Number and Size Distribution of Colorectal Adenomas under the Multistage Clonal Expansion Model of Cancer

Anup Dewanji[1], Jihyoun Jeon[2], Rafael Meza[2,3], E. Georg Luebeck[4]*

[1] Applied Statistics Division, Indian Statistical Institute, Kolkata, India, [2] Program in Biostatistics and Biomathematics, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America, [3] Department of Epidemiology, University of Michigan, Ann Arbor, Michigan, United States of America, [4] Program in Computational Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America

Abstract

Colorectal cancer (CRC) is believed to arise from mutant stem cells in colonic crypts that undergo a well-characterized progression involving benign adenoma, the precursor to invasive carcinoma. Although a number of (epi)genetic events have been identified as drivers of this process, little is known about the dynamics involved in the stage-wise progression from the first appearance of an adenoma to its ultimate conversion to malignant cancer. By the time adenomas become endoscopically detectable (i.e., are in the range of 1–2 mm in diameter), adenomas are already comprised of hundreds of thousands of cells and may have been in existence for several years if not decades. Thus, a large fraction of adenomas may actually remain undetected during endoscopic screening and, at least in principle, could give rise to cancer before they are detected. It is therefore of importance to establish what fraction of adenomas is detectable, both as a function of when the colon is screened for neoplasia and as a function of the achievable detection limit. To this end, we have derived mathematical expressions for the detectable adenoma number and size distributions based on a recently developed stochastic model of CRC. Our results and illustrations using these expressions suggest (1) that screening efficacy is critically dependent on the detection threshold and implicit knowledge of the relevant stem cell fraction in adenomas, (2) that a large fraction of non-extinct adenomas remains likely undetected assuming plausible detection thresholds and cell division rates, and (3), under a realistic description of adenoma initiation, growth and progression to CRC, the empirical prevalence of adenomas is likely inflated with lesions that are not on the pathway to cancer.

Citation: Dewanji A, Jeon J, Meza R, Luebeck EG (2011) Number and Size Distribution of Colorectal Adenomas under the Multistage Clonal Expansion Model of Cancer. PLoS Comput Biol 7(10): e1002213. doi:10.1371/journal.pcbi.1002213

Editor: Marek Kimmel, Rice University, United States of America

Received March 7, 2011; **Accepted** August 14, 2011; **Published** October 13, 2011

Funding: This work was supported by National Institutes of Health Grant RO1 CA107028-05A1. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: gluebeck@fhcrc.org

Introduction

Adenomatous polyps (or adenomas) in the large intestine are considered benign precursors of colorectal cancer (CRC) and both clinical and molecular evidence suggest that they may sojourn for many years before turning into cancer [1,2]. For this reason, adenomas are considered a primary intervention target if detected and removed before they become malignant. However, questions remain regarding the significance of their histopathology, molecular signatures, as well as their number and sizes in average risk individuals. Since endoscopic screening for neoplastic lesions is generally limited by macroscopic detection thresholds (of the order of a few mm in caliper size), a large fraction of adenomas may actually be missed, especially if the bulk of adenomas is too small for detection. Potentially, such "occult" adenomas could give rise to cancer before they are detected by endoscopy. Here we use a biologically-based model of colorectal carcinogenesis, which has previously been fitted to the age-specific incidence of CRC, to compute the number and size distributions of adenomas. Of particular interest is the fraction of detectable adenomas, as functions of age, detection threshold and the underlying cell kinetics in the adenomas.

The underlying multistage clonal expansion (MSCE) model for CRC upon which our results are based explicitly considers the initiation, promotion and malignant conversion of adenomas [3–8]. According to this model, adenomas arise from normal colonic stem cells that suffer at least two rare rate-limiting events. We interpret these events as the biallelic inactivation of a tumor suppressor gene, in particular the APC tumor suppressor gene, which is the gene responsible for familial adenomatous polyposis (FAP), and which is frequently mutated in colorectal neoplasia [9]. The inactivation of APC is understood to occur in colonic crypts (the fundamental proliferative unit in the colon) whose stem cells have previously acquired a mutation at one of the two APC alleles. Because the process of adenoma formation may involve additional genes (such as KRAS), we extend the model framework to accommodate additional rate-limiting mutations for the initiation of an adenoma and generalize the mathematical derivation of their number and size distribution accordingly. However, there is both clinical and experimental evidence that the number of requisite rate-limiting events or mutations for adenoma initiation is small. Once a stem cell is initiated in this model, it is free to proliferate. The basic version of our CRC model assumes that adenoma initiation occurs when the remaining wild-type copy of the APC tumor suppressor gene is deleted or mutated in a stem cell of a (pre-initiated) APC+/− colonic crypt. In a more realistic model, which is supported by recent experimental findings in the murine system [10], we also model the transient amplification of

Author Summary

The adenomatous polyp (or adenoma) is considered the common precursor lesion for colorectal cancer (CRC). Although the natural history of adenomas is well-characterized in terms of their histopathology and (epi)genomic changes, little is known about their dynamics in the stage-wise progression from the first appearance of an adenoma to its conversion to malignant cancer. By the time adenomas become endoscopically detectable (i.e., are in the range of 1–2 mm in diameter), adenomas are already comprised of hundreds of thousands of cells. A large fraction of adenomas may therefore remain undetected during screening and, in spite of their small (subthreshold) size, could give rise to cancer prior to being detected. It is therefore of importance to establish what fraction of adenomas is detectable, both as a function of the age at screening for colorectal neoplasia and the size (threshold) above which adenomas can be detected reliably. Here we derive mathematical expressions for the distribution of adenoma number and sizes based on a recently developed stochastic model for CRC, which has previously been calibrated and validated against age-specific CRC incidence data.

APC$-/-$ stem cells prior to their clonal expansion, effectively adding a stage to the initiation process [10–12].

The theoretical results derived here are complemented by model predictions for the adenoma size distributions and their (age-specific) prevalence based on parameter estimates obtained previously from fitting cancer incidence data. Since not all biological model parameters can be directly estimated from incidence data alone (non-identifiability issue), we explore the sensitivity of our findings by varying unknown parameters, such as the cell division rate of initiated stem cells, within their plausible ranges. In spite of the model uncertainties and the lack of precise clinical data on adenoma number and sizes, a biologically based approach that is broadly consistent with the pathogenesis of CRC makes it possible to explore more rationally the impact of risk factors and interventions on adenoma development and cancer progression.

Model

The Multistage Clonal Expansion (MSCE) Model

First we briefly review the MSCE model for CRC and then introduce the notation for the relevant stochastic processes involved in the formation of adenomas. We have previously derived expressions for the number and size distribution of non-extinct pre-malignant clones in the context of the two-stage clonal expansion (TSCE) model [13]. This model assumes that the clones develop from a (deterministic) source of progenitor cells via a non-homogeneous Poisson process. An important extension of this work was put forward by Dewanji et al. [14] for the size distribution of a random sum of Poisson-generated (pre-malignant) clones, which corresponds to a generalized Luria-Delbrück (GLD) distribution for mutant colonies. A hallmark of this distribution is a long tail reflecting large fluctuations of the total (mutant) population size. A further extension derived expressions for the number and size distributions of pre-malignant clones conditioning on observations from individuals who have not previously been diagnosed with CRC [6].

Adenoma development. For colorectal adenomas, the MSCE model assumes that adenomas arise within colonic crypts maintained by immortal stem cells that have accumulated $K(\geq 1)$ requisite (epi)genetic pre-initiation events. In this model, adenomas are allowed to be multi-focal since the initiation process, starting from the time when the Kth pre-initiation event has occurred in a stem cell, is a point process representing the continuous generation of initiated cells from the K-stage cell, the progeny of which is free to undergo clonal expansion. In this picture, pre-initiated stem cells are blocked from clonal expansion; however, they are allowed to divide asymmetrically to generate daughter cells that will (ultimately) undergo terminal differentiation (see Figure 1). This is described in more detail below. To give an example, for a 3-step initiation model ($K=2$), adenomas arise from crypts whose stem cell population sustains two consecutive hits (e.g., the inactivation of both alleles of the APC gene). In this case initiating events occur when the crypt stem cells suffer a third event.

