Joint Analyses of Longitudinal Measurements of Physiological Indices and Cancer Incidence Rates

Konstantin G. Arbeev[*1], Dora Il'yasova[2], Svetlana V. Ukraintseva[1,2], Igor Akushevich[1], Alexander M. Kulminski[1], Liubov S. Arbeeva[1], Irina V. Culminskaya[1], Deqing Wu[1], Anatoliy I. Yashin[1,2]

[1]Center for Population Health and Aging, Duke University, Durham, NC, 27708-0408, USA

[2]Duke Cancer Institute, Duke University, Durham, NC, 27710, USA

[*] **Corresponding author:** Dr. Konstantin G. Arbeev, Center for Population Health and Aging, Duke University, Trent Hall, Room 002, Box 90408, Durham, NC, 27708-0408, USA. Tel.: +1-919-668-2707; fax: +1-919-684-3861; e-mail: konstantin.arbeev@duke.edu

Abstract

Background: Different mechanisms regulating age-dynamics of physiological variables may affect the risk of onset of cancer. The impact of such mechanisms can be evaluated indirectly from longitudinal measurements of physiological variables.

Methods: We applied the proportional hazards model and the stochastic process model of aging to data on ages at onset of cancer (all sites but skin) and longitudinal measurements of hematocrit in the Framingham Heart Study (original cohort).

Results: Analyses using the proportional hazards model showed a marginally significant effect of hematocrit on the risk of onset of cancer (all sites but skin) in the Framingham original cohort (estimate of regression parameter: -0.021 ($p=0.04$); hazard ratio (95% confidence interval) for a change of 10% in hematocrit: 0.808 (0.659; 0.991)) suggesting a generally negative effect of lower hematocrit values on the risk of cancer development. Analyses using the stochastic process model revealed non-symmetric and age-dependent U-shapes of incidence rates as a function of hematocrit in both sexes. We found statistically significant ($p<0.0001$) age related decline in adaptive capacity associated with hematocrit regulation indicating that at older ages more time is needed for adjusting this value.

Conclusions: The value and longitudinal dynamics of hematocrit are associated with the risk of onset of cancer. The aging related decline in adaptive capacity and resistance to stresses, as well as accumulation of allostatic load are the factors which contribute to longitudinal dynamics of this variable.

Key words: cancer, hematocrit, Framingham Heart Study, stress resistance, adaptive capacity, physiological norm, allostatic load, age trajectory

1. Introduction

Longitudinal data on aging, health, and longevity provide information on individual age trajectories of different physiological variables. Such trajectories can be analyzed in connection with mortality/morbidity risks using appropriate statistical tools. Associations between selected characteristics and the risks can be evaluated using standard statistical approaches (such as the Cox proportional hazards model), which can be useful at the initial step of analyses when the presence of the effects needs to be identified. However, such methods generally may not be appropriate for studying biological mechanisms relating the observed age trajectories of physiological variables and risks of death or diseases, because these methods ignore existing knowledge about regularities of aging-related changes in an organism and underlying biological mechanisms and concepts of aging available in the literature.

The stochastic process model (SPM) of aging [1] serves as a useful tool for analyzing respective mechanisms and their relation to the mortality/morbidity risks from longitudinal data on age trajectories of physiological variables and data on ages at onset of disease or ages at death. This approach has been applied in different settings to analyzing data on longitudinal measurements of different physiological variables (such as blood glucose, body mass index, cholesterol, diastolic blood pressure, hematocrit, pulse pressure, and pulse rate) in relation to risks of death or onset of "unhealthy life" [2-5]. It was also applied to the "indices of cumulative deficits" (which proved to be a useful method for analyses of a wide spectrum of information in relation to health- and aging-related changes and represents a better characteristic of the aging phenotype than chronological age) and mortality risks [6, 7], and analyses of trajectories of medical costs in relation on mortality risks [8].

In this paper we use this approach to investigate mechanisms linking age trajectories of hematocrit and risks of onset of cancer. For this purpose, we apply the model by Yashin et al. [1] to analyze relationship between the risk of onset of cancer (all sites but skin) and longitudinal measurements of hematocrit in participants of the Framingham Heart Study (original cohort). We evaluate different components of the aging process, such as the decline in resistance to stresses and adaptive capacity, and accumulation of allostatic load, from age trajectories of hematocrit and data on onset of cancer, and show how these processes can influence the risk of onset of cancer.

2. Data and Method

2.1. Framingham Heart Study (FHS) data

The FHS Original Cohort consists of 5,209 respondents (nearly all are Caucasians, 46% male) aged 28-62 years at baseline and residing in Framingham, Massachusetts, between 1948 and 1951, and who had not yet developed overt symptoms of cardiovascular disease or suffered a heart attack or stroke [9, 10]. The study continues to the present with biennial examinations (30 exams to date; data from exams 1-26 were available for this study) that include detailed medical history, physical exams, and laboratory tests. Examination of participants, including an interview, physical examination, and laboratory tests, has been taken biennially. The FHS Original Cohort has been followed for more than 60 years (information on about 55 years of follow-up was available for this study) for the occurrence of diseases (including cancer) and death through surveillance of hospital admissions, death registries, and other available sources. The FHS cancer dataset contains only primary cancers (ICD-O behavior codes of 2 or 3) and does not include metastatic cancers (ICD-O behavior code of 6). Only first occurrence of site and cell type is included in this data set. In this study we used data on first occurrence of cancer (all

sites but skin; i.e., all WHO ICD-O topography codes except 173) from the follow-up data to calculate the age at onset of cancer for FHS Original Cohort participants. Longitudinal measurements of hematocrit from exams 4-9 and 11-20 were used in this study.

The dataset available for this study contained information on 5,079 participants of the Original FHS cohort (2,785 females; 2,294 males). We excluded from analyses individuals with onset of cancer (all sites but skin) before the entry into the FHS and individuals for whom measurements of hematocrit were not available in any exam. The resulting sample of 4,611 individuals (2,546 females; 2,065 males) was used in analyses of the stochastic process model described below. There were 784 participants (384 females; 400 males) who had onset of cancer occurring within two years since the last observation of hematocrit, or at some point between two observations of hematocrit, and respective ages at onset were calculated for these individuals. Individuals without the occurrence of cancer within two years (which is the average period between the exams in the original FHS cohort) since the last observation of hematocrit were censored at respective ages, or at the latest ages for which information on cancer was available, whichever were the earliest. Any measurements of hematocrit after the ages at onset of cancer or ages at censoring were not analyzed in the model.

2.2. Statistical Analysis: The Model Describing Age Dynamics of Physiological Variables and Risks of Cancer and Its Application to FHS Data

We started with analyses of cancer incidence using the proportional hazards model with measurements of hematocrit considered as a time-dependent covariate. The models were adjusted for sex and age at the first exam. We also performed analyses stratified by sex. These calculations have been performed using PROC PHREG in SAS/STAT 9.2.

