Research Paper

Methods for joint modeling of longitudinal omics data and time-to-event outcomes: applications to lysophosphatidylcholines in connection to aging and mortality in the Long Life Family Study

Konstantin G. Arbeev[1], Olivia Bagley[1], Svetlana V. Ukraintseva[1], Alexander Kulminski[1], Eric Stallard[1], Michaela Schwaiger-Haber[2,3,4], Gary J. Patti[2,3,4], Yian Gu[5,6,7,8], Anatoliy I. Yashin[1], Michael A. Province[9]

[1]Biodemography of Aging Research Unit, Social Science Research Institute, Duke University, Durham, NC 27708, USA
[2]Department of Chemistry, Washington University in St. Louis, St. Louis, MO 63130, USA
[3]Department of Medicine, Washington University in St. Louis, St. Louis, MO 63130, USA
[4]Center for Metabolomics and Isotope Tracing at Washington University in St. Louis, St. Louis, MO 63130, USA
[5]Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA
[6]G.H. Sergievsky Center, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA
[7]Department of Neurology, Vagelos College of Physicians and Surgeons, Columbia University, and the New York Presbyterian Hospital, New York, NY 10032, USA
[8]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY 10032, USA
[9]Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110, USA

Correspondence to: Konstantin G. Arbeev; **email:** ka29@duke.edu
Keywords: lysophosphatidylcholines, aging, mortality, longitudinal omics, repeated measurements
Received: August 2, 2024 **Accepted:** May 16, 2025 **Published:** May 27, 2025

ABSTRACT

Studying the relationships between longitudinal changes in omics variables and event risks requires specific methodologies for joint analyses of longitudinal and time-to-event outcomes. We applied two such approaches (joint models [JM], stochastic process models [SPM]) to longitudinal metabolomics data from the Long Life Family Study, focusing on the understudied associations of longitudinal changes in lysophosphatidylcholines (LPCs) with mortality and aging-related outcomes. We analyzed 23 LPC species, with 5,066 measurements of each in 3,462 participants, 1,245 of whom died during follow-up. JM analyses found that higher levels of the majority of LPC species were associated with lower mortality risks, with the largest magnitude observed for LPC 15:0/0:0 (hazard ratio: 0.71, 95% CI (0.64, 0.79)). SPM applications to LPC 15:0/0:0 revealed that the JM association reflects underlying aging-related processes: a decline in robustness to deviations from optimal LPC levels, higher equilibrium LPC levels in females, and the opposite age-related changes in the equilibrium and optimal LPC levels (declining and increasing, respectively), which lead to increased mortality risks with age. Our results support LPCs as biomarkers of aging and related decline in biological robustness, and call for further exploration of factors underlying age-related changes in LPC in relation to mortality and diseases.

INTRODUCTION

Contemporary longitudinal studies on humans started collecting repeated measurements of various omics (e.g., metabolomics, proteomics) data for study participants. The availability of other types of information on the participants, such as follow-up data on mortality and the onset of diseases, genetic markers, questionnaires, repeated measures of health-related biomarkers, etc., provides extensive opportunities to study complex relations of individual age-trajectories of omics variables with risks of diseases and mortality, in connection to various genetic and non-genetic factors. However, this abundance of information and opportunities comes along with many methodological challenges related to the analyses of such massive data. One particular complication deals with the inherent complexity of analyzing trajectories of health-related variables (repeated measurements of omics variables provide a good example of such) in relation to time-to-event outcomes. The Cox model with time-dependent covariates [1] is the conventional approach traditionally used for joint analyses of time-to-event data and repeated measurements of covariates. However, it is well known that it has certain limitations: ignoring measurement errors or biological variation of covariates and using their observed "raw" values as time-dependent covariates in the Cox model may lead to biased estimates and incorrect inferences [2–4], especially when covariates are measured at sparse examinations or with a long-time interval before an outcome event. This applies to analyzing repeated omics measurements in relation to time-to-event outcomes as well. Even though relevant biostatistical methods, known as joint models (JM) [4, 5], have found broad applications in different research areas, their use in the analyses of longitudinally measured omics data is still limited to a few small-sample proteomics studies [6–8].

One particular class of models for joint analyses of longitudinal and time-to-event outcomes, the stochastic process model (SPM), has been developed in the biodemographic literature based on the mathematical foundations laid out in [9–11]. Recent developments in SPM methodology merged the statistical rigor of the general approach with the biological soundness of specific assumptions built into its structure [12–15] (see [16] for a non-technical introduction to SPM). This brings the biological content to the model structure, making such models particularly appealing for research on aging. The main advantage of using SPM for research on aging is that it allows disentangling a general association between the longitudinal and time-to-event outcomes that can be found using JM into several components representing specific aging-related characteristics embedded in the model. This allows researchers not only to evaluate mortality or incidence rates but also to estimate age-related changes in the mechanism of homeostatic regulation of biological variables, the age-related decline in adaptive capacity and stress resistance, effects of allostatic adaptation, and allostatic load. This provides a more detailed perspective on the impact of the longitudinal changes of the respective variables on the risk of the modeled events in the context of aging. Despite broad applications of SPM to different outcomes and biomarkers (see, e.g., [13, 17–23]), to date, there have been no applications of SPM to analyses of longitudinal omics measurements in relation to time-to-event outcomes.

In this paper, we fill these gaps and apply both JM and SPM to longitudinal measurements of metabolomics collected from participants of the Long Life Family Study (LLFS) [24]. To illustrate applications of the approaches, we focus on a particular class of lipid metabolites, lysophosphatidylcholines (LPCs), which have been actively discussed in the literature in relation to cardiovascular, infectious, and neurodegenerative diseases, and tested as potential early markers of Alzheimer's disease and accelerated aging [25–31]. Overall, the literature suggests (see, e.g., the recent review [32]) that the reported LPC findings are somewhat contradictory because most of the recent studies, in contrast to older ones, found lower LPC levels to be associated with unfavorable outcomes such as mortality. In addition, the longitudinal changes of LPC in relation to mortality and aging-related outcomes remain understudied. Here, we aimed to test general associations of different LPC species with total (all-cause) mortality in the LLFS using JM and to investigate how such general associations can be decomposed into relations of the mortality risk with different aging-related characteristics (such as robustness, resilience, age-specific norms, and allostatic trajectories [16]), and whether such relationships/ characteristics differ by sex.

