Published in final edited form as:

Math Biosci. 2009 April ; 218(2): 88–97. doi:10.1016/j.mbs.2008.12.007.

Studying Health Histories of Cancer: A New Model Connecting Cancer Incidence and Survival

Anatoli I. Yashin, Igor Akushevich, Konstantin Arbeev, Lucy Akushevich, Alexander Kulminski, and Svetlana Ukraintseva

Department of Sociology, Duke University, 002 Trent Hall, Box 90408 Durham, NC 27708-0408, USA

Abstract

The results of recent experimental and epidemiological studies provide evidence on the connection between carcinogenesis, cancer progression, and aging. Existing models, however, are traditionally focused only on one of these aspects of health deterioration. In this paper, we derive a new model of cancer, which describes the connection between the age at disease onset, the duration of disease, and life span of respective individuals. The model combines ideas used in the two hits model of carcinogenesis with those used in the Le Bras multistate model of aging with constant transition intensities. The model is used in the joint analyses of the U.S. demographic mortality data and SEER data for selected cancers. The results show that the developed approach is capable of explaining links among health history data and provides useful insights on mechanisms of cancer occurrence, disease progression, other aging-related changes, and mortality. Further developments of this model are discussed.

Keywords

aging; cancer; case fatality; duration of disease; Le Bras model; cancer mortality

Introduction

Traditional epidemiological studies of cancer often approximate age patterns of the incidence and mortality rates by appropriate parametric functions. Although such formal description is certainly convenient from the demographic point of view, it does not allow for biological interpretation of the respective parameters, and, therefore, has a limited use beyond the demographic applications. The idea of describing cancer hazard rates in terms of biologically interpretable parameters resulted in the development of the research area called the “mathematical modeling of cancer” or “cancer modeling.” Comprehensive reviews on this area can be found in refs. [1–7]. Many important aspects of initiation and development of different types of cancer have been mathematically described during more than fifty years of cancer modeling. The properties of biological mechanisms involved in cancer initiation and development have been investigated by fitting respective models to experimental data [8–13]. The models, however, do not allow for studying the connection between cancer and aging

Corresponding Author: Igor Akushevich, Department of Sociology, Duke University, 002 Trent Hall, Box 90408, Durham, NC 27708-0408, Phone: (919)668-2715, Fax: (919)684-3861, e-mail: E-mail: igor.kushevich@duke.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

detected in recent molecular biological, epidemiological, and experimental studies [14–18]. To take such connections into account, new models of cancer are needed. In this paper, we suggest a new approach to modeling health history events associated with cancer and aging. The events include the onset of cancer, death from causes different from cancer, and death of individuals with and without cancer. Besides, this approach allows for taking into account unobservable precancerous changes in critical tissues, as well as deterioration in individuals' health associated with the aging-related processes. The approach extends the Le Bras model (LBM) of aging [19] used earlier in the analyses of mortality data in heterogeneous populations (see [20] and references therein). An attractive feature of this model is that all functional forms for the population characteristics can be derived analytically (this analytical derivation was independently verified using the computer algebra software MAPLE 9.00). The derived model was tested by fitting it to the combined data set consisting of the Surveillance, Epidemiology and End Results (SEER) dataset on cancer incidence and case-fatality data in the U.S. [21] and the U.S. overall mortality data taken from the Human Mortality Database (HMD) [22].

Description of the Standard Le Bras Model (LBM)

Le Bras [19] suggested a multistate model with age-independent transitions which can be used for describing the age pattern of the mortality rate at the interval of human aging. The model assumes that at age x an individual can be in one of the infinite number of health states $S_0, S_1, \dots, S_n, \dots$ corresponding to different levels of functioning (Fig. 1). In any state S_i , a person faces a hazard of death and a hazard of moving to state S_{i+1} where the chances of survival are lower and chances of health deterioration are higher than those in the S_i -th state. Transitions from S_i -th state to S_{i-1} -th state are not possible. This means that this model excludes possibility of recovery from aging-related health deterioration. This assumption is completely justified for aging-related changes [23]. Assume that all individuals in a cohort start from the state S_0 . This means that all newborns in the cohort have the same initial chances of survival. Let λ_0 and μ_0 be the transition rates from state S_0 to state S_1 , and to death, respectively. For the S_i -th state, let these transition rates be $\lambda_0 + i\lambda$, and $\mu_0 + i\mu$. Then the survival function in such a system is:

$$S(x) = e^{-(\lambda_0 + \mu_0)x} \left(\frac{\lambda + \mu}{\mu + \lambda e^{-(\lambda + \mu)x}} \right)^{\frac{\lambda_0}{\lambda}}$$

and the mortality rate is:

$$\bar{\mu}(x) = \mu_0 + \frac{\mu \lambda_0 [1 - e^{-(\lambda + \mu)x}]}{\mu + \lambda e^{-(\lambda + \mu)x}}.$$

Although transition intensities in this model do not depend on age, they do increase from one state to the next. The advantage of this model is that it describes a multistate process of aging and mortality with infinite number of states using four parameters. Yashin et al. [20] showed that the mortality curve derived from the modified LBM gives a nice fit to the mortality data at the entire interval of aging. Moreover, the parametric form of the mortality rate coincides with that calculated for the Gamma-Makeham frailty model. It turns out that this approach can be extended to capture different aspects of cancer initiation and development, as well as mortality for individuals with and without cancer.

Population Model of Carcinogenesis

General scheme and notations

The diagram in Fig. 2 illustrates a further extension of the Le Bras model. The standard LBM is a particular case of that shown in Fig. 2, which occurs when considering the lower line of boxes and respective transitions excluding transitions to the upper states.

Horizontal transitions in the lower line of boxes (*H*-line) are interpreted as the process of individual aging. Individuals taken from any box from this line are free of cancer or pre-cancer, so we will refer to individuals from this line of boxes as individuals in a “healthy” stage *H*, and the respective health states will be denoted as H_i , $i = 0, 1, 2, \dots$

Vertical transitions to the upper states from the *H*-line are interpreted as unobserved transfers to a precancerous state. This is in accordance with the two-stage hypothesis of development of carcinogenesis [24–28]. The precancerous states are denoted by P_{jn} , where the first index j characterizes the functional state in the lowest line from which pre-cancer was originated, and the second index n reflects the individual aging state in the pre-cancer individuals in the same manner as in the healthy states.

The vertical transitions from the *P*-lines to the states marked by M_{kjni} (the malignant stage) are interpreted as an onset of cancer. Each malignant state is described by four indices. The pairs jn denote the state of origin in precancerous lines and the last index i reflects the individual aging state in k -th line with malignant states (*M*-line). The index k is used to show that each line with malignant states (*M*-lines) originated from the state in the precancerous line during a certain age interval $(x_0 + k\Delta_x, x_0 + (k + 1)\Delta_x)$ marked by k . It means that the cohort is formed in the first malignant state during a certain time period. Beyond this time period, no new cases come to the malignant line. For example, if one-year time periods for collecting new malignant cases are considered, the malignant line corresponding to $k = 0$ (and respective j and n characterizing the parental precancerous state) is formed during the time period $(x_0, x_0 + 1)$. Then, for $x = x_0 + 1$, this line unhooks from the connection to the precancerous state P_{jn} and only dynamic changes with this cohort, as described by the standard Le Bras model, occur. New cancer cases occurring in the next time period (i.e., $(x_0 + 1, x_0 + 2)$) in the state P_{jn} appear to the next malignant line with $k = 1$. This approach allows us to reflect the typical situation that the age of onset is observed and, therefore, to describe the dynamics of malignant lines with different ages at onset by different sets of parameters, i.e., transition rates. Thus, the scheme in Fig. 2 corresponds to the k -th age period $(x_0 + k\Delta_x, x_0 + (K + 1)\Delta_x)$, where x_0 is the age of cohort formation, and Δ_x is the length of the age interval ($\Delta_x = 1$ year in our analysis), and k runs over all age periods. The dynamics of a particular malignant line is presented in Fig. 3. The up and down transitions correspond to the mortality caused by cancer and other causes. Such an approach results in a combined model simultaneously describing the rates of the total mortality, cancer-specific mortality, and cancer incidence.

Similarly to the standard LBM, the number of states in each line is infinite. This is an advantage of the model because it is not necessary to introduce an artificial parameter corresponding to the number of states in each line. Dealing with an infinite number of states requires generalization of the technique of summation over these infinite states originally used in the standard LBM.