Such analyses, however, do not take into account that different values of hematocrit may

minimize the risk of cancer at different ages. More subtle analyses of relationship between the risk of cancer and longitudinal dynamics of hematocrit used the stochastic process model of aging which incorporates several major concepts of aging [1, 4]. A discrete-time version of the model (with values of a physiological variable evaluated at one-year age intervals using linear approximation of respective observations in the adjacent FHS exams) was applied to data on onset of cancer (all sites but skin) in females and males in the FHS original cohort. The details of the likelihood estimation procedure are given in [1]. Below we provide specifications of the versions of the models used in this study.

The age dynamics of a physiological variable (i.e., hematocrit in our applications) with age is described by the following stochastic differential equation [1, 4]:

$$dY_t = a(t)(Y_t - f_1(t))\,dt + b(t)\,dW_t, \quad Y_0\,. \tag{1}$$

Here Y_t is the value of a physiological variable at age t. The function $f_1(t)$ describes the effect of allostatic adaptation, i.e., the trajectory that the physiological variable is forced to follow by homeostatic forces in the presence of external disturbances described by a Wiener process W_t (which is independent of the initial normally distributed value Y_0). The strength of homeostatic forces is characterized by the negative feedback coefficient $a(t)$: larger values of this function correspond to faster return of the trajectory of the physiological variable to the allostatically prescribed values $f_1(t)$. Therefore, the decline in the absolute value of this function with age represents the decline in adaptive (homeostatic) capacity with age ("homeostenosis") which has been shown to be an important characteristic of aging [11-14].

The diffusion coefficient $b(t)$ in (1) was modeled constant ($b(t) = \sigma_1$). Initial values of a physiological variable were assumed normal, $Y_{t_0} \sim N(f_1(t_0), \sigma_0)$, where the parameter σ_0 is estimated from the data. We used a linear approximation of the decline in adaptive capacity with

age, i.e., the feedback coefficient $a(t)$: $a(t) = a_Y + b_Y t$ (with $a_Y < 0$ and $b_Y \geq 0$). Description of $f_1(t)$ is provided below.

The concept of allostatic adaptation implicitly assumes the notion of "deviation from the norm," that is, the "normal" state of an organism corresponding to "optimal" functioning in terms of minimizing respective risk (e.g., onset of a disease). The studies of how persistent external unfavorable conditions get "under the skin" of affected person increasing his/her susceptibility to diseases and death [15, 16] provide evidence that many such conditions affect set-points of physiological homeostasis changing physiological balance from the "normal" to "abnormal" state. This means that the trajectory of a physiological variable that an organism is forced to follow under the persistent external disturbances ($f_1(t)$) may be different from the "optimal" trajectory minimizing the risk (i.e., the trajectory which the homeostatic regulation would force to follow in the absence of external disturbances), which we denote $f(t)$. The difference between $f_1(t)$ and such "optimal" trajectory (which can be interpreted as age-specific "physiological norm") provides the measure of the *allostatic load*.

Specification of the functional form of the cancer incidence rate as a function of age and the value of a physiological variable (hematocrit in our case), which we will denote $\mu(t, Y_t)$, deserves a special attention. It is clear by definition of the "optimal" age trajectory $f(t)$ that deviations of a physiological variable from respective "optimal" levels for each age (represented by this function $f(t)$) increase individual's chances to develop the disease. Different studies observed U- or J- shape of the risks as functions of various physiological variables [17-27]. Thus, it may be argued based on these observations, that a quadratic function can model dependence of the risk on deviations of trajectories of physiological variable Y_t from the optimal trajectory $f(t)$ [2-4, 6, 8, 28]. Nevertheless, the actual functional form of $\mu(t, Y_t)$ is unknown.

Thus using different formulas for description of $\mu(t,Y_t)$ can be helpful to check whether different specifications of the functional form of $\mu(t,Y_t)$ lead to the same conclusions [6, 7] or to select the best-fitting model.

We used two specifications for the hazard rate (i.e., cancer incidence rate in our applications). First, we used the quadratic hazard as in [1, 4]:

$$\mu(t,Y_t) = \mu_0(t) + (Y_t - f(t))^2 \mu_1(t). \quad (2)$$

Here $\mu_0(t)$ is the "residual" or "baseline" hazard that represents the incidence rate which would be observed if the physiological variable Y_t followed the "optimal trajectory" represented by the function $f(t)$. It models the effect of other factors (such as the senescence process) that impact the incidence rate. The non-negative multiplier $\mu_1(t)$ in the quadratic part of the hazard characterizes sensitivity of the risk function (incidence rate) to deviations of a physiological variable from the "optimal" function $f(t)$. This multiplier can be interpreted in terms of the "robustness," or "vulnerability," component of stress resistance. When the value of this function increases (i.e., the U-shape of the risk narrows), an organism becomes more vulnerable to deviations from the "normal" state caused by external disturbances (because the same magnitude of deviation from the "optimal" trajectory results in a larger increase in the risk). An increase in this index with age (that is, the decline in stress resistance) can be considered as a manifestation of the senescence process [6, 29-34]. We considered the non-symmetric U-shape of the hazard, i.e., we assumed that the "price" for deviations of the age trajectory of physiological variable to the left (i.e., to smaller values) and to the right (i.e., to larger values) from the "optimal" trajectory can be different. We specified $\mu_1(t)$ as a linear function of age: $\mu_1(t) = \mu_{11}(t)$, if $Y_t \le f(t)$, where $\mu_{11}(t) = a_{\mu_{11}} + b_{\mu_{11}} t$, and $\mu_1(t) = \mu_{12}(t)$, if $Y_t > f(t)$, where

$\mu_{12}(t) = a_{\mu_{12}} + b_{\mu_{12}} t$.

We used two different specifications of the baseline hazard $\mu_0(t)$: the logistic (gamma-Gompertz) function $\mu_0(t) = \mu_0^0(t) \Big/ (1 + \sigma_2^2 \int_0^t \mu_0^0(u)du)$, where $\mu_0^0(t) = a_{\mu_0} e^{b_{\mu_0} t}$ (the version of the model with this baseline hazard is denoted as Model 1 in the text), and the gamma-Weibull function: $\mu_0(t) = \mu_0^0(t) \Big/ (1 + \sigma_2^2 \int_0^t \mu_0^0(u)du)$, where $\mu_0^0(t) = \frac{b_{\mu_0}}{(a_{\mu_0})^{b_{\mu_0}}} t^{b_{\mu_0}-1}$ (Model 2). The specific choices for the baseline hazard are motivated by observations that cancer incidence rates decelerate or even decline at advanced ages [32, 35-39].