RESULTS

Applications of the basic JM

Summary of results

Table 1 shows results of applications of the basic JM [4, 5, 33] (see Eqs. 1–2 in the section Joint models: General specifications in Materials and Methods) to measurements of LPC species and mortality data in the LLFS. The table presents the values of the association parameter (α in Eq. 1) (columns Alpha) for the respective metabolites in the survival sub-model, along with corresponding hazard ratios (for a unit increase in

Table 1. Results of applications of joint models to measurements of LPC species and mortality data in the LLFS: Estimates of the association parameter for the metabolite in the survival sub-model.

Metabolite	Total		Females		Males	
	Alpha	HR (95% CI)	Alpha	HR (95% CI)	Alpha	HR (95% CI)
LPC 0:0/16:0	−0.185	**0.831 (0.737, 0.938)**	−0.177	**0.838 (0.707, 0.994)**	−0.228	**0.796 (0.665, 0.952)**
LPC 0:0/16:1	−0.036	0.965 (0.849, 1.097)	0.005	1.005 (0.828, 1.220)	−0.046	0.955 (0.803, 1.135)
LPC 0:0/18:0	−0.03	0.970 (0.883, 1.066)	0.005	1.005 (0.880, 1.147)	−0.062	0.940 (0.818, 1.080)
LPC 0:0/18:1	−0.129	**0.879 (0.780, 0.990)**	−0.082	0.921 (0.782, 1.084)	−0.191	**0.826 (0.701, 0.974)**
LPC 0:0/18:2	−0.209	**0.811 (0.715, 0.920)**	−0.071	0.931 (0.777, 1.116)	−0.307	**0.736 (0.616, 0.879)**
LPC 0:0/20:3	−0.151	**0.860 (0.767, 0.963)**	−0.134	0.875 (0.752, 1.020)	−0.165	0.848 (0.716, 1.004)
LPC 0:0/20:4	−0.161	**0.851 (0.766, 0.945)**	−0.158	**0.854 (0.731, 0.997)**	−0.159	**0.853 (0.739, 0.985)**
LPC 0:0/22:6	−0.194	**0.824 (0.736, 0.922)**	−0.257	**0.773 (0.657, 0.909)**	−0.139	0.870 (0.741, 1.021)
LPC 14:0/0:0	−0.185	**0.831 (0.716, 0.965)**	−0.15	0.861 (0.701, 1.058)	−0.224	**0.799 (0.642, 0.995)**
LPC 15:0/0:0	−0.341	**0.711 (0.640, 0.790)**	−0.322	**0.725 (0.626, 0.840)**	−0.383	**0.682 (0.582, 0.798)**
LPC 16:0/0:0	−0.189	**0.828 (0.743, 0.923)**	−0.19	**0.827 (0.708, 0.967)**	−0.203	**0.816 (0.697, 0.955)**
LPC 16:1/0:0	−0.052	0.949 (0.839, 1.074)	−0.005	0.995 (0.826, 1.197)	−0.067	0.935 (0.793, 1.101)
LPC 17:0/0:0	−0.195	**0.823 (0.736, 0.921)**	−0.194	**0.824 (0.704, 0.965)**	−0.203	**0.816 (0.692, 0.962)**
LPC 18:0/0:0	0.022	1.022 (0.935, 1.117)	0.053	1.054 (0.932, 1.193)	−0.015	0.985 (0.865, 1.121)
LPC 18:1/0:0	−0.202	**0.817 (0.729, 0.916)**	−0.159	0.853 (0.727, 1.002)	−0.256	**0.774 (0.658, 0.910)**
LPC 18:2/0:0	−0.255	**0.775 (0.682, 0.880)**	−0.126	0.882 (0.734, 1.060)	−0.371	**0.69 (0.577, 0.824)**
LPC 18:3/0:0	−0.113	0.893 (0.747, 1.068)	−0.004	0.996 (0.798, 1.244)	−0.241	**0.786 (0.631, 0.980)**
LPC 20:2/0:0	−0.196	**0.822 (0.729, 0.927)**	−0.144	0.866 (0.733, 1.024)	−0.248	**0.780 (0.678, 0.896)**
LPC 20:3/0:0	−0.25	**0.779 (0.689, 0.881)**	−0.211	**0.810 (0.687, 0.955)**	−0.305	**0.737 (0.621, 0.874)**
LPC 20:4/0:0	−0.173	**0.841 (0.754, 0.938)**	−0.137	0.872 (0.746, 1.020)	−0.218	**0.804 (0.692, 0.934)**
LPC 20:5/0:0	−0.268	**0.765 (0.676, 0.866)**	−0.3	**0.741 (0.625, 0.878)**	−0.222	**0.801 (0.678, 0.945)**
LPC 22:5/0:0	−0.243	**0.784 (0.704, 0.873)**	−0.248	**0.780 (0.661, 0.919)**	−0.226	**0.798 (0.695, 0.917)**
LPC 22:6/0:0	−0.223	**0.800 (0.718, 0.892)**	−0.264	**0.768 (0.653, 0.903)**	−0.205	**0.815 (0.702, 0.946)**

Alpha – estimates of the association parameter for the longitudinal variable (metabolite) in the survival sub-model of the joint model (parameter α in Eq. 1); HR – hazard ratios (for a unit increase in transformed metabolite values) computed from the association parameters; 95% CI – 95% confidence intervals for HRs (highlighted in bold are cases where confidence intervals do not contain one). The joint models were estimated using R-package *JM*.

transformed metabolite values) and their 95% confidence intervals (CI) (columns HR (95% CI)) estimated from the JM adjusted for the covariates indicated in the section Joint models: Specific versions used in applications in Materials and Methods. For the majority of LPC species (18 out of 23) in the total sample, the estimates of the association parameter α are negative and CI for respective HR do not contain one. As explained in the section Joint models: General specifications in Materials and Methods, this means that the increase in the levels of these metabolites reduces mortality risk. Sex-specific analyses revealed similar observations for 10 out of 23 LPCs in females and 17 out of 23 LPCs in males.