Pre-initiation events. There are two distinct biological ways for pre-initiation events to occur at the cellular level. A pre-initiation event may occur in a stage $k(k=1,\cdots,K)$ stem cell via an asymmetric stem cell division in which a mutation occurs in one of the two daughter stem cells. In other words, a stem cell in the kth pre-initiation stage may divide asymmetrically to yield two daughter stem cells, one in stage k and the other in stage $(k+1)$ which acquires a new mutation (Figure 1A). This process is mathematically modeled as a Poisson process with intensity rate $\mu_k(\cdot)$. In this case, both daughters are retained in the stem cell compartment. The other possibility is represented by another kind of asymmetric stem cell division which causes one daughter stem cell to acquire a mutation leading to a transition of this cell to stage $(k+1)$, while the other daughter is committed to differentiation (Figure 1. B). For historical reasons, we refer to this event as an Armitage-Doll (AD) type transition [15].

For both cases we assume that a stage $k(\leq K)$ stem cell is not yet initiated and therefore lacks the potential for clonal expansion via symmetric cell divisions. However, once a stem cell enters stage $(K+1)$, it is considered initiated and free to undergo clonal expansion. Furthermore, all stem cells in the pre-initiation stages are assumed immortal.

Multi-focal nature of an adenoma. In the context of the MSCE model for CRC, an adenoma consists of the collection of all initiated (stage $K+1$) cells that derive from a single stage K progenitor cell. This definition, while not unique, is consistent with the assumption that the stage K progenitor may be subject to transient amplification, which is represented by frequent Poisson 'emissions' of initiated (stage $K+1$) cells that are free to undergo independent clonal expansions resulting in the formation of multiple sub-clones (Figure 1C).

We call this ensemble of sub-clones an *adenoma* or *adenomatous polyp*. Since information on adenoma number and sizes is typically obtained via screening of individuals who have not previously been diagnosed with CRC, we will also derive the results of the model conditioning on no prior clinical detection of CRC. For this purpose, we assume that the last rate-limiting event (with rate $\mu_{K+1}(\cdot)$) in the MSCE model, which is usually associated with the malignant conversion of an initiated cell, represents detection of a clinical cancer as in Jeon et al. [6].

Basic notation. Suppose there are $K(\geq 1)$ pre-initiation stages. We refer to the kth pre-initiation event as a P_k-mutation and the cells, which have gone through this P_k-mutation, as P_k-cells $(k=1,\cdots,K)$. For completeness, we refer to the normal stem cells as P_0-cells. The generation of P_k-cells from one P_{k-1}-cell can be modeled through a non-homogeneous PP with rate $\mu_{k-1}(\cdot)$ per cell. Alternatively, it can be modeled as a direct transition of a P_{k-1}-cell into a P_k-cell, the AD-type transition referred to above, with hazard rate $\mu_{k-1}(\cdot)$ and density $f_{k-1}(\cdot)$ for the waiting time of

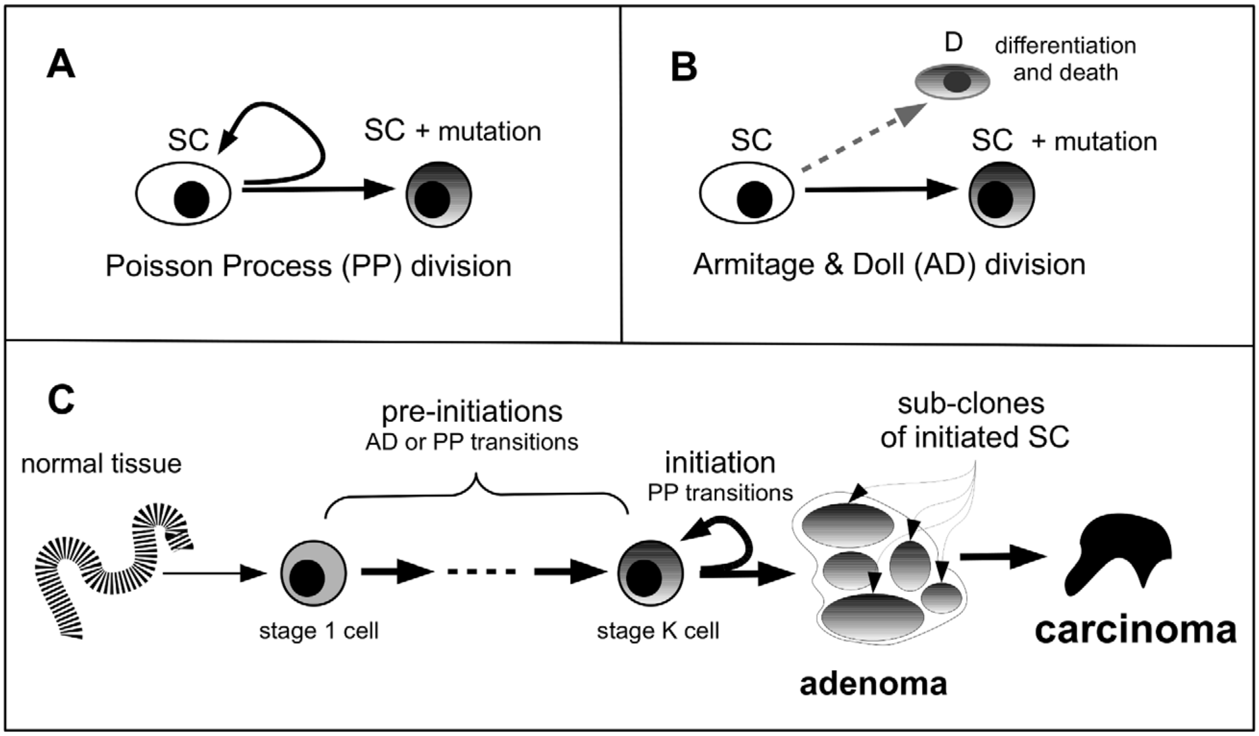

Figure 1. Multistage clonal expansion (MSCE) model. SC: stem cell, D: differentiated cell.
doi:10.1371/journal.pcbi.1002213.g001

a P_{k-1}-cell before the P_k-mutation takes place. For the AD type transition, it is assumed that the functions $\mu_{k-1}(\cdot)$ and $f_{k-1}(\cdot)$ depend only on the time since the P_{k-1}-mutation occurred. If $\mu_{k-1}(\cdot)$ is a constant, then $f_{k-1}(\cdot)$ is the density of an exponential distribution. Once a P_K-cell is formed (by means of a P_K-mutation), it generates initiated cells (i.e., P_{K+1}-cells) according to a non-homogeneous PP with rate $\mu_K(\cdot)$ and the initiated cells grow according to a linear birth and death process with rates $\alpha(\cdot)$ and $\beta(\cdot)$, respectively. Note, the initiation of P_{K+1}-cells and their clonal expansion recapitulates the two-stage clonal expansion (TSCE) model for which explicit solutions for the number and size distributions have been derived [13]. Figure 1.C illustrates the MSCE model for carcinogenesis. The pre-initiation events can be either PP-type or AD-type transitions. However, the first step in the MSCE model represents the successive (random) generation of P_1 mutant stem cells over time and throughout the normal tissue, which is assumed to be very large in size - about 10^8 normal stem cells in colon and rectum combined. Hence, without loss of generality, the MSCE model assumes that the arrivals of the first mutations are of the PP-type.

