide interval estimates. When applying frequentist maximum likelihood approaches, lack of model identifiability can hamper the ability to obtain interval estimates.

We have not discussed estimation based on direct Monte Carlo simulation of the likelihood, which could be implemented using embedded simulation to obtain estimates of data-likelihood parameters used in calculating likelihoods. Such a Monte Carlo approach could provide maximum likelihood estimates of MSM parameters. But because the functional form of the likelihood is not readily available, the information matrix also is not readily available, and thus Monte Carlo estimates cannot directly provide measures of precision. In any case, asymptotic precision likely is not relevant for MSM estimation when only limited calibration data are available.

The accuracy of MSM models depends on appropriate modeling of disease processes. Between-model comparisons (Berry et al. 2005) offer one approach to model assessment. Model checking based on both calibration and validation data also provides information on model adequacy. How closely the model should calibrate to observed data is unclear, especially when calibration data are variable and may provide conflicting infor-

mation. Prediction intervals from our MSM cover 15 of the 16 SEER cancer rates considered to be key calibration points, as well as 12 of 14 other calibration data points. Is this enough? It depends on how modelers trade-off concerns about possibly overparameterizing and overfitting calibration data relative to the importance of exactly replicating observed or expected results. We lean toward model parsimony. For complex disease processes, even a relatively parsimonious model may require the inclusion of parameters that are not well informed by available data.

The data available to inform MSMs often span decades, in this case ranging from SEER cancer incidence data from 1975–1979 to colonoscopy data collected in 2002–2003. Much of the information needed to calibrate the model parameters became available only after the advent of colonoscopy. We used all available data and assumed no secular trends in MSM parameters. Other assumptions are possible; for example, we could make stronger prior assumptions about model parameters and restrict our calibration data to specific date ranges, or could make assumptions about secular trends in adenoma parameters and use all available data.

Once MSM parameters are calibrated to reproduce known or expected results, MSMs may be used to examine the impact of potential interventions, including the cost-effectiveness of interventions (Ness et al. 2000; Will et al. 2001; Vogelaar et al. 2006). These results may in turn influence public policy decisions. Because of their potential policy impact, MSM assumptions and calibration methods should be clearly described to allow for their critical evaluation. Ideally, point estimates of effects should be accompanied by measures of precision so that policy makers can assess the strength of evidence models provide.

APPENDIX: CONVERGENCE OF THE APPROXIMATE MARKOV CHAIN MONTE CARLO

Here we show that our proposed MCMCC approach converges to the target distribution, $h(\theta|y)$, by showing that the detailed balance condition holds asymptotically, as $m \to \infty$, when $\hat{g}(\theta)$ is a maximum likelihood estimator. The detailed balance condition is given by $h(\theta|y)\alpha(\theta,\theta^*) = h(\theta^*|y)\alpha(\theta^*,\theta)$. Detailed balance of a Markov chain is a sufficient, but not necessary, condition for reversibility and convergence to $h(\theta|y)$ (Casella and Robert 1999). Detailed balance for the MH algorithm is easy to confirm when $g(\theta)$ is known.

When r is estimated, the probability of transition from θ to θ^*, unconditional on $\hat{r}$, equals $P(\hat{r} > 1) + E(\hat{r}|\hat{r} \le 1)P(\hat{r} \le 1)$, where the expectation results from integration over the distribution of $\hat{r}$. Note that $E(\hat{\alpha}|\hat{r} > 1) = 1$ and $E(\hat{\alpha}|\hat{r} \le 1) = E(\hat{r}|\hat{r} \le 1)$. When r is estimated, the asymptotic behavior of the MCMC depends on the distribution of the estimators, $\hat{r}$ and $1/\hat{r}$. Next, we show that $\hat{r}$ and $1/\hat{r}$ are asymptotically Normally distributed.

