Supplemental Information (SI): Impact of Tumor Progression on Cancer Incidence Curves

Mathematical Methods

We briefly summarize the mathematical development for the MSCE model (Figure 1). Tumor initiation is modeled as a two-hit Poisson process. The two hits may represent inactivation of both alleles of a tumor suppressor gene or any two other rate limiting mutations sufficient to provide a growth advantage and loss of homeostatic control. After the first hit, cells are considered pre-initiated, and assigned a label P^*. After a second hit, a pre-initiated cell becomes initiated, or premalignant, and is assigned a label P. The first clonal expansion process represents clonal growth or extinction of P cells through a birth-death-migration process with cell division rate α_P, cell death/differentiation rate β_P, and a per cell malignant transformation rate μ_2 for malignant cells, each assigned a label M. The MSCE model assumes a second clonal expansion of M cells occurs through a birth-death-detection process with cell division rate α_M, cell death/differentiation rate β_M, and malignant transformation rate ρ per cell per year. The simplified version of the MSCE model used in previous studies [1–4] (Figure 1, elevated dashed lines) assumes a lag-time which represents the time from first M cell to time of cancer diagnosis.

Backward Kolmogorov equations for the MSCE model hazard, h_{MSCE}

$N(t)$ = number of renewing normal stem cells in a tissue at time t

$P^*(t)$ = number of pre-initiated cells at time t

$P(t)$ = number of premalignant (initiated) cells at time t

$M(t)$ = number of malignant (preclinical) cells (prior to detection) at time t

$C(t)$ = number of cancer cells (after detection) at time t

For clinical detection, we will employ the following indicator function, $D(t)$,

$$D(t) = \begin{cases} 0 \text{ if no cancer detected clinically by time } t \\ 1 \quad \text{otherwise} \quad \text{ie. } C(\tau) > 0 \text{ for some } \tau \leq t \end{cases}$$

Beginning with a single pre-initiated (P^*), premalignant (P), or malignant (M) cell at time τ, the probability distribution for number of cells at time $t \geq \tau$, is represented by the generating function $\Phi_{_{P^*}}$, $\Phi_{_{P^*}}$, or $\Phi_{_M}$, respectively, with

$$\Phi_{_M}(y_3, z; \tau, t) = E[y_3^{M(t)} z^{D(t)} | M(\tau) = 1, D(\tau) = 0] \tag{1}$$

$$\Phi_{_P}(y_2, y_3, z; \tau, t) = E[y_2^{P(t)} y_3^{M(t)} z^{D(t)} | P(\tau) = 1, M(\tau) = 0, D(\tau) = 0] \tag{2}$$

$$\Phi_{_{P^*}}(y_1, y_2, y_3, z; \tau, t) = E[y_1^{P^*(t)} y_2^{P(t)} y_3^{M(t)} z^{D(t)} | P^*(\tau) = 1, P(\tau) = 0, M(\tau) = 0, D(\tau) = 0] \tag{3}$$

And the generating function for the entire process Ψ starting from normal cells is the following

$$\Psi(y_1, y_2, y_3, z; \tau, t) = E[y_1^{P^*(t)} y_2^{P(t)} y_3^{M(t)} z^{D(t)} | P^*(\tau) = 0, P(\tau) = 0, M(\tau) = 0, D(\tau) = 0] \tag{4}$$

$$= \sum_{i,j,k,l} y_1^i y_2^j y_3^k z^l P(i, j, k, l) \tag{5}$$

where $P(i, j, k, l) = \Pr[P^*(t) = i, P(t) = j, M(t) = k, D(t) = l | P^*(\tau) = 0, P(\tau) = 0, M(\tau) = 0, D(\tau) = 0]$,

and $l = \{0, 1\}$. For constant rates and number of normal stem cells $N(t) = N$, the generating functions satisfy the following Kolmogorov backward equations

$$\begin{aligned} \frac{\partial \Phi_{_M}(y_3, z; \tau, t)}{\partial \tau} &= -\alpha_{_M} \Phi_{_M}^2(y_3, z; \tau, t) - \beta_{_M} \\ &\quad - z\rho \Phi_{_M}(y_3, z; \tau, t) + [\alpha_{_M} + \beta_{_M} + \rho] \Phi_{_M}(y_3, z; \tau, t) \\ \frac{\partial \Phi_{_P}(y_2, y_3, z; \tau, t)}{\partial \tau} &= -\alpha_{_P} \Phi_{_P}^2(y_2, y_3, z; \tau, t) - \beta_{_P} \\ &\quad + [\alpha_{_P} + \beta_{_P} + \mu_2] \Phi_{_P}(y_2, y_3, z; \tau, t) - \mu_2 \Phi_{_P}(y_2, y_3, z; \tau, t) \Phi_{_M}(y_3, z; \tau, t) \end{aligned}$$

$$\begin{aligned}\frac{\partial \Phi_{P*}(y_1, y_2, y_3, z; \tau, t)}{\partial \tau} &= -\mu_1 \Phi_{P*}(y_1, y_2, y_3, z; \tau, t)[\Phi_P(y_2, y_3, z; \tau, t) - 1] \\ \frac{\partial \Psi(y_1, y_2, y_3, z; \tau, t)}{\partial \tau} &= -\mu_0 N \Psi(y_1, y_2, y_3, z; \tau, t)[\Phi_{P*}(y_1, y_2, y_3, z; \tau, t) - 1]\end{aligned}$$

In the main text, we wished to solve for the overall survival function (for cancer detection), starting at time 0, which in our notation is

$$S_{MSCE}(t) = 1 - P_{MSCE}(t) = \Pr[D(t) = 0 | P^{*}(0) = 0, P(0) = 0, M(0) = 0, D(0) = 0]$$

$$= \Psi(1, 1, 1, 0; 0, t)$$

where $P_{MSCE}(t)$ is the probability of a cancer detection at time t,

$$P_{MSCE}(t) = \Pr[D(t) = 1 | P^{*}(0) = 0, P(0) = 0, M(0) = 0, D(0) = 0]$$

We will here denote $\Phi_M(1, 0; \tau, t) \equiv \Phi_M(\tau, t)$, $\Phi_P(1, 1, 0; \tau, t) \equiv \Phi_P(\tau, t)$, $\Phi_{P*}(1, 1, 1, 0; \tau, t) \equiv \Phi_{P*}(\tau, t)$, and $\Psi(1, 1, 1, 0; \tau, t) \equiv \Psi(\tau, t)$. and a dot designating a first derivative with respect to t. The hazard function, i.e., the rate at which cancer is detected in individuals who have not been diagnosed before, is given by

$$h_{MSCE}(t) = -\frac{\dot{S}_{MSCE}(t)}{S_{MSCE}(t)} = -\frac{\dot{\Psi}(0, t)}{\Psi(0, t)} = -\frac{d}{dt} \ln[\Psi(0, t)] \tag{6}$$

For fixed t, this boundary value system of coupled PDEs can be converted into an initial value problem (IVP) with the change of variables $u = t - \tau$, where u is the “running” time. This redefinition and equations hereafter follow the method used by Crump et al. 2005 [5]. Define the following variables for the new IVP: $Y_1(u, t) = \Phi_M(\tau, t), Y_2(u, t) = \dot{\Phi}_M(\tau, t), Y_3(u, t) = \Phi_P(\tau, t), Y_4(u, t) = \dot{\Phi}_P(\tau, t), Y_5(u, t) = \Phi_{P*}(\tau, t), Y_6(u, t) = \dot{\Phi}_{P*}(\tau, t), Y_7(u, t) = \Psi(\tau, t), Y_8(u, t) = -\dot{\Psi}(\tau, t)/\Psi(\tau, t)$ with corresponding ICs $Y_1(0, t) = Y_3(0, t) = Y_5(0, t) = Y_7(0, t) = 1$,

$Y_4(0,t) = Y_6(0,t) = Y_8(0,t) = 0$, and $Y_2(0,t) = -\rho$.

$$\frac{dY_1(u,t)}{du} = \beta_{_M} - (\alpha_{_M} + \beta_{_M} + \rho)Y_1(u,t) + \alpha_{_M} Y_1^2(u,t) \tag{7}$$

$$\frac{dY_2(u,t)}{du} = 2\alpha_{_M} Y_1(u,t)Y_2(u,t) - (\alpha_{_M} + \beta_{_M} + \rho)Y_2(u,t) \tag{8}$$

$$\frac{dY_3(u,t)}{du} = \beta_{_P} + \mu_2 Y_1(u,t)Y_3(u,t) - (\alpha_{_P} + \beta_{_P} + \mu_2)Y_3(u,t) + \alpha_{_P} Y_3^2(u,t) \tag{9}$$

$$\frac{dY_4(u,t)}{du} = 2\alpha_{_P} Y_3(u,t)Y_4(u,t) + \mu_2(Y_4(u,t)Y_1(u,t) + Y_3(u,t)Y_2(u,t)) - (\alpha_{_P} + \beta_{_P} + \mu_2)Y_4(u,t) \tag{10}$$

$$\frac{dY_5(u,t)}{du} = \mu_1 Y_5(u,t)(Y_3(u,t) - 1) \tag{11}$$

$$\frac{dY_6(u,t)}{du} = \mu_1(Y_6(u,t)Y_3(u,t) - Y_6(u,t) + Y_5(u,t)Y_4(u,t)) \tag{12}$$

$$\frac{dY_7(u,t)}{du} = \mu_0 N Y_7(u,t)(Y_5(u,t) - 1) \tag{13}$$

$$\frac{dY_8(u,t)}{du} = -\mu_0 N Y_6(u,t) \tag{14}$$

These 8 ODEs can be solved numerically to obtain

$$h_{_{MSCE}}(t) = Y_8(t,t) \quad \text{and} \quad S_{_{MSCE}}(t) = Y_7(t,t)$$

Approximation of the MSCE model: Derivation of μ_2^{eff}, t_{lag}, T_1, T_1^{eff}, T_2

