Age-specific incidence of cancer: Phases, transitions, and biological implications

Rafael Meza*†‡, Jihyoun Jeon*, Suresh H. Moolgavkar*§, and E. Georg Luebeck*†‡¶

*Program in Biostatistics and Biomathematics and ¶Program in Computational Biology, Fred Hutchinson Cancer Research Center 1100 Fairview Avenue North, Seattle, WA 98109-1024; and §Exponent, Inc., 15375 SE 30th Place, Bellevue, WA 98007

Edited by Alfred G. Knudson, Jr., Fox Chase Cancer Center, Philadelphia, PA, and approved August 18, 2008 (received for review February 04, 2008)

The observation that the age-specific incidence curve of many carcinomas is approximately linear on a double logarithmic plot has led to much speculation regarding the number and nature of the critical events involved in carcinogenesis. By a consideration of colorectal and pancreatic cancers in the Surveillance Epidemiology and End Results (SEER) registry we show that the log-log model provides a poor description of the data, and that a much better description is provided by a multistage model that predicts two basic phases in the age-specific incidence curves, a first exponential phase until the age of ≈60 followed by a linear phase after that age. These two phases in the incidence curve reflect two phases in the process of carcinogenesis. Paradoxically, the early-exponential phase reflects events between the formation (initiation) of premalignant clones in a tissue and the clinical detection of a malignant tumor, whereas the linear phase reflects events leading to initiated cells that give rise to premalignant lesions because of abrogated growth/differentiation control. This model is consistent with Knudson's idea that renewal tissue, such as the colon, is converted into growing tissue before malignant transformation. The linear phase of the age-specific incidence curve represents this conversion, which is the result of recessive inactivation of a gatekeeper gene, such as the APC gene in the colon and the CDKN2A gene in the pancreas.

colorectal | pancreatic | multistage carcinogenesis | neoplastic progression | Knudson's "two-hit" hypothesis

The precise shape of the age-specific incidence of various cancers, especially of nonembryonal solid tumors, and what information can be gleaned from their behavior, is still subject to scientific debate. A widely held view, put forward independently by Muller (1) and Nordling (2) and which reflects the basis of the Armitage–Doll model (3), conceives the stepwise progression of normal cells to cancer as a multistage process involving a number of rate-limiting (epi)genetic events. When viewed at the population level, this assumption uniquely defines the mathematical shape of the age-specific incidence of a cancer, also reflecting the assumed number of rate-limiting events. Indeed, at some level of mathematical approximation (see, e.g., ref. 4), the sequential nature of such a multistep process imposes a power-law behavior, that is, the age-specific incidence of cancers that arise as a consequence of several rate-limiting genomic alterations is predicted to increase with a power of age that is one less than the number of events necessary for malignant transformation. Although it is generally recognized that the carcinogenic process is more complicated and possibly punctuated by selection of advantageous mutations and clonal expansions (5), the qualitative power-law behavior of the age-specific cancer incidence is still considered a reasonable approximation for many cancers and continues to be invoked to argue for or against the importance of specific biological events in carcinogenesis (e.g., 6–14). However, this assumption remains largely untested, mathematically and statistically, despite the availability of high-quality cancer data to explore its adequacy.

How much can be gleaned from cancer incidence data for the purpose of modeling carcinogenesis? The answer to this question clearly depends on the level of biological detail one hopes to capture. Perhaps the most challenging issue in discerning biological mechanisms from population level data is the fact that the observed incidence of a cancer can be fraught with significant secular trends due to changes in lifestyle (smoking, alcohol consumption, diet, and exercise), environmental factors, and changing screening and surveillance practices (15). The incidence of colorectal cancer, for example, is showing a remarkable reduction following the increased adoption of colonoscopies, sigmoidoscopies, and fecal occult blood tests in the 1970s and 1980s (16, 17), whereas increasing trends in obesity (or body mass index) may well explain increasing rates of colon cancer in younger (postwar) cohorts (18). To disentangle the effects of age, period, and cohort on cancer incidence, a number of mathematical and computational approaches have been suggested (15, 19–21), but are rarely invoked.

Despite these difficulties in identifying the natural course of the age-specific incidence, free of secular trends and factors, several firm predictions concerning the behavior of the age-specific incidence can be made, predictions based on mathematical analyses, sound statistical criteria, and biological plausibility. Here, we make the case that neoplastic progression that is initiated by the loss of a tumor suppressor (or "gatekeeper") gene in two rare rate-limiting events necessarily leads to a linearly increasing age-specific incidence above an age that is characteristic for the timescale of neoplastic progression, a behavior that reflects the steadily increasing risk of malignant transformations in premalignant clones whose incidence increases linearly in the target organ. This finding is in contrast with the multistage carcinogenesis model view (alluded to above), claiming a power-law (age) behavior that reflects the number of rate-limiting events in the carcinogenic process (via its power), and with the two-stage clonal expansion model, which assumes that initiation occurs after a single rate-limiting event (22–24) and does not predict a linearly increasing age-specific incidence except for very young ages.

Our finding is consistent with Knudson's two-hit oncogenesis idea, applied to the recessive formation of a slowly expanding neoplasm in a tissue normally protected from uncontrolled cell proliferation by a tumor suppressor or gatekeeper gene (25–27). Here, by using colorectal and pancreatic cancer as examples, we show that it is exactly this early "tumor initiation" process that leads to the linear (not "log-log linear") increase of the age-specific cancer incidence above midage. Our claim is supported by molecular studies on tumor suppressor inactivation, growth rates of premalignant neoplasms, and the analyses of colorectal cancer (CRC) and pancreatic cancer (PC) incidence data described here.