Second, we used the stochastic process model with the exponential form of $\mu(t, Y_t)$:

$$\mu(t, Y_t) = \mu_0(t) \exp\{\mu_{11}(t)(f(t) - Y_t) I(Y_t \le f(t)) + \mu_{12}(t)(Y_t - f(t)) I(Y_t > f(t))\}, \quad (3)$$

where *I*(.) is an indicator function, which equals 1 if the inequality in the parentheses is true, and 0 otherwise, and the functions $\mu_{11}(t)$ and $\mu_{12}(t)$ are specified above. Two different baseline hazards were used: gamma-Gompertz (Model 3) and gamma-Weibull (Model 4) hazards (see formulas above).

To represent the "optimal" trajectory *f*(*t*) in the models, we calculated the average age trajectory (in 5-year age groups, from ages 40-44 to 90+) of hematocrit in individuals who survived until advanced ages (≥ 90 for females, ≥ 85 for males) without developing cancer. These empirical trajectories were then fitted by cubic polynomials and these fitted trajectories were used as the "optimal" trajectories *f*(*t*) in the models (see Fig. 1).

Fig. 1 is about here

We assumed that the "allostatic trajectory" $f_1(t)$ (which we denote the "mean allostatic state") is related to the "optimal" trajectory *f*(*t*) as follows: $f_1(t) = f(t) + \Delta_{f_1}$. The absolute value

of parameter Δ_{f_1} represents the measure of "allostatic load" (averaged across all ages).

In each model, we tested statistical hypotheses about factors and mechanisms affecting the dynamic properties of the age trajectories of hematocrit and their relation to risk of cancer. All such hypotheses were tested using the likelihood ratio test. For example, to test the hypothesis about the decline in adaptive capacity with age in Model 1, we estimated the likelihood functions in the "general" model with such a decline (that is, with a linear $a(t)$ as specified above) and in the "restricted" model without such a decline (i.e., with $b_Y = 0$), where all other functions (except $a(t)$) are specified similarly in both models, and then applied the likelihood ratio test. As Models 1-4 are not nested, we compared different models to define the best-fitting model using the Akaike Information Criterion (AIC) [40]. All statistical analyses of the stochastic process model (the likelihood optimization and the statistical tests) have been performed using Optimization and Statistical Toolboxes in MATLAB R2010a.

3. Results

Analyses using the proportional hazards model showed a marginally significant effect of hematocrit on the risk of onset of cancer (all sites but skin) in the Framingham original cohort (estimate of regression parameter: -0.021; p=0.04; hazard ratio (95% confidence interval) for a unit change in hematocrit: 0.979 (0.959; 0.999); hazard ratio for a change of 10% in hematocrit: 0.808 (0.659; 0.991)). Analyses stratified by sex did not show a significant effect, although the estimates of the regression parameter were still negative and the absolute value was smaller for females. This suggests a generally negative effect of lower hematocrit values on the risk of cancer development. Such analyses, however, do not take into account that different values of hematocrit may minimize the risk at different ages. More subtle analyses of relationship between

the risk of cancer and longitudinal dynamics of hematocrit that use the stochastic process model [1] are presented in Table 1 and Figs. 2-3.

Estimates of parameters of the baseline incidence rate, $\mu_0(t)$, the multipliers $\mu_{11}(t)$ and $\mu_{12}(t)$ in the quadratic (or exponential) parts of the hazard, the age-specific adaptive capacity $a(t)$, and other parameters of Models 1-4 are given in Table 1. The last column ("Diff. AIC") represents the difference between AIC for respective models and the model with minimal AIC (separately for females and males). It shows that for females the best fitting model is Model 2 (i.e., the model with the quadratic hazard and the gamma-Weibull baseline hazard), with Model 1 close behind. For males, the best fitting model is Model 3 (i.e., the model with the exponential hazard and the gamma-Gompertz baseline hazard) and Model 4 gives a slightly worse fit. The estimates of different components in Models 1-4 are described below.

Table 1 is about here

Fig. 2 is about here

All four models consistently show that the absolute value of the feedback coefficient in the equation for the age dynamics of physiological variable ($a(t)$) tends to decline with age (see panel B in Fig. 2 and Table 1). This decline (which is interpreted as the decline in adaptive capacity [1, 4, 28, 41]) is highly significant in all models (the null hypotheses on no decline, i.e., $b_Y = 0$, are rejected in all models with p<0.0001, see Table 1, column b_Y). The patterns of decline virtually coincide in females and males.

Null hypotheses on zero quadratic (or exponential) part of the hazard, i.e., $\mu_{11}(t) = \mu_{12}(t) = 0$, are rejected in all models (p<0.0001 for Models 3 and 4 for males and p<0.01 in all other cases). All four models reject at different significance levels (from p<0.0001 to p<0.01) the null hypotheses on zero multiplier $\mu_{11}(t)$, see column $a_{\mu_{11}}$ in Table 1. This means

that deviations to smaller than optimal values of hematocrit significantly increase the risk of cancer development in both females and males. The null hypotheses on zero multiplier $\mu_{12}(t)$ are rejected at 0.05 level in all models but only for males (see column $a_{\mu_{12}}$ in Table 1). This means that deviations to larger than optimal values of hematocrit significantly increase the risk of cancer development in males but not in females. The results on age dependence of multipliers $\mu_{11}(t)$ and $\mu_{12}(t)$ are mixed and not significant for females (see columns $b_{\mu_{11}}$ and $b_{\mu_{12}}$ in Table 1) and for the right parts in males (see column $b_{\mu_{12}}$ in Table 1). Therefore, no reliable conclusions can be made on the age dependence in these cases. The results for age-dependence of $\mu_{11}(t)$ in males are consistent in all models (see column $b_{\mu_{11}}$ in Table 1). All estimates are positive and significant at 0.05 level in Models 1, 2, and 4 and marginally significant in Model 3 (p=0.056). Fig. 2 (panels C and D) illustrates the patterns of $\mu_{11}(t)$ and $\mu_{12}(t)$ for females and males in the "optimal" models (i.e., Model 2 for females and Model 3 for males).

In all models, the estimates of parameter Δ_{f_1} are different from zero and are higher in females than in males. The null hypotheses on $\Delta_{f_1} = 0$ (i.e., that "mean allostatic state" $f_1(t)$ coincides with the "physiological norm" $f(t)$) are rejected (p<0.0001) for females in all models (see column Δ_{f_1} in Table 1).

Fig. 2 (panel A) shows that for both sexes the pattern of the baseline incidence rate decelerates with age (or even slightly declines with age for females). The null hypotheses on no decline in baseline hazards at old ages (i.e., $\sigma_2 = 0$) is rejected at various levels of significance in Models 1 and 3 for females and Models 1, 3 and 4 for males (see column σ_2 in Table 1).