Illustrative example: Association of LPC 15:0/0:0 with mortality risk

The strongest association in the total sample in terms of the point estimate was observed for LPC 15:0/0:0 (α = −0.341, HR = 0.711), indicating a 28.9% reduction in the risk of the event (death) happening for each unit increase in (transformed) LPC 15:0/0:0 levels. Similarly, LPC 15:0/0:0 showed the lowest HR among all LPCs in sex-specific analyses (HR = 0.725 in females, HR = 0.682 in males), corresponding to a 27.5% and a 31.8% reduction in the risk of death for each unit increase in LPC 15:0/0:0 in females and males, respectively. This metabolite was selected for additional analyses illustrating different specifications of JM and more detailed investigation of its association with different aging-related characteristics embedded in the structure of SPM; see below. Supplementary Figure 1 displays diagnostic plots assessing the goodness-of-fit and assumptions of JM in applications to LPC 15:0/0:0. Supplementary Figure 1A shows random behavior of standardized marginal residuals around zero (with 94.47% of values lying within the (−1.96, 1.96) interval), validating the assumptions for the within-subjects covariance structure

Table 2. Results of applications of joint models with shared random effects (JM-SRE) to measurements of LPC 15:0/0:0 and mortality data in the LLFS: Estimates of the association parameters for the random intercepts and random slopes of the metabolite in the survival sub-model.

Model	Sex	Variable	α_0 (α_1)	HR	95% CI for HR	SD of Variable
int	F+M	b_{0i}	**−0.436**	**0.786**	**(0.728, 0.840)**	**0.553**
intslope	F+M	b_{0i}	**−0.452**	**0.775**	**(0.726, 0.846)**	**0.563**
intslope	F+M	b_{1i}	−0.765	0.998	(0.994, 1.004)	0.002
int	F	b_{0i}	**−0.428**	**0.783**	**(0.709, 0.858)**	**0.572**
intslope	F	b_{0i}	**−0.408**	**0.778**	**(0.694, 0.865)**	**0.615**
intslope	F	b_{1i}	−0.084	0.999	(0.973, 1.041)	0.011
int	M	b_{0i}	**−0.461**	**0.785**	**(0.711, 0.854)**	**0.525**
intslope	M	b_{0i}	**−0.439**	**0.780**	**(0.698, 0.855)**	**0.568**
intslope	M	b_{1i}	−0.238	0.997	(0.956, 1.006)	0.011

Model – type of joint model (int – random intercept of LPC in survival sub-model, intslope – random intercept and slope of LPC in survival sub-model), see section Joint models: General specifications; Variable – b_{0i}: random intercept of the metabolite, b_{1i}: random slope of the metabolite; α_0 (α_1) – estimates of the regression parameters for b_{0i} (b_{1i}) in respective models; HR – hazard ratios for an increase by a standard deviation of Variable; 95% CI for HR – respective 95% confidence intervals for HRs; SD of Variable – standard deviation of Variable. Highlighted in bold are cases where confidence intervals do not contain one. The JM were estimated using R-package *joineR*. LPC values were transformed (see section Data).

in the longitudinal part of JM. The Cox-Snell residuals plot (Supplementary Figure 1B) also shows the overall good fit of the survival sub-model of JM.

Applications of JM with shared random effects (JM-SRE)

Summary of results

The results of applications of the general JM described in the previous section established the associations of the LPC species with mortality risk. Here, we describe the results of applications of a different type of JM, JM-SRE [34–36] (see Eqs. 3–5 in the section Joint models: General specifications in Materials and Methods), with individual intercepts and slopes of (transformed) LPCs that provide further insight into the relationships between the age-related changes of the metabolites and mortality risk. Table 2 presents the results of applications of the models with individual intercepts (rows with "int" in the column Model, see Eq. 4) and individual intercepts and slopes (rows with "intslope" in the column Model, see Eq. 5) of LPC 15:0/0:0 in the total (rows with "F+M" in the column Sex) and sex-specific samples (rows with "F" and "M" in the column Sex for females and males, respectively).

The estimates of the regression parameter α_0 for the random intercept b_{0i} in the hazard rate (see Eqs. 4–5) are negative (see the values in the column α_0 (α_1) in rows with b_{0i} in the column Variable), indicating that larger baseline levels of LPC 15:0/0:0 are associated with reduced mortality risks (after adjusting for relevant covariates, see the section Joint models: Specific versions used in applications in Materials and Methods). This observation holds in both models ("int" and "intslope") and in total and sex-specific analyses. Therefore, similar conclusions about the reduction of the risk of death for increasing baseline LPC 15:0/0:0 levels can be made for the model with individual intercept and slope and for females and males.

Similarly, there are negative estimates of the regression parameter α_1 for the random slope b_{1i} in the hazard rate in the random intercept and slope model (see Eq. 5) (see the values in the column α_0 (α_1) in rows with b_{1i} in the column Variable), but the associations were not significant. The negative estimates of this parameter suggest that an increase in the individual slope of LPC 15:0/0:0 might be associated with reduced mortality risk (in the model adjusting for the covariates indicated in the section Joint models: Specific versions used in applications in Materials and Methods). Supplementary Table 1 presents the results of applications of JM-SRE to all 23 LPC species in the total sample. As this table shows, similar associations were observed for many other LPC species, and LPC 15:0/0:0 still has the largest effect size among all LPCs, as in the basic JM analyses presented in the previous section.