Size and detection of an adenoma. Suppose we observe, for an individual at a particular time t, the number of detectable adenomas, $N(t)$, and their sizes (in terms of the number of initiated or P_{K+1}-cells in each adenoma), $\{Y_i(t), i=1,\cdots,N(t)\}$. We assume that an adenoma is detectable with probability one if its size is greater than a fixed threshold y_0. In the following sections, we derive the joint distribution of $\{N(t),(Y_i(t),i=1,\cdots,N(t))\}$ for different values of K and different assumptions regarding the type of pre-initiation process (i.e., PP or AD). For $K=2$, we essentially consider the model recommended by Luebeck & Moolgavkar [12] for CRC and used by Jeon et al. [6] for evaluating screening strategies for adenomas in colon and rectum. The latter study provided an efficient approach to simulate the natural history of CRC by recognizing that the size of the adenomas, given the arrival time of a P_K-cell, follows a GLD distribution as derived by Dewanji et al. [14].

Let s_K be the time of a particular P_K-mutation, that is, the arrival time for a progenitor cell (a P_K-cell) for an adenoma. Then this progenitor cell will generate initiated cells ($P_{K+1}-$cells) by a Poisson process, and each initiated cell forms a sub-clone of $P_{K+1}-$cells via a clonal expansion. Note, $Y(s_K,t)$ is the (random) sum of all sub-clones of initiated cells ($P_{K+1}-$cells) that were generated from the P_K progenitor stem cell which was born at time $s_K<t$.

In particular, when the initiation rate $\mu_K(\cdot)$, the cell division rate $\alpha(\cdot)$, and the cell death rate $\beta(\cdot)$ are constants, the GLD distribution reduces to a Negative Binomial distribution (Eq. (3.34) in [6]), given by

$$\begin{aligned} p_n(s_K,t) &\equiv P[Y(s_K,t)=n|Y(s_K,s_K)=0] \\ &= \binom{\mu_K/\alpha+n-1}{\mu_K/\alpha-1}(1-\alpha\zeta)^{\mu_K/\alpha}(\alpha\zeta)^n, n\geq 0, \end{aligned} \tag{1}$$

where $Y(s_K,t)$ denotes the size of the adenoma at time t, given the time of P_K-mutation s_K, and

$$\zeta=\frac{e^{(\alpha-\beta)(t-s_K)}-1}{\alpha e^{(\alpha-\beta)(t-s_K)}-\beta}.$$

Under these assumptions, the sub-clones that lead to this (multi-focal) distribution arise from initiated P_{K+1}-cells that expand clonally by following a linear birth-death process. For this reason, the results derived here are readily applied to the situation when the sub-clones are also identified clinically. Conditioning this distribution on adenomas that occur in cancer-free individuals, i.e., individuals who have not had a prior occurrence of CRC, we have (see Eq.(3.30) in [6])

$$p_n^*(s_K,t) \equiv P[Y(s_K,t)=n|Z(s_K,t)=0, Y(s_K,s_K)=0] \\ = \begin{pmatrix} \mu_K/\alpha+n-1 \\ \mu_K/\alpha-1 \end{pmatrix} (1-\alpha\zeta_c)^{\mu_K/\alpha}(\alpha\zeta_c)^n, n \geq 0, \tag{2}$$

where

$$\zeta_c = \frac{e^{-p(t-s_K)} - e^{-q(t-s_K)}}{(q+\alpha)e^{-p(t-s_K)} - (p+\alpha)e^{-q(t-s_K)}}, \tag{3}$$

$$p = \frac{1}{2}(-\alpha+\beta+\mu_{K+1} - \sqrt{(\alpha+\beta+\mu_{K+1})^2 - 4\alpha\beta}), \tag{4}$$

$$q = \frac{1}{2}(-\alpha+\beta+\mu_{K+1} + \sqrt{(\alpha+\beta+\mu_{K+1})^2 - 4\alpha\beta}), \tag{5}$$

and $Z(s_K,t)$ is the indicator variable for clinical detection of cancer at time t with P_K-cells born at time s_K.

Number and Size Distribution for $K=1$

For $K=1$ we have only one pre-initiation event, defined by a P_1-mutation, and P_2-cells are the initiated cells. As mentioned in the previous section, the P_1-mutation follows a PP formulation. Let $Y(s_1,t)$ denote the size of the adenoma at time t with the first pre-initiation (P_1-mutation) time s_1. The distribution of $Y(s_1,t)$ is given by the GLD distribution previously derived by Dewanji et al. [14], for the process originating at time s_1 and involving the initiation rate $\mu_K(\cdot)=\mu_1(\cdot)$ and the birth and death rates of the initiated cells given by $\alpha(\cdot)$ and $\beta(\cdot)$, respectively.

As mentioned before, we assume that the generation of P_1-cells follows a non-homogeneous PP with rate $\mu_0(\cdot)X(\cdot)$, where $X(\cdot)$ gives the deterministic growth curve for the normal stem cells in the tissue. Let $M(t)$ be the number of first pre-initiations (P_1-mutations) by time t and let $s_{1j}, j=1,\cdots,M(t)$ be the occurrence times of these P_1-mutations. Also, write $N(s_{1j},t)=I(Y(s_{1j},t)>y_0)$, where $I(\cdot)$ is the indicator function. That is, $N(s_{1j},t)=1$, if the corresponding adenoma is detectable at time t, and 0 otherwise. Then, the number of detectable adenoma $N(t)$ can be written as a filtered Poisson process

$$N(t) = \sum_{j=1}^{M(t)} N(s_{1j},t).$$

The probability generating function (PGF) of $N(t)$ can be written as

$$\psi^{(1)}(u;t) = \exp\left[\int_0^t \mu_0(s_1)X(s_1)\{\psi(u;s_1,t)-1\}ds_1\right], \tag{6}$$

where $\psi(u;s_1,t)$ is the PGF of the binary variate $N(s_1,t)$ with success probability

$$p_1^{(1)}(s_1,t) = P[N(s_1,t)=1] = P[Y(s_1,t)>y_0|Y(s_1,s_1)=0].$$

Note that $p_1^{(1)}(s_1,t)$ is the probability that a P_1-mutation taking place at time s_1 results in a detectable adenoma at time t. This probability can be obtained from the distribution of $Y(s_1,t)$. For constant parameters, this reduces to, using (1) with $K=1$, $p_1^{(1)}(s_1,t) = 1 - \sum_{n=0}^{y_0} p_n(s_1,t)$.

Adenoma prevalence. It follows that the number of P_1-cells which lead to detectable adenomas at time t follows a non-homogeneous PP with rate $\mu_0(s_1)X(s_1)p_1^{(1)}(s_1,t)$, with $s_1 \leq t$, and $N(t)$ has a Poisson distribution with mean $\int_0^t \mu_0(s_1)X(s_1)p_1^{(1)}(s_1,t)ds_1$. Since the adenoma prevalence is defined as the probability of at least one detectable adenoma at age t, it is given by

$$1 - P[N(t)=0] = 1 - \exp\left[-\int_0^t \mu_0(s_1)X(s_1)p_1^{(1)}(s_1,t)ds_1\right]. \tag{7}$$

Detection probability and size distribution of adenomas. The probability of detecting an adenoma of size $Y(t)>y_0$ at age t is given by

$$P[Y(t)>y_0] = \sum_{i=y_0+1}^{\infty} \int_0^t \frac{\mu_0(s_1)X(s_1)P[Y(s_1,t)=i|Y(s_1,s_1)=0]}{\int_0^t \mu_0(s_1)X(s_1)ds_1} ds_1. \tag{8}$$

For constant parameters, this reduces to $\frac{1}{t}\int_0^t p_1^{(1)}(s_1,t)ds_1$. Similarly, the size distribution of a detectable adenoma at age t is given by

$$P[Y(t)=y|Y(t)>y_0] = \frac{\int_0^t \mu_0(s_1)X(s_1)P[Y(s_1,t)=y|Y(s_1,s_1)=0]ds_1}{\int_0^t \mu_0(s_1)X(s_1)p_1^{(1)}(s_1,t)ds_1}. \tag{9}$$