Fig. 3 displays contour plots of hazard rate ($\mu(t,Y_t)$), relative risk ($RR(t,Y_t)=\mu(t,Y_t)/\mu_0(t)$, where $\mu_0(t)$ is the baseline rate) and the absolute increase in the risk compared to the baseline ($\mu(t,Y_t)-\mu_0(t)$) as functions of ages (t) and values of hematocrit at these ages (Y_t) estimated from the best-fitting models (i.e., Model 2 for females and Model 3 for males. Fig. 3A and 3B show clear U-shapes of hazard rates with non-symmetric left and right parts for both females and males. Both sexes have a substantial increase in cancer risk at smaller then "optimal" values of hematocrit, with a peak at ages 80-85 for females and about 90 for males. At the oldest old ages, however, the same low values of hematocrit result in smaller cancer risk than that at the ages 80-90. Deviations to the larger than "optimal" values of hematocrit in females do not produce a substantial increase in the risk, which is only observed at the oldest old ages. For males, the pattern is reversed – at younger ages deviations to larger values increase the risk but the effect diminishes at the oldest old ages. This reflects the patterns of estimates of the quadratic (or exponential) terms in the hazard (see Table 1 and Fig. 1). The absolute increase in the risk compared to the baseline ($\mu(t,Y_t)-\mu_0(t)$, Figs. 3E and 3F) shows the same pattern as those of the hazard rates in Fig. 3A and 3B. The relative risk also shows non-symmetric U-shapes in both sexes (Figs. 3C and 3D). The shape of relative risk resembles that of the absolute increase for males (except that the maximal relative risk for larger than "optimal" values of hematocrit is concentrated at younger ages whereas the maximal absolute increase is at the ages around 80). For females, the picture is similar for the relative risk and the absolute increase in the risk for larger than "optimal" values of hematocrit. For smaller than "optimal" values of hematocrit, however, the maximal absolute increase is reached at ages about 80, whereas the largest relative risk is observed at younger ages (where the baseline hazard is the smallest).

Fig. 3 is about here

4. Discussion

In this paper we investigated the relationship between the age dynamics of hematocrit and risk of cancer using data on ages at onset of cancer (all sites but skin) and longitudinal measurements of hematocrit in the Framingham Heart Study (original cohort). Analyses using the proportional hazards model suggested a generally negative effect of lower hematocrit values on the risk of cancer development. We also investigated how different mechanisms regulating the age dynamics of physiological variables (e.g., hematocrit in our applications) may affect the risk of onset of cancer applying the stochastic process model of aging [1]. These analyses revealed non-symmetric and age-dependent U-shapes of incidence rates as a function of hematocrit in both sexes, as discussed below.

We found that deviations to smaller than optimal values of hematocrit significantly increase the risk of cancer development in both females and males and that deviations to larger than optimal values of hematocrit significantly increase the risk of cancer development in males but not in females.

The decline in stress resistance is an important characteristic of the aging process [6, 29-34] which can lead to development of aging-related diseases and death. Available longitudinal studies typically contain very limited (if any at all) information on external disturbances affecting individuals during their life course and the intensities and magnitudes of persistent external stresses that affect an organism's functioning are generally unknown. Therefore, the direct estimation of such "stresses" from the data is not possible. Our approach makes possible indirect estimation of decline in stress resistance associated with deviations of physiological variables from the "optimal" trajectories minimizing the risks of disease development or death.

In this paper, we revealed age-dependence of the multiplier $\mu_{11}(t)$ for males. This means that the width of the left part (i.e., smaller than optimal values) of the U-shape of the risk (as a function of hematocrit) is getting narrower with age and the range of values of hematocrit corresponding to a "tolerable" increase in the risk is also getting narrower with age. Hence the "price" for the same magnitude of deviation to smaller than "optimal" values (in terms of an absolute increase in the risk of onset of cancer compared to the baseline level at that age) becomes higher for males at older ages. The higher values of this function ($\mu_{11}(t)$) at old ages suggest that males have lower resistance to stresses measured in terms of increase in the risk of onset of cancer due to deviation of hematocrit to smaller than "optimal" values at respective ages. The results also indicate that the right part of the U-shape of the risk (associated to larger than normal deviations of hematocrit with age) does not narrow significantly with age. This suggests that the "price" (i.e., an increase in the risk of cancer development) for the deviations to larger than "optimal" values of hematorcit does not change significantly with age in both sexes.

The decline in adaptive capacity is an important feature of aging ([11-14]) which may contribute to development of aging-related diseases and death. However, direct measurements of adaptive capacity are typically lacking in available longitudinal studies of aging, health, and longevity. Therefore, one needs to rely on some indirect methods to evaluate this component from the data. The use of the feedback coefficient in the equation for the age dynamics of physiological variable in our model allows us to evaluate this from the data because the absolute value of this feedback coefficient characterizes the adaptive capacity, see [1, 4, 28, 41]. In our applications to onset of cancer and hematocrit, we found that the decline in adaptive capacity with age is highly significant in all models and the patterns of decline virtually coincide in females and males. This result indicates that at older ages more time is needed for the trajectory

of hematocrit to approach the one that the organism tends to follow (i.e., "mean allostatic state" $f_1(t)$), compared to younger ages.

In all models, the estimates of parameter Δ_{f_1} are different from zero and are higher in females that in males. Non-zero estimates indicate that the processes of compensatory adaptation and remodeling regulating the age dynamics of hematocrit force its age trajectories to follow the curves which do not tend to minimize the risk of onset of cancer. Persistent deviations from the "norm" characterize the effects of allostatic adaptation and the magnitudes of such deviations can be associated with components of allostatic load leading to increased chances of onset of cancer. Note also that for this to be true, the values of the quadratic (or exponential) part of the hazard should be non-zero, which was always the case in our applications. The higher values of this parameter for females indicate that in females the trajectory of hematocrit tends to deviate further from the "optimal" trajectory minimizing the risk of onset of cancer.

For both sexes the pattern of the baseline incidence rate decelerates with age (or even slightly declines with age for females). This pattern corresponds to the patterns of cancer incidence rate at old ages observed in different countries and at different times [32, 36, 42]. Such decelerated patterns of cancer incidence rates at old ages may reflect the contribution of basal process of aging in the body which is manifested in slowdown of metabolism, proliferative response and information processing with age [32, 35]. Note also that the baseline rates for females and males intersect around ages 55-60 in both models, with female rates being higher before these ages, whereas after these ages female rates become lower than those of males. Such universal pattern of male/female cancer incidence rates is observed in different countries and time periods, and one possible explanation of such stable behavior involves the difference in ontogenetic components of aging between males and females [43].

In sum, the analyses using the stochastic process model showed that such aging-related processes as decline in adaptive capacity and resistance to stresses, and accumulation of allostatic load may contribute to an increase in the risk of onset of cancer with age. The results indicated the presence of substantial gender differences in these processes, which may contribute to the difference in the shape of the sex-specific patterns of cancer incidence rates. The underlying determinants of such differences (which may be of genetic or non-genetic origin) require additional studies.