Author contributions: R.M. and E.G.L. designed research; R.M., J.J., and E.G.L. performed research; R.M. and E.G.L. contributed new reagents/analytic tools; R.M., J.J., and E.G.L. analyzed data; and R.M., S.H.M., and E.G.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

†R.M. and E.G.L. contributed equally to this work.

‡To whom correspondence may be addressed. E-mail: rmeza@fhcrc.org or gluebeck@fhcrc.org.

This article contains supporting information online at www.pnas.org/cgi/content/full/0801151105/DCSupplemental.

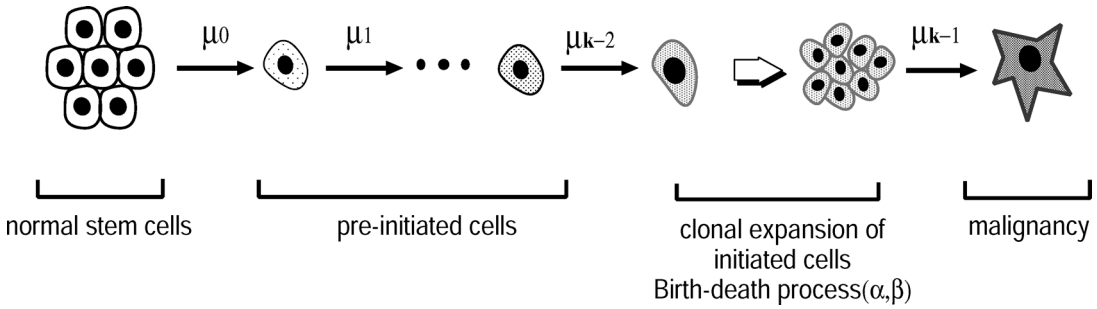

Fig. 1. Multistage clonal expansion (MSCE) model. Pictorial representation of the MSCE model. Cells arrive in the first preinitiation stage according to a Poisson process with intensity $\mu_0 X$, where X is the number of normal susceptible cells. A cell in the ith preinitiation stage can divide into one ith preinitiated cell and one $(i+1)$th preinitiated cell with rate μ_i, for $i = 1, \cdots, k-3$. A $(k-2)$th preinitiated cell can divide into one $(k-2)$th preinitiated cell and one initiated cell with rate μ_{k-2}. Once a cell is initiated, it expands clonally via a birth–death–mutation process with rates α, β, μ_{k-1}, respectively.

Results

Theoretical Results. Our main finding concerns the role of the first rate-limiting steps toward initiation of a premalignant lesion in carcinogenesis and their impact on the "hazard function," which is estimated by the observed cancer incidence. Here, we define the hazard function as the rate at which malignant cancers occur in individuals that have not previously developed that cancer. The premalignant lesion is understood to represent a neoplasm that undergoes persistent (and possibly very slow) clonal expansion and evolution toward malignant transformation. We will provide simple mathematical expressions, and a direct "graphical" determination based on the observed incidence curves that provide numerical estimates of the effective transit or sojourn time of (nonextinct) premalignant lesions beginning with their birth and ending in their transformation to cancer.

Our results are based on a class of multistage models that allow for the sequential accumulation of a specific number of mutations (or epigenetic events) prior to initiation of a premalignant lesion (or "preneoplasm"). The most parsimonious model, within this class, consistent with the incidence of CRC in Surveillance Epidemiology and End Results (SEER) and with what is known about the pathogenesis of microsatellite-stable CRC, is a four-stage model that posits two rare rate-limiting events and a high rate of asymmetric stem cell divisions (possibly reflecting transient amplification in a crypt) for the initiation of an adenomatous polyp (21).[‖] This model is also consistent with molecular evidence showing that adenomas frequently show biallelic inactivation of the APC gene (function) (28–32). The mathematical treatment of the model, in particular, the derivation of the (cancer) survival and hazard function, is standard and has been described previously [e.g., see Refs. 21–23, 33, and supporting information (SI) *Appendix*].

However, the minimal (not necessarily optimal or best-fitting) model still consistent with the idea that tumor initiation requires biallelic inactivation of a tumor suppressor locus (reminiscent of Knudson's recessive oncogenesis hypothesis) (25), is a three-stage model (see Fig. 1). Once initiation occurs, clonal expansion of the initiated progenitor cell may proceed until the clone undergoes a malignant transformation that gives rise to cancer. To keep this discussion simple and to avoid unnecessary mathematical arguments, we ignore the lag time between the malignant transformation event and appearance of clinically detectable cancer. This time is considered short compared with the dwell time of a premalignant lesion. For the three-stage model, then, we have the following parameters: the number of susceptible stem cells, X, the mutation rate of the first hit at the tumor suppressor locus, μ_0, the mutation rate of the second copy of the tumor suppressor gene, μ_1, the premalignant cell division rate, α, the cell death or differentiation rate, β, and the malignant transformation rate, μ_2. As we show next, the behavior of the hazard function of this three-stage model, $h_3(t)$, exhibits four distinct phases that, in time-reversed order from older to younger ages, reveal increasingly more aspects of the carcinogenic process. However, only two of these phases can typically be observed in cancer incidence data. Mathematical details and proofs supporting our findings are given in the accompanying *SI Appendix*, including similar results for the four-stage model.