Illustrative example: Results for the random intercept model in the total sample

The first row in Table 2 presents results for the random intercept ("int") model applied to the total sample (F+M). It shows that the HR per standard deviation (SD) of an individual intercept in the random intercept

model is 0.786. This means a 21.4% reduction in mortality risk for each 0.553 (see the value in the column SD of Variable) increase in transformed LPC 15:0/0:0 levels at the baseline age.

Applications of SPM

Summary of SPM results

In this section, we summarize the results of applications of SPM that further decompose the associations between LPC 15:0/0:0 and mortality risk. These applications consider different aging-related components embedded in the structure of SPM, providing additional details in the context of the aging process. Supplementary Tables 2, 3 present the results of testing of various null hypotheses (H0s) in applications of SPM to measurements of (transformed) LPCs and mortality in the LLFS metabolomics sample. Supplementary Table 4 contains estimates of parameters in the main (unrestricted) model. Supplementary Text Stochastic process models: Interpretation and illustration of components, parameters, and related null hypotheses and Supplementary Figures 2–9 provide detailed descriptions and illustrations of SPM components and parameters, as well as interpretations of related H0s. Below, we provide several illustrative examples describing the results of applications of SPM to LPC 15:0/0:0, referring to rows corresponding to LPC 15:0/0:0 in Supplementary Tables 2–4.

Illustrative example: U-shape of mortality as a function of LPC 15:0/0:0

This example illustrates the results presented in columns Qzero and QnoT in Supplementary Table 2, column QnoC in Supplementary Table 3 for LPC 15:0/0:0, and parameters shown in section Q (t, c) in Supplementary Table 4 for this metabolite. The key hypothesis to test before interpreting the SPM results is whether the term in the quadratic part of the hazard (Eq. 7), Q (t, c), is zero. If this were the case, there would be no quadratic part in the hazard, and the age trajectories of LPC 15:0/0:0 would be unrelated to the mortality risk (Supplementary Figure 2B). Rejection of H0: $Q(t, c) = 0$ (column Qzero in Supplementary Table 2) thus justifies the use of SPM in analyses of LPC 15:0/0:0 and mortality. As Supplementary Figure 10B illustrates, there is a non-zero quadratic part in the hazard, which adds up to the baseline level μ_0 (t, c) to produce the resulting shape of the total mortality rate as a function of age and LPC 15:0/0:0 shown in Supplementary Figure 10A. Rejection of H0: Q (t, c) = Q (c) (column QnoT in Supplementary Table 2) and a positive value of the parameter b_Q (see the respective column in Supplementary Table 4) indicate that the U-shape of the mortality risk as a function of LPC 15:0/0:0 narrows with age (see Supplementary Text, paragraph *"The quadratic hazard term Q (t, c)"* and Supplementary Figure 2). This means that, as individuals grow older, they become more vulnerable to deviations of LPC 15:0/0:0 from the optimal trajectory f_0 (t, c) because the same deviation of LPC 15:0/0:0 from f_0 (t, c) results in a larger increase in the mortality risk at older ages compared to younger ages. Figure 1A displays the trajectories of Q (t, c) for females (c = 0) and males (c = 1), which increase with age as $b_Q > 0$. As Figure 1A shows, the trajectory of Q (t, c) for males lies below that of females as the respective coefficient shown in column β_Q in Supplementary Table 4 is negative. However, H0: Q (t, c) = Q (t) (column QnoC in Supplementary Table 3) was not rejected. Therefore, this difference is not statistically significant, so the width of U-shape of the hazard (as a function of LPC 15:0/0:0) does not depend on sex, i.e., there is no sex difference in vulnerability to deviations of trajectories of LPC 15:0/0:0 from the optimal levels f_0 (t, c).

Illustrative example: Feedback and volatility coefficients of LPC 15:0/0:0

This example discusses the results shown in column AnoT in Supplementary Table 2, columns AnoC and BnoC in Supplementary Table 3 for LPC 15:0/0:0, and respective parameters from sections a (t, c) and b (t, c) in Supplementary Table 4. The H0 about the feedback coefficient a (t, c), H0: a (t, c) = a (c) (column AnoT in Supplementary Table 2), was not rejected. The coefficient b_Y is nearly zero (see the respective column in Supplementary Table 4) so that a (t, c) does not change with age as Figure 1B illustrates. This means that there is no age-related decline in biological resilience related to deviations of trajectories of LPC 15:0/0:0 from the equilibrium levels f_1 (t, c) (Supplementary Text, paragraph *"The (negative) feedback coefficient a (t, c)"*). The coefficient β_Y is negative (see the respective column in Supplementary Table 4), so the absolute value of a (t, c) is larger in males (Figure 1B); however, this sex difference is not statistically significant as H0: a (t, c) = a (t) (column AnoC in Supplementary Table 3) was not rejected. In addition, we found that the volatility of LPC 15:0/0:0 was similar in females and males as H0: b (t, c) = b (t) (column BnoC in Supplementary Table 3) was not rejected. Figure 1C displays the values of the volatility coefficient b (t, c) for females (c = 0) and males (c = 1), showing a slightly higher volatility of LPC 15:0/0:0 in males (see the positive value of β_W in the respective column in Supplementary Table 4), see also Supplementary Text, paragraph *"The volatility coefficient b (t, c)"*.

Illustrative example: Equilibrium trajectories of LPC 15:0/0:0

This example presents the results shown in column F1noT in Supplementary Table 2 and column F1noC in

Supplementary Table 3 for LPC 15:0/0:0, and corresponding parameters from section f_1 (t, c) in Supplementary Table 4. Rejection of H0: f_1 (t, c) = f_1 (c) (column F1noT in Supplementary Table 2) indicates that the equilibrium values of LPC 15:0/0:0 change with age (see also Supplementary Text, paragraph *"The equilibrium trajectory f_1 (t, c)"*). The estimated value of the slope parameter in f_1 (t, c) (parameter b_{f_1}) is negative (see the respective column in Supplementary Table 4), so the equilibrium trajectory of LPC 15:0/0:0 declines with age. Figure 1D shows the trajectories of f_1 (t, c) for females (c = 0) and males (c = 1), illustrating the higher equilibrium levels in females (see the negative value of β_{f_1} in the respective column in Supplementary Table 4 and the rejected H0: f_1 (t, c) = f_1 (t) in column F1noC in Supplementary Table 3).