Funding

The research reported in this paper was supported by the National Institute on Aging (grant numbers R01AG030612 and R01AG032319). 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.

Acknowledgements

The Framingham Heart Study (FHS) is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the FHS Investigators. This manuscript was prepared using a limited access dataset obtained from the NHLBI and does not necessarily reflect the opinions or views of the FHS or the NHLBI.

References

1. Yashin AI, Arbeev KG, Akushevich I, *et al.* Stochastic model for analysis of longitudinal data on aging and mortality. Mathematical Biosciences 2007;208(2):538-551.

2. Yashin AI, Ukraintseva SV, Arbeev KG, *et al.* Maintaining physiological state for exceptional survival: What is the normal level of blood glucose and does it change with age? Mechanisms of Ageing and Development 2009;130(9):611-618.

3. Yashin AI, Arbeev KG, Akushevich I, *et al.* Exceptional survivors have lower age trajectories of blood glucose: lessons from longitudinal data. Biogerontology 2010;11(3):257-265.

4. Arbeev KG, Ukraintseva SV, Akushevich I, *et al.* Age trajectories of physiological indices in relation to healthy life course. Mechanisms of Ageing and Development 2011;132(3):93-102.

5. Yashin AI, Arbeev KG, Ukraintseva SV, *et al.* Patterns of Aging Related Changes on the Way to 100: An Approach to Studying Aging, Mortality, and Longevity from Longitudinal Data. In. *2011 Living to 100 Monograph. Society of Actuaries Monograph M-LI11-1*. Schaumburg, IL: Society of Actuaries; 2011.

6. Yashin AI, Arbeev KG, Kulminski A, *et al.* Health decline, aging and mortality: how are they related? Biogerontology 2007;8(3):291-302.

7. Yashin AI, Arbeev KG, Kulminski A, *et al.* Cumulative index of elderly disorders and its dynamic contribution to mortality and longevity. Rejuvenation Research 2007;10(1):75-86.

8. Yashin AI, Arbeev KG, Kulminski A, *et al.* What age trajectories of cumulative deficits and medical costs tell us about individual aging and mortality risk: Findings from the NLTCS-Medicare data. Mechanisms of Ageing and Development 2008;129(4):191-200.

9. Dawber TR, Meadors GF, Moore FE. Epidemiological approaches to heart disease: The Framingham Study. American Journal of Public Health 1951;41(3):279-286.

10. Dawber TR. *The Framingham Study: The Epidemiology of Atherosclerotic Disease*. Cambridge, MA: Harvard University Press; 1980.

11. Troncale JA. The aging process: Physiologic changes and pharmacologic implications. Postgraduate Medicine 1996;99(5):111-114, 120-122.

12. Lund J, Tedesco P, Duke K, *et al.* Transcriptional profile of aging in C-elegans. Current Biology 2002;12(18):1566-1573.

13. Hall DM, Xu L, Drake VJ, *et al.* Aging reduces adaptive capacity and stress protein expression in the liver after heat stress. Journal of Applied Physiology 2000;89(2):749-759.

14. Rankin MM, Kushner JA. Adaptive beta-cell Proliferation Is Severely Restricted With Advanced Age. Diabetes 2009;58(6):1365-1372.

15. Sterling P, Eyer J. Allostasis: A New Paradigm to Explain Arousal Pathology. In: Fisher S, Reason J, (eds). *Handbook of Life Stress, Cognition and Health*. New York: John Wiley & Sons; 1988, 629-649.

16. McEwen BS. Allostasis and allostatic load: Implications for neuropsychopharmacology. Neuropsychopharmacology 2000;22(2):108-124.

17. Witteman JCM, Grobbee DE, Valkenburg HA, *et al.* J-shaped relation between change in diastolic blood pressure and progression of aortic atherosclerosis. Lancet 1994;343(8896):504-507.

18. Allison DB, Faith MS, Heo M, *et al.* Hypothesis concerning the U-shaped relation between body mass index and mortality. American Journal of Epidemiology 1997;146(4):339-349.

19. Boutitie F, Gueyffier F, Pocock S, *et al.* J-shaped relationship between blood pressure and mortality in hypertensive patients: New insights from a meta-analysis of individual-patient data. Annals of Internal Medicine 2002;136(6):438-448.

20. Kulminski AM, Arbeev KG, Kulminskaya IV, *et al.* Body mass index and nine-year mortality in disabled and nondisabled older U.S. Individuals. Journal of the American Geriatrics Society 2008;56(1):105-110.

21. Kuzuya M, Enoki H, Iwata M, *et al.* J-shaped relationship between resting pulse rate and all-cause mortality in community-dwelling older people with disabilities. Journal of the American Geriatrics Society 2008;56(2):367-368.

22. Mazza A, Zamboni S, Rizzato E, *et al.* Serum uric acid shows a J-shaped trend with coronary mortality in non-insulin-dependent diabetic elderly people. The CArdiovascular STudy in the ELderly (CASTEL). Acta Diabetologica 2007;44(3):99-105.

23. Okumiya K, Matsubayashi K, Wada T, *et al.* A U-shaped association between home systolic blood pressure and four-year mortality in community-dwelling older men. Journal of the American Geriatrics Society 1999;47(12):1415-1421.

24. Protogerou AD, Safar ME, Iaria P, *et al.* Diastolic blood pressure and mortality in the elderly with cardiovascular disease. Hypertension 2007;50(1):172-180.

25. Troiano RP, Frongillo EA, Sobal J, *et al.* The relationship between body weight and mortality: A quantitative analysis of combined information from existing studies. International Journal of Obesity 1996;20(1):63-75.

26. Yashin AI, Ukraintseva SV, De Benedictis G, *et al.* Have the oldest old adults ever been frail in the past? A hypothesis that explains modern trends in survival. Journals of

Gerontology. Series A, Biological Sciences and Medical Sciences 2001;56(10):B432-B442.

27. van Uffelen JGZ, Berecki-Gisolf J, Brown WJ, *et al.* What Is a Healthy Body Mass Index for Women in Their Seventies? Results from the Australian Longitudinal Study on Women's Health. Journals of Gerontology Series A Biological Sciences and Medical Sciences 2010;65(8):844-850.

28. Arbeev KG, Akushevich I, Kulminski AM, *et al.* Genetic model for longitudinal studies of aging, health, and longevity and its potential application to incomplete data. Journal of Theoretical Biology 2009;258(1):103-111.

29. Strehler BL, Mildvan AS. General theory of mortality and aging. Science 1960;132(3418):14-21.

30. Strehler B. *Time, Cells, and Aging*. London: Academic Press; 1962.

31. Semenchenko GV, Khazaeli AA, Curtsinger JW, *et al.* Stress resistance declines with age: analysis of data from a survival experiment with Drosophila melanogaster. Biogerontology 2004;5(1):17-30.