Likelihood for the number and size of detectable adenomas. Using the properties described above, it is straightforward to show that the joint probability (or likelihood L_1) of having n detectable adenomas with sizes $y_i, i=1,\cdots,n$ at age t, i.e, $\{N(t)=n,(Y_i(t)=y_i, i=1,\cdots,n)\}$ is

$$L_1 = P[N(t)=n] \\ \times \prod_{i=1}^n \left\{\int_0^t \left[\frac{\mu_0(s_1)X(s_1)p_1^{(1)}(s_1,t)}{\int_0^t \mu_0(s_1)X(s_1)p_1^{(1)}(s_1,t)ds_1} \frac{P[Y(s_1,t)=y_i|Y(s_1,s_1)=0]}{p_1^{(1)}(s_1,t)}\right]ds_1\right\} \\ \propto \exp\left[-\int_0^t \mu_0(s_1)X(s_1)p_1^{(1)}(s_1,t)ds_1\right] \prod_{i=1}^n \left\{\int_0^t \mu_0(s_1)X(s_1)P[Y(s_1,t)=y_i|Y(s_1,s_1)=0]ds_1\right\}. \tag{10}$$

Extension to observations in individuals with no prior CRC. Here we derive analogous results for cancer-free individuals, i.e., individuals who have not developed CRC by time t. To this end, we require the conditional probability $p_1^{(1*)}(s_1,t)$ that a P_1-mutation taking place at time s_1, results in a detectable adenoma at time t prior to developing CRC. Hence, we need to compute the conditional probability $p_1^{(1*)}(s_1,t) = P[Y(s_1,t)>y_0|Z(s_1,t)=0, Y(s_1,s_1)=0]$. For constant parameters, using (2), $p_1^{(1*)}(s_1,t)$ can be calculated as $1 - \sum_{n=0}^{y_0} p_n^*(s_1,t)$.

Adenoma prevalence among individuals with no prior CRC. As before, the number of detectable adenoma at time t in a cancer-free individual, $N^*(t)$, follows a Poisson distribution with mean given by $\int_0^t \mu_0(s_1)X(s_1)S_2(s_1,t)p_1^{(1*)}(s_1,t)ds_1$, where the two-stage survival function $S_2(s_1,t)$ represents the probability that a P_1-cell born at time s_1 does not lead to CRC by time t. Thus, the adenoma prevalence conditioned on cancer-free is given by

$$1 - \exp\left[-\int_0^t \mu_0(s_1)X(s_1)S_2(s_1,t)p_1^{(1*)}(s_1,t)ds_1\right]. \tag{11}$$

For constant parameters, this two-stage survival function has been derived previously, i.e.

$$S_2(s_1,t)=\left(\frac{q-p}{qe^{-p(t-s_1)}-pe^{-q(t-s_1)}}\right)^{\mu_1/\alpha},$$

where p and q are defined in (4) and (5) with $K=1$ (see [6,12] for details).

Detection probability, size distribution, and likelihood function for adenomas in individuals with no prior CRC. Here we provide analogous expressions for the detection probability (8) and size distribution (9), but properly conditioned on no occurrence of prior CRC. The probability of detecting an adenoma at age t with size $y>y_0$, conditioned on no prior CRC, can be written as

$$\begin{aligned}&P[Y(t)>y_0|Z(t)=0]\\&=\sum_{i=y_0+1}^{\infty}\int_0^t\frac{\mu_0(s_1)X(s_1)S_2(s_1,t)P[Y(s_1,t)=i|Z(s_1,t)=0,Y(s_1,s_1)=0]}{\int_0^t\mu_0(s_1)X(s_1)S_2(s_1,t)ds_1}ds_1\\&=\frac{\int_0^t\mu_0(s_1)X(s_1)S_2(s_1,t)p_1^{(1*)}(s_1,t)ds_1}{\int_0^t\mu_0(s_1)X(s_1)S_2(s_1,t)ds_1}.\end{aligned}\tag{12}$$

Similarly, for the size distribution of detectable adenomas (i.e., their sizes exceeding the threshold y_0) at age t, conditioned on no prior CRC, we find

$$\begin{aligned}&P[Y(t)=y|\,Y(t)>y_0,Z(t)=0]\\&=\frac{\int_0^t\mu_0(s_1)X(s_1)S_2(s_1,t)P[Y(s_1,t)=y|Z(s_1,t)=0,Y(s_1,s_1)=0]ds_1}{\int_0^t\mu_0(s_1)X(s_1)S_2(s_1,t)p_1^{(1*)}(s_1,t)ds_1}.\end{aligned}\tag{13}$$

Finally, following the derivation of (10), the joint distribution of the number and sizes of detectable adenomas, in a cancer-free individual, can be written as

$$\begin{aligned}&L_1^*\propto\exp\left[-\int_0^t\mu_0(s_1)X(s_1)S_2(s_1,t)p_1^{(1*)}(s_1,t)ds_1\right]\\&\times\prod_{i=1}^{n}\left\{\int_0^t\mu_0(s_1)X(s_1)S_2(s_1,t)P[Y(s_1,t)=y_i|Z(s_1,t)=0,Y(s_1,s_1)=0]ds_1\right\}.\end{aligned}\tag{14}$$

Number and Size Distribution for $K=2$

Here the two pre-initiation events (P_1- and P_2-mutations) precede initiation and growth of initiated P_3-cells into sub-clones. Let $Y(s_1,s_2,t)$ denote the size of the adenoma at time t with the corresponding P_1- and P_2-mutations taking place at times s_1 and s_2, respectively. Note that the distribution of $Y(s_1,s_2,t)$ is given by the GLD distribution originating at time s_2 and involving the initiation rate $\mu_K(\cdot)=\mu_2(\cdot)$ and the birth and death rates of the initiated or P_3-cells, given by $\alpha(\cdot)$ and $\beta(\cdot)$, respectively. We derive explicit expressions for the number and size distributions for the case when both P_1- and P_2-mutations are of PP-type. The case when the P_2-mutations are of AD-type is described in the online supplement (Text S1).

As before, the number of detectable adenomas at time t can be written as

$$N(t)=\sum_{j=1}^{M(t)}N^{(1)}(s_{1j},t),\tag{15}$$

where $N^{(1)}(s_{1j},t)$ is the number of detectable adenomas that emerged from a P_1-cell born at time $s_{1j}\le t$. Then, as in (6), the PGF of $N(t)$ can be expressed by

$$\psi^{(2)}(u;t)=\exp\left[\int_0^t\mu_0(s_1)X(s_1)\left\{\psi^{(1)}(u;s_1,t)-1\right\}ds_1\right],\tag{16}$$

where $\psi^{(1)}(u;s_1,t)$ is the PGF of $N^{(1)}(s_1,t)$. Using the Lemma and eq. (10) of Dewanji et al. [14], we have further

$$\begin{aligned}P_0(t)&=P[N(t)=0]=\psi^{(2)}(0;t)\\&=\exp\left[-\int_0^t\mu_0(s_1)X(s_1)\left\{1-P[N^{(1)}(s_1,t)=0]\right\}ds_1\right],\end{aligned}\tag{17}$$

and for $n\ge1$,

$$P_n(t)=P[N(t)=n]=\sum_{i=0}^{n-1}\frac{n-i}{n}P_i(t)q_{n-i}^{(1)}(t),\tag{18}$$

where, for $l\ge1,q_l^{(1)}(t)=\int_0^t\mu_0(s_1)X(s_1)P[N^{(1)}(s_1,t)=l]ds_1$.