Let $G = (g(\theta), g(\theta^*))$ and $\widehat{G}_m = (\hat{g}_m(\theta), \hat{g}_m(\theta^*))$. Assume that θ and θ^* are fixed but arbitrary, and given θ and θ^*, $\hat{g}_m(\theta)$ and $\hat{g}_m(\theta^*)$ are independent. As $m \to \infty$, $\sqrt{m}, \widehat{G}_m$ converges in law to Normal(G, V), where V is a diagonal 2×2 matrix with diagonal terms $v_{11} = \text{var}(y|g(\theta))$ and $v_{22} = \text{var}(y|g(\theta^*))$. If $\dot{r}(G) = (\partial r/\partial g_i(\theta), \partial r/\partial g_i(\theta^*))$ is continuous in a neighborhood of G then, given $\hat{g}_m(\theta)$ and $\hat{g}_m(\theta^*)$, the sequence $\sqrt{m}\, r(\widehat{G}_m)$ converges in law to a Normal distribution centered at $r(G)$, with variance

$$r^2\left[\left(\frac{\partial f(y|g(\theta))/\partial g(\theta)}{f(y|g(\theta))}\right)^2 \text{var}(y|g(\theta)) + \left(\frac{\partial f(y|g(\theta^*))/\partial g(\theta^*)}{f(y|g(\theta^*))}\right)^2 \text{var}(y|g(\theta^*))\right]$$

(Ferguson 1996). Similarly, for large m, $1/r(\widehat{G}_m)$ is approximately Normally distributed with mean $1/r(G)$.

Given the asymptotic Normality of $\hat{r}$ and $1/\hat{r}$, we can show that detailed balance holds for MCMCC. Suppose that $r \le 1$; then

$$h(\theta|y)E(\hat{\alpha}(\theta,\theta^*)|r \le 1) = h(\theta|y)[P(\hat{r} > 1|r \le 1) + P(\hat{r} \le 1|r \le 1)E(\hat{r}|r \le 1 \,\&\, \hat{r} \le 1)].$$

As $m \to \infty$, $E(\hat{r}|r \le 1 \,\&\, \hat{r} \le 1) \to r$. Using $h(\theta|y) = \frac{1}{r}h(\theta^*|y)$, it is easy to show that

$$h(\theta|y)E(\hat{\alpha}(\theta,\theta^*)|r \le 1) = h(\theta^*|y)\left[1 + \left(\frac{1}{r} - 1\right)P(\hat{r} > 1|r \le 1)\right].$$

For fixed r such that $0 < r + \epsilon < 1$ for some $\epsilon > 0$,

$$\lim_{m\to\infty}\left(\frac{1}{r} - 1\right)P(\hat{r} > 1|r \le 1) = 0.$$

At the upper boundary,

$$\lim_{r\to 1}\left(\frac{1}{r} - 1\right)P(\hat{r} > 1|r \le 1) = 0,$$

because $\lim_{r\to 1}(\frac{1}{r} - 1) \to 0$ and $\lim_{r\to 1} P(\hat{r} > 1|r \le 1) \to 0.5$. At the lower boundary,

$$\lim_{r\to 0}\left(\frac{1}{r} - 1\right)P(\hat{r} > 1|r \le 1) = 0,$$

by l'Hôpital's rule. Thus, for $m \to \infty$, $h(\theta|y)E(\hat{\alpha}(\theta,\theta^*)|r \le 1) \to h(\theta^*|y)$ for $0 \le r \le 1$.