32. Ukraintseva SV, Yashin AI. Individual aging and cancer risk: How are they related? Demographic Research 2003;9(8):163-196.

33. Robb EL, Page MM, Stuart JA. Mitochondria, cellular stress resistance, somatic cell depletion and lifespan. Current Aging Science 2009;2(1):12-27.

34. Yashin AI, Akushevich IV, Arbeev KG, *et al.* Insights on aging and exceptional longevity from longitudinal data: novel findings from the Framingham Heart Study. Age 2006;28(4):363-374.

35. Ukraintseva SV, Yashin AI. How individual age-associated changes may influence human morbidity and mortality patterns. Mechanisms of Ageing and Development 2001;122(13):1447-1460.

36. Arbeev KG, Ukraintseva SV, Arbeeva LS, *et al.* Mathematical models for human cancer incidence rates. Demographic Research 2005;12(10):237-271.

37. Ukraintseva S, Sloan F, Arbeev K, *et al.* Increasing rates of dementia at time of declining mortality from stroke. Stroke 2006;37(5):1155-1159.

38. Ukraintseva SV, Arbeev KG, Akushevich I, *et al.* Trade-Offs Between Cancer and Other Diseases: Do They Exist and Influence Longevity? Rejuvenation Research 2010;13(4):387-396.

39. Rockwood K, Awalt E, MacKnight C, *et al.* Incidence and outcomes of diabetes mellitus in elderly people: report from the Canadian Study of Health and Aging. Canadian Medical Association Journal 2000;162(6):769-772.

40. Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control 1974;AC-19(6):716-723.

41. Yashin AI, Arbeev KG, Akushevich I, *et al.* The Quadratic Hazard Model for Analyzing Longitudinal Data on Aging, Health, and the Life Span. Physics of Life Reviews 2012:in press.

42. Harding C, Pompei F, Lee EE, *et al.* Cancer suppression at old age. Cancer Research 2008;68(11):4465-4478.

43. Arbeev KG, Ukraintseva SV, Arbeeva LS, *et al.* Difference between male and female cancer incidence rates: How can it be explained? In: Nikulin M, Commenges D, Huber C,

(eds). *Probability, Statistics and Modelling in Public Health*. New York: Springer-Verlag; 2006, 12-22.

Tables:

Table 1: Estimates of parameters of models 1-4 (see the text) applied to data on incidence of cancer (all sites but skin) and longitudinal measurements of hematocrit in female (F) and male (M) participants of the Framingham Heart Study (original cohort)

Sex	Model	Baseline Hazard ($\mu_0(t)$)			Multiplier in Quadratic (or Exponential) Part in Hazard ($\mu_1(t)$)				Adaptive Capacity ($a(t)$)		Other Parameters			ln L	Diff. AIC
		a_{μ_0}	b_{μ_0}	σ_2	$a_{\mu_{11}}$	$b_{\mu_{11}}$	$a_{\mu_{12}}$	$b_{\mu_{12}}$	a_Y	b_Y	σ_0	σ_1	Δ_{f_1}		
F	1	0.23	0.106	3.04^{*}	$-0.194^{\S}$	0.205	-0.636	0.159	-0.259	$1.920^{\dagger}$	3.12	1.72	$0.18^{\dagger}$	-65532.64	0.07
	2	91.94	5.507	2.11	$-0.185^{\S}$	0.203	-0.635	0.159	-0.259	$1.920^{\dagger}$	3.12	1.72	$0.18^{\dagger}$	-65532.61	0.00
	3	0.33	0.100	3.12^{*}	$0.121^{\#}$	-0.046	-0.081	0.204	-0.259	$1.920^{\dagger}$	3.12	1.72	$0.18^{\dagger}$	-65534.67	4.13
	4	95.56	5.070	2.07	$0.126^{\#}$	-0.054	-0.081	0.202	-0.259	$1.920^{\dagger}$	3.12	1.72	$0.18^{\dagger}$	-65534.63	4.05
M	1	0.06	0.120	$1.73^{\#}$	$-2.458^{\#}$	0.614^{*}	-1.130^{*}	0.330	-0.260	$1.957^{\dagger}$	3.18	1.82	0.08	-51086.85	9.09
	2	87.59	7.558	1.24	$-2.482^{\#}$	0.620^{*}	-0.939^{*}	0.296	-0.260	$1.957^{\dagger}$	3.18	1.82	0.08	-51087.53	10.44
	3	0.08	0.115	$1.88^{\S}$	$-0.133^{\dagger}$	0.333	0.191^{*}	-0.182	-0.260	$1.957^{\dagger}$	3.18	1.82	0.08	-51082.31	0.00
	4	89.59	7.109	1.33^{*}	$-0.134^{\dagger}$	0.336^{*}	0.191^{*}	-0.182	-0.260	$1.957^{\dagger}$	3.18	1.82	0.08	-51082.89	1.17

Notes:

1) **ln L** – logarithm of the likelihood function;
2) **Diff. AIC** – difference between AIC for respective models and the model with minimal AIC (separately for females and males);
3) The estimates of some parameters are rescaled for better visibility in the table: a_{μ_0} are multiplied by 10^4 in Models 1 and 3; b_Y are multiplied by 10^3 in all models; $a_{\mu_{1j}}$, j = 1, 2, are multiplied by 10^4 in Models 1 and 2; $b_{\mu_{1j}}$, j = 1, 2, are multiplied by 10^5 in Models 1 and 2 and by 10^2 in Models 3 and 4;
4) The symbols after the numbers in the following columns of Table 1 denote p-values (evaluated by the likelihood ratio test) for different null hypotheses tested for respective models:
   Column "σ_2": null hypothesis – Gompertz (or Weibull) baseline hazard rates $\mu_0(t)$, i.e., no decline in baseline hazard at old ages (restriction on parameters: $\sigma_2 = 0$);
   Columns "$a_{\mu_{1j}}$": null hypothesis – zero left (j = 1) or right (j = 2) parts of $\mu_1(t)$ ($a_{\mu_{1j}} = b_{\mu_{1j}} = 0$);
   Columns "$b_{\mu_{1j}}$": null hypothesis – age-independent left (j = 1) or right (j = 2) parts of $\mu_1(t)$ ($b_{\mu_{1j}} = 0$);
   Column "b_Y": null hypothesis – no aging-related decline in the adaptive capacity $a(t)$ ($b_Y = 0$);

Column "Δ_{f_1}": null hypothesis – "mean allostatic state" $f_1(t)$ coincides with "physiological norm" $f(t)$ ($\Delta_{f_1} = 0$).

The symbols in these columns denote: †: $p < 0.0001$; §: $0.0001 \le p < 0.001$; #: $0.001 \le p < 0.01$; *: $0.01 \le p < 0.05$, for respective null hypotheses. The absence of symbols after the numbers in these columns means that respective p-values exceed 0.05. Note that all other columns in the table, except the columns mentioned above, are not used to represent information on testing any null hypotheses and therefore they do not contain any symbols.