Illustrative example: Optimal trajectories of LPC 15:0/0:0

This example illustrates the results presented in column F0noT in Supplementary Table 2 and column F0noC in Supplementary Table 3 for LPC 15:0/0:0, as well as parameters shown in section f_0 (t, c) in Supplementary Table 4 for this metabolite. We found that the optimal LPC 15:0/0:0 levels also change with age (H0: f_0 (t, c) = f_0 (c) was rejected, see column F0noT in Supplementary Table 2). This means that the U-shape of the quadratic part in the hazard shifts with age, and so does the minimal mortality level as a function of LPC 15:0/0:0 at specific ages (see also Supplementary Text, paragraph *"The optimal trajectory f_0 (t, c)"*). The slope of f_0 (t, c) (parameter b_{f_0}) is positive (see the respective column in Supplementary Table 4), so the optimal LPC 15:0/0:0

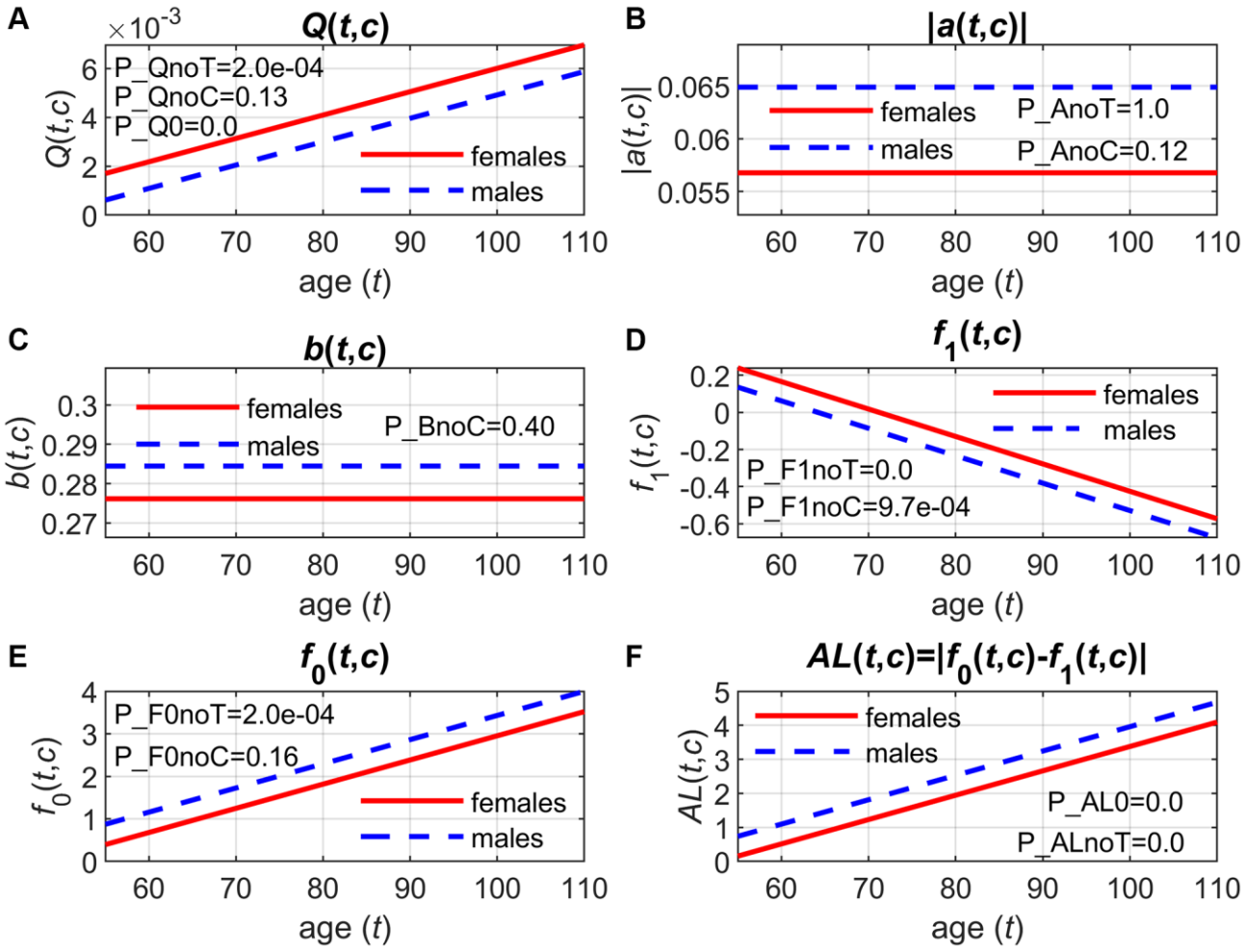


Figure 1. Applications of stochastic process models to measurements of LPC 15:0/0:0 and mortality data in the LLFS: Estimates of different components of the model. (**A**) quadratic hazard term (Q(t, c)); (**B**) adaptive capacity (|a(t, c)|); (**C**) volatility coefficient (b(t, c)); (**D**) equilibrium trajectory (f_1 (t, c)); (**E**) optimal trajectory (f_0 (t, c)); (**F**) measure of allostatic load (AL(t, c) = |f_0 (t, c)–f_1 (t, c)|); p-values shown on the graphs are for different null hypotheses (H0): H0: Q(t, c) = Q(c) (P_QnoT); H0: Q(t, c) = Q(t) (P_QnoC); H0: Q(t, c) = 0 (P_Q0); H0: a(t, c) = a(c) (P_AnoT); H0: a(t, c) = a(t) (P_AnoC); H0: b(t, c) = b(t) (P_BnoC); H0: f_1(t, c) = f_1(c) (P_F1noT); H0: f_1(t, c) = f_1(t) (P_F1noC); H0: f_0(t, c) = f_0(c) (P_F0noT); H0: f_0(t, c) = f_0(t) (P_F0noC); H0: f_1(t, c) = f_0(t, c), i.e., AL(t, c) = 0 (P_AL0); H0: f_1(t, c) = f_1(c) and f_0(t, c) = f_0 (c), i.e., AL(t, c) = AL(c) (P_ALnoT). LPC values were transformed (see Data).