Transition rates

Four types of the transition rates are considered in the model. They correspond to effects of aging (marked by λ 's), transitions between carcinogenesis stages (marked by γ 's), and mortalities caused by cancer (δ 's) and non-cancer (μ).

In all stages (i.e., in healthy, precancerous, and malignant stages), the process of a functional decline is represented by the hazard rates of a general form $\lambda_0 + j\lambda + n\bar{\lambda} + i\lambda_k$, where j , n , and i are numbers of steps passed in healthy, precancerous, k -th malignant lines. Thus, the rates in states H_i , P_{ji} , and M_{kjni} are equal to $\lambda_0 + i\lambda$, $\lambda_0 + j\lambda + i\bar{\lambda}$, and $\lambda_0 + j\lambda + n\bar{\lambda} + i\lambda_k$, respectively. For conciseness, we use additional notations: $\lambda_j = \lambda_0 + j\lambda$ (in precancerous lines in Fig. 2) and $\lambda_{0k} = \lambda_0 + j\lambda + n\bar{\lambda}$ (in malignant lines in Fig. 3). Note that in such an approach there are no discontinuities in the rates for transfers between different lines: the rate can change only by λ , $\bar{\lambda}$, or λ_k at the respective line. We consider different parameters for malignant lines corresponding to different k -th age periods. This will allow us to test alternative hypotheses considering whether these parameters are different or not. Keeping in mind that the survival function in the standard LBM is defined through the ratios of λ_0 and λ , it is convenient to introduce the following notation:

$$\xi = \frac{\lambda_0}{\lambda}, \quad \xi_0 = \frac{\lambda_0}{\bar{\lambda}}, \quad \bar{\xi} = \frac{\lambda}{\bar{\lambda}}, \quad \xi_{0k} = \frac{\lambda_0}{\lambda_k}, \quad \xi_k = \frac{\lambda}{\lambda_k}, \quad \bar{\xi}_k = \frac{\bar{\lambda}}{\lambda_k}.$$

Only three of them are independent, e.g., $\xi_0 = \xi\bar{\xi}$, $\xi_{0k} = \xi\bar{\xi}_k$, and $\xi_k = \bar{\xi}\bar{\xi}_k$.

The model is adapted for considering two types of mortality transitions: those caused or not caused by cancer. The rules of change in the non-cancer-specific mortality transitions are exactly the same as those for the functional decline, i.e., a general form of these transition rates is $\mu_0 + j\mu + n\bar{\mu} + i\mu_k$, and they are $\mu_0 + i\mu$ for the healthy line, $\mu_0 + j\mu + i\bar{\mu}$ for the precancerous lines, and $\mu_0 + j\mu + n\bar{\mu} + i\mu_k = \mu_{0k} + i\mu_k$ for the malignant lines. The cancer-specific mortality appears only in the malignant lines and has a general form $\delta_{0kjn} + i\delta_k$. If the cause of death is not measured, then all δ 's have to be assigned to zero and μ 's define the total mortality rate.

The last type of rates describes transfers between the lines. The transfer from the healthy line to the precancerous line is not observed. Its rate is $\gamma_0 + i\gamma$, i.e., it depends on a parental state as above. The transfer from the precancerous lines to the malignant lines corresponds to the observed age at onset. The general form of this rate is $\bar{\gamma}_0 + i\bar{\gamma}$. At the beginning of the precancerous lines, the rate is $\bar{\gamma}_0$, i.e., we assume that it is independent of a parental state in the healthy lines.

Master equations for all states

Let all individuals in a cohort start from the healthy state H_0 at the age x_0 . The differential equation for the probability of being in this state is

$$\frac{d}{dx}P_0(x) = -(\lambda_0 + \gamma_0 + \mu_0)P_0(x), \quad P(x_0) = 1.$$

The solution of this equation is: $P_0(x) = \exp(-\Lambda_0(x - x_0))$, where $\Lambda_0 = \lambda_0 + \gamma_0 + \mu_0$. For the states with $i \geq 1$, the differential equations are

$$\frac{d}{dx}P_i(x) = -(\lambda_0 + \gamma_0 + \mu_0 + i(\lambda + \gamma + \mu))P_i(x) + (\lambda_0 + (i-1)\lambda)P_{i-1}(x), \quad P_i(x_0) = 0.$$

The solution of this infinite system of differential equations is:

$$P_i(x) = \frac{e^{-\Lambda_0(x-x_0)} \left[1 - e^{-\Lambda(x-x_0)} \right]^i}{i! \Lambda^i} \prod_{m=1}^i (\lambda_0 + (m-1)\lambda), \quad (1)$$

where $\Lambda = \lambda + \gamma + \mu$. If to adopt the usual convention that $\prod_{i=1}^0 \bullet = 1$, the solution is valid for $i = 0$ as well. For zero values of γ_0 and γ , the solution reduces to that of the standard LBM (see e.g., [20] and references therein).

For the precancerous lines, the differential equations take the form

$$\frac{d}{dx} P_{j0}(x) = -(\lambda_j + \bar{\gamma}_0 + \mu_j) P_{j0}(x) + (\gamma_0 + j\gamma) P_j(x),$$

and

$$\frac{d}{dx} P_{ji}(x) = -(\lambda_j + \bar{\gamma}_0 + \mu_j + i(\bar{\lambda} + \bar{\gamma} + \bar{\mu})) P_{ji}(x) + (\lambda_j + (i-1)\bar{\lambda}) P_{j,i-1}(x),$$

for $i \geq 1$. Finally, for the malignant states they are

$$\frac{d}{dx} P_{kjm0}(x) = -(\lambda_{0k} + \delta_{0k} + \mu_{0k}) P_{kjm0}(x) + (\bar{\gamma}_0 + n\bar{\gamma}) P_{jn}(x) \text{In}(x_0 + k\Delta_x \leq x \leq x_0 + (k+1)\Delta_x),$$

where the indicator function *In* reflects the fact that, according to our assumptions, the malignant line is attached to the precancerous line during a limited age period. For $i \geq 1$, the equations are

$$\frac{d}{dx} P_{kjni}(x) = -(\lambda_{0k} + \delta_{0k} + \mu_{0k} + i(\lambda_k + \gamma_k + \mu_k)) P_{kjni}(x) + (\lambda_{0k} + (i-1)\lambda_k) P_{kjni-1}(x).$$

At the beginning (i.e., $x=x_0$), there are no individuals in any of the precancerous or malignant states, i.e., $P_{ji}(x_0) = 0$ and $P_{kjni}(x_0) = 0$. Recall that four subscripts in the functions describing probabilities of being in the malignant line reflect the age at onset (k), the parental compartment of the precancerous line (j and n), and compartments of the malignant line (i).

The remarkable property of such description is that the solution of all three systems of equations (i.e., for the healthy, precancerous, and malignant lines) can be obtained in terms of a “building block” model.

Building block model

Let us consider the following situation which will be regarded as a building block for all models we deal with:

$$\begin{aligned}\frac{d}{dx}P_0(x) &= -(\lambda_0 + \gamma_0 + \mu_0)P_0(x) + f(x), \\ \frac{d}{dx}P_i(x) &= -(\lambda_0 + \gamma_0 + \mu_0 + i(\lambda + \gamma + \mu))P_i(x) + (\lambda_0 + (i-1)\lambda)P_{i-1}(x), \quad i \geq 1,\end{aligned}$$

and for all states (i.e., $i \geq 0$), the initial conditions are $P_i(x_0) = 0$. The solution is

$$P_i(x) = \frac{1}{\Lambda^i i!} \left[\prod_{m=1}^i (\lambda_0 + (m-1)\lambda) \right] \int_{x_0}^x du f(u) e^{-\Lambda_0(x-u)} (1 - e^{-\Lambda(x-u)})^i. \quad (2)$$

The details of calculations are given in Appendix. The probabilities are defined by the following four parameters: λ_0 , λ , $\Lambda_0 = \mu_0 + \lambda_0 + \gamma_0$, and $\Lambda = \mu + \lambda + \gamma$. Thus, to apply the results of the model to a specific case, one has to specify these four parameters and a function $f(u)$. This function models the influx of individuals to the system.

Below we consider specific cases corresponding to the probabilities of being in the healthy, precancerous, and malignant lines.