From the backward Kolmogorov equations for the 4-stage MSCE model derived in the previous section we have

$$\frac{\partial \Phi_{_P}(\tau,t)}{\partial \tau} = -\alpha_{_P}\Phi_{_P}^2(\tau,t) - \beta_{_P} + [\alpha_{_P} + \beta_{_P} + \mu_2]\Phi_{_P}(\tau,t) - \mu_2\Phi_{_P}(\tau,t)\Phi_{_M}(\tau,t)$$

$$\Rightarrow \frac{\partial \Phi_{_P}(\tau,t)}{\partial \tau} = -\alpha_{_P}\Phi_{_P}^2(\tau,t) - \beta_{_P} + [\alpha_{_P} + \beta_{_P} + \mu_2(1 - \Phi_{_M}(\tau,t))]\Phi_{_P}(\tau,t) \tag{15}$$

This has the same form as the backward equation for the 3-stage MSCE-1 model with the mutation rate μ_2 from MSCE-1 assuming a time-dependent form, i.e,

$$\mu_2 \to \mu_2(1 - \Phi_{_M}(\tau, t)) = \mu_2(1 - S_{_M}(u)) \tag{16}$$

where $u = t - \tau$ and $S_{_M}(u)$ is the "survival function" for cancer detection of a preclinical cancer clone a time u since the clone was born (ie. τ is the time the first malignant cell of the clone was transformed). This is defined as (see full derivation using Kolmogorov forward equations in Luebeck et al. 2002 [1])

$$S_{_M}(u) = 1 + \frac{1}{\alpha_{_M}} \frac{p_{_M} q_{_M} e^{-p_{_M} u} - q_{_M} p_{_M} e^{-q_{_M} u}}{q_{_M} e^{-p_{_M} u} - p_{_M} e^{-q_{_M} u}} \tag{17}$$

$$p_{_M} = \frac{1}{2}(-(\alpha_{_M} - \beta_{_M} - \rho) - \sqrt{(\alpha_{_M} - \beta_{_M} - \rho)^2 + 4\alpha_{_M}\rho}) \tag{18}$$

$$q_{_M} = \frac{1}{2}(-(\alpha_{_M} - \beta_{_M} - \rho) + \sqrt{(\alpha_{_M} - \beta_{_M} - \rho)^2 + 4\alpha_{_M}\rho}) \tag{19}$$

Let p_∞ be the probability that a single preclinical clone initiated at time τ eventually becomes detected:

$$p_\infty = \lim_{u \to \infty} (1 - S_{_M}(u)) \approx 1 - \frac{\beta_{_M}}{\alpha_{_M}}$$

i.e., as $u \to \infty$, $S_{_M}(u)$ approaches approximately $\beta_{_M}/\alpha_{_M}$, which is the probability of extinction of the preclinical clone. From this we have that $\frac{1 - S_M(u)}{p_\infty} \to 1$ as $u \to \infty$. Hence, the u dependent mutation rate $\mu_2(1 - S_{_M}(u))$ is bounded between 0 and approximately $\mu_2(1 - \beta_{_M}/\alpha_{_M}) = \mu_2 p_\infty \equiv \mu_2^{eff}$, which takes into account the non-extinction of small preclinical clones before they are detected.

Thus, for fixed t, the PDE from Eq. (15) that we wish to integrate over $[0, t]$ can be written as

$$\frac{\partial \Phi_{_P}(\tau, t)}{\partial \tau} = -\alpha_{_P} \Phi_{_P}^2(\tau, t) - \beta_{_P} + \left[\alpha_{_P} + \beta_{_P} + \mu_2^{eff} \left(\frac{1 - S_{_M}(u)}{p_\infty}\right)\right] \Phi_{_P}(\tau, t) \tag{20}$$

with the initial condition $\Phi_{_P}(t, t) = 1$.

Assuming $(\alpha_{_M} - \beta_{_M}) \gg (\alpha_{_P} - \beta_{_P})$, the function $\frac{1 - S_M(u)}{p_\infty}$ has a steep ascent from near lower bound of 0 to upper bound of 1, with an inflection point u^* at $\frac{1-S_M(u^*)}{p_\infty} = \frac{1}{2} + O(\rho)$. For small ρ, u^* can be approximated as the following

$$\frac{d^2 S_M(u)}{du^2} = 0 \;\Rightarrow\; u^* = -\frac{\ln(-p_M/q_M)}{p_M - q_M} = -\frac{\ln(\alpha_M \rho/(\alpha_M - \beta_M)^2)}{\alpha_M - \beta_M} + O(\rho)$$

We now aim to show that this time u^* is approximately the mean sojourn time of a surviving preclinical tumor. As mentioned in the main text, the main effect of a preclinical tumor progression (the second clonal expansion in the MSCE model) on cancer incidence is a delay (or lag) of time for a cancer clone to grow from a single malignant cell into a detectable tumor, conditional that it doesn't become extinct. Let T be the random sojourn time from transformation of a single malignant cell to tumor detection. Again since $S_M(u) \to \beta_M/\alpha_M$, the proper cumulative distribution function for T is given by

$$\Pr[T \le u] = \frac{1 - S_M(u)}{p_\infty} \to 1 \quad \text{as } u \to \infty$$

Since T only takes non-negative values, we can use the following fact

$$E[T] = \int_0^\infty \Pr[T \ge u] du = \int_0^\infty 1 - \frac{1 - S_M(u)}{p_\infty} du$$

Then this mean sojourn time of a surviving tumor is defined as

$$T_2 \equiv E[T] = \int_0^\infty \left(1 - \frac{1 - S_M(u)}{p_\infty}\right) du = -\frac{\ln\left(\frac{q_M/(-p_M)}{1+q_M/(-p_M)}\right)}{(-p_M)} = -\frac{\ln(\alpha_M \rho/(\alpha_M - \beta_M)^2)}{\alpha_M - \beta_M} + O(\rho). \tag{21}$$

Therefore, the point of inflection u^* of the function $\frac{1-S_M(u)}{p_\infty}$ equals approximately T_2 when ρ is very small. We will continue by approximating $\left(\frac{1-S_M(u)}{p_\infty}\right)$ in Eq. (20) by a piecewise constant function on $[0, t]$ such that

$$\frac{1 - S_M(u)}{p_\infty} = \begin{cases} 0 & \text{if } u \in [0, t_{lag}) \\ 1 & \text{if } u \in [t_{lag}, t] \end{cases} \tag{22}$$

where $t_{lag} < t$ will be set equal to T_2. First, we rewrite our ODE for fixed t as before in terms of $u = t - \tau$ and solve on the first interval $[0, t_{lag}]$ with initial condition $\Phi_P(0,t) = 1$:

$$\begin{aligned}
\frac{d\Phi_P(u,t)}{du} &= -\alpha_P \Phi_P^2(u,t) - \beta_P + \left[\alpha_P + \beta_P + \mu_2^{eff} \cdot 0\right] \Phi_P(u,t) \\
\Rightarrow \frac{d\Phi_P(u,t)}{du} &= -\alpha_P \Phi_P^2(u,t) - \beta_P + \left[\alpha_P + \beta_P\right] \Phi_P(u,t) \\
&\Rightarrow \Phi_P(u,t) = 1 \quad \text{for all } u \in [0, t_{lag})
\end{aligned}$$

Next, on $[t_{lag}, t]$, we are solving for a shifted 2-stage survival function with initial condition $\Phi_P(t_{lag}, t) = \Phi_P(0, t - t_{lag}) = 1$. (See Luebeck et al. 2002 for solution details [1])

$$\begin{aligned}
\frac{d\Phi_P(u,t)}{du} &= -\alpha_P \Phi_P^2(u,t) - \beta_P + \left[\alpha_P + \beta_P + \mu_2^{eff}\right] \Phi_P(u,t) \\
\Rightarrow \Phi_P(u,t) &= \left(\frac{q_P - p_P}{q_P e^{-p_P(u - t_{lag})} - p_P e^{-q_P(u - t_{lag})}} \right)^{\mu_1/\alpha_P} \quad \text{for } u \in [t_{lag}, t]
\end{aligned}$$

with

$$\begin{aligned}
p_P &= \frac{1}{2}(-(\alpha_P - \beta_P - \mu_2^{eff}) - \sqrt{(\alpha_P - \beta_P - \mu_2^{eff})^2 + 4\alpha_P \mu_2^{eff}}), && (23) \\
q_P &= \frac{1}{2}(-(\alpha_P - \beta_P - \mu_2^{eff}) + \sqrt{(\alpha_P - \beta_P - \mu_2^{eff})^2 + 4\alpha_P \mu_2^{eff}}). && (24)
\end{aligned}$$

The mathematical approximation for the MSCE model takes the following form

$$h_{MSCE}(t) \approx h_{MSCE-1}^{eff}(t) = \mu_0 X \left(1 - \left(\frac{q_P - p_P}{q_P e^{-p_P(t - t_{lag})} - p_P e^{-q_P(t - t_{lag})}} \right)^{\mu_1/\alpha_P} \right), \qquad (25)$$

with $\mu_2^{eff} \equiv \mu_2(1 - \beta_M/\alpha_M)$ and $t_{lag} \equiv T_2$

Therefore, we can compute an analytical approximation to the MSCE model hazard function for incidence by using the form of the hazard of the MSCE-1 model, $h_{MSCE-1}(t)$ (See Jeon et al. 2008 for full derivation not given here) and replacing μ_2 with $\mu_2^{eff} \equiv \mu_2(1-\beta_M/\alpha_M)$ and time t with $t - t_{lag} \equiv t - T_2$. This equivalency can be seen from the derivation above containing the solution of $\Phi_P(u,t)$ on $[t_{lag}, t]$ with guaranteed survival from $[0, t_{lag}]$. This approximation implicitly assumes that, conditional on non-extinction, the first surviving malignancy in the tissue is eventually the first to be observed, described further in the main text. Alongside this altered, effective mutation rate, the time shift by the mean sojourn time of the first surviving tumor, T_2 will push the MSCE-1 type hazard function to the right to coincide with the actual MSCE hazard curve. Figure 2 in the main text displays the comparison between this analytical approximation and the numerical MSCE hazard solution.