Asymptotic phase of the hazard function. For times much larger than $(\mu_1 p_\infty)^{-1}$, where p_∞ denotes the asymptotic probability of nonextinction of an initiated clone (which is also the probability that a clone becomes malignant), we derive the approximation

$$h_3(t) \approx \mu_0 X, \qquad [1]$$

which means that asymptotically (for very large times) the hazard function approaches a constant, which is the rate at which stem cells acquire the first mutation in the tissue. However, because $(\mu_1 p_\infty)^{-1} \gg$ *human lifetime* (for typical locus-specific mutation rates), this asymptotic phase will in reality never be reached, and therefore, $\mu_0 X$ cannot be identified from incidence data.

Linear phase (mid to old age). For mid to old ages, the hazard function increases linearly with age and can be approximated by

$$h_3(t) \approx \mu_0 X \mu_1 p_\infty (t - T_s), \qquad [2]$$

where T_s denotes the mean (or effective) sojourn time of the premalignant neoplasm, that is, the mean duration from the birth of a premalignant clone to its eventual development into a malignant tumor, conditional on nonextinction. In particular (see *SI Appendix*),

$$T_s \approx -\frac{\ln(\alpha\mu_2/(\alpha-\beta)^2)}{\alpha-\beta}. \qquad [3]$$

This approximation is valid for ages t between T_s and $(\mu_1 p_\infty)^{-1}$. Thus, the extrapolation of the linear phase of the hazard function across the time–age axis allows for the identification and direct estimation of the effective sojourn time of the premalignant clones (Figs. 2 and 3). This formula of T_s is in general agreement with the result given by Herrero-Jimenez *et al.* (34) (equation 30 therein) in their analysis of colon cancer mortality in the United States.

[‖] The SEER database does not differentiate between microsatellite-unstable (MSI) and microsatellite-stable CRCs. MSI CRCs constitute a minority (≈10–15%) of the CRCs in the population and the fact that we do not consider them explicitly in our models should not affect our conclusions.

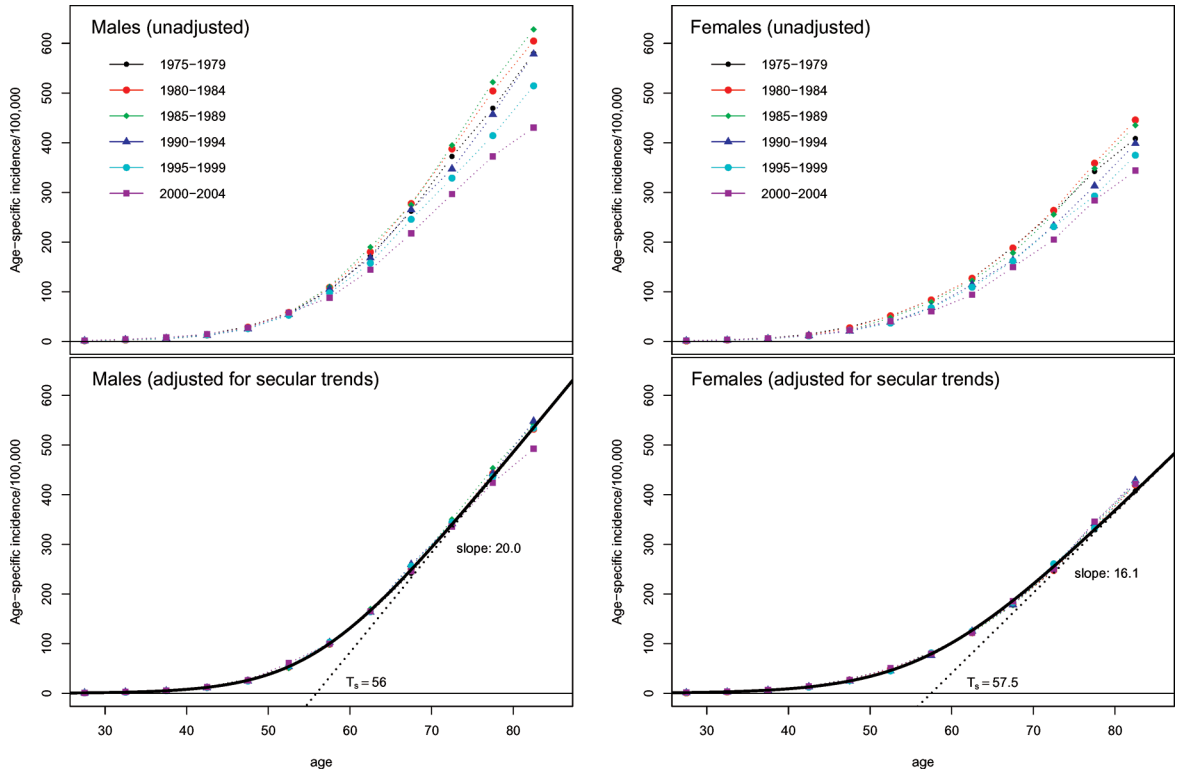

Fig. 2. Colorectal cancer incidence. (*Upper*) SEER CRC incidence. (*Lower*) CRC incidence adjusted for secular trends (using estimated calendar year and birth-cohort effects from the three-stage model fit). Solid line: three-stage hazard. The slope of the linear phase of the hazard and the mean sojourn time of premalignant lesions can be determined directly from the adjusted incidence data.

Polynomial and exponential phase (birth to mid age). For times $t < T_s$, the hazard function can be approximated by

$$h_3(t) \approx \frac{\mu_0 X \mu_1}{\alpha} \xi [\exp((\alpha - \beta)t) - (\alpha - \beta)t - 1], \qquad [4]$$

where $\xi \approx \alpha\mu_2/(\alpha - \beta)^2$.