level shifts to larger values with age. Figure 1E illustrates this by displaying the optimal LPC 15:0/0:0 levels f_0 (t, c) for females (c = 0) and males (c = 1). Supplementary Figure 10A, 10B present 3D plots of the mortality rate and the quadratic part in the hazard as a function of age and LPC 15:0/0:0, showing the shifting patterns of the U-shape of the quadratic part and the minimal mortality rate (as a function of LPC 15:0/0:0) at specific ages. The estimate of the parameter β_{f_0} is positive (see the respective column in Supplementary Table 4), which corresponds to higher optimal LPC 15:0/0:0 levels in males (Figure 1E); however, this sex difference is not significant (see column F0noC in Supplementary Table 3).

Illustrative example: The gap between equilibrium and optimal trajectories of LPC 15:0/0:0

This example discusses the results shown in columns ALzero and ALnoT in Supplementary Table 2. As Figure 1D, 1E reveal, the equilibrium and optimal trajectories of LPC 15:0/0:0 have opposite directions of change with age (declining vs. increasing), so the absolute value of the difference between these trajectories increases with age (see Figure 1F). This gap between f_1 (t, c) and f_0 (t, c), and its increase with age, are significant (see columns ALzero and ALnoT in Supplementary Table 2). This results in an additional mortality risk if LPC 15:0/0:0 is at the equilibrium level, compared to the minimal risk at respective ages given by the baseline mortality μ_0 (t, c), and this additional mortality "load" increases with age (see also the last paragraph in Supplementary Text).

Summary of SPM findings for LPC 15:0/0:0

In brief, we found that:

1. The U-shape of mortality as a function of LPC 15:0/0:0 narrows with age, making older individuals more vulnerable to deviations of LPC 15:0/0:0 concentrations from the trajectory of its optimal values.
2. Equilibrium trajectories of LPC 15:0/0:0 decline with age.
3. Females have higher equilibrium levels of LPC 15:0/0:0 than males.
4. The optimal values of LPC 15:0/0:0 that minimize the mortality risk increase with age.
5. There is a gap between the optimal and equilibrium trajectories, and this gap increases with age.

Sensitivity analyses

Supplementary Table 5 presents estimates of JM using the familial bootstrap approach [37]. There was one case where the 95% CI for HR in the main calculations (Table 1) did not contain 1.0, but the HR range in the familial bootstrap included 1.0 (highlighted in yellow in Supplementary Table 5). There were three opposite cases (highlighted in grey in Supplementary Table 5). Thus, in most cases, the sensitivity analysis confirmed the results shown in Table 1. In particular, LPC 15:0/0:0 still showed the strongest association with mortality (e.g., HR = 0.713, range: (0.643, 0.778) in the combined females + males analyses).

DISCUSSION

This work is the first application of two approaches dealing with joint modeling of longitudinal and time-to-event outcomes (JM and SPM) to a large-scale metabolomics study that collected repeated measurements of metabolomics from thousands of participants. These approaches allow for statistically rigorous analyses of repeated measurements of omics data jointly with time-to-event outcomes, avoiding common pitfalls of traditional tools that ignore biological variability/measurement errors in longitudinal outcomes and informative missingness arising due to attrition from mortality (a common situation in aging research) [2, 3, 16]. The basic JM [4] used in our applications allows establishing general associations of longitudinal omics variables with time-to-event outcomes by including the "true" values (i.e., the difference between the observed value and the error term, see Eq. 2) of the variable in the hazard rate and computing respective hazard ratios. The JM-SRE version captures associations between longitudinal and time-to-event outcomes by a latent Gaussian process [34–36], providing different specifications of associations that include individual intercepts and slopes of omics variables in the hazard rate (Eqs. 3–5). These are derived from the random part of the longitudinal sub-model of JM, representing individual characteristics after adjustment for covariates (in the fixed part of the longitudinal sub-model) and the error term. Such models expand analyses by the basic JM and quantify the relations between these individual characteristics of omics trajectories and time-to-event outcomes, e.g., computing hazard ratios for a unit increase in an individual intercept or slope. The SPM digs deeper into the relations between longitudinal changes of omics variables and time-to-event outcomes, decomposing the associations observed in JM into several components representing relevant aging-related characteristics. These characteristics include biological/physiological norms ("sweet spots" [38–40]), allostatic (equilibrium) trajectories and allostatic load, as well as age-related decline in adaptive response to deviations from equilibrium trajectories and age-related increase in vulnerability to deviations from the norms, which represent a decline in biological/physiological robustness and resilience, considered key manifestation of aging

[41]. SPM analyses thus can shed more light on relations between age trajectories of omics variables and time-to-event outcomes in the context of aging.

Our applications of JM to data on repeated measurements of different LPC species and mortality in the LLFS found that, for many LPCs, larger levels were associated with reduced mortality risk (or, equivalently, lower levels were associated with increased mortality risk) in the total sample as well as in separate analyses of females and males. This confirms recent results that reported associations of lower LPC levels with unfavorable health outcomes, including mortality (see, e.g., reviews in [25, 32]). For example, in the study of patients with sepsis [42], the non-survival group had significantly lower levels of LPCs 16:0, 17:0, and 18:0 compared to the survival group. In the study involving acute-on-chronic liver failure patients [43], those who died had lower LPC levels than survivors. Decreased LPC levels were significantly associated with increased mortality in bacterial community-acquired pneumonia patients [44]. Reduced LPC levels were associated with poor prognosis (including mortality) in individuals with acute liver failure [45]. All these prior publications reported findings in small samples from specific groups (patients with different diseases/conditions). To the best of our knowledge, our work is the first study that confirmed associations of LPC species with total (all-cause) mortality in a large longitudinal study with thousands of participants and repeated metabolomics measurements.