The state probabilities for the *H*-line

The four parameters which have to be specified in the model for the *H*-line are λ_0 , λ , $\Lambda_0 = \mu_0 + \lambda_0 + \gamma_0$, and $\Lambda = \mu + \lambda + \gamma$. Since individuals enter the *H*-line only at time x_0 , the function $f(x)$ can be represented in terms of the Dirac delta function as $f(x) = \delta(x - x_0)$. Then, the distribution of individuals in the healthy line is given by

$$P_i(x) = \frac{e^{-\Lambda_0(x-x_0)} (1 - e^{-\Lambda(x-x_0)})^i}{\Lambda^i i!} \left[\prod_{m=1}^i (\lambda_0 + (m-1)\lambda) \right] \quad (3)$$

and its functional form coincides with (1) as expected. Thus, the model for the healthy line reproduces the standard LBM in the limit $\gamma, \gamma_0 \rightarrow 0$, i.e., when there are no transitions to the precancerous lines or when the transition to the precancerous states and mortality are considered as a generalized event characterizing the drop out of the system.

The state probabilities for the precancerous lines

Consider a precancerous line with j -th parental compartment of the healthy line. The four parameters specifying the building block model (λ_0 , λ , Λ_0 , and Λ) have the meaning of the rate of functional decline in the j th compartment, the rate of change in this rate, the rate of leaving the j th state, and the rate of change in this rate, respectively. For a sub-cohort of individuals leaving j -th compartment of the *H*-line for the respective precancerous lines, these parameters are specified as $\lambda_j = \lambda_0 + j\lambda$, $\bar{\lambda}$, $\bar{\Lambda}_0 = \mu_0 + j\mu + \lambda_0 + j\lambda + \bar{\gamma}_0$, and $\bar{\Lambda} = \mu + \bar{\mu} + \bar{\lambda} + \bar{\gamma}$. The flux of individuals entering the j -th precancerous line is continuous and is modeled as $f(x) = (\gamma_0 + j\gamma)P_j(x)$. The distribution of individuals in the j -th precancerous line is obtained by using these specifications in (3):

$$P_{ji}(x) = \frac{\gamma_0 + j\gamma}{\bar{\Lambda}^i i!} \left[\prod_{m=1}^i (\lambda_j + (m-1)\bar{\lambda}) \right] \int_{x_0}^x du P_j(u) e^{-\bar{\Lambda}_0(x-u)} (1 - e^{-\bar{\Lambda}(x-u)})^i.$$

Using the explicit equation (3) for the probability $P_j(x)$ in the j -th health parental compartment, we obtain

$$P_{ji}(x) = \frac{\gamma_0 + j\gamma}{i! \bar{\Lambda}^i j!} \left[\prod_{m=1}^i (\lambda_j + (m-1)\bar{\lambda}) \right] \left[\prod_{m=1}^j (\lambda_0 + (m-1)\lambda) \right] \times \int_{x_0}^x du e^{-\bar{\Lambda}_0(x-u)} e^{-\Lambda_0(u-x_0)} (1 - e^{-\Lambda(u-x_0)})^j (1 - e^{-\bar{\Lambda}(x-u)})^i, \quad (4)$$

and, finally,

$$P_{ji}(x) = e^{-\Lambda_0(x-x_0)} \frac{\gamma_0 + j\gamma}{i! j!} \left[\prod_{m=1}^i (\xi_0 + j\bar{\xi} + m - 1) \right] \left[\prod_{m=1}^j (\xi + m - 1) \right] \int_{x_0}^x du Q_1^i Q_2^j,$$

$$\text{where } Q_1 = \frac{\bar{\lambda}}{\bar{\Lambda}} \left(1 - e^{-\bar{\Lambda}(x-u)} \right) \text{ and } Q_2 = \frac{\lambda}{\Lambda} \left(1 - e^{-\Lambda(u-x_0)} \right) e^{-(\mu+\lambda)(x-u)}.$$

The state probabilities for the malignant lines

Specifics of constructing the model for probabilities in the malignant lines is that the flux of individuals entering each malignant line is nonzero during a particular time period Δ_x , e.g., $\Delta_x = 1$ year. As we discussed before, this is related to the fact that the age at onset is typically observed, and it is reasonable to separate respective cohorts in our consideration. Thus, the function $f(x)$ has to be specified in terms of an indicator function (or any other appropriate specification, e.g., a window function, a product of two Heaviside step functions) that constraints our specification to a particular age region. Here the index k runs over all regions with a specific age of onset: $[x_0 + k\Delta_x, x_0 + (k+1)\Delta_x]$. Then the function $f(x)$ is specified as

$$f(x) = (\bar{\gamma}_0 + n\bar{\gamma}) P_{jn}(x) \text{In}(x_0 + k\Delta_x \leq x < x_0 + (k+1)\Delta_x) \equiv (\bar{\gamma}_0 + n\bar{\gamma}) P_{jn}(x) \text{In}_k(x).$$

The four parameters specifying the building block model (λ_0 , λ , Λ_0 , and Λ) for the k n j -th malignant line become $\lambda_{0k} = \lambda_j + n\bar{\lambda} = \lambda_0 + j\lambda + n\bar{\lambda}$, $\lambda_k = \lambda_0$, $\Lambda_{0k} = \mu_{0k} + \lambda_{0k} \delta_0$ (where $\mu_{0k} = \mu_{0k} + j\mu + n\mu$), and $\Lambda_k = \mu_k + \lambda_k + \delta_k$. Thus, the distribution of individuals in the k j n -th malignant line is obtained in the form:

$$P_{kjni}(x) = \frac{\bar{\gamma}_0 + n\bar{\gamma}}{\Lambda_k^i i!} \left[\prod_{m=1}^i (\lambda_{0k} + (m-1)\lambda_k) \right] \int_{x_0}^x du P_{jn}(u) \text{In}_k(u) e^{-\Lambda_{0k}(x-u)} (1 - e^{-\Lambda_k(x-u)})^i.$$

Assuming that the age interval defined by the indicator function is sufficiently narrow, the integral can be estimated for a certain mean point x_k inside the region (in practice, a midpoint value of the interval can be chosen for x_k):

$$P_{kjni}(x) = \frac{\bar{\gamma}_0 + n\bar{\gamma}}{\Lambda_k^i i!} \left[\prod_{m=1}^i (\lambda_{0k} + (m-1)\lambda_k) \right] P_{jn}(x_k) e^{-\Lambda_{0k}(x-x_k)} (1 - e^{-\Lambda_k(x-x_k)})^i \Delta_x.$$

Using the explicit form of $P_{jn}(x_k)$ from (4), we finally obtain

$$P_{kjni}(x) = e^{-\Lambda_0(x-x_0)} \frac{\Delta_x(\bar{\gamma}_0+n\bar{\gamma})(\gamma_0+j\gamma)}{i!j!n!} \left[\prod_{m=1}^i (\xi_{0k} + j\xi_k + n\bar{\xi}_k + m - 1) \right] \times \\ \times \left[\prod_{m=1}^n (\xi_0 + j\bar{\xi} + m - 1) \right] \left[\prod_{m=1}^j (\xi + m - 1) \right] Q_i^i \int_{x_0}^{x_k} du Q_j^j Q_n^n.$$

All Q 's are defined as

$$Q_i = \frac{\lambda_k}{\Lambda_k} (1 - e^{-\Lambda_k(x-x_k)}), Q_j = \frac{\lambda}{\Lambda} (1 - e^{-\Lambda(u-x_0)}) e^{-(\lambda+\mu)(x-u)}, Q_n = \frac{\bar{\lambda}}{\bar{\Lambda}} (1 - e^{-\bar{\Lambda}(x_k-u)}) e^{-(\bar{\lambda}+\bar{\mu})(x-x_k)}.$$

The integral variable u in the expressions for P_{jn} and P_{kjni} has the meaning of the age at a latent event transferring an individual to a precancerous line. This event is unobservable and, therefore, the expressions contain an integral over this variable.