Similarly,

$$h(\theta^*|y)E(\hat{\alpha}(\theta^*,\theta)|r \le 1) = h(\theta^*|y)\left[P\left(\frac{1}{\hat{r}} \ge 1\middle|\frac{1}{r} \ge 1\right) + P\left(\frac{1}{\hat{r}} < 1\middle|\frac{1}{r} \ge 1\right)E\left(\frac{1}{\hat{r}}\middle|\frac{1}{\hat{r}} < 1 \,\&\, \frac{1}{r} \ge 1\right)\right].$$

As $m \to \infty$, $E(\frac{1}{\hat{r}}|\frac{1}{\hat{r}} < 1 \,\&\, \frac{1}{r} \ge 1) \to 1$, so that

$$\lim_{m\to\infty} E(\hat{\alpha}(\theta^*,\theta)|r \le 1) = \lim_{m\to\infty}[P(\hat{r}^{-1} \ge 1|r^{-1} \ge 1) + P(\hat{r}^{-1} < 1|r^{-1} \ge 1)] = 1$$

and $\lim_{m\to\infty} E(\hat{\alpha}(\theta^*,\theta)|r \le 1) = h(\theta^*|y)$. Thus, when $r \le 1$, as $m \to \infty$, $h(\theta|y)E(\hat{\alpha}(\theta,\theta^*)|r \le 1) = h(\theta^*|y)E(\hat{\alpha}(\theta^*,\theta)|r \le 1)$ and detailed balance holds. Similar arguments show that detailed balance holds when $r \ge 1$.

[Received September 2008. Revised October 2008.]

REFERENCES

American Cancer Society (2008), "Cancer Facts & Figures," available at *http://www.cancer.org/docroot/stt/content/stt_1x_cancer_facts_and_figures_2008.asp*

Berry, D. A., Cronin, K. A., Plevritis, S. K., Fryback, D. G., Clarke, L., Zelen, M., Mandelblatt, J. S., Yakovlev, A. Y., Habbema, J. D. F., and Feuer, E. J. (2005), "Effect of Screening and Adjuvant Therapy on Mortality From Breast Cancer," *New England Journal of Medicine*, 353, 1884–1892.

Blatt, L. J. (1961), "Polyps of the Colon and Rectum: Incidence and Distribution," *Diseases of the Colon and Rectum*, 4, 277–282.

Bombi, J. A. (1988), "Polyps of the Colon in Barcelona, Spain," *Cancer*, 61, 1472–1476.

Brooks, S. P., and Gelman, A. (1998), "Alternative Methods for Monitoring Convergence of Iterative Simulations," *Journal of Computational and Graphical Statistics*, 7, 434–455.

Casella, G., and Robert, C. (1999), *Monte Carlo Statistical Methods*, New York: Springer Verlag.

CDC (2001), "Trends in Screening for Colorectal Cancer—United States, 1997 and 1999," *Morbidity and Mortality Weekly Report*, 50, 162–166.

Chapman, I. (1963), "Adenomatous Polypi of Large Intestine: Incidence and Distribution," *Annals of Surgery*, 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.

Chia, Y. L., Salzman, P., Plevritis, S. K., and Glynn, P. W. (2004), "Simulation-Based Parameter Estimation for Complex Models: A Breast Cancer Natural History Modelling Illustration," *Statistical Methods in Medical Research*, 13, 507–524.

Christen, J. A., and Fox, C. (2005), "Markov Chain Monte Carlo Using an Approximation," *Journal of Computational and Graphical Statistics*, 14, 795–810.

Church, J. M. (2004), "Clinical Significance of Small Colorectal Polyps," *Diseases of the Colon and Rectum*, 47, 481–485.

Croyle, R. (2006), "Director's Update: CISNET Offers Powerful New Tools for Cancer Control," *NCI Cancer Bulletin*, 3, 1–2.

de Koning, H. J., Draisma, G., Fracheboud, J., and de Bruijn, A. (2006), "Overdiagnosis and Overtreatment of Breast Cancer: Microsimulation Modeling Estimates Based on Observed Screen and Clinical Data," *Breast Cancer Research*, 8, 202–206.

Eide, T. J., and Stalsberg, H. (1978), "Polyps of the Large Intestine in Northern Norway," *Cancer*, 42, 2839–2848.