Analogous to the derivation for this mean sojourn time of a malignant clone conditioned on survival, we can compute other sojourn times, as presented in Table 1. The first, T_1, refers to the mean time from the initiation of a premalignant cell to the appearance of the first malignant cell from this premalignant cell's progeny. The survival function for this process $S_P(u)$ corresponds to the usual 3-stage MSCE-1 model with malignant transformation rate μ_2, defined analogously to Eq. (17) with premalignant parameters. Next, T_1^{eff} refers to the mean time from the initiation of a premalignant cell until, after it clonally expands, the first transformation of a malignant, preclinical cell conditional that this malignant cell's clone will survive. The survival function for this process $S_P^{eff}(u)$ corresponds to the usual 3-stage MSCE-1 model with malignant transformation rate $\mu_2^{eff} = \mu_2 p_\infty$ (see Eq. (17)), which as discussed above, guarantees the non-extinction of small preclinical clones before they are detected. Therefore, similar to Eq. (21), we compute these two other mean sojourn times as follows

$$T_1 = \int_0^\infty \left(1 - \frac{1 - S_P(u)}{p_\infty}\right) du \approx -\frac{\ln(\alpha_P \mu_2/(\alpha_P - \beta_P)^2)}{\alpha_P - \beta_P}. \tag{26}$$

$$T_1^{eff} = \int_0^\infty \left(1 - \frac{1 - S_P^{eff}(u)}{p_\infty}\right) du \approx -\frac{\ln(\alpha_P \mu_2^{eff}/(\alpha_P - \beta_P)^2)}{\alpha_P - \beta_P}. \tag{27}$$

CRC males	λ x 10^{-4}	g_P	g_M	μ_2^{eff} x 10^{-6}
α_P [2, 50]	[2.13, 2.13]	[0.162, 0.162]	[2.58, 2.59]	[0.15, 3.76]
α_M [25,100]	[2.13, 2.13]	[0.162, 0.162]	[2.52, 2.67]	[0.73, 0.77]
ρ [10^{-6},10^{-8}]	[2.13, 2.13]	[0.162, 0.162]	[2.08, 3.07]	[0.75, 0.76]

CRC females	λ x 10^{-4}	g_P	g_M	μ_2^{eff} x 10^{-6}
α_P [2, 50]	[1.57, 1.57]	[0.149, 0.149]	[2.04, 2.05]	[0.32, 8.09]
α_M [25,100]	[1.57, 1.57]	[0.149, 0.149]	[1.92, 2.17]	[1.61, 1.62]
ρ [10^{-6},10^{-8}]	[1.57, 1.57]	[0.149, 0.149]	[1.61, 2.46]	[1.61, 1.63]

Table S1: Sensitivity analysis to test the effect of uncertainty in the cell division rates α_P and α_M, and in the detection rate ρ on the estimated MSCE model parameters.

Parameter estimates, correlations, and uncertainty

Fits to SEER incidence data In the following we describe the fits of SEER-9 incidence data of CRC, GaC, PaC, and EAC using both the MSCE model and its approximation (referred to as MSCE-1) [6]. A comparison shows that the estimated lag-time parameters for the MSCE-1 approximation are in generally good agreement with corresponding estimates for T_2 in the MSCE model assuming biologically plausible values for the cell division rates of premalignant cells ($\alpha_P \in [2, 50]$ per year) and malignant cells ($\alpha_M \in [25, 100]$ per year, Wilson et al., 1993 [7]) and for the observation event rate $\rho \in [10^{-6}, 10^{-8}]$ per cell per year. See Table S1 for the results of the sensitivity analysis of $\lambda, g_P, g_M, \mu_2^{eff}$ on choices for α_P, α_M, ρ.

In general, the exact solution of the MSCE hazard function with two consecutive clonal expansions may be difficult to distinguish from a model with a single clonal expansion plus a time lag. This also means that the parameters of the model in Figure 1, in particular, the parameters that pertain to preclinical cancer progression, may be difficult to estimate from cancer incidence data alone. Additional assumptions such as equality of the mutation rates μ_0, μ_1 and μ_2, replacing some of the parameters with measured values, or inclusion of screening data on the number and sizes of preclinical tumors, are required to estimate all relevant model parameters. For EAC females, we fixed the parameter μ_2^{eff} to the MCMC estimate obtained for males (4.65×10^{-6}). This choice stabilized the estimation and yielded a sex ratio of 2:1 for the BE rate ν consistent with the male-to-female BE prevalence ratios seen in epidemiological studies [8,9].

Figures S1 (a-h) show scatterplots of the Markov Chain Monte Carlo (MCMC) samples drawn from the posterior distributions of the MSCE model parameters for CRC, GaC, PaC, and EAC using uniform prior distributions defined on positive intervals $(0, C]$ with finite but large upper limits C. We ran 8 independent chains with 10000 cycles each. For EAC, we doubled the number of cycles for each run. All runs were started with the parameters set at (or near) their respective maximum likelihood estimates (MLEs) and appeared to converge rapidly after a short 1000 cycle burn-in period. The pairwise scatterplots also show the locations/values of the parameters associated with the 10 best likelihood values (red marks) found in each MCMC batch. These values are generally very close to the MLEs when the likelihood maximization yielded stable parameter estimates. The parameter T_2, defined in the previous section, is the mean sojourn time of a malignancy. It is not a free parameter but a composite quantity defined by Eq. (21).

Birth cohort and calendar year trends

We used a modified age-period-cohort (APC) approach to adjust the incidences for secular trends. While the period and cohort effects are modeled non-parametrically, the age effect follows the hazard function of the MSCE model and is therefore parametrically constrained. This finesses a non-trivial identifiably problem of the APC approach and allows us to separate cleanly age, period and cohort effects (see Luebeck and Moolgavkar, 2002; Meza et al., 2008, 2010 [1,3,4]). Briefly, the APC ansatz assumes an incidence function [4]

$$I_{bc}(t) = \Theta_b \; \Theta_c \; h_{_{MSCE}}(t) \tag{28}$$

where Θ_b and Θ_c are coefficients that modify the MSCE-2 hazard $h_{_{MSCE}}(t)$.

The calendar year coefficients c_j and birth cohort coefficients $b_{_i}$ were estimated via maximum likelihood jointly with all other model parameters with the exception of $\alpha_{_M}$ and ρ, which were kept fixed (see Figure 1 and text). Figures S2 a and b show the estimated birth cohort coefficients (left panel) and calendar year coefficients (right panel). The birth cohort 1920-1924 was used as a reference and anchored to 1. Similarly, the calendar year 1985 was used as a reference, i.e., $c_{1985} = 1$ for GaC and PaC. For EAC we set $c_{1975} = 1$ because of the very strong increase of the incidences with calendar year. For CRC no calendar year reference was necessary since the observed period effects are likely due to colon screening which rarely occurs in people 50 years or younger. Thus, CRC had an internal calendar year anchoring as the coefficients c_j were set to 1 for all ages less than 54 and 57 for females and males, respectively.

For CRC and GaC there appears to be an increase of the incidence with younger cohorts (after 1955), however the confidence intervals are very wide and the effect highly uncertain. In contrast, with the exception of PaC, which remains relatively flat over the past 30 years, we find strong and significant increases of EAC incidence with calendar year (6-7 fold, for males and 4 fold for females) and moderate but significant decreases for CRC and GaC. Secular trends for CRC are discussed in detail in Meza et al. (2010) [4].

Figures S3 (a-h) show age-specific incidences of CRC, GaC, PaC and EAC in SEER-9 from 1975-2008 by gender and 5 year calendar periods. The left panels show the raw (unadjusted) incidences whereas the right panels show period and cohort adjusted incidence curves as well as the MSCE model prediction. For CRC, GaC and PaC both the exponential and the linear behavior of the adjusted curves stand out. For EAC, the exponential-linear behavior is masked by the presence of an additional step in the model which represents a tissue conversion from normal squamous epithelium to Barrett's metaplasia.

References

1. Luebeck EG, Moolgavkar SH. Multistage carcinogenesis and the incidence of colorectal cancer. Proc Natl Acad Sci U S A 2002;99:15095-100.

2. Jeon J, Luebeck EG, Moolgavkar SH. Age effects and temporal trends in adenocarcinoma

of the esophagus and gastric cardia (United States). Cancer Causes Control 2006;17:971-81.

3. Meza R, Jeon J, Moolgavkar SH, Luebeck EG. Age-specific incidence of cancer: Phases, transitions, and biological implications. Proc Natl Acad Sci U S A 2008;105:16284-9.

4. Meza R, Jeon J, Renehan AG, Luebeck EG. Colorectal cancer incidence trends in the United States and United kingdom: evidence of right- to left-sided biological gradients with implications for screening. Cancer Res. 2010;70:5419-29.