Observable quantities

When all probabilities of being in a specific state are determined, the functions representing the quantities observable in demographic and epidemiologic data can be constructed by calculating the respective sums over states. For example, the survival function is obtained as a sum over all states in the model:

$$S(x) = \sum_{j=0}^{\infty} P_j(x) + \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} P_{ji}(x) + \sum_{k, x_k < x} \sum_{j=0}^{\infty} \sum_{n=0}^{\infty} \sum_{i=0}^{\infty} P_{kjni}(x) = S_h + S_p + \sum_{k, x_k < x} S_m.$$

The survival function of the healthy life (defined here as the life with no cancer) is evaluated through the sums over the states in the healthy and precancerous lines: $S_h + S_p$. Here we define the following sums:

$$S_h = \sum_{j=0}^{\infty} P_j(x), S_p = \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} P_{ji}(x), S_m = \sum_{j=0}^{\infty} \sum_{n=0}^{\infty} \sum_{i=0}^{\infty} P_{kjni}(x).$$

In addition, the sums of probabilities multiplied by an index will be needed:

$$S_{ih} = \sum_{i=1}^{\infty} iP_i(x), S_{ip} = \sum_{j=0}^{\infty} \sum_{i=1}^{\infty} iP_{ji}(x), S_{jp} = \sum_{j=1}^{\infty} \sum_{i=0}^{\infty} jP_{ji}(x), \\ S_{im} = \sum_{j=0}^{\infty} \sum_{n=0}^{\infty} \sum_{i=1}^{\infty} iP_{kjni}(x), S_{nm} = \sum_{j=0}^{\infty} \sum_{n=1}^{\infty} \sum_{i=0}^{\infty} nP_{kjni}(x), S_{jm} = \sum_{j=1}^{\infty} \sum_{n=0}^{\infty} \sum_{i=0}^{\infty} jP_{kjni}(x).$$

The incidence and mortality rates can be obtained as derivatives of the respective survival function. It is more efficient, however, to construct them directly as explained below.

The incidence rate is defined as a fraction of new cancer cases occurring within a time unit (e.g., one year). In terms of the model's notation, this is the ratio of the number of individuals

entering the malignant lines per time unit to the total number of individuals in the healthy and precancerous states:

$$I(x) = \frac{\sum_{j=0}^{\infty} \sum_{i=0}^{\infty} (\bar{\gamma}_0 + i\bar{\gamma}) P_{ji}(x)}{\sum_{j=0}^{\infty} P_j(x) + \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} P_{ji}(x)} = \frac{\bar{\gamma}_0 S_p + \bar{\gamma} S_{ip}}{S_h + S_p}.$$

The mortality in the k -th malignant line (i.e., the line which corresponds to the cohort of individuals with age at cancer onset detected during the k -th age period) is defined as

$$\mu_{ck}(x) = \frac{\sum_{j=0}^{\infty} \sum_{n=0}^{\infty} \sum_{i=0}^{\infty} (\mu_{0k} + \delta_{0k} + i(\mu_k + \delta_k)) P_{kjni}(x)}{\sum_{j=0}^{\infty} \sum_{n=0}^{\infty} \sum_{i=0}^{\infty} P_{kjni}(x)} = \mu_0 + \delta_{0k} + \frac{\mu S_{jm} + \bar{\mu} S_{nm} + (\mu_k + \delta_k) S_{im}}{S_m}.$$

For the mortality from the healthy and precancerous states, we have

$$\mu_{hp}(x) = \frac{\sum_{i=0}^{\infty} P_i(x)(\mu_0 + i\mu) + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} P_{ji}(x)(\mu_0 + j\mu + i\bar{\mu})}{\sum_{i=0}^{\infty} P_i(x) + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} P_{ji}(x)} = \mu_0 + \frac{\mu S_{ih} + \mu S_{jp} + \bar{\mu} S_{ip}}{S_h + S_p}.$$

The total mortality is the sum of contributions from the healthy, precancerous, and malignant lines,

$$\mu_t(x) = \mu_0 + \frac{\mu S_{ih} + \mu S_{jp} + \bar{\mu} S_{ip} + \sum_{k, x_k < x} \delta_{0k} S_m + \mu S_{jm} + \bar{\mu} S_{nm} + (\mu_k + \delta_k) S_{im}}{S(x)}.$$

For the total mortality and the mortality in the k -th malignant line the contribution of the cancer-specific mortality can be extracted by keeping the terms with δ_{0k} and δ_k

Calculation of infinite sums

The remarkable property of the model is that the infinite sums defined above can be analytically calculated using the well-known formula

$$\sum_{i=0}^{\infty} \frac{a^i}{i!} \prod_{m=1}^i [b + (m-1)] = (1-a)^{-b}, \quad (5)$$

and the other two which can be obtained by differentiation over a :

$$\sum_{i=1}^{\infty} i \frac{a^i}{i!} \prod_{m=1}^i [b+(m-1)] = ab(1-a)^{-b-1}, \quad \sum_{i=1}^{\infty} i^2 \frac{a^i}{i!} \prod_{m=1}^i [b+(m-1)] = ab(1+ab)(1-a)^{-b-2}. \quad (6)$$

Using these formulae, we immediately obtain the sums of probabilities for the healthy lines:

$$S_h = \sum_{j=0}^{\infty} P_j(x) = e^{-\Lambda_0(x-x_0)}(1-Q_h)^{-\xi}, \quad S_{ih} = \sum_{j=0}^{\infty} iP_j(x) = e^{-\Lambda_0(x-x_0)}\xi Q_h(1-Q_h)^{-\xi-1},$$

$$\text{where } Q_h = \frac{\lambda}{\Lambda} (1 - e^{-\Lambda(x-x_0)}).$$

Double sums involving probabilities of being in the precancerous lines have to be calculated in a certain order. Because of the structure of the products in the expression for $P_{ji}(x)$, the sum over i has to be calculated first. This results in

$$\sum_{i=0}^{\infty} P_{ji}(x) = e^{-\Lambda_0(x-x_0)} \frac{\gamma_0 + j\gamma}{j!} \left[\prod_{m=1}^j (\xi + m - 1) \right] \int_{x_0}^x du Q_2^j (1 - Q_1)^{-\xi_0 - j\bar{\xi}},$$

$$\sum_{i=0}^{\infty} iP_{ji}(x) = e^{-\Lambda_0(x-x_0)} \frac{\gamma_0 + j\gamma}{j!} (\xi_0 + j\bar{\xi}) \left[\prod_{m=1}^j (\xi + m - 1) \right] \int_{x_0}^x du Q_2^j Q_1 (1 - Q_1)^{-\xi_0 - j\bar{\xi} - 1}.$$

Both these results are obtained by the direct calculation using eqs. (5,6). Alternatively, the second equality becomes obvious if one takes into account that it is a formal first derivative over Q_1 subsequently multiplied by Q_1 . Then denoting $\bar{Q}_2 = Q_2(1-Q_1)^{-\bar{\xi}}$ and making simple combinations in the above formulae, we perform summations over j :

$$S_p = \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} P_{ji}(x) = e^{-\Lambda_0(x-x_0)} \int_{x_0}^x du (1 - \bar{Q}_2)^{-\xi} (1 - Q_1)^{-\xi_0} \left[\gamma_0 + \gamma \xi \frac{\bar{Q}_2}{1 - \bar{Q}_2} \right],$$

$$S_{ip} \equiv \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} iP_{ji}(x) = e^{-\Lambda_0(x-x_0)} \xi_0 \int_{x_0}^x du Q_1 (1 - \bar{Q}_2)^{-\xi} (1 - Q_1)^{-\xi_0 - 1} \times$$

$$\times \left[\left(\gamma_0 + \gamma \xi \frac{\bar{Q}_2}{1 - \bar{Q}_2} \right) \left(1 + \frac{\bar{Q}_2}{1 - \bar{Q}_2} \right) + \gamma \frac{\bar{Q}_2}{(1 - \bar{Q}_2)^2} \right].$$

One can cross check that the abovementioned derivative rule still works, i.e.,

$$Q_1 \frac{\partial S_p}{\partial Q_1} = S_{ip}.$$

Triple sums S_m , S_{im} , S_{nm} , and S_{jm} have to be also calculated in a specific order, i.e., over i , over n , and then over j . We illustrate the calculation details for S_m and then discuss how the results for the remaining sums can be obtained. We calculated these equations analytically and verified the results with MAPLE 9.00 software.