Etzioni, R., Leggler, J. M., Feuer, E. J., Merrill, R. M., Cronin, K. A., and Hankey, B. F. (1999), "Cancer Surveillance Series: Interpreting Trends in Prostate Cancer—Part III: Quantifying the Link Between Population Prostate-Specific Antigen Testing and Recent Declines in Prostate Cancer Mortality," *Journal of the National Cancer Institute*, 91, 1033–1039.

Ferguson, T. S. (1996), *A Course in Large Sample Theory*, New York: Chapman & Hall.

Feuer, E. J., Etzioni, R., Cronin, K. A., and Mariotto, A. (2004), "The Use of Modeling to Understand the Impact of Screening on US Mortality: Examples From Mammography and PSA Testing," *Statistical Methods in Medical Research*, 13, 421–442.

Garrett, E. S., and Zeger, S. L. (2000), "Latent Class Model Diagnostics," *Biometrics*, 56, 1055–1067.

Gelman, A., and Rubin, D. (1992), "Inference From Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457–511.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian Data Analysis*, New York: Chapman & Hall.

Gille, U., and Salomon, F. V. (2000), "Brain Growth in Mallards, Pekin and Muscovy Ducks (Anatidae)," *Journal of Zoology*, 252, 399–404.

Gustafson, P. (2005), "On Model Expansion, Model Contraction, Identifiability and Prior Information: Two Illustrative Scenarios Involving Mismeasured Variables," *Statistical Science*, 20, 111–140.

Hardcastle, J. D., Chamberlain, J. O., Robinson, M. H., Moss, S. M., Amar, S. S., Balfour, T. W., James, P. D., and Mangham, C. (1996), "Randomised Controlled Trial of Faecal-Occult-Blood Screening for Colorectal Cancer," *Lancet*, 348, 1472–1477.

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," *Journal of the National Cancer Institute*, 82, 1769–1772.

Howe, H. L., Wingo, P. A., Thun, M. J., Ries, L. A., Rosenberg, H. M., Feigal, E. G., and Edwards, B. K. (2001), "Annual Report to the Nation on the Status of Cancer (1973 Through 1998), Featuring Cancers With Recent Increasing Trends," *Journal of the National Cancer Institute*, 93, 824–842.

Howe, H. L., Wu, X., Ries, L. A., Cokkinides, V., Ahmed, F., Jemal, A., Miller, B., Williams, M., Ward, E., Wingo, P. A., Ramirez, A., and Edwards, B. K. (2006), "Annual Report to the Nation on the Status of Cancer, 1975–2003, Featuring Cancer Among U.S. Hispanic/Latino Populations," *Cancer*, 107, 1711–1742.

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," *New England Journal of Medicine*, 343, 169–174.

——— (2002), "Results of Screening Colonoscopy Among Persons 40 to 49 Years of Age," *New England Journal of Medicine*, 346, 1781–1785.

Janoschek, A. (1957), "Das reaktionskinetische Grundgesetz und seine Beziehungen zum Wachstums- und Ertragsgesetz," *Statistische Vierteljahresschrift*, 10, 25–37.

Jemal, A., Thun, M. J., Ries, L. A., Howe, H. L., Wier, H. K., Center, M. W., Ward, E., Wu, X. C., Eheman, C., Anderson, R., Ajani, U. A., Kohler, B., and Edwards, B. K. (2008), "Annual Report to the Nation on the Status of Cancer, 1975–2005, Featuring Trends in Lung Cancer, Tobacco Use, and Tobacco Control," *Journal of the National Cancer Institute*, 100, 1672–1694.

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.

Kiefer, J., and Wolfowitz, J. (1952), "Stochastic Estimation of the Maximum of a Regression Function," *Annals of Mathematical Statistics*, 23, 462–466.

Kronborg, O., Fenger, C., Olsen, J., Jorgensen, O. D., and Sondergaard, O. (1996), "Randomised Study of Screening for Colorectal Cancer With Faecal-Occult-Blood Test," *Lancet*, 348, 1467–1471.