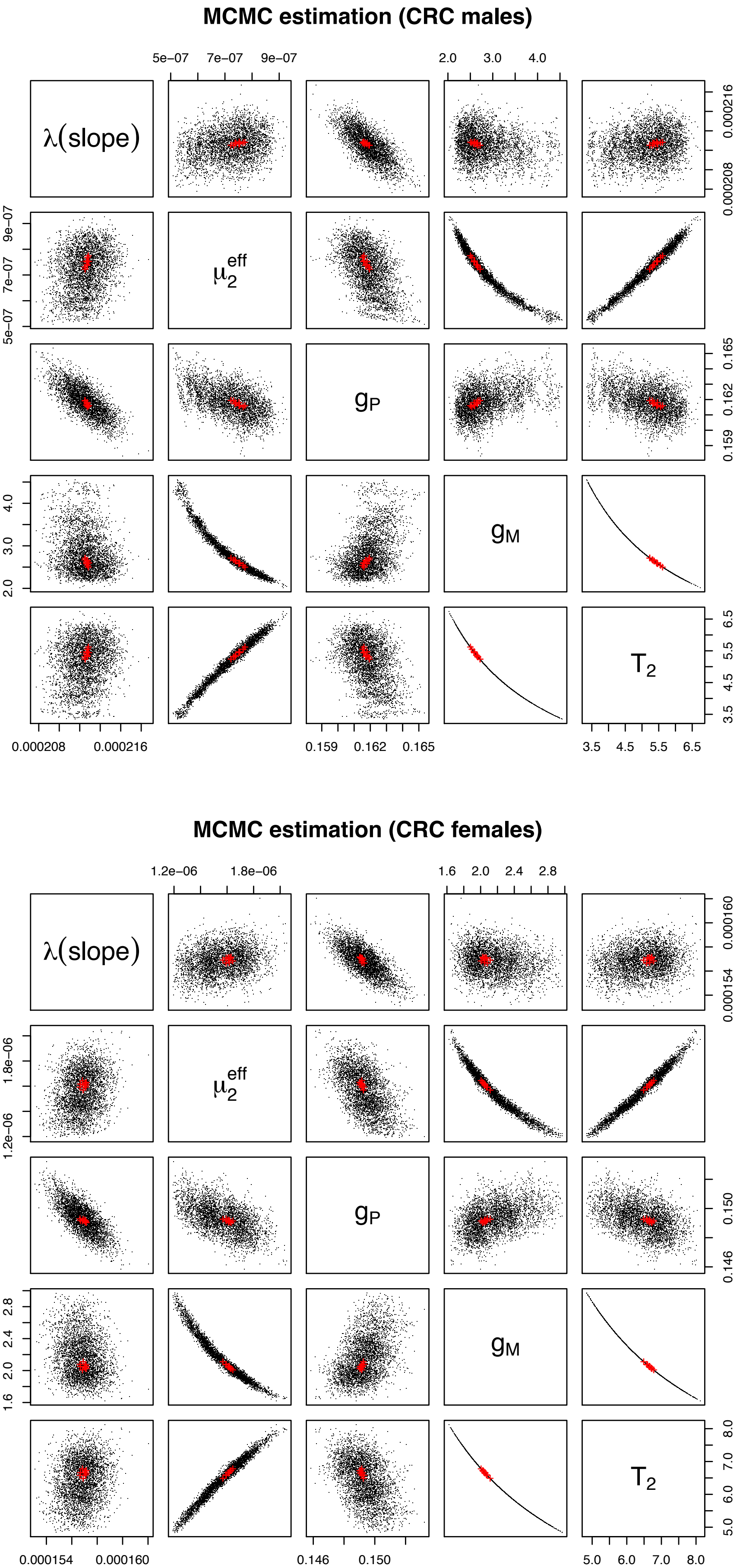
5. Crump CS, Subramaniam RP, Van Landingham CB. A Numerical Solution to the Non-homogeneous Two-Stage MVK Model of Cancer. Risk Analysis 2005;25:921-926.

6. Surveillance, Epidemiology, and End Results (SEER) Program SEER*Stat Database: Incidence -SEER 9 Regs Limited-Use, Nov 2009 Sub (1973-2007) (Katrina/Rita Population Adjustment) - Linked To County Attributes - Total U.S., 1969-2007 Counties. April 2010, based on the November 2009 submission ed: National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch. [last accessed on 2012 Mar 13] Available from: http://seer.cancer.gov/.

7. Wilson MS, West CML, Wilson GD, Roberts SA, James RD, Schofield PF. Intra-tumoral heterogeneity of tumour potential doubling times (T_{pot}) in colorectal cancer. Br. J. Cancer 1993;68:501-6.

8. Cook MB, Wild CP, Forman D. A Systematic Review and Meta-Analysis of the Sex Ratio for Barretts Esophagus, Erosive Reux Disease, and Nonerosive Reux Disease. Am. Journal of Epidemiology 2005;162:1050-61.

9. Falk GW, Prashanthi NT, Richter JE, Connor JT, Wachsberger DM. Barretts Esophagus in Women: Demographic Features and Progression to High-Grade Dysplasia and Cancer. Clinical Gastro. and Hep. 2005;3:1089-94.



MCMC estimation (CRC males)

MCMC estimation (CRC females)

Figure S1 a and b: Scatterplots of the Markov Chain Monte Carlo (MCMC) samples drawn from the posterior distribution of the model parameters for male (a) and female (b) CRC incidence in SEER-9. For further details, see text.

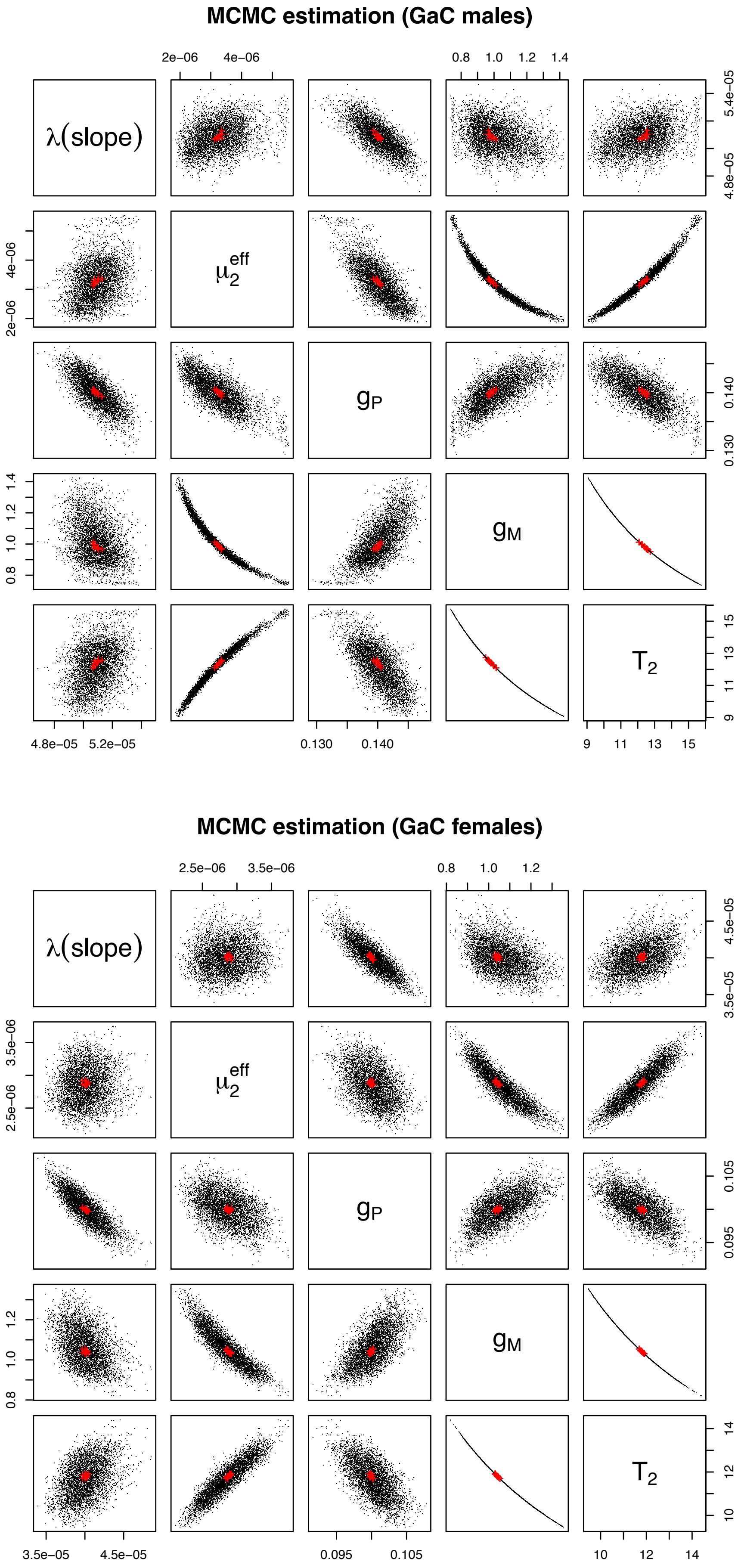
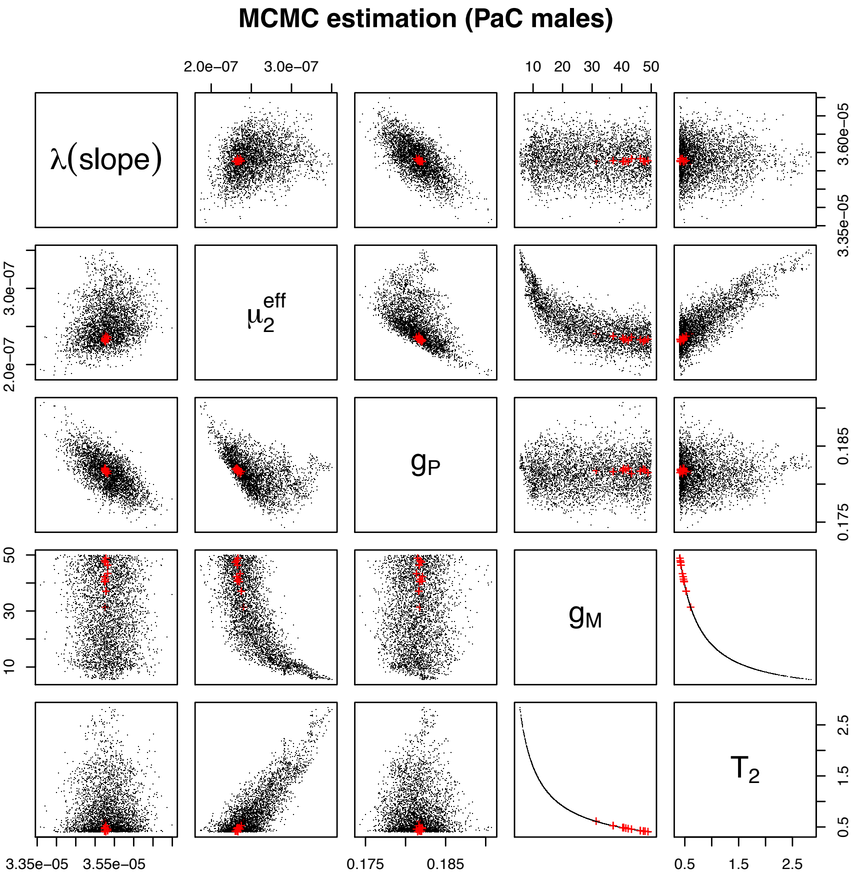


Figure S1 c and d: Same as S1 a and b, but for GaC incidence.