The first summation gives

$$\sum_{i=0}^{\infty} P_{kjni}(x) = e^{-\Lambda_0(x-x_0)} \frac{\Delta_x(\bar{\gamma}_0 + n\bar{\gamma})(\gamma_0 + j\gamma)}{j!n!} \left[\prod_{m=1}^n (\xi_0 + j\bar{\xi} + m - 1) \right] \left[\prod_{m=1}^j (\xi + m - 1) \right] \times \\ \times (1 - Q_i)^{-\xi_{0k} - j\xi_k - n\bar{\xi}_k} \int_{x_0}^{x_k} du Q_j^j Q_n^n.$$

It is reasonable to introduce auxiliary functions by combining all bases under the power of n $\bar{Q}_j = Q_j(1 - Q_i)^{-\xi_k}$ and $\bar{Q}_n = Q_n(1 - Q_i)^{-\xi_k}$. Then, summation over n results in

$$\sum_{n=0}^{\infty} \sum_{i=0}^{\infty} P_{kjni}(x) = e^{-\Lambda_0(x-x_0)} \frac{\Delta_x(\gamma_0 + j\gamma)}{j!} \left[\prod_{m=1}^j (\xi + m - 1) \right] \times \\ \times (1 - Q_i)^{-\xi_{0k}} \int_{x_0}^{x_k} du \bar{Q}_j^j (1 - \bar{Q}_n)^{-\xi_0 - j\bar{\xi}} \left(\bar{\gamma}_0 + \bar{\gamma}(\xi_0 + j\bar{\xi}) \frac{\bar{Q}_n}{1 - \bar{Q}_n} \right).$$

Similarly, defining $\bar{\bar{Q}}_j = \bar{Q}_j(1 - \bar{Q}_n)^{-\bar{\xi}}$ and simplifying the final expression by defining additional structural variables we obtain

$$S_m = \sum_{j=0}^{\infty} \sum_{n=0}^{\infty} \sum_{i=0}^{\infty} P_{kjni}(x) = e^{-\Lambda_0(x-x_0)} \Delta_x(1 - Q_i)^{-\xi_{0k}} \int_{x_0}^{x_k} du (1 - \bar{Q}_n)^{-\xi_0} (1 - \bar{\bar{Q}}_j)^{-\bar{\xi}} (\bar{g}_0 g_0 + \bar{g} g),$$

where

$$g_0 = \gamma_0 + \xi g, \quad \bar{g}_0 = \bar{\gamma}_0 + \bar{g}, \quad g = \frac{\gamma \bar{\bar{Q}}_j}{1 - \bar{\bar{Q}}_j}, \quad \bar{g} = \frac{\xi_0 \bar{\gamma} \bar{Q}_n}{(1 - \bar{Q}_n)(1 - \bar{\bar{Q}}_j)}.$$

The sums S_{im} , S_{nm} , and S_{jm} can be obtained similarly. The calculation is straightforward though a little tedious. The practical implementation of this calculation requires generalizing formulae (5,6) to include higher degrees of summation variables i , n , and j . An alternative way is based on employing the derivative rule, i.e.,

$$S_{jm} = Q_j \frac{\partial S_m}{\partial Q_j}, \quad S_{im} = Q_i \frac{\partial S_m}{\partial Q_i}, \quad S_{nm} = Q_n \frac{\partial S_m}{\partial Q_n}.$$

We performed both ways of calculation and found the analytical coincidence of the results for these sums that provides a cross check of the obtained results. The explicit results for S_{im} , S_{nm} , and S_{jm} are

$$\begin{aligned}
S_{im} &= \sum_{j=0}^{\infty} \sum_{n=0}^{\infty} \sum_{i=1}^{\infty} i P_{kjni}(x) = e^{-\Lambda_0(x-x_0)} \Delta_x Q_i (1-Q_i)^{-\xi_{0k}-1} \int_{x_0}^{x_k} du (1-\bar{Q}_n)^{-\xi_0-1} (1-\bar{Q}_j)^{-\xi-1} \times \\
&\times \left(\left(\xi_{0k} \bar{g}_0 + \bar{g} \left(\xi_k \bar{Q}_j + \bar{\xi}_k (1-\bar{Q}_j) \right) \right) (g_0+g) + \xi_k (1+\xi) \bar{g} g \right), \\
S_{im} &= \sum_{j=0}^{\infty} \sum_{n=0}^{\infty} \sum_{i=1}^{\infty} i P_{kjni}(x) = e^{-\Lambda_0(x-x_0)} \Delta_x (1-Q_i)^{-\xi_{0k}} \int_{x_0}^{x_k} du \bar{Q}_n (1-\bar{Q}_n)^{-\xi_0-1} (1-\bar{Q}_j)^{-\xi-1} \times \\
&\times \left(\left(\xi_{0k} \bar{g}_0 + \bar{g} \left(\xi_k \bar{Q}_j + \bar{Q}_n^{-1} (1-\bar{Q}_j) \right) \right) (g_0+g) + \bar{\xi} (1+\xi) \bar{g} g \right), \\
S_{jm} &= \sum_{j=1}^{\infty} \sum_{n=0}^{\infty} \sum_{i=0}^{\infty} j P_{kjni}(x) = e^{-\Lambda_0(x-x_0)} \Delta_x (1-Q_i)^{-\xi_{0k}} \int_{x_0}^{x_k} du \bar{Q}_j (1-\bar{Q}_n)^{-\xi_0} (1-\bar{Q}_j)^{-\xi-1} \times \\
&\times \left(\xi \bar{g}_0 \left(g_0 + \bar{Q}_j^{-1} g \right) + \bar{g} (g_0 + (1+\xi)g) + \bar{g} g \bar{Q}_j^{-1} \right).
\end{aligned}$$

An important property of the model is that all functions calculated above depend on a finite number of parameters which can be estimated from the data on health histories of cancer.

Data Preparation and Parameter Estimation

Data used for the parameter estimation include the SEER and the HMD. The SEER [21] began in 1973 and covered approximately 14% of the U.S. population. Further expansion of this Register increased coverage to approximately 26%. The information collected about each case of cancer diagnosis includes the patient's demographic characteristics, date of diagnosis, data about up to 10 diagnosed cancer cases (e.g., histology, stage, and grade), type of surgical treatment and radiation therapy, recommended or provided within 4 months of diagnosis, follow-up of vital status, and the cause of death, if applicable. As a complement to the SEER Register data, there are files with population counts for one-year intervals for calendar year, age, sex and race groups. The HMD [22] includes calculated death rates and life tables by age, time, and sex, along with all of the raw data (vital statistics, census counts, population estimates) used in computing these quantities. Data are presented in a variety of formats with regard to age groups and time periods.

Data used for the parameter estimation include i) age-specific mortality rates extracted from the HMD, ii) age-specific incidence rates for each cancer extracted from the SEER, and iii) time-after-onset-specific mortality rates extracted from the SEER data as well. Methods of parameter estimation discussed below are specified for this dataset.

Statistical uncertainties have to be additionally balanced for the parameter estimation when data from different datasets are used. We recalculated statistical errors for the rates taking into account the requirement that person-years for the age-specific mortality and incidence rates are equal. The simplest estimations can be constructed using nonlinear least-squared (LS) methods. Let r_i be estimates of all respective rates in all considered age intervals with standard errors σ_i . Then, since these estimates can be considered as uncorrelated, LS-parameter estimations, $\{\theta\}$, are obtained by minimization of the following functional:

$$\sum \frac{(f_i(x, \{\theta\}) - r_i)^2}{\sigma_i^2} \rightarrow \min,$$

where the summation is performed over all rates and all considered age intervals.

Alternatively, the maximum likelihood estimates can be also constructed. The model developed above has a cohort structure. Therefore, data with balanced statistical uncertainties for rates that are obtained by combining two datasets, can be obtained by simulation of an artificial cohort which includes healthy individuals and a series of sub-cohorts for individuals with onsets between x and $x + \Delta_x$. The likelihood is constructed as

$$L = \prod_{n=0}^{n_{\max}} I(x_n)^{NI_n} \mu(x_n)^{NHD_n} (1 - I(x_n) - \mu(x_n))^{N_n - NI_n - NHD_n} \times \prod_{m=0}^{m_{\max}} \mu_c(y_m, x_n)^{NMD_{nm}} (1 - \mu_c(y_m, x_n))^{N_{mn} - NMD_{nm}},$$

where N_n is the number of healthy individuals at the beginning of n -th age interval, i.e., $(x_0 + n\Delta_x, x_0 + (n+1)\Delta_x)$; NI_n is the number of individuals with onset at n -th age interval; NHD_n is the number of individuals who died in this interval from the healthy stages; NMD_{nm} is the number of individuals with onset at n -th age interval and who died in m -th age interval; N_{mn} is the number of survivors at the beginning of m -th age interval among individuals with onset

at n -th age interval, i.e., $N_{mn} = NI_n - \sum_{k=0}^{m-1} NMD_{nk}$ for $m > 0$ and $N_{mn} = NI_n$ for $m = 0$; $n_{\max}$ is the total number of age intervals and $m_{\max}$ is the number of age intervals spent in malignant states. The latter is defined by corresponding n and a maximal cohort age, i.e., $m_{\max} = n_{\max} - n$. Notations x_n and y_m are used for mean values of corresponding time intervals. Note that data on mortality from cancer can also be used. Respective likelihood function has to be modified accordingly.