Again,

$$N^{(1)}(s_1,t)=\sum_j N(s_1,s_{2j},t),\tag{19}$$

where this sum is over all the P_2-mutations by time t that occurred at times $s_{2j}\le t$ and which emerged from a P_1-cell that was born at time s_1. Here, $N(s_1,s_2,t)=1$ if the adenoma originated from the P_1-cell born at time s_1 and the P_2-cell born at time s_2 is detectable at time t and 0 otherwise; that is, $N(s_1,s_2,t)=I(Y(s_1,s_2,t)>y_0)$. Note that $Y(s_1,s_2,t)=Y(s_2,t)$, since $Y(s_1,s_2,t)$ does not depend on s_1. Therefore, as in the previous section, $N^{(1)}(s_1,t)$ is a filtered PP with the PGF $\psi^{(1)}(u;s_1,t)$, similar to that in (6). Also, $N^{(1)}(s_1,t)$ is a non-homogeneous PP with rate $\mu_1(s_2)p_2^{(1)}(s_2,t)$, for $s_1\le s_2\le t$, and for fixed t, is a Poisson variate with mean $\int_{s_1}^t\mu_1(s_2)p_2^{(1)}(s_2,t)ds_2$, where

$$p_2^{(1)}(s_2,t)=P[N(s_1,s_2,t)=1]=P[Y(s_2,t)>y_0|\,Y(s_2,s_2)=0].\tag{20}$$

This probability can be obtained from the distribution of $Y(s_2,t)$ given in (1).

Adenoma prevalence. $p_2^{(1)}(s_2,t)$ has the same form as that of $p_1^{(1)}(s_1,t)$ but with $K=2$. Thus, using (17), the adenoma prevalence is calculated by

$$1-\exp\left[-\int_0^t\mu_0(s_1)X(s_1)\left\{1-\exp\left[-\int_{s_1}^t\mu_1(s_2)p_2^{(1)}(s_2,t)ds_2\right]\right\}ds_1\right].\tag{21}$$

The distribution of $N(t)$ can now be obtained by using (18), and the expected number of detectable adenomas can be readily obtained using (16), i.e.

$$E[N(t)]=\frac{\partial}{\partial u}\psi^{(2)}(u;t)\bigg|_{u=1}=\int_0^t\mu_0(s_1)X(s_1)\int_{s_1}^t\mu_1(s_2)p_2^{(1)}(s_2,t)ds_2ds_1.\tag{22}$$

Detection probability and size distribution of adenomas. The probability of detecting an adenoma of size $Y(t)>y_0$ at age t is given by

$$P[Y(t)>y_0] = \sum_{i=y_0+1}^{\infty} \int_0^t \frac{\mu_0(s_1)X(s_1)\int_{s_1}^t \mu_1(s_2)P[Y(s_2,t)=i|\,Y(s_2,s_2)=0]ds_2}{\int_0^t \mu_0(s_1)X(s_1)\int_{s_1}^t \mu_1(s_2)ds_2 ds_1} ds_1. \tag{23}$$

For constant pre-initiation rates μ_0 and μ_1, and a constant normal stem cell number X, this reduces to $2\int_0^t s_2 p_2^{(1)}(s_2,t)ds_2/t^2$. Similarly, the size distribution of a detectable adenoma at age t is simply given by

$$P[Y(t)=y|\,Y(t)>y_0] = \frac{\int_0^t \mu_0(s_1)X(s_1)\int_{s_1}^t \mu_1(s_2)P[Y(s_2,t)=y|\,Y(s_2,s_2)=0]ds_2 ds_1}{\int_0^t \mu_0(s_1)X(s_1)\int_{s_1}^t \mu_1(s_2)p_2^{(1)}(s_2,t)ds_2 ds_1}. \tag{24}$$

Likelihood for the number and size of detectable adenomas. Let $M^s(t)$ be the number of 'special' P_1-mutations that lead to at least one detectable adenoma at time t. Because of the filtered-Poisson-process nature of the generation of the adenomas, the occurrence of such special P_1-mutations follows a PP with rate $\mu_0(s_1)X(s_1)p_1^{(2)}(s_1,t)$, for $s_1 \le t$, where $p_1^{(2)}(s_1,t)$ is the probability that a P_1-mutation that occurred at time s_1 leads to at least one detectable adenoma by time t. Using the distribution of $N^{(1)}(s_1,t)$, we have

$$\begin{aligned} p_1^{(2)}(s_1,t) &= 1 - P[N^{(1)}(s_1,t)=0] \\ &= 1 - \exp\left[-\int_{s_1}^t \mu_1(s_2)p_2^{(1)}(s_2,t)ds_2\right]. \end{aligned} \tag{25}$$

We now turn to the joint distribution of having n detectable adenomas with sizes $y_i, i=1,\cdots,n$ at age t, i.e, $\{N(t)=n,(Y_i(t)=y_i,i=1,\cdots,n)\}$, when $n>0$. First, let $N_i(t)$ denote the number of detectable adenomas arising out of the ith 'special' P_1-mutation with sizes $Y_{ij}(t),=1,\cdots,N_i(t)$. Clearly, $\sum_{i=1}^{M^s(t)} N_i(t)=N(t)$. Then, given $M^s(t)=m>0$, the events $\{N_i(t),Y_{ij}(t),j=1,\cdots,N_i(t)\}$, for $i=1,\cdots,m$, are independent and identically distributed with

$$\begin{aligned} P[N_i(t)=n_i,\ Y_{ij}(t)=y_{ij},\, j=1,\cdots,n_i|m] &= \left[\int_0^t \mu_0(s_1)X(s_1)p_1^{(2)}(s_1,t)ds_1\right]^{-1} \\ \times \int_0^t \mu_0(s_1)X(s_1)p_1^{(2)}(s_1,t) &\frac{e^{-\int_{s_1}^t \mu_1(s_2)p_2^{(1)}(s_2,t)ds_2}\left(\int_{s_1}^t \mu_1(s_2)p_2^{(1)}(s_2,t)ds_2\right)^{n_i}/n_i!}{p_1^{(2)}(s_1,t)} \\ \times \prod_{j=1}^{n_i}\left\{\int_{s_1}^t \frac{\mu_1(s_2)p_2^{(1)}(s_2,t)}{\int_{s_1}^t \mu_1(s_2)p_2^{(1)}(s_2,t)ds_2}\right. &\left.\frac{P[Y(s_2,t)=y_{ij}|\,Y(s_2,s_2)=0]}{p_2^{(1)}(s_2,t)}ds_2\right\}ds_1. \end{aligned} \tag{26}$$

Therefore, the joint probability of $\{M^s(t)=m\}$ and $\{N_i(t)=n_i, Y_{ij}(t)=y_{ij},\, j=1,\cdots,n_i\}$ for $i=1,\cdots,m$, is given by

$$\begin{aligned} &L(m,n_i,y_{ij},j=1,\cdots,n_i,i=1,\cdots,m) \\ &= \left(m!\prod_{i=1}^m n_i!\right)^{-1}\exp\left[-\int_0^t \mu_0(s_1)X(s_1)p_1^{(2)}(s_1,t)ds_1\right] \\ &\quad\times \prod_{i=1}^m\left(\int_0^t \mu_0(s_1)X(s_1)e^{-\int_{s_1}^t \mu_1(s_2)p_2^{(1)}(s_2,t)ds_2}\right. \\ &\qquad\left.\times \prod_{j=1}^{n_i}\left\{\int_{s_1}^t \mu_1(s_2)P[Y(s_2,t)=y_{ij}|\,Y(s_2,s_2)=0]ds_2\right\}ds_1\right). \end{aligned} \tag{27}$$

In order to derive the joint distribution of $\{N(t)=n,(Y_i(t)=y_i,i=1,\cdots,n)\}$, using (27), let $N=\{1,\cdots,n\}$ denote the set of labels for the n clones. Then, summing over all possible m and the possible partitions of N with m subsets, the likelihood is given by

$$L_2 = \sum_{m=1}^n \sum_{(I_1,\cdots,I_m)} L(m,n_i,y_{ij},j\in I_i,i=1,\cdots,m),$$

where $I_1,\cdots,I_m$ are disjoint subsets of N with $\bigcup_{i=1}^m I_i=N$, and the second sum is over all such $(I_1,\cdots,I_m)$ with $n_i=|I_i|$, the number of labels in I_i. If n is not very large, this sum is not difficult to work with.