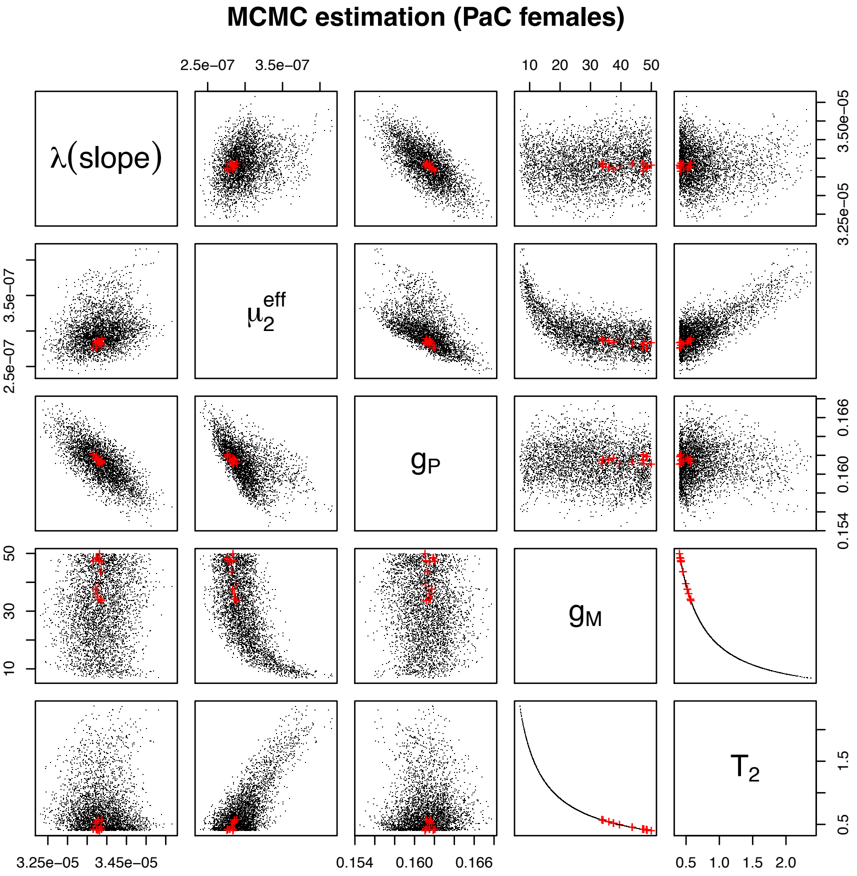
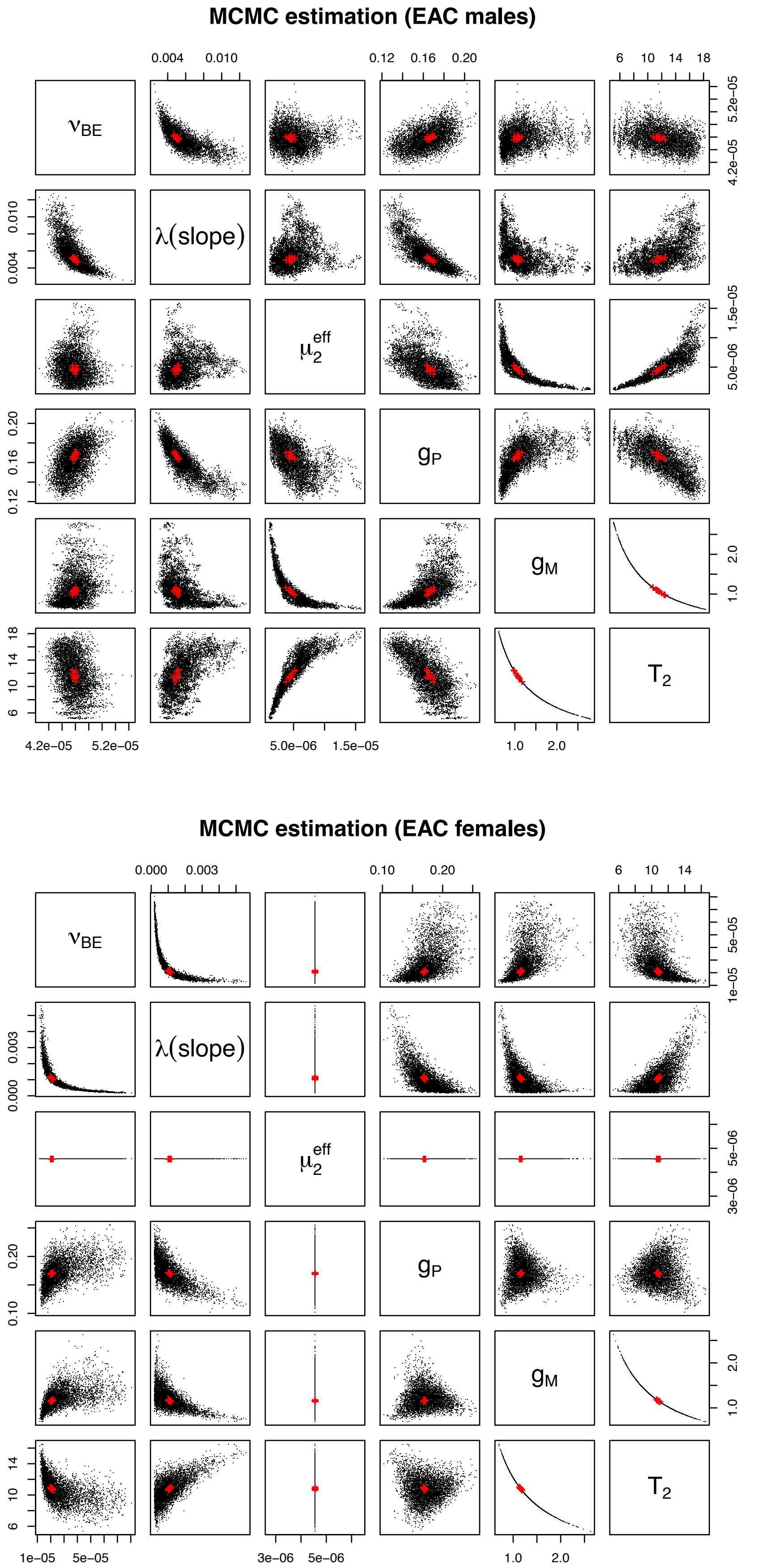


Figure S1 e and f: Same as S1 a and b, but for PaC incidence.


MCMC estimation (EAC males)

MCMC estimation (EAC females)

Figure S1 g and h: Same as S1 a and b, but for EAC incidence.

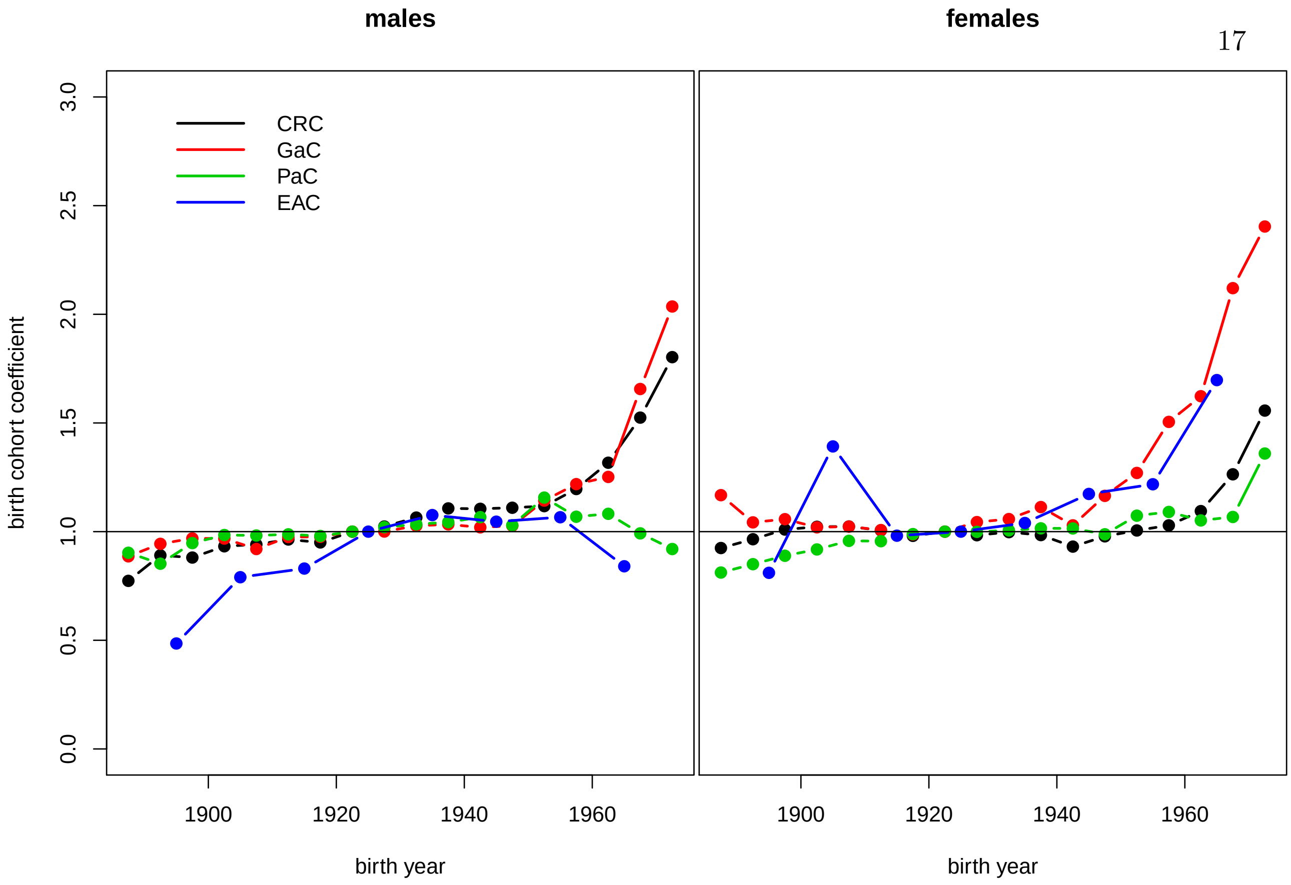


Figure S2 a: Maximum likelihood estimates for the birth-cohort coefficients. See Eq.(28).