Cancer Models for Specific Sites: Discussion of the Results

To illustrate the capacity of the model in description of health history data, we evaluated the model's parameters for lung, colon, and skin cancers using data on the US male population in years 1973–2003, based on SEER and HMD data. These data include: i) the total mortality that is approximately equal to the mortality from non-malignant lines, ii) incidence rates of each type of cancer, and iii) the mortality rate from malignant lines, i.e., after onset of cancer (after cancer diagnosis). Five age patterns are fitted in each example: 55-year (from 30 to 85) age pattern of non-cancer mortality, 55-year (also from 30 to 85) age pattern of cancer incidence, and three 20-year patterns after onset in 65–69, 70–74, and 75–79. Totally, 17 parameters are fitted to describe these patterns simultaneously. Ten of them are designed to describe non-cancer mortality and incidence patterns. They include four parameters as in the standard LBM: μ_0 , μ , λ_0 , and λ . Additional six parameters include two rates of changes in rates in the precancerous lines: $\bar{\mu}$ and $\bar{\lambda}$, and four parameters responsible for unobservable and observable transferring between the malignant stages: γ_0 , γ , $\bar{\gamma}_0$, and $\bar{\gamma}$. One parameter x_0 corresponds to the beginning of time of the first latent event in the cohort. Only two additional parameters are used to describe mortality after onset at specific age: λ_k and μ_k . In this example, we do not distinguish between causes of death, therefore only a sum of μ_k and δ_k is identifiable (we use μ_k for their sum and interpret it as the parameter responsible for the total mortality). Also we set all δ_{0k} . This provides the continuity in mortality rates even in case of the transfer through an observed age at onset of cancer.

The parameter estimates are presented in Figure 4 and Table 1. Generally, the quality of fit is satisfactory and the obtained parameter estimates are interpretable. We would like to emphasize the two important features of estimated hazard rates. First, the model is capable of describing the effects of a leveling off and even a decline in the cancer incidence. Such age pattern is detected for many cancer sites [29]. Traditionally, such a shape of hazard rate at late ages was described using ideas of hidden heterogeneity (unobserved frailty) in susceptibility to disease

[30–34]. The model developed in this paper describes such an effect without additional assumptions about hidden random variables affecting the cancer risk. This is probably because the numerous states in the model explicitly represent such heterogeneity. Second, the model can describe the decline in mortality in cancer patients after onset of cancer. This remarkable property of the model is manifested in the framework of the multistage concept of carcinogenesis, which admits only its progressive development. This again is due to many functional and health states involved in description of cancer initiation and development. Individuals, which contract the disease from different precancerous states, carry their accumulated mortality rates. The deaths in an early time after onset occur for individuals contracting the disease from precancerous states with large numbers j and n and, therefore, having a larger current mortality rate. Capturing these two effects (i.e., the decline in the cancer incidence rate at advanced ages and the decline in the mortality rate during the earlier time since onset of cancer) by the model which assumes only progressive deterioration of the health state is its distinguishing and attractive feature. To our knowledge, there are no other models capable of analyzing the entire health history data and capturing age patterns of these two important hazard rates.

All parameter estimates are significantly different from zero with the exceptions of $\bar{\lambda}$ for the lung cancer model and γ_0 and/or γ for some other cancer models. Four parameters contributing to the rates at healthy lines are similar to those obtained in analyses of the standard LBM [20]. Since these parameters are used in the incidence rate model as well, their values do not coincide for different cancer models. Note that this is also the reflection of mutual correlation between parameters describing effects of carcinogenesis and aging, which may have an important biological meaning. The difference in the estimates of parameters λ in different lines allows for speculating about the rates of functional decline or the “aging rate” occurring under different conditions. For example, such a decline is going faster in precancerous (P) lines rather than in the healthy (H) line, though the magnitude of the effect (given by a difference between λ and $\bar{\lambda}$) is at the level of several percents. Although the difference between parameter estimates is significant for some models and sometimes has clear biological interpretation, one should be careful in making use of quantitative values of these estimates. One reason for this is the possibility of bias in these estimates. The sources of possible bias, as well as their effects on interpretation have to be carefully investigated before making biologically meaningful conclusions. Note that the main purpose of this paper is to show that the new model is capable of describing the entire disease-history data. The further developments of this model will allow for reducing the bias and making model more interpretable.

One of the fundamental questions in the studies of aging is whether the aging rate can be affected by health conditions. The difference in λ and μ estimated for different types of cancer may indicate that different cancers affect the aging rate differently. Comparison of λ_k for different ages at onset shows that there is a certain deceleration of aging with higher ages. Parameters describing the rate of mortality increase are also in a reasonable range. Parameters for the precancerous line look overestimated. This effect can be artificial and due to different shapes of mortality and incidence age patterns. To provide simultaneous fit for both patterns, the model has to respectively distribute individuals differently in healthy and precancerous states. In the current set-up in which only one precancerous stage is assumed, the model can do it by artificially increasing the parameters $\bar{\mu}$ (and, perhaps, $\bar{\gamma}_0$) and decreasing $\bar{\lambda}$. This is a clear indication to the fact that the real number of precancerous stages and, therefore, the number of unobserved transfers, can be larger than that in the considered version of the model. An addition of the second level of latent precancerous stages would allow for making this redistribution. Note that inclusion of additional precancerous stages is straightforward. It will add two or three new parameters and will require extensive analytical and/or numerical calculations. Note that such addition will allow for the use of SEER data on disease stages,

which are currently ignored in most models of cancer. This will improve identifiability of the extended model and make it more realistic.

The quality of fit is good but not ideal. The latter is because of many sources of hidden heterogeneity in data capable of producing bias in parameter estimates. Among them are i) different stages of diagnosed cancer with different shapes of incidence rates and forthcoming survival, ii) different histological forms of cancer, iii) different race effects and effects of genetic predisposition, iv) different contributions of environmental exposure, v) different causes of death of individuals after onset of cancer, and vi) different cohort effects due to time trends coming from the progress in medical technologies, screening investigations and variety of clinical interventions. In future developments, several of these effects can be taken into account by considering more homogeneous age patterns (e.g., race-, stage-specific rates), respective auxiliary models (e.g., effects of cohorts, progress in medical technologies), and using additional sources of data (e.g., information on cause of death, or on living area and information on maps with environmental exposure).

Discussion and Conclusion

The traditional LBM of mortality and aging exploits the Markov property of the process describing transitions between functional states. The respective probabilities of being in selected functional states satisfy an infinite system of linear differential equations, which admit an analytical solution. Introduction of the additional lines of states into the model violates the Markov assumption: the transition rates from the unhealthy lines may depend on time spent in these states. Fortunately, the model remains mathematically tractable: the functional changes (horizontal transitions in Fig. 2) in the sub-cohorts of individuals who contracted a disease at age $x, x+1, x+2, \dots$ are described similarly to those in the original LBM. These sub-cohorts differ in the initial conditions: in contrast to the traditional LBM in which all individuals begin their life in the box H_0 , the individuals in the sub-cohorts contracted diseases at different ages are distributed with respect to functional states.

The model parameters can be estimated using available data. For the U.S., the SEER data and demographic mortality data can be combined. The SEER data set contains important information about age at onset of different types of cancer, as well as data about the age at death for individuals with cancer. Since the SEER data represent different cohorts of individuals, one can evaluate time trends in the respective hazard rates and associate them with time trends in the model parameters.

An important advantage of the new model is that it captures all transitions associated with dynamics of selected disease. This allows for evaluating the age pattern of prevalence of the disease (the “burden” of disease). This characteristic is important for studying disease-associated medical costs, as well as effects of population health on economics and society as a whole because staying in the unhealthy state reduces individuals’ productivity and diminishes his/her quality of life. The model can be extended to allow for evaluating consequences of alternative preventive and treatment strategies: the first will affect parameters of incidence rate (γ ’s), the second will deal with parameters of case-fatality rates such as μ_{65-69} , or δ , if data on cause-specific mortality are used.