Legends to Figures:

Fig. 1: Average age trajectories (±s.e.) of hematocrit for individuals from the Framingham Heart Study (original cohort) who survived until advanced ages (≥ 90 for females, ≥ 85 for males) without developing cancer (all sites but skin) and their fit by cubic polynomials. Note: values of hematocrit measured after the onset of cancer are excluded from these calculations.

Fig. 2: Estimates of the baseline hazard rate ($\mu_0(t)$, panel **A**), adaptive capacity (|*a*(*t*)|, panel **B**), and the multiplier $\mu_1(t)$ in the quadratic (or exponential) part of the hazard (the "left side," $\mu_{11}(t)$, panel **C**, corresponding to deviations to the lower values from the "optimal" trajectory *f*(*t*), and the "right side," $\mu_{12}(t)$, panel **D**, corresponding to deviations to the higher values from *f*(*t*)) in Models 2 (for females) and 3 (for males) applied to data on onset of cancer (all sites but skin) and longitudinal measurements of hematocrit in participants of the Framingham Heart Study (original cohort). Note that the multipliers $\mu_1(t)$ in Models 2 and 3 have different scales. The scales for Model 2 (females) are shown on the left side of panels **C** and **D**, and the scales for Model 3 (males) are shown on the right side of these panels.

Fig. 3: Contour plots of hazard rate ($\mu(t,Y_t)$), relative risk ($RR(t,Y_t) = \mu(t,Y_t)/\mu_0(t)$, where $\mu_0(t)$ is the baseline rate) and the absolute increase in the risk compared to the baseline ($\mu(t,Y_t) - \mu_0(t)$) as functions of ages (*t*) and values of hematocrit at these ages (Y_t) estimated from the stochastic process models applied to data on incidence of cancer (all sites but skin) and longitudinal measurements of hematocrit in participants of the Framingham Heart Study

(original cohort): (**A**): hazard rates for females in Model 2 (see the text); (**B**): hazard rates for males in Model 3 (see the text); (**C**): relative risk for females in Model 2; (**D**): relative risk for males in Model 3; (**E**): the absolute increase in the risk compared to the baseline for females in Model 2; (**F**): the absolute increase in the risk compared to the baseline for males in Model 3. The dotted lines show the "optimal" trajectories $f(t)$.

Figures:

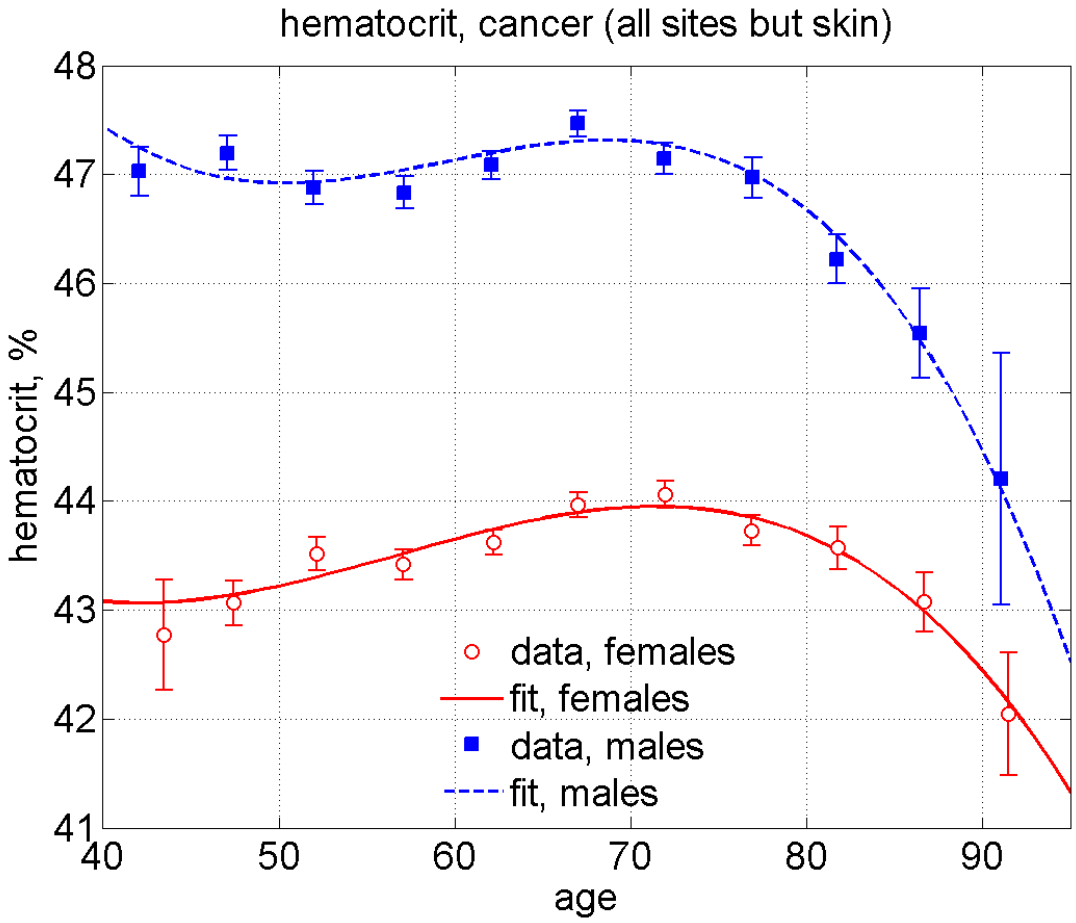

Fig. 1

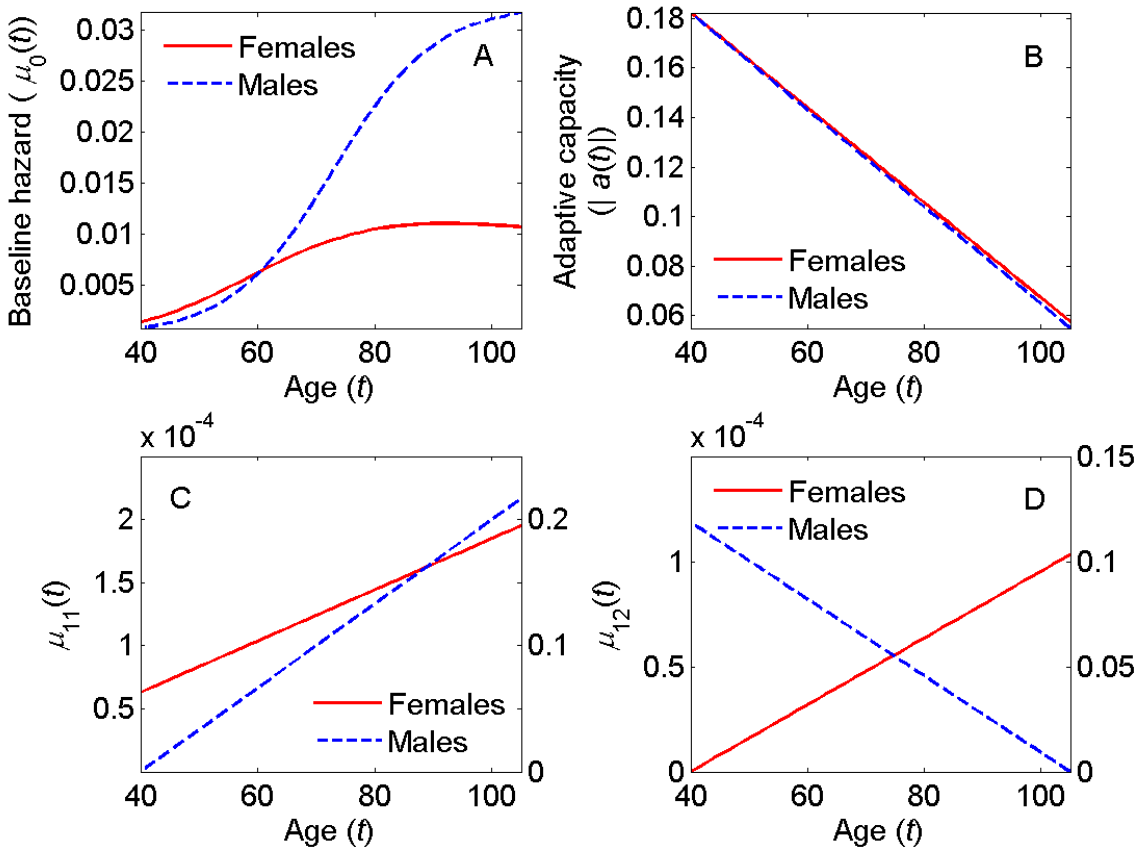
A
Females
Males
Baseline hazard ($\mu_0(t)$)
0.03
0.025
0.02
0.015
0.01
0.005
40
60
80
100
Age (t)
B
Adaptive capacity
(| a(t)|)
0.18
0.16
0.14
0.12
0.1
0.08
0.06
Females
Males
Age (t)
C
x 10^{-4}
$\mu_{11}(t)$
2
1.5
1
0.5
0.2
0.1
0
Females
Males
Age (t)
D
x 10^{-4}
$\mu_{12}(t)$
1
0.5
0
0.15
0.1
0.05
Females
Males
Age (t)

Fig. 2

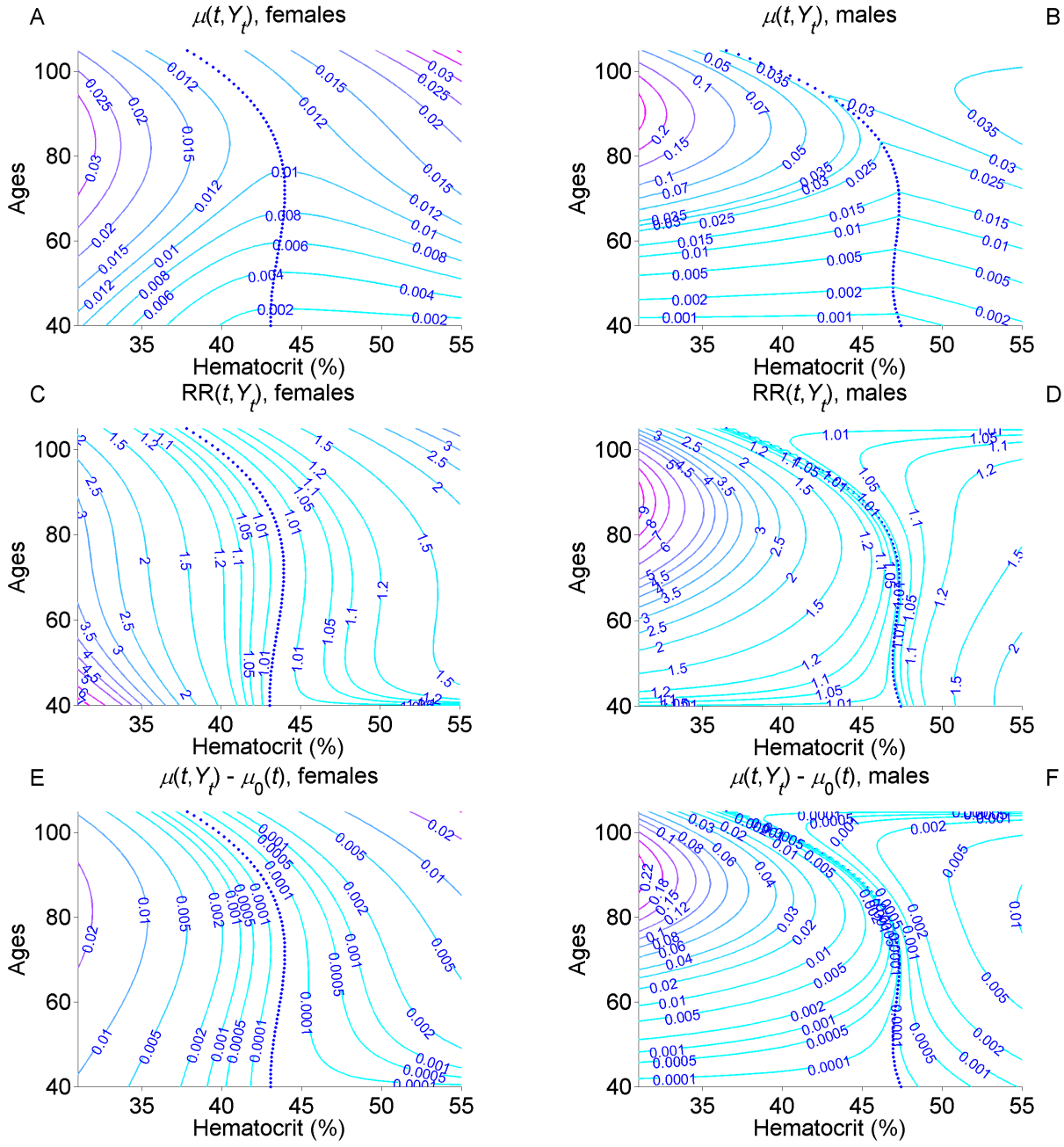
A
$\mu(t,Y_t)$, females
B
$\mu(t,Y_t)$, males
C
RR(t,Y_t), females
D
RR(t,Y_t), males
E
$\mu(t,Y_t) - \mu_0(t)$, females
F
$\mu(t,Y_t) - \mu_0(t)$, males
Ages
Hematocrit (%)

Fig. 3