The exponential term dominates this expression for times $> (\alpha - \beta)^{-1}$ and the hazard rises exponentially with rate $(\alpha - \beta)$, the growth rate of premalignant neoplasms. For times much smaller than $(\alpha - \beta)^{-1}$, the hazard is essentially quadratic, consistent with the Armitage and Doll approximation to the solution of a three-stage model without clonal expansion

$$h_3(t) \approx \frac{1}{2}\mu_0 X \mu_1 \mu_2 t^2. \qquad [5]$$

There is clinical and experimental evidence that suggests that the growth rate of benign human neoplasms, $(\alpha - \beta)$, such as intestinal polyps, is very small, consistent with clonal doubling times of several years. However, for growth rates >0.1 per year, the quadratic (Armitage and Doll) phase occurs mainly during the first decade of life so that this phase may be difficult to discern for sporadic cancers in a population that may include susceptible individuals and individuals with heritable forms of the cancer.

Data Analysis Results. We fitted the three-stage, four-stage, and Armitage–Doll models to the incidence of CRC and PC in the SEER database, while adjusting for period and cohort effects (see *Methods*). Figs. 2 and 3 summarize our results for the analyses of CRC and PC incidence, respectively. These figures correspond to the three-stage model fits. The four-stage model yields almost identical results. Both cancer types are analyzed by gender and for the largest ethnic category. For more details, see *Methods*. Numerical results, that is, maximum likelihood estimates (MLEs) of the relevant biological parameters (for the three- and four-stage models), in particular, the slopes of the linear phase of the age-specific incidence and the estimated sojourn time, T_s, are provided in the *SI Appendix*, Table 1 (CRC), Table 2 (PC), and Table 3. The Akaike Information Criteria (AIC) of the three- and four-stage models, relative to the Armitage–Doll model, are also provided in the *SI Appendix*, Tables 1 and 2. The AICs indicate a huge improvement in fit when using the three- and four-stage models instead of the Armitage–Doll model. Fig. 2 *Upper* (*Left* and *Right*) shows the empirical incidence of CRC averaged across 5-year calendar year periods (1975–1979, 1980–1984, …, 2000–2004) for females (*Right*) and males (*Left*). Fig. 2 *Lower* (*Left* and *Right*) shows the incidence curves (normalized to 1975) after adjusting for period and cohort effects (as discussed in the *Methods*), together with the corresponding three-stage model hazard (thick dark line) and the linear-phase approximation (dotted line). Fig. 3 is organized the same way, but for PC.

Whereas CRC shows substantial period effects (data not shown), reflecting what are believed to be changes in screening and intervention practices, PC shows little variation with period or birth cohort. However, the incidence of PC among males reveals a downward trend with increasing period, possibly reflecting the role of smoking prevalence as a risk factor for PC (35, 36).** Apart from these differences in secular trends, the main difference in the adjusted age-specific incidence curves between CRC and PC are the slopes of the linear phase. For the three-stage model, the slope is given approximately by $\mu_0 X \mu_1 \times$ (probability a premalignant clone does not go extinct). The slope for PC is only 1/5.1 of that for CRC among males, and 1/5.8 of that for CRC among females. Qualitatively, the two models

**The inferred secular trends using the three-stage model are very close to those inferred from the four-stage model and similar to those obtained by using the Armitage–Doll model.