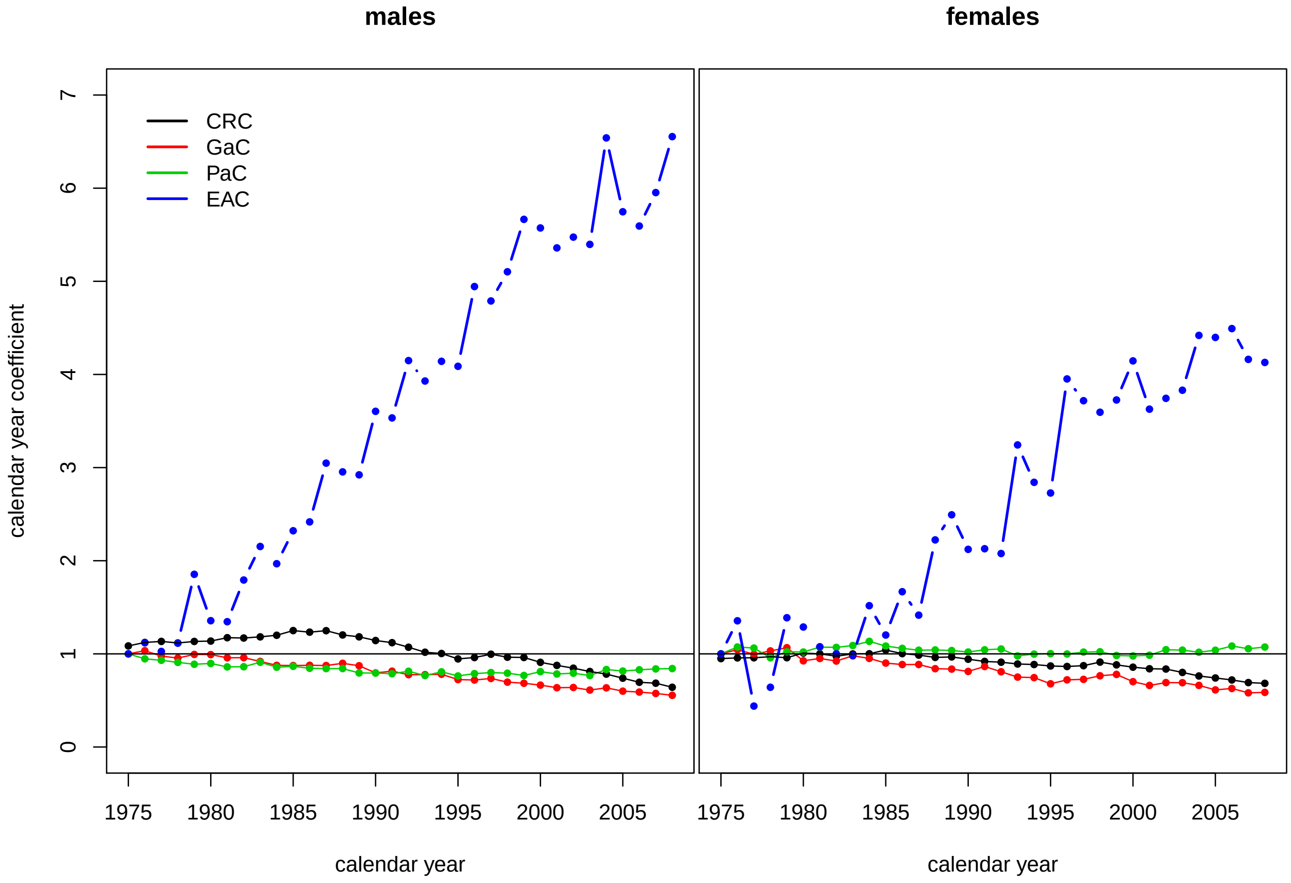


Figure S2 b: Maximum likelihood estimates for the calendar-year coefficients. See Eq.(28).

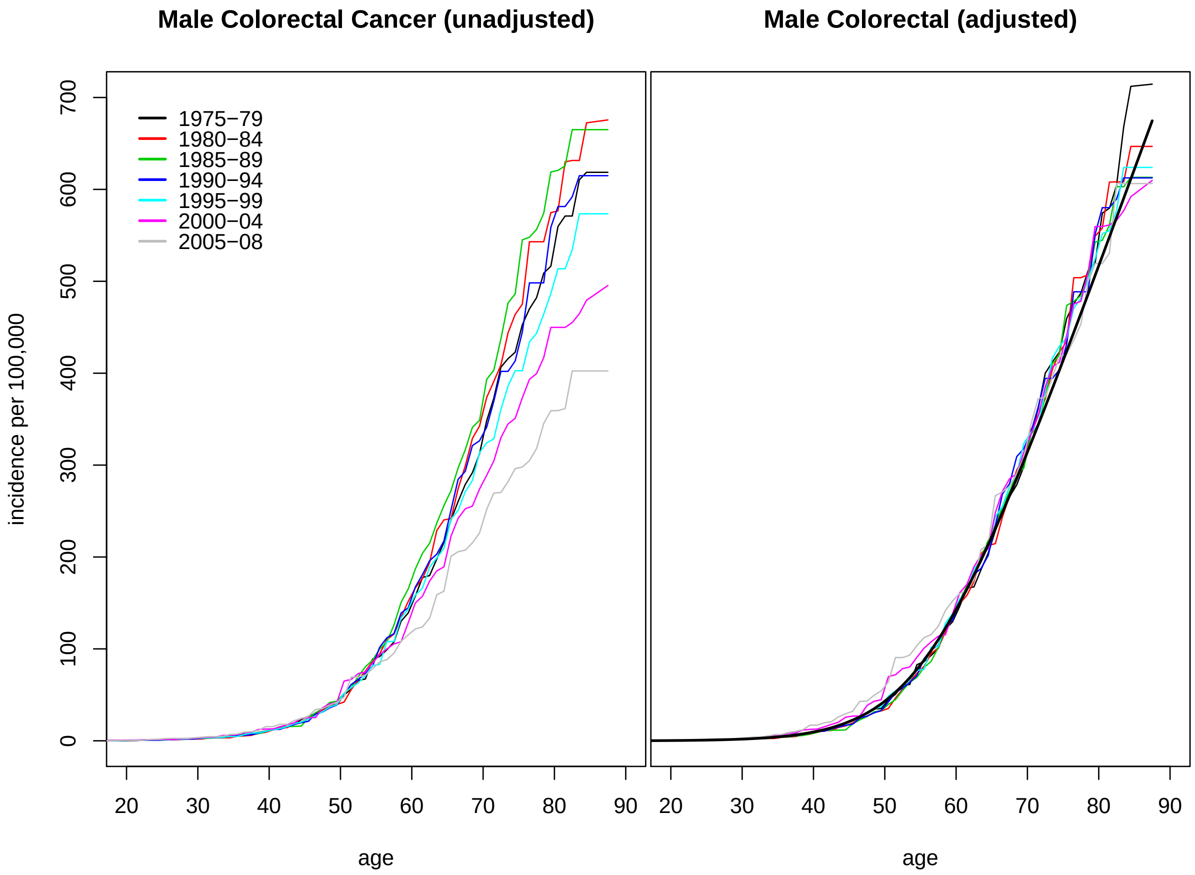


Figure S3 a: Left panel: unadjusted CRC incidences among males from SEER-9 by 5 year calendar periods; right panel: adjusted incidences using the estimated calendar-year and birth-cohort coefficients, and the MSCE model fit (thick black line).

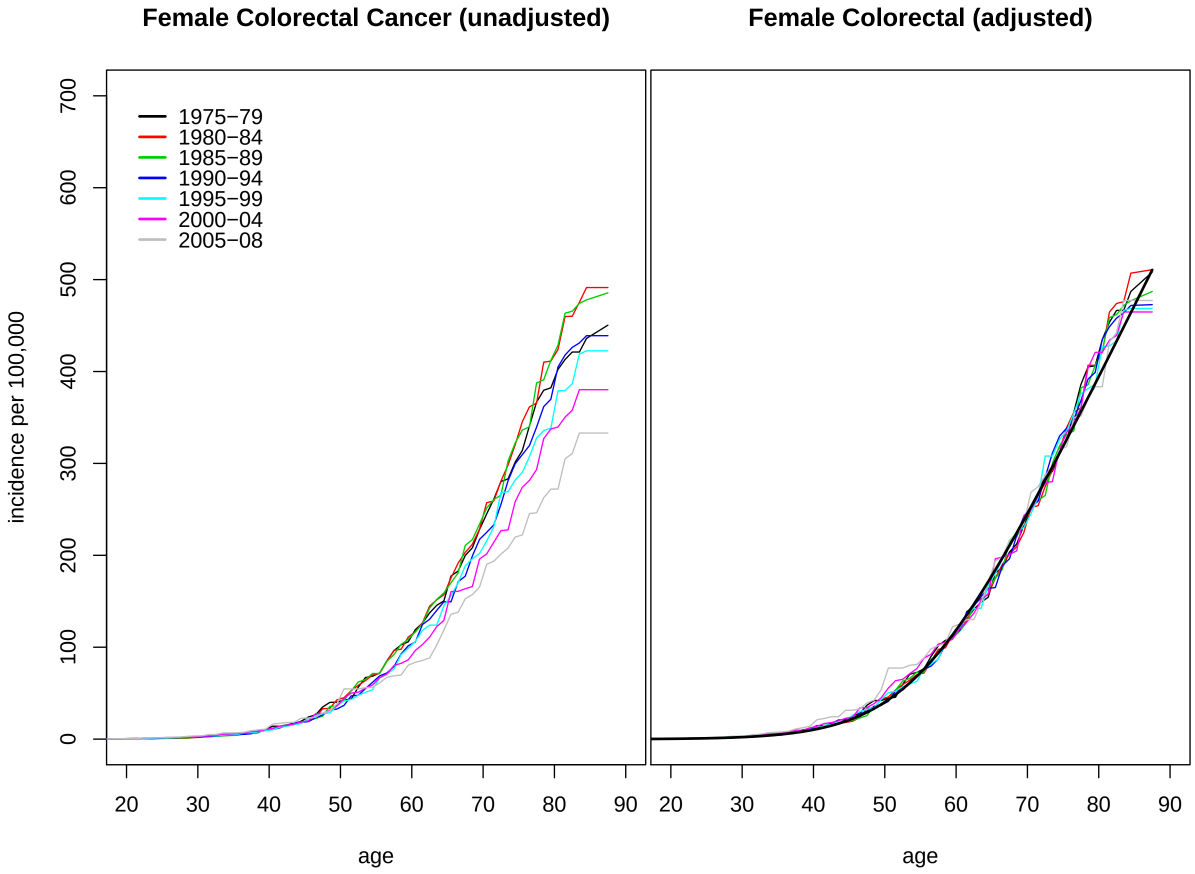


Figure S3 b: Same as S3 a, but for CRC females.