The model allows for studying a possible trade-off between two lines of organisms’ defense against a disease. The first is responsible for its vulnerability (robustness) and affects the age of disease onset. The second deals with organism’s resilience, i.e., its ability of coping with disease when it strikes. It affects chances of recovery, as well as duration of staying in an unhealthy state until death. If the resource specified for the defense against the disease is common for both strategies then the trade-off having genetic, epigenetic, or other origin is

possible. The presence of such a trade-off adds an important component of population heterogeneity into the model: some individual organisms sacrifice robustness for the sake of resilience and vice-versa. The model allows for testing hypotheses about such heterogeneity, as well as about the possibility of trade-off between these two defense strategies.

Note that although a decline in functioning is described similarly in sick and healthy individuals, the parameters of respective transitions may differ. For example, the presence of disease can speed up the horizontal transitions (increase the rate of functional decline or the rate of aging). This hypothesis can be tested from the data using the likelihood ratio test. Different types of cancer (or other diseases) may affect these parameters differently. This may have an important biological meaning: the strength of connection between cancer and aging may depend on the type of cancer.

Susceptibility to cancer may correlate (positively or negatively) with susceptibility to other diseases [35]. The model can easily be extended to include incidence of two or more other chronic conditions. The model can also describe the accumulation of deficits during the aging process. The data on such accumulation are recently investigated in a number of studies [36–40].

The existence of genetic connection between cancer and aging, revealed in a number of recent experimental studies using laboratory animals (see [41,42] and references in them), can be tested using human population data. Note that although studies on cancer using laboratory rodents deal with a limited number of cancer sites [43], the molecular-biological data suggest that the connection between cancer and aging may involve more cancers. Moreover, our analyses showed that cancer may interact with other diseases [35]. These connections may be described using versions of the model investigated above.

Acknowledgments

This work was supported by NIA/NIH grants 1R01AG032319, 5R01-AG-028259, 5R01-AG-027019 and PO1-AG-008761. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Aging or the National Institutes of Health.

References

1. Tan, WY. Stochastic Models of Carcinogenesis. Marcel Dekker, Inc; NY: 1991.
2. Moolgavkar, S.; Krewski, D.; Schwarz, M. Mechanisms of carcinogenesis and biologically based models for estimation and prediction of risk. In: Moolgavkar, S.; Krewski, D.; Zeise, L.; Cardis, E.; Møller, H., editors. Quantitative Estimation and Prediction of Human Cancer Risks, Scientific publications No. 131. International Agency for Research on Cancer; Lyon: 1999. p. 179-237.
3. van Leeuwen IM, Zonneveld C. From exposure to effect: a comparison of modeling approaches to chemical carcinogenesis. *Mutat Res* 2001;489(1):17–45. [PubMed: 11673088]Review. Erratum in: *Mutat Res*, 511(1): 87, 2002.
4. Tan, W. Stochastic Models with Applications to Genetics, Cancers, AIDS, and other Biomedical Systems. World Scientific; London: 2002.
5. Heidenreich WF, Luebeck EG, Hazelton WD, Paretzke HG, Moolgavkar SH. Multistage models and the incidence of cancer in the cohort of atomic bomb survivors. *Radiat Res* 2002;158(5):607–14. [PubMed: 12385638]
6. Arbeev KG, Ukraintseva SV, Arbeeva LS, Yashin AI. Mathematical Models for Human Cancer Incidence Rates. *Demographic Research* 2005;12(10):237–271.
7. Akushevich, I.; Veremeeva, G.; Kulminski, A.; Ukraintseva, S.; Arbeev, K.; Akleev, AV.; Yashin, AI. New Perspectives in Modeling of Carcinogenesis Induced by Ionizing Radiation. The 13th International Congress of Radiation Research; San Francisco, California. July 08–12; 2007. p. 246Abstract PS4175. In abstract book

8. Little MP. Generalisations of the two-mutation and classical multi-stage models of carcinogenesis fitted to the Japanese atomic bomb survivor data. *J Radiol Prot* 1996;16(1):7–24.
9. Little MP, Haylock RG, Muirhead CR. Modelling lung tumour risk in radon-exposed uranium miners using generalizations of the two-mutation model of Moolgavkar, Venzon and Knudson. *Int J Radiat Biol* 2002;78(1):49–68. [PubMed: 11747553]
10. Brugmans MJ, Rispens SM, Bijwaard H, Laurier D, Rogel A, Tomasek L, Tirmarche M. Radon-induced lung cancer in French and Czech miner cohorts described with a two-mutation cancer model. *Radiat Environ Biophys* 2004;43(3):153–163. [PubMed: 15316819]
11. Heidenreich WF, Tomasek L, Rogel A, Laurier D, Tirmarche M. Studies of radon-exposed miner cohorts using a biologically based model: comparison of current Czech and French data with historic data from China and Colorado. *Radiat Environ Biophys* 2004;43(4):247–256. [PubMed: 15645313]
12. Luebeck EG, Moolgavkar SH. Multistage carcinogenesis and the incidence of colorectal cancer. *Proc Natl Acad Sci U S A* 2002;99(23):15095–15100. [PubMed: 12415112]
13. Gregori G, Hanin L, Luebeck G, Moolgavkar S, Yakovlev A. Testing goodness of fit for stochastic models of carcinogenesis. *Math Biosci* 2002;175(1):13–29. [PubMed: 11779625]
14. Campisi J. Cancer and ageing: rival demons? *Nat Rev Cancer* 2003;3(5):339–349. [PubMed: 12724732]
15. Ukraintseva, SV.; Yashin, AI. Individual aging and cancer risk: How are they related?; Demographic Research. 2003. p. 163-196.<http://www.demographic-research.org/>
16. Ukraintseva SV, Yashin AI. Opposite Phenotypes of Cancer and Aging Arise from Alternative Regulation of Common Signaling Pathways. *Ann N Y Acad Sci* 2003;1010:489–492. [PubMed: 15033776]
17. Ukraintseva SV, Yashin AI. Treating cancer with embryonic stem cells: Rationale comes from aging studies. *Frontiers in Bioscience* 2005;10:588–595. [PubMed: 15569600]
18. Campisi J. Hot Topics Aging and Cancer Cell Biology, 2008. *Aging Cell*. 2008 Mar 10;Epub ahead of print
19. Le Bras H. Lois de Mortalite at Age Limite. *Population* 1976;31:655–692.
20. Yashin AI, Vaupel JW, Iachine IA. A duality in aging: the equivalence of mortality models based on radically different concepts. *Mech Ageing Dev* 1994;74(1–2):1–14. [PubMed: 7934200]
21. Surveillance, Epidemiology, and End Results (SEER). 2007. Program (www.seer.cancer.gov) Limited-Use Data (1973–2004), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2007, based on the November 2006 submission.
22. Human Mortality Database. Max Planck Institute for Demographic Research (Germany). University of California; Berkeley (USA): Available at www.mortality.org or www.humanmortality.de (data downloaded on January 2008)
23. Gavrilov, LA.; Gavrilova, NS. The Biology of life Span: A Quantitative Approach. New York: Harwood Academic Publisher; 1991.
24. Armitage P, Doll R. A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. *Br J Cancer* 1957;11(2):161–169. [PubMed: 13460138]
25. Knudson AG. Mutation and cancer: Statistical study of retinoblastoma. *Proc Natl Acad Sci USA* 1971;68:820–823. [PubMed: 5279523]
26. Moolgavkar SH, Venzon DJ. Two-event models for carcinogenesis: Incidence curves for childhood and adult tumors. *Math Biosci* 1979;47:55–77.
27. Moolgavkar SH, Knudson AG. Mutation and cancer: A model for human carcinogenesis. *J Natl Cancer Inst* 1981;66:1037–52. [PubMed: 6941039]
28. UNSCEAR 2000 published. United Nations Scientific Committee on the Effects of Atomic Radiation. *Health Phys* 2001;80(3):291.
29. Ries, LAG.; Devesa, SS. Cancer incidence, mortality, and patient survival in the United States. In: Schottenfeld, D.; Fraumeni, JF., Jr, editors. *Cancer epidemiology and prevention*. Oxford University Press; New York: 2006. p. 139-173.
30. Trussell J, Richards T. Correcting for Unmeasured Heterogeneity in Hazard Models Using the Heckman - Singer Procedure. *Sociological Methodology* 1985;15:242–276.
31. Aalen OO. Phase type distributions in survival analysis. *Scand J Stat* 1995;22:447–463.