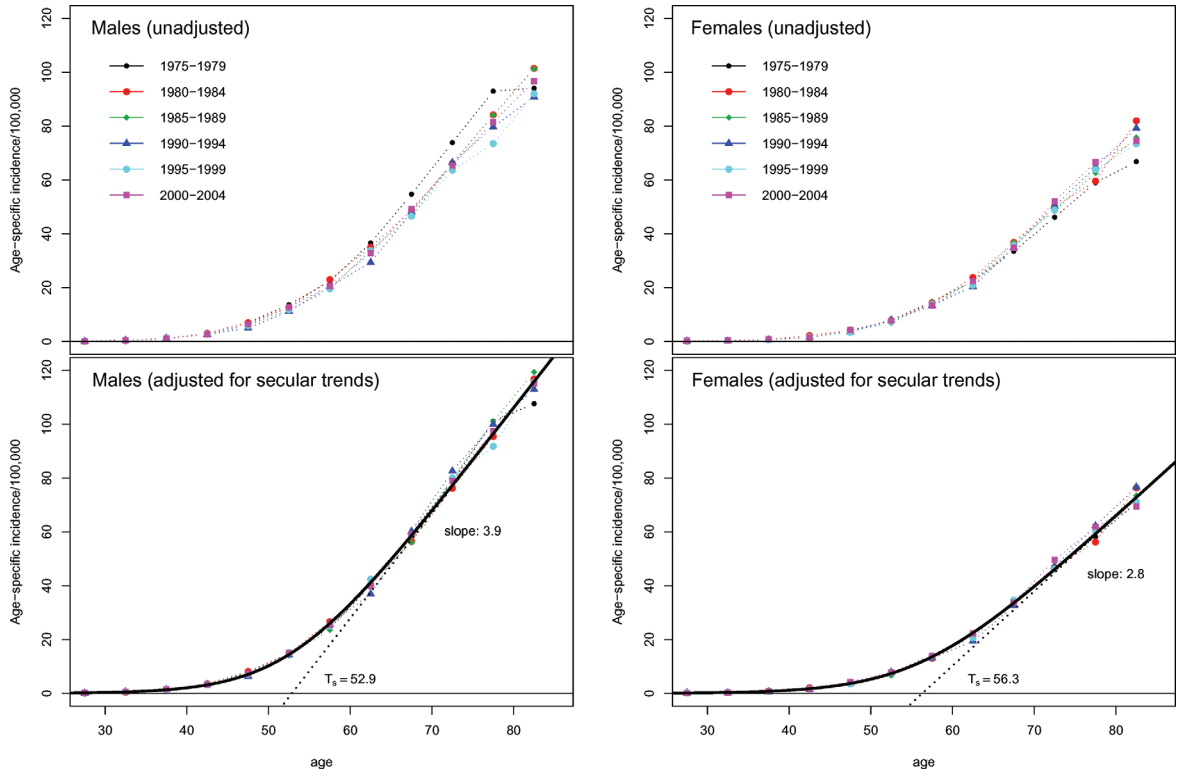

Fig. 3. Pancreatic cancer incidence. (*Upper*) SEER PC incidence. (*Lower*) PC incidence adjusted for secular trends (using estimated calendar year and birth-cohort effects from the three-stage model fit). Solid line: three-stage hazard. The slope of the linear phase of the hazard and the mean sojourn time of premalignant lesions can be determined directly from the adjusted incidence data.

used for this demonstration (i.e., the three- and four-stage clonal expansion model) fit the data equally well, but for CRC the four-stage model yields a significantly better AIC (see *SI Appendix*, Table 1) and therefore is statistically superior.

Discussion

The age-specific incidence of both CRC and PC, despite their distinct pathogenesis, show a remarkable similarity when adjusted for secular trends (see Figs. 2 and 3). The incidence curves for both cancers essentially follow an exponential growth phase with similar growth parameters, $\alpha - \beta$, followed by a mostly linear increase for ages >60 years. This behavior was also reported by Herrero-Jimenez *et al*. (34) in their analysis of CRC mortality and is expected (see *Theoretical Results*) for multistage carcinogenesis models that require at least two rate-limiting events prior to the clonal expansion of premalignant lesions. For CRC, the obvious precursor lesion is the adenomatous polyp. Adenomatous polyps, in colon and rectum, are considered the main precursor lesion for colorectal adenocarcinoma and are targets for cancer screening, intervention, and prevention (37, 38). Molecular evidence also suggests that a large majority of colorectal tumors carry APC mutations or loss of heterozygosity at the APC locus, consistent with biallelic inactivation of the APC gene (function) (29, 32, 39, 40). The situation for PC is less clear. However, ductal pancreatic atypia in the form of pancreatic intraepithelial neoplasias (PanINs) have been identified in biopsies as putative precursor lesions on the pathway to PC (41) and PanINs frequently show biallelic inactivation of the CDKN2A (p16) gene.

Our analysis of the multistage clonal expansion model shows how a stage-wise progression toward cancer, although limited here to only one stage of clonal expansion, maps onto distinct phases of the age-specific incidence (or hazard) function. These phases recapitulate the multistage process in reverse, with late-stage events having a discernible impact on the hazard function mainly early in life, clonal expansion (or promotion) mainly effecting the exponential behavior of the hazard function in midlife, whereas the early (pre)initiation events almost exclusively control the shape of the hazard function later in life. Specifically, as time goes to infinity the hazard functions associated with the models shown in Fig. 1 all approach a constant value. For $k > 2$, that is, for more than two stages, they all approach the asymptote $\mu_0 X$, a limit that is unlikely to be reached in a person's lifetime for typical (locus-specific) mutation rates and values for X, the number of susceptible stem cells, in the hundred thousands or millions. For this reason the parameter $\mu_0 X$ is not estimable from incidence data. However, for ages below $1/\mu_1$ but larger than the sojourn time of the lesion that arises from a stem cell that has suffered the first two oncogenic events, the hazard function essentially behaves linearly. The sojourn time of this lesion, however, depends on whether further rate-limiting events are required for malignant transformation, on its net growth once initiation has taken place, and on the rate of malignant transformation (see *Results*).

Both the incidence of PC and CRC are well described by three-stage models that posit premalignant (neoplastic) lesions that require two rate-limiting events for their initiation. For CRC, a four-stage model that assumes an additional (nonmutational) high-frequency event fits the data somewhat better (see *SI Appendix*, Table 1). However, this additional high-frequency process only introduces a relatively short transient quadratic departure from the linear phase of the hazard function and changes little in the main behavior of the hazard function and the estimates of sojourn times. Our finding that two very different cancers, with known pathologic and pathogenetic differences, appear to share such similarity in the analytic features of their respective age-specific incidence is noteworthy. Both cancers exhibit an exponential phase that can be attributed to uniform promotion, that is, the constant slow clonal expansion of initiated (premalignant) clones. There appears to be no hint in the data

that multiple expansive stages, with disparate growth rates, are involved. Strikingly, the expansion rate parameters, $\alpha - \beta$, are very similar for these two cancers (in the range of 0.14–0.18 per year for the three-stage model, and 0.14–0.19 per year for the four-stage model; see *SI Appendix*, Tables 1 and 2). Because the sojourn time T_s depends mainly on the inverse of $\alpha - \beta$ and only logarithmically on the malignant transformation rate μ_{k-1} (see *Results*), the estimates of T_s are also very similar for the two cancers (*SI Appendix*, Table 3). For colon, adenoma-to-carcinoma sojourn times on the order of 10–20 years have been surmised from clinical data (42, 43) and more recently from comparative sequencing (44), but such estimates typically ignore the duration of occult growth that starts with a single initiated cell and ends with the detection of an observable (millimeter size) adenoma. Whether the similarity of the estimated sojourn times in colon and pancreas is a general feature of epithelial cancers or mere coincidence for the two particular cancers studied here remains to be seen. However, estimates of $\alpha - \beta$ in lung (45, 46), breast (W. D. Hazelton, personal communication), and Barrett's esophagus (47) suggest that clonal expansion of the earliest epithelial neoplasms is uniformly slow, ensuing with rates not too different from those found here (between 0.1 and 0.2 per year).