Extension to observations in individuals with no prior CRC. Analogous results can be obtained for observations in individuals who have not developed clinical CRC by time t by directly conditioning the Poisson rates and using conditional size distributions for $Y(s_2,t)$. For example, the relevant joint distribution of adenoma number and sizes (i.e., (27)) can be obtained by the following substitutions:

$$p_2^{(1)}(s_2,t) \rightarrow p_2^{(1*)}(s_2,t) = P[Y(s_2,t)>y_0|Z(s_2,t)=0, Y(s_2,s_2)=0],$$

$$\mu_0(s_1)X(s_1) \rightarrow \mu_0(s_1)X(s_1)S_3(s_1,t),$$

$$\mu_1(s_2) \rightarrow \mu_1(s_2)S_2(s_2,t),$$

and $P[Y(s_2,t)=y_{ij}|\,Y(s_2,s_2)=0]$ is replaced by

$$P[Y(s_2,t)=y_{ij}|Z(s_2,t)=0, Y(s_2,s_2)=0].$$

Note, $S_3(s_1,t)$ and $S_2(s_2,t)$ represent the respective probabilities that a P_1-cell born at time s_1 and a P_2-cell born at time s_2 do not give rise to CRC by time t. In other words, they are the tumor survival functions of a 3-stage and 2-stage carcinogenesis model, respectively. For constant parameters, these survival functions are as following (see [6,12] for the derivation):

$$\begin{aligned} S_2(s_2,t) &= \left(\frac{q-p}{qe^{-p(t-s_2)}-pe^{-q(t-s_2)}}\right)^{\mu_2/\alpha}, \\ S_3(s_1,t) &= \exp\left[-\int_0^{t-s_1} \mu_1\left(1-\frac{q-p}{qe^{-ps}-pe^{-qs}}\right)^{\mu_2/\alpha} ds\right], \end{aligned} \tag{28}$$

where p and q are defined in (4) and (5) with $K=2$ (see [6,12] for details). Thus, the conditional expression for adenoma prevalence is given by

$$1-\exp\left[-\int_0^t \mu_0(s_1)X(s_1)S_3(s_1,t)\left\{1-\exp\left[-\int_{s_1}^t \mu_1(s_2)S_2(s_2,t)p_2^{(1*)}(s_2,t)ds_2\right]\right\}ds_1\right]. \quad (29)$$

The expected number of detectable adenomas conditioned on no prior CRC can be obtained through analogous replacements in (22).

Detection probability and size distribution for adenomas in individuals with no prior CRC. As before for the case with $K=1$, we also provide analogous expressions for the detection probability (23) and size distribution (24), but properly conditioned on no previous occurrence of CRC. The probability of detecting an adenoma at age t with size $y>y_0$, becomes

$$\begin{aligned}
&P[Y(t)>y_0|Z(t)=0]\\
&=\sum_{i=y_0+1}^{\infty}\int_0^t \frac{\mu_0(s_1)X(s_1)S_3(s_1,t)\int_{s_1}^t \mu_1(s_2)S_2(s_2,t)P[Y(s_2,t)=i|Z(s_2,t)=0,Y(s_2,s_2)=0]ds_2}{\int_0^t \mu_0(s_1)X(s_1)S_3(s_1,t)\int_{s_1}^t \mu_1(s_2)S_2(s_2,t)ds_2ds_1}ds_1 \quad (30)\\
&=\frac{\int_0^t \mu_0(s_1)X(s_1)S_3(s_1,t)\int_{s_1}^t \mu_1(s_2)S_2(s_2,t)p_2^{(1*)}(s_2,t)ds_2ds_1}{\int_0^t \mu_0(s_1)X(s_1)S_3(s_1,t)\int_{s_1}^t \mu_1(s_2)S_2(s_2,t)ds_2ds_1},
\end{aligned}$$

and for the size distribution of detectable adenomas (i.e., their sizes exceeding the threshold y_0) at age t, conditioned on no prior CRC, we have

$$\begin{aligned}
&P[Y(t)=y|Z(t)=0,Y(t)>y_0]\\
&=\frac{\int_0^t \mu_0(s_1)X(s_1)S_3(s_1,t)\int_{s_1}^t \mu_1(s_2)S_2(s_2,t)P[Y(s_2,t)=y|Z(s_2,t)=0,Y(s_2,s_2)=0]ds_2ds_1}{\int_0^t \mu_0(s_1)X(s_1)S_3(s_1,t)\int_{s_1}^t \mu_1(s_2)S_2(s_2,t)p_2^{(1*)}(s_2,t)ds_2ds_1}. \quad (31)
\end{aligned}$$

Number and Size Distribution for General K

The notation introduced in the previous section is easily generalized to $K>2$. The random variable $Y(s_1,\cdots,s_K,t)$ denotes the size of the adenoma at time t with the corresponding P_1-, $\cdots$,P_K-mutation times $s_1,\cdots,s_K$, respectively. The distribution of $Y(s_1,\cdots,s_K,t)$ is, as before, given by the GLD distribution with time origin at s_K and with initiation rate $\mu_K(\cdot)$. Initiated P_{K+1}-cells divide or die with rates $\alpha(\cdot)$ and $\beta(\cdot)$, respectively. Importantly, $Y(s_1,\cdots,s_K,t)$ depends on s_K alone (i.e., $Y(s_1,\cdots,s_K,t)=Y(s_K,t)$), and the distribution is given by (1) for constant parameters.

Various combinations of AD-type and PP-type generations for the K pre-initiations are possible, but the formulation of the likelihood becomes more complicated. The special cases when all pre-initiations are of PP-type or AD-type are covered in the online supplement (Text S1).

Results

The derived expressions allow us to readily predict both observable and unobservable numbers of pre-malignant tumors in a tissue. Such predictions are helpful in validating cancer models using intermediate endpoints on precursor lesions, in particular their number and sizes. Furthermore, being able to predict the unobserved portion of precursor lesions is of clinical relevance for early detection and cancer prevention. Here, we illustrate the utility of the derived expressions using the example of colonic adenomas. Specifically, we present the predicted size distribution of adenomas and the age-specific adenoma prevalence, i.e., the probability of finding at least one observable adenoma in an individual as a function of age. Since population-level screening is typically performed on asymptomatic individuals, we also condition on individuals having not developed cancer in the tissue of interest at the time of observation.

The predictions presented here are for $K=2$ as described above. The underlying CRC model for cancer incidence is the 4-stage model previously derived by Luebeck & Moolgavkar [12] and updated by Meza et al. [7]. The alternative model (PP for P_1- and AD for P_2-mutation for $K=2$ in the online supplement (Text S1)) yields very similar results (not shown). Importantly, not all biological parameters of the MSCE/CRC model are estimable from incidence data alone. For example, for the 4-stage model used here, only the parameters p,q, the product (slope parameter) $\mu_0X\mu_1$, and the ratio μ_2/α are identifiable. However, if the cell division rate of initiated cells, α, is known, all parameters of the model can be determined (assuming that the number of normal tissue stem cells, X, is known and that $\mu_0=\mu_1$). For our illustrations, we choose plausible values for the cell division rate α, but keep the values of $p,q,\mu_0X\mu_1$, and μ_2/α as estimated by Meza et al. [7]. This affords explicit computation of the adenoma number and size distribution without altering the fits of the model to the observed CRC incidence.

Figure 2 (left panel) shows the predicted size distribution of non-extinct adenoma without an imposed detection threshold (i.e., $y_0=0$) for the model with $K=2$. With constant parameters, both the unconditional and conditional (on no prior CRC development) size distributions of detectable adenoma are given by expressions (24) and (31), respectively.