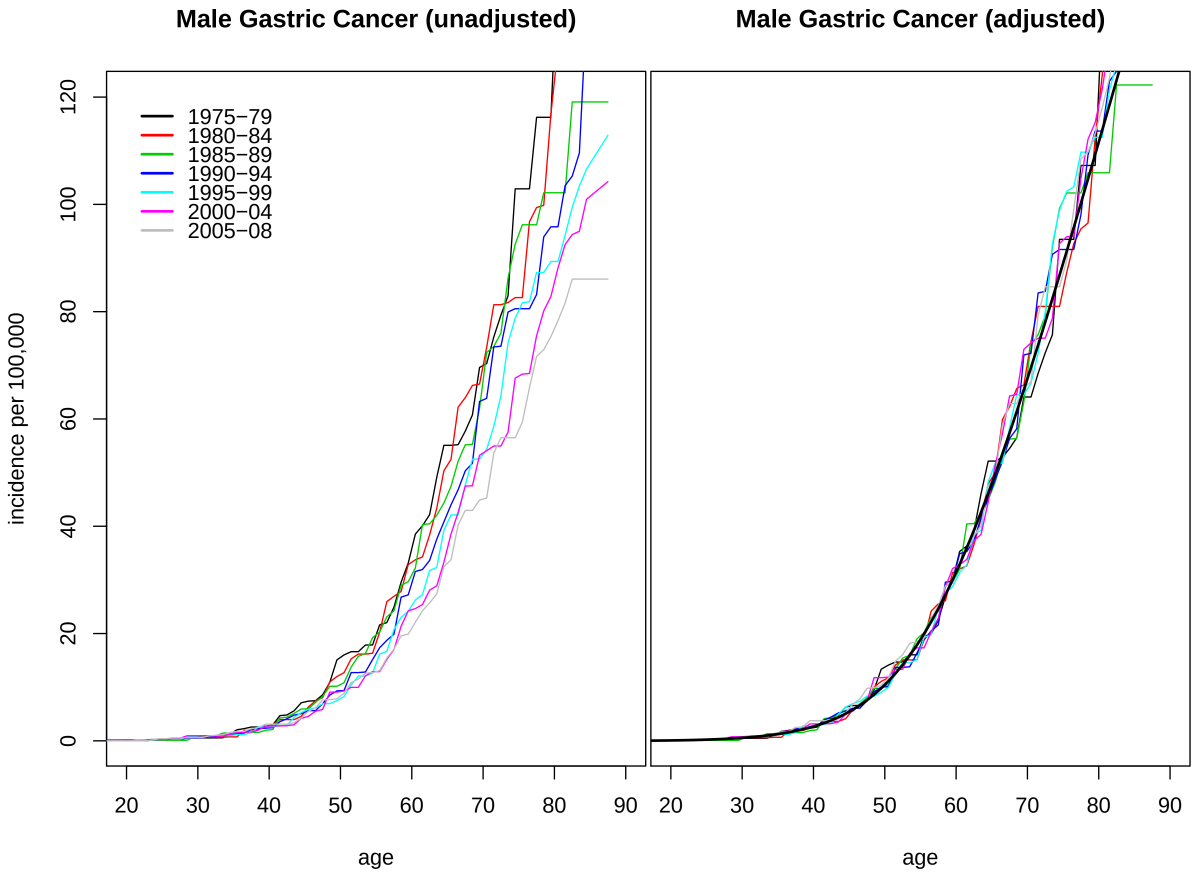


Figure S3 c: Same as S3 a, but for GaC males.

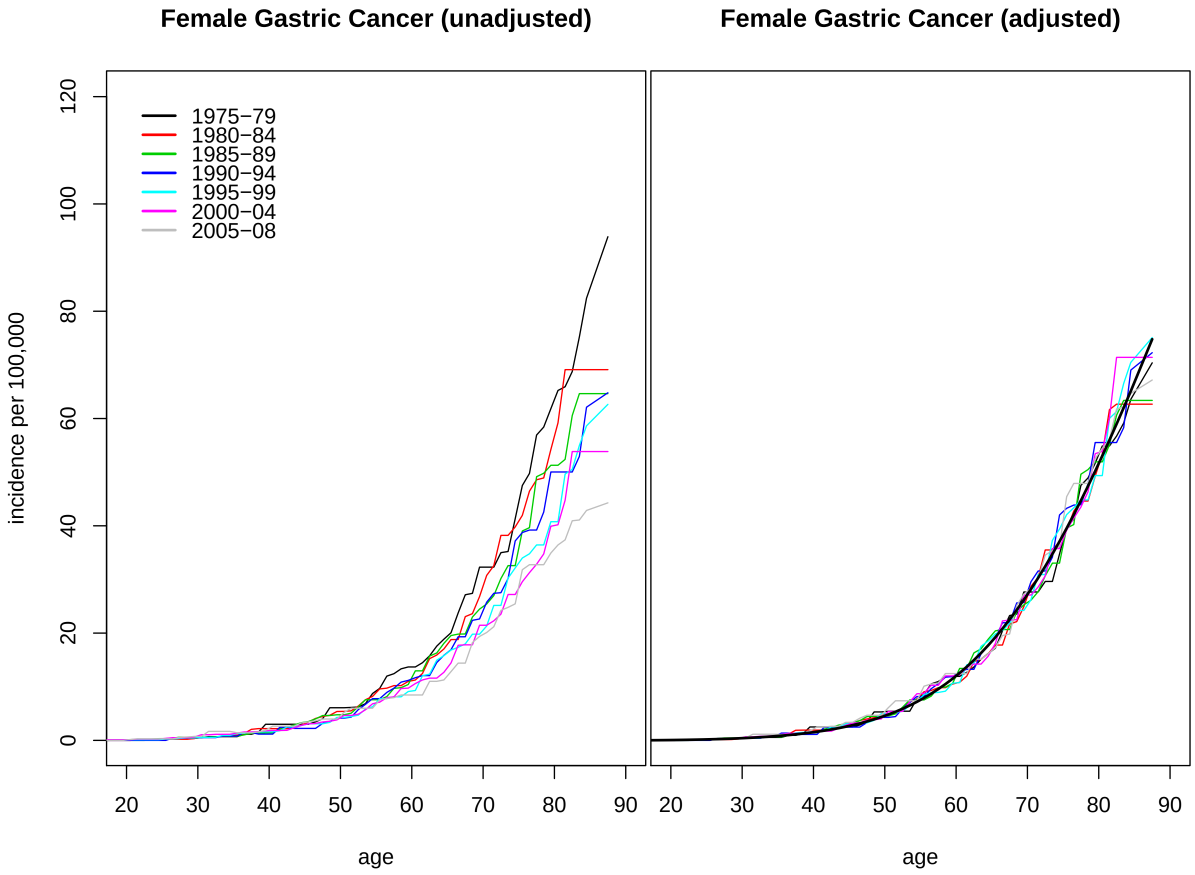


Figure S3 d: Same as S3 b, but for GaC females.

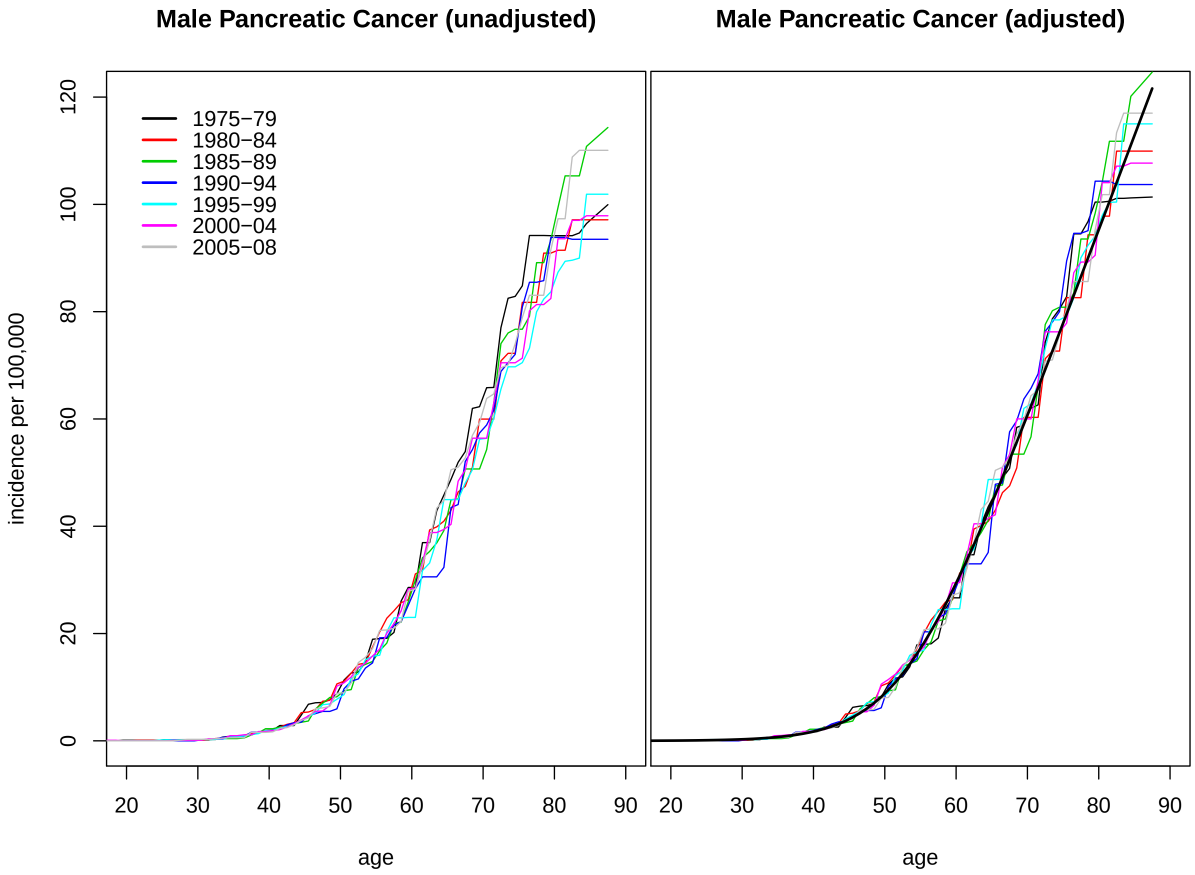


Figure S3 e: Same as S3 a, but for PaC males.