What then is the main difference in the age-specific incidence of CRC and PC? A comparison of the fitted (and adjusted) age-specific incidence curves (Figs. 2 and 3) and the slopes of their linear phases suggests that differences in the two-step initiation process that gives rise to a slow-growing precursor lesion are the main reason. The slope of the linear phase of the hazard function (essentially $\mu_0 X \mu_1$) can be interpreted as the curvature in the age-specific prevalence of precursor lesions in a population (see *SI Appendix*). Thus, the observed difference in the incidence of the two cancers (PC and CRC) could mainly arise from a difference in X, the number of susceptible stem cells, or from a difference in the mutation rates $\mu_{0(1)}$, or both. The human colon has $\approx 2 \times 10^7$ crypts (48), each maintained by a number of adult colonic stem cells. Nicolas *et al*. (49) have estimated the number of such stem cells to be between 8 and 20 cells, although their findings also allow larger numbers. Assuming equality of the first two mutational events and using the four-stage model, we estimate a mutation rate $\mu_0 = 0.7$–1.1×10^{-7} per year. This estimate is somewhat smaller than those given in Iwama (50) and Luebeck *et al*. (21), but consistent with those estimated by Herrero-Jimenez *et al*. (34, 51).

The low magnitude of our slope estimates therefore suggests that initiation is likely the sporadic (biallelic) inactivation of a gatekeeper gene such as APC for colon or CDKN2A (the gene that expresses p16INK4A) for pancreas, consistent with Knudson's two-hit oncogenesis hypothesis for tumor initiation rather than transformation. In particular, for PC, which appears more consistent with a three-stage clonal expansion model than a four-stage model (the AIC for males is only marginally better, but statistically inferior for females), initiation of a slowly expanding precursor lesion in two rare rate-limiting steps is sufficient to explain the linear behavior of its age-specific incidence above age 60.

The broad view that "multistage carcinogenesis" is merely a stepwise accumulation of specific (epi)genetic aberrations in the genome of susceptible stem cells, and the notion that this accumulation leads essentially to a log-log linear age-specific incidence function cannot be reconciled with the observed cancer incidence of two important cancers. Our analyses with the (exact) Armitage–Doll model not only provide very poor fits, but also lead to estimates of the number of rate-limiting events much higher than previously reported. For CRC we estimate 10–11 rate-limiting events, rather than the 6–7 often reported in the literature (however, see ref. 52). This discrepancy is likely a consequence of our using the exact solution for the Armitage–Doll hazard function rather than the aforementioned power-law approximation.

Our results also have bearing on the question whether or not genomic instability (as chromosomal instability, or CIN) is induced prior to the loss of control imposed by the gatekeeper gene (53, 54). For the colon, our estimates of the "slope" parameters, together with estimates of stem cell numbers reported in the literature, suggest that the induction of CIN is unlikely to occur prior to the first gatekeeper mutation, but could represent the second (rare) event before the second gatekeeper allele is lost. However, under this assumption, the "fast" third event predicted by our four-stage model would translate into a gatekeeper mutation frequency (under CIN) that is ≈ 1 per stem cell division (parameter μ_{k-2}/α; *SI Appendix*, Table 1), clearly too high for a mutational event, but not too high to represent the transient amplification of mutant (gatekeeper$-/-$) stem cells in the colonic crypt. Similar arguments against the occurrence of early CIN can be made for pancreas, a tissue that does not have crypts and for which the cancer incidence is better described by a three-stage clonal expansion model.

The results presented here are consistent with Knudson's suggestion (26) that renewal tissue is converted into growing tissue in the first stage of carcinogenesis. Moreover, our results suggest that this conversion occurs as a result of biological events equivalent to the biallelic inactivation of tumor suppressor genes. This model predicts that inheritance of the "first hit" leads to the appearance of a large number of benign lesions in the target tissue. Indeed, a number of clinical conditions, the phakomatoses, have been associated with mutations in specific tumor suppressor genes, for example, the APC gene in FAP, NF1/2 in neurofibromatosis 1/2, and VHL in Von Hippel Landau Syndrome.

We provide evidence (by using the SEER data for CRC and PC) that the age-specific incidence increases linearly for the majority of cases above age 60–65 and that, unless the population is heterogeneous in the rates of the first two mutations or in X (e.g., see ref. 55), this linearity would continue to hold beyond the attainable lifespan. At least for CRC and PC, the incidence above age 60 appears vastly more consistent with a linear age-specific cancer incidence than one that assumes log-log linearity. Arguments based on the later assumption may be erroneous and should be revisited. Although log-log linearity can be rejected on statistical grounds, the number of cancer pathways, in particular, those involved in the adenoma-carcinoma sequence, and the number of obligatory (epi)genetic events associated with these pathways, are presently unknown. Recent genomic evidence, however, indicates that this number is low (44) and that the very long sojourn time of an adenoma could be punctuated by multiple (epi)genomic sweeps or clonal expansions. However, exploratory fits with at least two such expansions for CRC and PC indicate that there is no clear statistical signature in the SEER data for additional clonal expansion stages. The fits we obtained did not yield a better description of the incidence of CRC and PC. This is by no means proof that additional expansions do not exist, but it does indicate that they may occur late in precursor or tumor development, perhaps at such short timescales that they should not be considered rate-limiting.