For sizes sufficiently large, the unconditional adenoma size distribution is roughly log-log-linear, while the conditional size distribution shows departures from this behavior for sizes above $\sim 100{,}000$ cells, i.e., when the risk of an adenoma-to-carcinoma transition increases more rapidly. This phenomenon is more noticeable when the cell division rate α is lower. Figure 2 (right panel) shows the probability of detecting an adenoma at age 70 as a function of the detection threshold y_0 for both unconditional (solid) and conditional (dashed) adenoma size distributions. Higher cell division rates (α) give rise to larger adenoma sizes and hence lead to higher detection probabilities even though the net cell proliferation rate ($\sim -p$) is approximately the same. For constant parameters, the unconditional and conditional detection probabilities are given by (23) and (30), respectively. This figure reveals that even for relatively small (i.e., sensitive) thresholds of a few thousand cells, many adenomas may go undetected. However, the precise proportion of detectable adenomas depends on the cell division rate α with higher values of α making detection more likely.

Figure 3 shows the predicted age-specific adenoma prevalence in asymptomatic individuals for both males and females and for the models with $K=1$ and $K=2$, as described by Meza et al. [7], including their dependence on the observation threshold y_0. The prevalence is defined as the probability of at least one detectable adenoma at age t, and is given by (11) for $K=1$ and (29) for $K=2$. In comparison with careful observations from autopsy studies [16], the model under-estimates these empirical data (represented by filled circles) unless one accepts a very small number of initiated stem cells to be observable. There are several explanations why the model-generated expected prevalence of adenoma might be lower than the clinical data would indicate (see the end of Discussion).

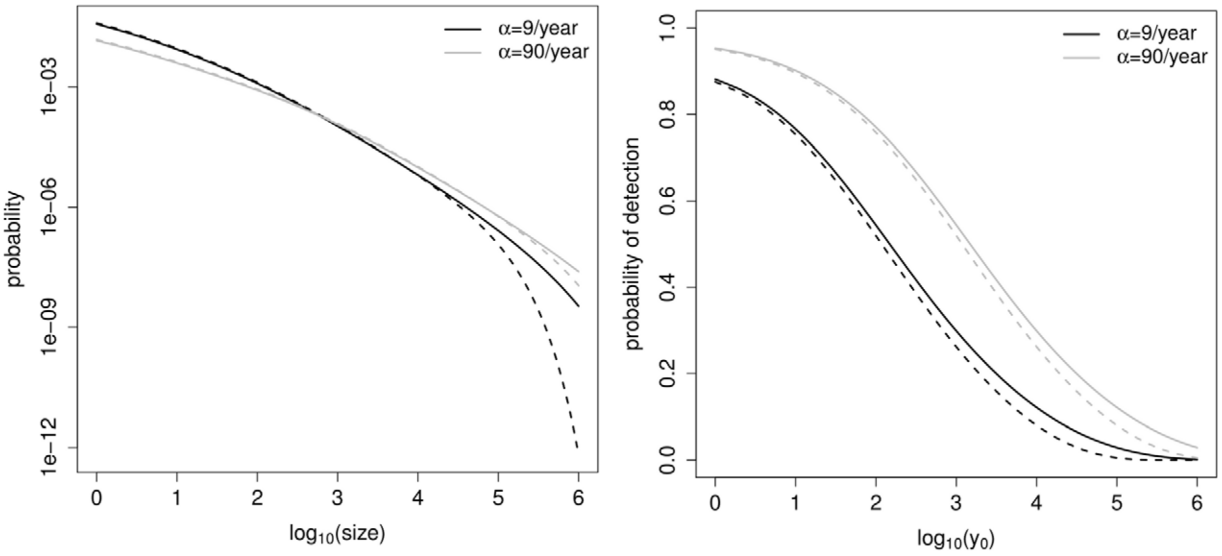

Figure 2. Size distribution of a detectable adenoma and probability of detection of an adenoma. Left panel: predicted unconditional (solid) and conditional (dashed) size distribution of a detectable adenoma at age 70 using the parameter estimates obtained by [7] for females in SEER with $K=2$ and $y_0=0$. The cell division rate of initiated cells, α, is assumed as either 9 or 90/year. Right panel: the probability of detection of an adenoma at age 70 as a function of the detection threshold y_0. Otherwise same as left panel.
doi:10.1371/journal.pcbi.1002213.g002

Discussion

We have previously derived number and size distributions of pre-malignant clones in the context of the two-stage clonal expansion (TSCE) model of carcinogenesis [13,17] and more recently established a formal connection of these results with fluctuation analyses based on the Luria-Delbrück distribution [14]. The mathematical tools derived in these papers were subsequently applied to the problem of screening for colorectal adenoma allowing for interventions resulting from their complete or incomplete removal [6]. These explorations, however, required time consuming computer simulations. In contrast, here we derive mathematical expressions that allow us to readily compute adenoma number and size distributions without simulation. These expressions can form the basis for computing the likelihood of adenoma data from screening studies involving sigmoidoscopies, colonoscopies and computed-tomographic colonographies, and thus are of significantly practical importance. Moreover the analytical form of the likelihood function allows for parameter estimation and likelihood-based hypothesis testing. Analyses of such data will be forthcoming and are the subject of a separate paper.

Our previous analyses of CRC incidence data suggest that K, the number of requisite pre-initiation mutations, is indeed small

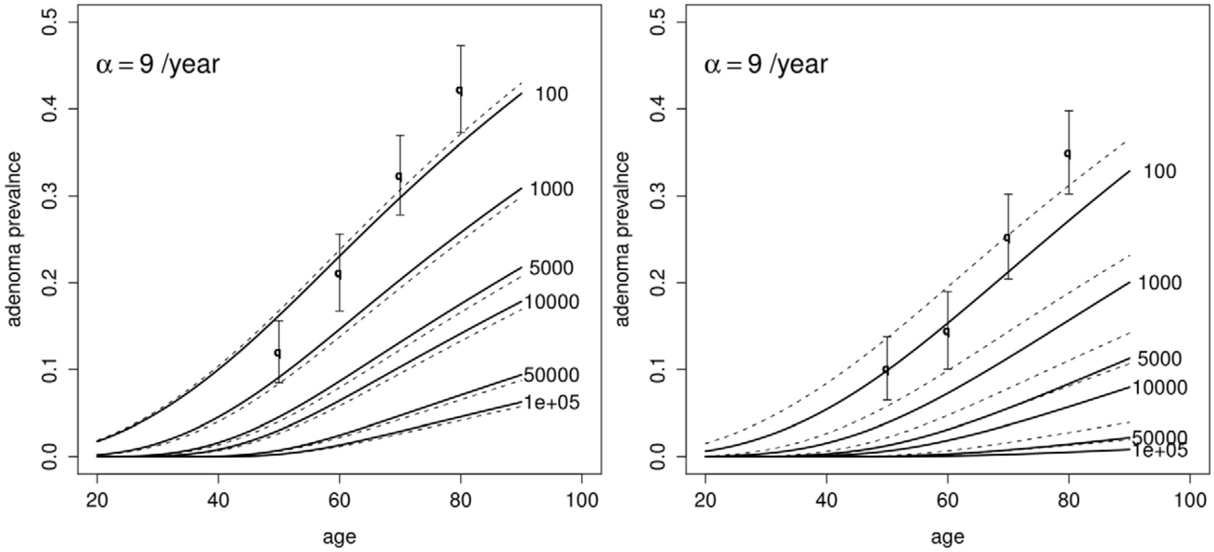

Figure 3. Adenoma prevalence. Predicted adenoma prevalence for both males (Left panel) and females (Right panel) as a function of age and various observation thresholds y_0 using the models with $K=2$ (solid lines) and $K=1$ (dashed lines). Empirical data from Clark et al. [16] in filled circles.
doi:10.1371/journal.pcbi.1002213.g003