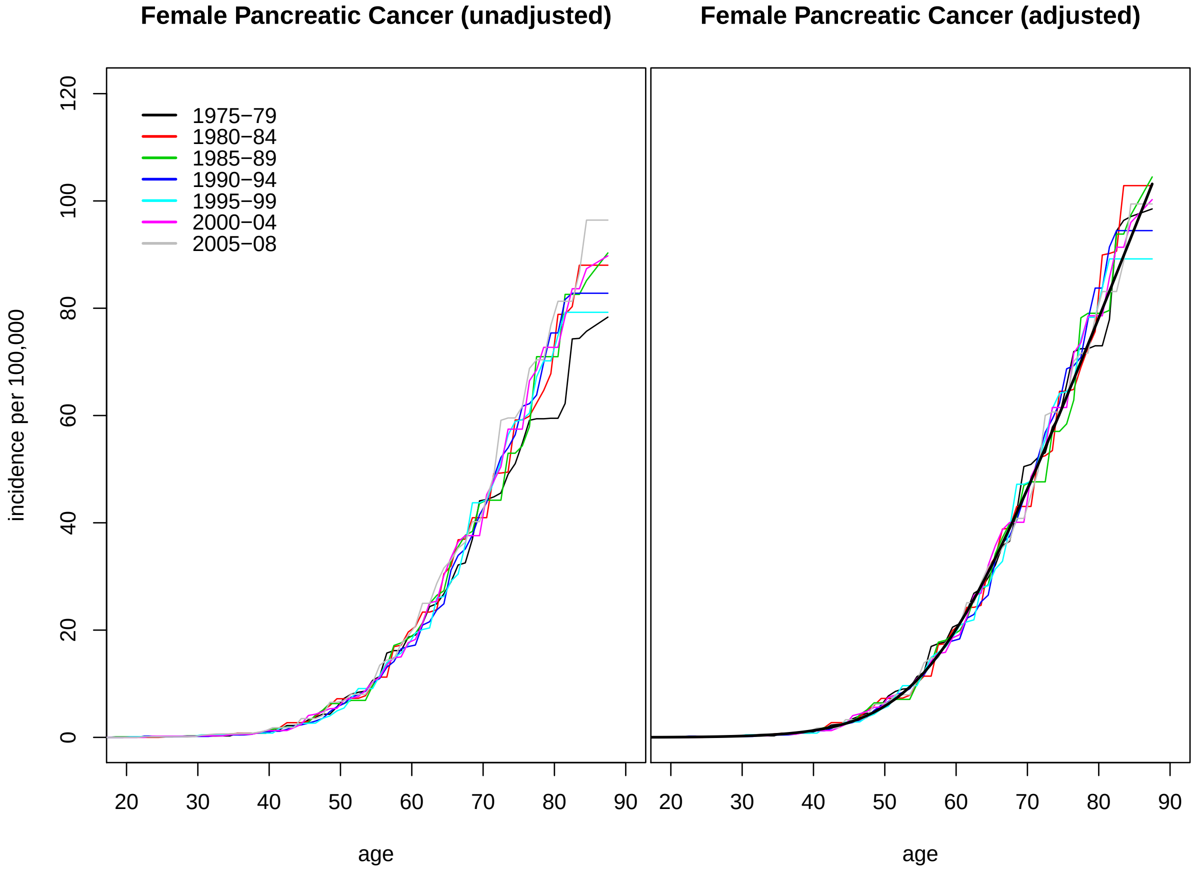


Figure S3 f: Same as S3 b, but for PaC females.

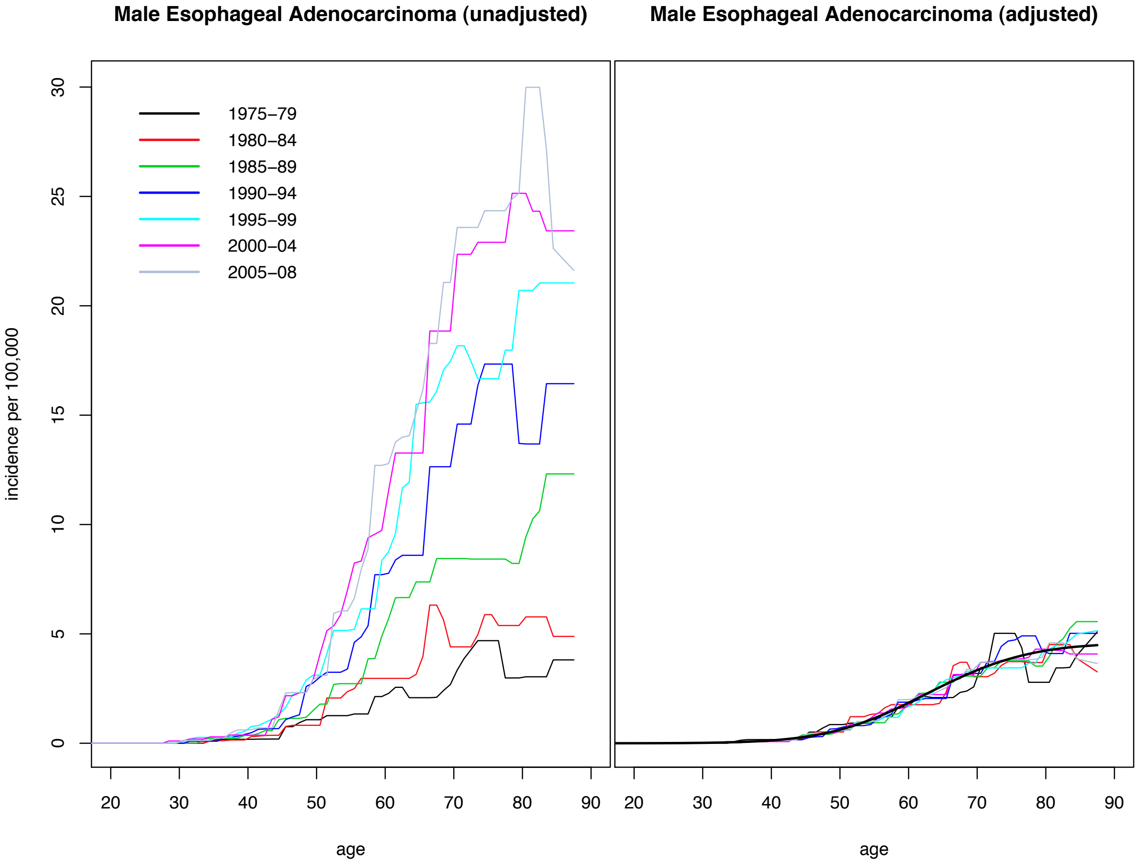


Figure S3 g: Same as S3 a, but for EAC males.

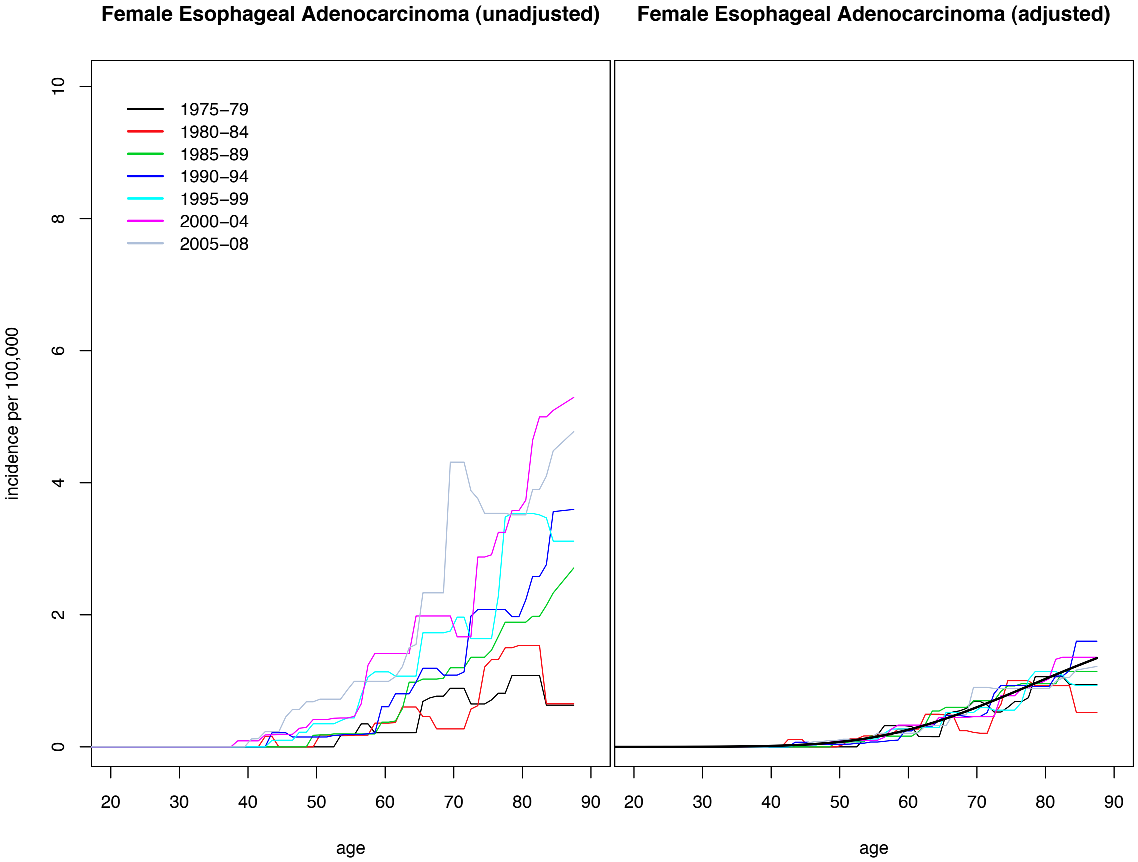


Figure S3 h: Same as S3 b, but for EAC females.