Methods

Adjustment of the Age-Specific Incidence for Secular Trends. We have previously presented likelihood-based analyses of the incidence of CRC in the SEER registry (years 1973–1996) by using multistage models (as shown in Fig. 1) to model parametrically the effect of age, whereas secular trends, that is, period and cohort effects, were modeled nonparametrically (see ref. 21). In brief, for a given age a, birth cohort b, and calendar year (period) c, the age-specific incidence was estimated by

$$I_c(a) = \theta_c \theta_b h_{\text{mult}}(a), \qquad [6]$$

where θ_c and θ_b are coefficients that modify the incidence function predicted by the multistage process $[h_{\text{mult}}(a)]$ allowing for nonspecific period and cohort trends. Conforming with the format the SEER data are distributed, we stratify the data in age groups (0–4, 5–9, …, 80–84, 85+ years) and into 32 calendar years (1973–2004). Age group 85+ was excluded for males because rapidly declining person years after age 85. We then fit three- and

four-stage clonal expansion models to the number of observed CRC cases [International Classification of Diseases, 9th Revision (ICD9): 153.0–154.1] and PC cases (ICD9: 157.0–157.3, 157.8–157.9) stratified by age group and calendar year. We obtain parameter estimates by maximizing the likelihood across all age-calendar strata assuming that the number of cases in each stratum is Poisson distributed with mean $I_{c_j}(a_i)PY_{ij}$, where PY_{ij} denotes the number of person years in age group i and calendar year j. Parameter estimates and confidence intervals are provided in *SI Appendix*, Tables 1 and 2.

ACKNOWLEDGMENTS. We thank Douglas Brash (Yale School of Medicine, New Haven, CT) for helpful comments. This work was supported by National Institutes of Health Grants RO1 CA107028 and RO1 CA119224.