SPM applications illustrated how the observed associations between the LPC species (taking as an example the variant with the strongest association, LPC 15:0/0:0) and mortality found in JM reflect underlying aging-related characteristics that shape the observed age trajectories of LPCs and their impact on mortality risk. In particular, we found that the U-shape of the mortality risk as a function of LPC 15:0/0:0 narrows with age, reflecting an aging-related decline in robustness to deviations of trajectories of LPC 15:0/0:0 from the optimal levels (that is, those minimizing the mortality risk at a given age). That is, the same magnitude of deviation at older ages leads to a larger increase in the risk than at younger ages. The estimated equilibrium levels of LPC 15:0/0:0 differ by sex, with females having higher (more favorable in terms of the mortality risk) levels than males. The equilibrium levels also decline with age, whereas the optimal levels show an increasing pattern with age. As a result, there is an increasing gap between the optimal and equilibrium levels, which leads to an increased mortality risk with age.

One particular advantage of SPM is that it allows evaluating optimal levels (or ranges [16]) of longitudinal outcomes (e.g., LPC species, as in our applications). Such levels/ranges derived from the model can take into account potential confounders and conceptualize the "optimum" as the levels minimizing the mortality risk (or risks of other events of interest, which, hypothetically, can differ). Such optimal levels do not necessarily coincide with average sex-specific levels for particular ages, as our SPM analyses of LPC 15:0/0:0 illustrate. Thus, SPM applications can expand and complement the ongoing efforts to compute reference values of metabolites for different ages and sexes [46] and can provide additional information that can be used in clinical decision-making processes.

Our SPM results are in line with other studies exploring the LPC-aging nexus. Recently, it was found [31] that higher levels of LPC species were associated with slower biological aging (expressed by two DNA methylation-based metrics). In particular, the LPC species with 15 carbons showed the strongest (negative) association with the biological aging metrics in that study. Lower baseline concentrations and faster declines in levels of several LPC species were associated with a faster decline in skeletal muscle mitochondrial function in longitudinal analyses [47]. Impaired mitochondrial oxidative capacity was previously found to be related to lower levels of several LPC species [48]. Older adults with dual declines in memory and speed showed the most extensive alterations (faster decline) in LPC metabolic profiles [49]. Anti-oxidative stress and anti-inflammatory responses have been suggested as potential biological mechanisms that can explain the observed associations of LPCs with slower biological aging [25, 31, 32]. However, as noted in a recent study [50], LPCs can exhibit opposite signatures, both anti-inflammatory and pro-inflammatory, so that their impact on health can be more ambiguous, with potentially pleiotropic or competing roles that may depend on physiological context, comorbidities, or other factors such as age. As our SPM applications indicate, sex can also be a significant factor contributing to various unobserved aging-related characteristics underlying the LPC trajectories and their relations to mortality. Analyzing sex differences in LPCs (and phospholipids in general) in relation to aging in longitudinal cohort studies is of considerable interest and importance because of the paucity of such studies, and, in particular, considering inconsistencies regarding sex differences in LPC levels during aging observed in prior research [51]. Impacts of other factors on the observed relations between LPC trajectories and mortality can be explored using the tools in this paper, including the genetic underpinnings of the relationships that can be evaluated using relevant tools [14, 52].

We acknowledge that our study has limitations. First, we used simple specifications for the models, e.g., linear

functions of age for SPM components. While versatile and flexible, such specifications do not allow exploring more complex non-linear age patterns of respective characteristics. We are limited in our choice by the current availability of repeated measurements of metabolomics in LLFS (up to two per individual). Second, the LLFS is predominantly (>99%) a white sample. Therefore, our findings need confirmation in other studies collecting data for other race/ethnicity groups. Third, we performed analyses of a single metabolite in one-dimensional JM and SPM. While multivariate JM and SPM are available [15, 53], their practical applications in analyses of samples similar in size to this study can be intractable. Relevant dimensionality reduction techniques (e.g., as in our prior works [54, 55]) can be used to mitigate this. Fourth, we used available tools developed for analyses of unrelated samples. Even though sensitivity analyses using the familial bootstrap confirmed the robustness of our results, development and validation of approaches handling relatedness among study participants can benefit future analyses of longitudinal omics data in family-based studies.

MATERIALS AND METHODS

Data

The Long Life Family Study (LLFS) [24] is a family-based, longitudinal study of healthy aging and longevity that enrolled participants at four field centers (three in the US: Boston, New York, Pittsburgh, and one in Denmark). The LLFS recruited 4,953 individuals from two-generational families selected for exceptional familial longevity based on the Family Longevity Selection Score [56]. The first in-person evaluation (Visit 1) was conducted in 2006–2009. The second in-person visit (Visit 2) of surviving participants from Visit 1 and newly enrolled participants was completed in 2014–2017. Visit 3 started in 2020 and is ongoing. The participants provided information on socio-demographic indicators, past and current medical conditions, medication use, and physical and cognitive functioning [24]. Annual telephone follow-ups were conducted to collect updates on participants' vital and health status. All reported deaths were adjudicated by an Adjudication Committee [24]. We used the August 19, 2024 release of the phenotypic LLFS data, with the latest recorded follow-up date on August 7, 2024. Baseline ages were validated using dates of birth from official documents in the US [57] and through the civil registration system in Denmark. Ages at censoring for those alive at the end of the follow-up period were determined from dates of birth and the last follow-up. Ages at death/censoring and an indicator of death were used as time-to-event outcomes in our applications of SPM. In JM applications, time since the baseline was used as the time variable due to the specifics of the JM software used in the analyses.