[7,12]. $K=1$ corresponds to a 'two-hit' model for initiation, in essence representing the biallelic inactivation of a tumor suppressor gene (Knudson's recessive oncogenesis) [7]. A model with $K=2$ may describe both the inactivation of a tumor suppressor gene as well as the activation of an oncogene [7,12]. Here, we also treat the case of general K, which can be viewed as a model for clonal evolution due to the tree-structure where the nodes represent immortal mutant stem cells that will give rise to specific sub-clones which may or may not be identified as such. We distinguish between two types of rate-limiting events, one that generates (potentially multiple) mutations via asymmetric cell division while preserving the progenitor stem cell from which the mutations arose (PP-type), and one that leads to a transition of a progenitor cell into one cell that acquires a new mutation (AD-type). The MSCE model used here assumes that all events that lead to initiated cells are PP-type. This is only a mild restriction since for rare events the PP-type emission is equivalent to a AD-type transition (see Figure 1). For frequent (high rate) events, the AD-type transition looses its rate-limiting nature and can be ignored, while the high rate (PP-type) process leads to the accumulation of multiple clones and thus has the potential to capture non-mutational events, such as the transient amplification of proliferative cells from resident stem cells in the colonic crypts. Once a stem cell is considered initiated, i.e., is of type P_{K+1}, we assume that it undergoes a stochastic birth-death process. This leads to the GLD distribution introduced in [14] for the adenoma size $Y(s_1,\cdots,s_K,t)$, which reduces to a Negative Binomial distribution for constant parameters. Note, however, our formalism is more general and can accommodate other growth models that do not result in a GLD size distribution for the initiated cell population emerging from a P_K progenitor cell [17,18].

Finally, we wish to comment on the predictions of the model for the age-specific adenoma prevalence (Figure 3). In comparison with the empirical data, our predictions appear too low. However, the discrepancy depends on what is assumed for the initiated stem cell division rate α and the detection threshold y_0. While the range of plausible values for α is limited by how fast initiated stem cell can cycle in the adenoma (unlikely more than 2–3 times a week), it is not clear what fraction of cells in an adenoma is truly at risk for malignant transformation [10]. Assuming that a 1 mm adenoma, the caliper size detection limit cited by Clark et al. [16], contains about 500,000 cells [19] and that only 1–10% of cells in an adenoma are tumor stem cells [10,20], y_0 may be as low as 5000 cells and therefore the discrepancy may be less dramatic. Alternatively, one might include pre-initiated cells in the adenoma size count. However, our assumption is that pre-initiated cells do not expand clonally, although they may increase in number as a result of multiple births of the same type of mutation from a single stem cell over time (via Poisson process emissions). Thus, since locus-specific mutations are rare (of the order of 10^{-6} to 10^{-7} per year), the contribution of pre-initiated cells to the overall number of cells in an adenoma is likely very small.

It is well-recognized that adenomas can be genetically diverse and differ widely in their neoplastic potential. Indeed, adenomas have been suggested to regress implying that there are adenomas that are not on the pathway to cancer [21], although regression may simply reflect the stochastic nature of tumor growth. A more intriguing possibility of resolving the discrepancy is that adenomas go through a growth-bottleneck (i.e., stagnancy) before they can become cancerous. In this scenario, adenomas might sojourn in a reservoir until an activating mutation or change in tumor microenvironment releases them from arrest [22,23]. Although incorporating this scenario into the MSCE model may be challenging, the framework presented here is independent of the particular dynamics of the initiated cells and the number of clonal expansions assumed.

Supporting Information

Text S1 Supplementary methods and results. A tabular glossary which summarizes our notation and succinctly defines the model parameters and terminology in use.
(PDF)

Acknowledgments

We thank our colleagues Dr. Hazelton and Dr. Moolgavkar for helpful discussions.

Author Contributions

Conceived and designed the experiments: AD JJ RM EGL. Analyzed the data: JJ EGL. Contributed reagents/materials/analysis tools: AD JJ. Wrote the paper: AD EGL. Drafted the paper: AD EGL. Revised the paper: JJ RM EGL.

References

1. Winawer S, Fletcher R, Rex D, Bond J, Burt R, et al. (2003) Colorectal cancer screening and surveillance: Clinical guidelines and rationaleupdate based on new evidence. Gastroenterology 124: 544–560.
2. Jones S, Chen WD, Parmigiani G, Diehl F, Beerenwinkel N, et al. (2008) Comparative lesion sequencing provides insights into tumor evolution. Proc Natl Acad Sci U S A 105: 4283–4288.
3. Luebeck EG, Moolgavkar SH (2002) Multistage carcinogenesis and the incidence of colorectal cancer. Proc Natl Acad Sci U S A 99: 15095–15100.
4. Meza R, Luebeck EG, Moolgavkar SH (2005) Gestational mutations and carcinogenesis. Math Biosci 197: 188–210.
5. Luebeck EG (2006) Multistage carcinogenesis: From intestinal stem cell to colon cancer in the population. In: Potten CS, Clarke RB, Wilson J, Renehan AG, eds. Tissue stem cells Taylor & Francis, Chapter 10. pp 215–228.
6. Jeon J, Meza R, Moolgavkar SH, Luebeck EG (2008) Evaluation of screening strategies for premalignant lesions using a biomathematical approach. Math Biosci 213: 56–70.
7. Meza R, Jeon J, Moolgavkar SH, Luebeck EG (2008) Age-specific incidence of cancer: Phases, transitions, and biological implications. Proc Natl Acad Sci U S A 105: 16284–16289.
8. Meza R, Jeon J, Renehan AG, Luebeck EG (2010) 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 70: 5419–5429.
9. Goss KH, Groden J (2000) Biology of the adenomatous polyposis coli tumor suppressor. J Clin Oncol 18: 1967–1979.
10. Barker N, Ridgway RA, van Es JH, van de Wetering M, Begthel H, et al. (2009) Crypt stem cells as the cells-of-origin of intestinal cancer. Nature 457: 608–611.
11. Potten CS (1998) Stem cells in gastrointestinal epithelium: numbers, characteristics and death. Philos Trans R Soc Lond B Biol Sci 353: 821–830.
12. Luebeck EG, Moolgavkar SH (2002) Multistage carcinogenesis and the incidence of colorectal cancer. Proc Natl Acad Sci U S A 99: 15095–15100.
13. Dewanji A, Venzon DJ, Moolgavkar SH (1989) A stochastic two-stage model for cancer risk assessment. ii. the number and size of premalignant clones. Risk Anal 9: 179–187.
14. Dewanji A, Luebeck EG, Moolgavkar SH (2005) A generalized luriadelbrck model. Math Biosci 197: 140–152.
15. Armitage P, Doll R (1954) The age distribution of cancer and a multi-stage theory of carcinogenesis. Br J Cancer 8: 1–12.
16. Clark JC, Collan Y, Eide TJ, Estve J, Ewen S, et al. (1985) Prevalence of polyps in an autopsy series from areas with varying incidence of large-bowel cancer. Int J Cancer 36: 179–186.
17. Luebeck EG, Moolgavkar SH (1991) Stochastic analysis of intermediate lesions in carcinogenesis experiments. Risk Anal 11: 149–157.
18. Tan WY (1986) A stochastic gompertz birth-death process. Stat Probabil Lett 4: 25–28.
19. Pinsky PF (2000) A multi-stage model of adenoma development. J Theor Biol 207: 129–143.
20. Boman BM, Huang E (2008) Human colon cancer stem cells: A new paradigm in gastrointestinal oncology. J Clin Oncol 26: 2828–2838.

21. Loeve F, Boer R, Zauber AG, Van Ballegooijen M, Van Oortmarssen GJ, et al. (2004) National Polyp Study data: evidence for regression of adenomas. Int J Cancer 111: 633–639.
22. Hahnfeldt P, Panigrahy D, Folkman J, Hlatky L (1999) Tumor development under angiogenic signaling. Cancer Res 59: 4770–4775.
23. Enderling H, Anderson AR, Chaplain MA, Beheshti A, Hlatky L, et al. (2009) Paradoxical dependencies of tumor dormancy and progression on basic cell kinetics. Cancer res 69: 8814–8821.