1. Muller HJ (1951) Radiation damage to the genetic material. *Sci Prog* 7:93–165, 481–493.
2. Nordling CO (1953) A new theory on cancer-inducing mechanism. *Br J Cancer* 7:68–72.
3. Armitage P, Doll R (1954) The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br J Cancer* 8:1–12.
4. Moolgavkar SH (1991) *Stochastic Models of Carcinogenesis*, eds Rao CR, Chakraborty R (Elsevier, New York), pp 372–393.
5. Nowell PC (1976) The clonal evolution of tumor cell populations. *Science* 194:23–28.
6. Renan MJ (1993) How many mutations are required for tumorigenesis? Implications from human cancer data. *Mol Carcinog* 7:139–146.
7. Pierce DA, Mendelsohn ML (1999) A model for radiation-related cancer suggested by atomic bomb survivor data. *Radiat Res* 152:642–654.
8. Frank SA (2005) Age-specific incidence of inherited versus sporadic cancers: A test of the multistage theory of carcinogenesis. *Proc Natl Acad Sci USA* 102:1071–1075.
9. Cairns J (2002) Somatic stem cells and the kinetics of mutagenesis and carcinogenesis. *Proc Natl Acad Sci USA* 99:10567–10570.
10. Frank SA, Nowak MA (2003) Developmental predisposition to cancer. *Nature* 422:494.
11. Loeb LA (2001) A mutator phenotype in cancer. *Cancer Res* 61:3230–3239.
12. Uys P, van Helden PD (2003) On the nature of genetic changes required for the development of esophageal cancer. *Mol Carcinog* 36:82–89.
13. Calabrese P, Tavare S, Shibata D (2004) Pretumor progression: Clonal evolution of human stem cell populations. *Am J Pathol* 164:1337–1346.
14. Hornsby C, Page KM, Tomlinson IP (2007) What can we learn from the population incidence of cancer? Armitage and Doll revisited. *Lancet Oncol* 8:1030–1038.
15. Holford TR (1991) Understanding the effects of age, period, and cohort on incidence and mortality rates. *Annu Rev Public Health* 12:425–457.
16. Chu KC, Tarone RE, Chow WH, Hankey BF, Ries LA (1994) Temporal patterns in colorectal cancer incidence, survival, and mortality from 1950 through 1990. *J Natl Cancer Inst* 86:997–1006.
17. Cress RD, Morris C, Ellison GL, Goodman MT (2006) Secular changes in colorectal cancer incidence by subsite, stage at diagnosis, and race/ethnicity, 1992–2001. *Cancer* 107:1142–1152.
18. Larsson SC, Wolk A (2007) Obesity and colon and rectal cancer risk: A meta-analysis of prospective studies. *Am J Clin Nutr* 86:556–565.
19. Clayton D, Schifflers E (1987) Models for temporal variation in cancer rates. I: Age-period and age-cohort models. *Stat Med* 6:449–467.
20. Clayton D, Schifflers E (1987) Models for temporal variation in cancer rates. II: Age-period-cohort models. *Stat Med* 6:469–481.
21. Luebeck EG, Moolgavkar SH (2002) Multistage carcinogenesis and the incidence of colorectal cancer. *Proc Natl Acad Sci USA* 99:15095–15100.
22. Moolgavkar SH, Venzon DJ (1979) Two-event models for carcinogenesis: Incidence curves for childhood and adult tumors. *Math Biosci* 47:55–77.
23. Moolgavkar SH, Luebeck G (1990) Two-event model for carcinogenesis: Biological, mathematical, and statistical considerations. *Risk Anal* 10:323–341.
24. Heidenreich WF, Luebeck EG, Moolgavkar SH (1997) Some properties of the hazard function of the two-mutation clonal expansion model. *Risk Anal* 17:391–399.
25. Knudson AG (1971) Mutation and cancer: Statistical study of retinoblastoma. *Proc Natl Acad Sci USA* 68:820–823.
26. Knudson AG (2000) Chasing the cancer demon. *Annu Rev Genet* 34:1–19.
27. Knudson AG (2001) Two genetic hits (more or less) to cancer. *Nat Rev Cancer* 1:157–162.
28. Bodmer WF, *et al*. (1987) Localization of the gene for familial adenomatous polyposis on chromosome 5. *Nature* 328:614–616.
29. Miyoshi Y, *et al*. (1992) Somatic mutations of the APC gene in colorectal tumors: Mutation cluster region in the APC gene. *Hum Mol Genet* 1:229–233.
30. Kinzler KW, *et al*. (1991) Identification of FAP locus genes from chromosome 5q21. *Science* 253:661–665.
31. Goss KH, Groden J (2000) Biology of the adenomatous polyposis coli tumor suppressor. *J Clin Oncol* 18:1967–1979.
32. Segditsas S, Tomlinson I (2006) Colorectal cancer and genetic alterations in the Wnt pathway. *Oncogene* 25:7531–7537.
33. Meza R, Luebeck EG, Moolgavkar SH (2005) Gestational mutations and carcinogenesis. *Math Biosci* 197:188–210.
34. Herrero-Jimenez P, Tomita-Mitchell A, Furth EE, Morgenthaler S, Thilly WG (2000) Population risk and physiological rate parameters for colon cancer. The union of an explicit model for carcinogenesis with the public health records of the United States. *Mutat Res* 447:73–116.
35. Moolgavkar SH, Stevens RG (1981) Smoking and cancers of bladder and pancreas: Risks and temporal trends. *J Natl Cancer Inst* 67:15–23.
36. Fuchs CS, *et al*. (1996) A prospective study of cigarette smoking and the risk of pancreatic cancer. *Arch Intern Med* 156:2255–2260.
37. Winawer SJ, *et al*. (2006) Guidelines for colonoscopy surveillance after polypectomy: A consensus update by the US Multi-Society Task Force on Colorectal Cancer and the American Cancer Society. *CA Cancer J Clin* 56:143–159.
38. Vogelaar I, *et al*. (2006) How much can current interventions reduce colorectal cancer mortality in the U.S.? Mortality projections for scenarios of risk-factor modification, screening, and treatment. *Cancer* 107:1624–1633.
39. Fearon ER, Vogelstein B (1990) A genetic model for colorectal tumorigenesis. *Cell* 61:759–767.
40. Groden J, *et al*. (1995) Response of colon cancer cell lines to the introduction of APC, a colon-specific tumor suppressor gene. *Cancer Res* 55:1531–1539.
41. Hruban RH, Goggins M, Parsons J, Kern SE (2000) Progression model for pancreatic cancer. *Clin Cancer Res* 6:2969–2972.
42. Winawer SJ, *et al*. (1997) Colorectal cancer screening: Clinical guidelines and rationale. *Gastroenterology* 112:594–642.
43. Ness RM, Holmes AM, Klein R, Dittus R (2000) Cost-utility of one-time colonoscopic screening for colorectal cancer at various ages. *Am J Gastroenterol* 95:1800–1811.
44. Jones S, *et al*. (2008) Comparative lesion sequencing provides insights into tumor evolution. *Proc Natl Acad Sci USA* 105:4283–4288.
45. Hazelton WD, Clements MS, Moolgavkar SH (2005) Multistage carcinogenesis and lung cancer mortality in three cohorts. *Cancer Epidemiol Biomarkers Prev* 14:1171–1181.
46. Meza R, Hazelton W, Colditz G, Moolgavkar S (2008) Analysis of lung cancer incidence in the nurses' health and the health professionals' follow-up studies using a multistage carcinogenesis model. *Cancer Causes Control* 19:317–328.
47. Jeon J, Luebeck EG, Moolgavkar SH (2006) Age effects and temporal trends in adenocarcinoma of the esophagus and gastric cardia (United States). *Cancer Causes Control* 17:971–981.
48. Potten CS, Booth C, Hargreaves D (2003) The small intestine as a model for evaluating adult tissue stem cell drug targets. *Cell Prolif* 36:115–129.
49. Nicolas P, Kim KM, Shibata D, Tavare S (2007) The stem cell population of the human colon crypt: Analysis via methylation patterns. *PLoS Comput Biol* 3:e28.
50. Iwama T (2001) Somatic mutation rate of the APC gene. *Jpn J Clin Oncol* 31:185–187.
51. Herrero-Jimenez P, *et al*. (1998) Mutation, cell kinetics, and subpopulations at risk for colon cancer in the United States. *Mutat Res* 400:553–578.
52. Heidenreich WF, Luebeck EG, Hazelton WD, Paretzke HG, Moolgavkar SH (2002) Multistage models and the incidence of cancer in the cohort of atomic bomb survivors. *Radiat Res* 158:607–614.
53. Komarova NL, Lengauer C, Vogelstein B, Nowak MA (2002) Dynamics of genetic instability in sporadic and familial colorectal cancer. *Cancer Biol Ther* 1:685–692.
54. Nowak MA, *et al*. (2002) The role of chromosomal instability in tumor initiation. *Proc Natl Acad Sci USA* 99:16226–16231.
55. Morgenthaler S, Herrero P, Thilly WG (2004) Multistage carcinogenesis and the fraction at risk. *J Math Biol* 49:455–467.