32. Aalen OO, Gjessing HK. Understanding the shape of the hazard rate: a process point of view. *Stat Sci* 2001;16(1):1–22.
33. Vaupel, JW.; Yashin, AI. MPIDR working paper. Rostock: 1999. Cancer rates over age, time, and place: insights from stochastic models of heterogeneous populations; p. 35(WP-1999–006). Internet: <<http://www.demogr.mpg.de/Papers/Working/wp-1999-006.pdf>>
34. Manton, KG.; Akushevich, I.; Kravchenko, JS. *Statistics for Biology and Health*. Springer; 2008. Cancer Mortality and Morbidity Patterns in the U.S. Population: An Interdisciplinary Approach.
35. Yashin AI, Ukraintseva SV, Akushevich I, Arbeev KG, Kulminski A, Akushevich L. Trade-off between Cancer and Aging: What Role Do Other Diseases Play? Evidence from Experimental and Human Population Studies. *Mech Ageing Dev*. 2008forthcoming.
36. Rockwood K, Mitnitski A. Frailty in relation to the accumulation of deficits. *J Gerontol A Biol Sci Med Sci* 2007;62(7):722–727. [PubMed: 17634318]
37. Yashin AI, Akushevich IV, Arbeev KG, Akushevich L, Ukraintseva SV, Kulminski A. Insights on aging and exceptional longevity from longitudinal data: novel findings from the Framingham Heart Study. *Age Dordr Neth* 2006;28(4):363–374.
38. Kulminski AM, Ukraintseva SV, Akushevich IV, Arbeev KG, Yashin AI. Cumulative index of health deficiencies as a characteristic of long life. *J Am Geriatr Soc* 2007;55(6):935–940. [PubMed: 17537097]
39. Kulminski A, Ukraintseva SV, Akushevich I, Arbeev KG, Land K, Yashin AI. Accelerated accumulation of health deficits as a characteristic of aging. *Exp Gerontol* 2007;42(10):963–970. [PubMed: 17601693]
40. Yashin AI, Arbeev KG, Kulminski A, Akushevich I, Akushevich L, Ukraintseva SV. What age trajectories of cumulative deficits and medical costs tell us about individual aging and mortality risk: Findings from the NLTCs-Medicare data. *Mech Ageing Dev* 2008;129(4):191–200. [PubMed: 18242665]
41. Campisi J. Cancer and aging: yin, yang, and p53. *Sci Aging Knowledge Environ* 2002;1:pe1. [PubMed: 14602974]
42. Vijg J, Campisi J. Puzzles, promises and a cure for ageing. *Nature* 2008;454:1065–71. [PubMed: 18756247]
43. Anisimov VN, Ukraintseva SV, Yashin AI. Cancer in experimental animals: Does it tell us about cancer in humans? *Nature Reviews Cancer* 5(10):807–19.

Appendix

Consider a system of differential equations:

$$\begin{aligned}\frac{d}{dx}P_0(x) &= -\Lambda_0 P_0(x) + f(x), \\ \frac{d}{dx}P_i(x) &= -(\Lambda_0 + i\Lambda)P_i(x) + L_i P_{i-1}(x), \quad i \geq 1,\end{aligned}$$

where $L_i = \lambda_0 + (i-1)\lambda$. Initial conditions for all states including 0-th are $P_i(0)=0$. To calculate $P_i(x)$, we note first that the solution of the heterogeneous differential equation

$$\frac{d}{dx}P_i(x) = -(\Lambda_0 + i\Lambda)P_i(x) + f_i(x)$$

with $P_i(0)=0$ can be represented in the form:

$$P_i(x) = \int_0^x f_i(u) e^{-(\Lambda_0 + i\Lambda)(x-u)} du.$$

Then, since $P_i(x)$ contributes to the heterogeneous part of differential equation for $P_{i+1}(x)$, a function $f_i(x)$ will contribute to $P_{i+1}(x)$ as a two-dimensional integral and so on. Thus, $P_i(x)$ can be represented in the form of $i + 1$ -dimensional integral as

$$P_i(x) = L_i \dots L_1 \int_0^x du_i \int_0^{u_i} du_{i-1} \dots \int_0^{u_1} du f(u) e^{-\Lambda_0(u_1-u)} \dots e^{-(\Lambda_0+(i-1)\Lambda)(u_i-u_{i-1})} e^{-(\Lambda_0+i\Lambda)(x-u_i)}.$$

The order of integration can be changed and exponents can be simplified:

$$P_i(x) = L_i \dots L_1 \int_0^x du f(u) e^{-\Lambda_0(x-u)} e^{-ix\Lambda} \bar{I}_i(u),$$

where

$$\bar{I}_i(u) = \int_u^x du_i \dots \int_u^{u_2} du_1 e^{-\Lambda(u_1+u_2+\dots+u_i)}.$$

Integrals over all variables except u can be calculated analytically:

$$\bar{I}_i(u) = \frac{1}{\Lambda^i i!} (e^{x\Lambda} - e^{u\Lambda})^i.$$

The result for $P_i(x)$ reads

$$P_i(x) = \frac{L_i \dots L_1}{\Lambda^i i!} \int_0^x du f(u) e^{-\Lambda_0(x-u)} (1 - e^{\Lambda(u-x)})^i.$$

Figure 1.
The standard Le Bras population model.

Figure 2.

The generalized Le Bras population multistate cancer model. There are three types of boxes: i) the lower line of boxes H_i , $i = 0, 1, 2, \dots$ denotes healthy states, ii) the boxes marked by P_{jn} , $j, n = 0, 1, 2, \dots$ are interpreted as precancerous states, and iii) the boxes describing states with diagnosed cancer are denoted by M with four indices (more detailed representation of a specific line of the M -boxes is in Figure 3). Arrows show possible transfers occurring with respective rates. Diamonds denote death states. Vertical dots between M -lines indicate subsequent cohorts of individuals with diagnosed cancer. Each cohort is formed in the lowest M -line among the column by the new cancer cases diagnosed at the time period $(x_0 + k\Delta_x, x_0 + (k+1)\Delta_x)$. At the end of each time period, a new M -line with the successive k is created for each column, and all other M -lines are shifted up. See also the text for detailed discussion of the design.

Figure 3.

The generalized Le Bras population cancer model: Detailed representation of the dynamics in a malignant states for a cohort formed at the time period $(x_0 + k\Delta_x, x_0 + (k+1)\Delta_x)$. Up and down transitions denote cancer-specific and non-cancer deaths.

Figure 4.

Rates for total mortality, incidence, and mortality after onset of cancer for three cancer site models: data (points) and model (line).

Table 1

Parameter estimates

Parameter	Units	Estimates for a specific cancer site		
		Lung	Colon	Skin
x_0	year	28.3	11.3	22.9
λ_0	year ⁻¹	0.318	0.485	9.31
λ	year ⁻¹	0.099	0.0948	0.0798
$\bar{\lambda}$	year ⁻¹	$3.83 \cdot 10^{-9}$	0.103	0.0799
μ_0	$10^{-2} \cdot \text{year}^{-1}$	0.157	0.0916	0.143
μ	$10^{-4} \cdot \text{year}^{-1}$	2.03	0.289	0.0297
$\bar{\mu}$	$10^{-4} \cdot \text{year}^{-1}$	817	4.95	0.384
γ_0	year ⁻¹	$1.32 \cdot 10^{-9}$	$3.54 \cdot 10^{-9}$	0.00382
$\bar{\gamma}_0$	year ⁻¹	231	19.7	0.00217
γ	$10^{-7} \cdot \text{year}^{-1}$	0.0717	0.0228	34.0
$\bar{\gamma}$	year ⁻¹	0.0410	0.000393	0
μ_{65-69}	$10^{-6} \cdot \text{year}^{-1}$	133	0.120	0.0705
λ_{65-69}	year ⁻¹	0.195	0.421	0.375
μ_{70-74}	$10^{-6} \cdot \text{year}^{-1}$	81.6	0.475	0.431
λ_{70-74}	year ⁻¹	0.240	0.363	0.286
μ_{75-79}	$10^{-6} \cdot \text{year}^{-1}$	28.1	0.707	1.04
λ_{75-79}	year ⁻¹	0.332	0.368	0.240