Leslie, A., Carey, F. A., Pratt, N. R., and Steele, R. J. C. (2002), "The Colorectal Adenoma–Carcinoma Sequence," *British Journal of Surgery*, 89, 845–860.

Lieberman, D. A., Weiss, D. G., Bond, J. H., Ahnen, D. J., Garewal, H., and Chejfec, G. (2000), "Use of Colonoscopy to Screen Asymptomatic Adults for Colorectal Cancer," *New England Journal of Medicine*, 343, 162–168.

Loeve, F., Boer, R., van Oortmarssen, G. J., van Ballegooijen, M., and Habbema, J. D. (1999), "The MISCAN-COLON Simulation Model for the Evaluation of Colorectal Cancer Screening," *Computers and Biomedical Research*, 32, 13–33.

Loeve, F., Boer, R., Zauber, A. G., van Ballegooijen, M., van Oortmarssen, G. J., Winawer, S. J., and Habbema, J. D. F. (2004), "National Polyp Study Data: Evidence for Regression of Adenomas," *International Journal of Cancer*, 111, 633–639.

Muto, T., Bussey, H. J. R., and Morson, B. C. (1975), "The Evolution of Cancer in the Colon and Rectum," *Cancer*, 36, 2251–2270.

Nelder, J. A., and Mead, R. (1965), "A Simplex Method for Function Minimization," *Computer Journal*, 7, 308–313.

Ness, R. N., Holmes, A. M., Klein, R., and Dittus, R. (2000), "Cost-Utility of One-Time Colonoscopic Screening for Colorectal Cancer at Various Ages," *The American Journal of Gastroenterology*, 95, 1800–1811.

Nusko, G., Mansmann, U., Altendorf-Hoffman, A., Groitl, H., Wittenkind, C., and Hahn, E. G. (1997), "Risk of Invasive Carcinoma in Colorectal Adenomas Assessed by Size and Site," *International Journal of Colorectal Surgery*, 12, 267–271.

Odom, S. R., Duffy, S. D., Barone, J. E., Ghevariya, V., and McClane, S. J. (2005), "The Rate of Adenocarcinoma in Endoscopically Removed Colorectal Polyps," *The American Surgeon*, 71, 1024–1026.

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., and Schindler, W. R. (2003), "Computed Tomographic Virtual Colonocopy to Screen for Colorectal Neoplasia in Asymptomatic Adults," *New England Journal of Medicine*, 349, 2191–2200.

Puvimanasinghe, J. P. A., Takkenberg, J. J. M., Edwards, M. B., Eijkemans, M. J. C., Steyerberg, E. W., van Herwerden, L. A., Taylor, K. M., Grunkemeir, G. L., Habbema, J. D. F., and Bogers, A. J. J. C. (2004), "Comparisons of Outcomes After Aortic Valve Replacement With a Mechanical Valve or a Bioprosthesis Using Microsimulation," *Heart*, 90, 1172–1178.

Ramsey, S. D., McIntosh, M., Etzioni, R., and Urban, N. (2000), "Simulation Modeling of Outcomes and Cost Effectiveness," *Hematology Oncology Clinics of North America*, 14, 925–938.

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.

Robbins, H., and Monro, S. (1951), "A Stochastic Approximation Method," *Annals of Mathematical Statistics*, 22, 400–407.

Roberts, G. O., and Sahu, S. K. (1997), "Updating Schemes, Correlation Structures, Blocking and Parameterization for the Gibbs Sampler," *Journal of the Royal Statistical Society, Ser. B*, 59, 291–317.

Rutter, C. M., Miglioretti, D. L., and Yu, O. (2007), "A Hierarchical Non-Homogeous Poisson Model for Meta-Analysis of Adenoma Counts," *Statistics in Medicine*, 26, 98–109.

Salomon, J. A., Weinstein, M. C., Hammitt, J. K., and Goldie, S. J. (2002), "Empirically Calibrated Model of Hepatitis C Virus Infection in the United States," *American Journal of Epidemiology*, 156, 761–773.

Shen, Y., and Parmigiani, G. (2005), "A Model-Based Comparison of Breast Cancer Screening Strategies: Mammograms and Clinical Breast Examinations," *Cancer Epidemiology Biomarkers and Prevention*, 14, 529–532.

Smith, A. F., and Roberts, G. O. (1993), "Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society, Ser. B*, 55, 3–23.

Stemmermann, G. N., and Yatani, R. (1973), "Diverticulosis and Polyps of the Large Intestine. A Necropsy Study of Hawaii Japanese," *Cancer*, 31, 1260–1270.

Stewart, S. L., King, J. B., Thompson, T. D., Friedman, C., and Wingo, P. A. (2004), "Cancer Mortality Surveillance–United States, 1990–2000," *MMWR Surveillance Summaries*, 53, 1–108.

Strul, H., Kariv, R., Leshno, M., Halak, A., Jakubowicz, M., Santo, M., Umansky, M., Shirin, H., Degani, Y., Revivo, M., Halpern, Z., and Arber, N. (2006), "The Prevalence Rate and Anatomic Location of Colorectal Adenoma and Cancer Detected by Colonoscopy in Average-Risk Individuals Aged 40–80 Years," *American Journal of Gastroenterology*, 101, 255–262.

Surveillance, Epidemiology, and End Results (SEER) Program (*www.seer.cancer.gov*) (2004), released April 2004, based on the November 2003 submission.

Szczepanski, W., Urban, A., and Wierzchowski, W. (1992a), "Colorectal Polyps in Autopsy Material. Part I. Adenomatous Polyps," *Patologia Polska*, 43, 79–85.

——— (1992b), "Colorectal Polyps in Autopsy Material. Part I. Adenomatous Polyps," *Patologia Polska*, 43, 79–85.

Tierney, L. (1994), "Markov Chanis for Exploring Posterior Distributions," *The Annals of Statistics*, 22, 1701–1762.

Towler, B., Irwig, L., Glasziou, P., Kewenter, J., Weller, D., and Silagy, C. (1998), "A Systematic Review of the Effects of Screening for Colorectal Cancer Using the Faecal Occult Blood Test, Hemoccult," *British Medical Journal*, 317, 559–565.

US Life Tables (2000), available at *http://www.cdc.gov/nchs/products/pubs/pubd/lftbls/life/1966.htm*.

van den Akker-van Marle, M. E., van Ballegooijen, M., van Oortmarssen, G. J., Boer, R., and Habbema, D. F. (2002), "Cost-Effectiveness of Cervical Cancer Screening: Comparison of Screening Policies," *Journal of the National Cancer Institute*, 94, 193–204.

Vogelaar, I., van Ballegooijen, M., Schrag, D., Boer, R., Winawer, S. J., Habbema, J. D. F., and Zauber, A. G. (2006), "How Much Can Current Interventions Reduce Colorectal Cancer Mortality in the US?" *Cancer*, 107, 1624–1633.

Will, B. P., Berthelot, J. M., Nobrega, K. M., Flanagan, W., and Evans, W. K. (2001), "Canada's Population Health Model (POHEM): A Tool for Performing Economic Evaluations of Cancer Control Interventions," *European Journal of Cancer*, 37, 1797–1804.

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.

Wingo, P. A., Cardinez, C. J., Landis, S. H., Greenlee, R. T., Ries, L. A., Anderson, R. N., and Thun, M. J. (2003), "Long-Term Trends in Cancer Mortality in the United States, 1930–1998," *Cancer*, 97, 3133–3275.

Yamaji, Y., Mitsushima, T., Yoshida, H., Watabe, H., Okamoto, M., Wada, R., Ikuma, H., Kawabe, T., and Omata, M. (2006), "The Malignant Potential of Freshly Developed Colorectal Polyps According to Age," *Cancer Epidemiology Biomarkers and Prevention*, 15, 2418–2421.