We used batch 6 (released on October 25, 2023) of LLFS metabolomics data, which provides information on 188 lipid metabolites measured longitudinally in the LLFS participants at Visits 1 and 2. Plasma samples were first processed by using solid-phase extraction kits with both aqueous and organic solvents [58]. Extracted metabolites were then analyzed with liquid chromatography/mass spectrometry (LC/MS). To assess lipid metabolites, reversed-phase chromatography was used in combination with an Agilent 6545 quadrupole time-of-flight mass spectrometer at Washington University in St. Louis. A combination of different tools was used to remove background, annotate adducts, and identify compounds [59–61]. Missing values were imputed using the half-minimum approach (i.e., zeros were replaced by half of the minimum value) [62]. Profiling was performed in batches of approximately 90 samples. Batch correction was accomplished by using a random forest-based batch correction algorithm [63], which outperformed other approaches in the lipid metabolite data [58]. The metabolites were annotated by using standardized names from RefMet, version 07/2023 [64]. We used 23 available lysophosphatidylcholines (LPCs) as the longitudinal outcomes in the analyses described below: LPC 0:0/16:0, LPC 0:0/16:1, LPC 0:0/18:0, LPC 0:0/18:1, LPC 0:0/18:2, LPC 0:0/20:3, LPC 0:0/20:4, LPC 0:0/22:6, LPC 14:0/0:0, LPC 15:0/0:0, LPC 16:0/0:0, LPC 16:1/0:0, LPC 17:0/0:0, LPC 18:0/0:0, LPC 18:1/0:0, LPC 18:2/0:0, LPC 18:3/0:0, LPC 20:2/0:0, LPC 20:3/0:0, LPC 20:4/0:0, LPC 20:5/0:0, LPC 22:5/0:0, LPC 22:6/0:0. Each metabolite was analyzed separately in one-dimensional models. Intensity values were transformed using the inverse-normal transformation [65] (INT) before use in the models.

In total, the LLFS metabolomics sample contains 6,776 measurements of the metabolites (4,221 in Visit 1 and 2,555 in Visit 2). The characteristics of the LLFS metabolomics sample are presented in Supplementary Table 6. Table 3 describes the analytic sample obtained after removal of records with missing information on covariates (see Notes under Supplementary Table 6), which comprised 3,462 participants with 5,066 measurements of each metabolite (3,142 in Visit 1 and 1,924 in Visit 2, with 1,858 participants having one measurement and 1,604 participants with two measurements); 1,245 participants died during the follow-up period.

Joint models: General specifications

The joint models for longitudinal and time-to-event outcomes (or simply joint models, JM) jointly estimate

Table 3. Characteristics of the Long Life Family Study metabolomics subsample used in analyses.

Characteristics	Field center				Total sample
	BU	NY	PT	DK	
Number of families	218	233	210	77	555
Number of participants at any visit	995	601	862	1,004	3,462
Number of participants at visit 1	947	530	809	856	3,142
Number of participants at visit 2	508	313	448	655	1,924
Number (%) of deceased participants	351 (35.3%)	249 (41.4%)	317 (36.8%)	328 (32.7%)	1,245 (36.0%)
Follow-up period (years) (mean ± SD (range))	10.9 ± 4.7 (0.28, 18.22)	10.6 ± 4.3 (0.72, 18.41)	11.2 ± 4.6 (0.29, 18.50)	12.2 ± 5.1 (0.55, 17.93)	11.3 ± 4.8 (0.28, 18.50)
Age at baseline (mean ± SD (range))	69.7 ± 15.6 (32, 104)	73.1 ± 15.4 (24, 101)	70.4 ± 15.5 (38, 104)	69.2 ± 13.9 (38, 104)	70.3 ± 15.2 (24, 104)
Whites (%)	99.6%	98.8%	99.5%	99.5%	99.4%
Females (%)	55.6%	53.7%	55.3%	55.2%	55.1%
Low educated participants (below high school) (%)	5.8%	6.5%	6.8%	28.2%	12.7%
Smokers (smoked >100 cigarettes in lifetime) (%)	41.7%	45.4%	36.5%	49.1%	43.2%
APOE ε4 allele carriers (%)	15.3%	19.1%	18.0%	26.9%	20.0%
Medication use: angina (%)	34.7%	31.8%	32.8%	25.7%	31.1%
Medication use: anti-diabetic (%)	6.7%	5.8%	9.0%	6.0%	6.9%
Medication use: anti-hypertensive (%)	52.5%	54.9%	55.1%	47.9%	52.2%
Medication use: lipid-lowering (%)	35.3%	45.3%	41.4%	24.0%	35.3%
No intense physical activity at baseline (%)	70.3%	66.4%	75.3%	75.8%	72.4%
SPPB total score at baseline (mean ± SD (range))	9.9 ± 3.0 (0, 12)	9.6 ± 2.9 (1, 12)	10.0 ± 2.8 (1, 12)	10.1 ± 3.0 (1, 12)	10.0 ± 2.9 (0, 12)
Prevalence of major diseases at baseline (%)	70.5%	71.2%	70.1%	68.9%	70.0%
BMI at baseline (mean ± SD (range))	27.5 ± 5.0 (17, 57)	26.5 ± 4.0 3 (17, 41)	27.7 ± 5.2 (17, 52)	26.4 ± 4.4 (13, 54)	27.1 ± 4.8 (13, 57)

(a) The numbers shown in "Number of deceased participants" and "Follow-up period" correspond to the LLFS data release used in this paper (see Data); (b) "Prevalence of major diseases" includes prevalence of any of the following diseases: heart disease, stroke, lung disease, cancer, hypertension, and diabetes. Abbreviations: BMI: body mass index; BU: Boston; DK: Denmark; NY: New York; PT: Pittsburgh; SD: standard deviation; SPPB: Short Physical Performance Battery.

the parameters of the longitudinal and time-to-event outcomes in the estimation procedure. We used the basic form of JM [4, 5] as implemented in the R-package *JM* [33]. In our applications, JM jointly estimates the parameters of the longitudinal trajectories of LPCs and mortality rates. The JM consists of two sub-models. The survival sub-model of JM represents the mortality rate as a function of a metabolite and other covariates:

$$h_i(t|M_i(t), w_i) = h_0(t)\exp\{\gamma^T w_i + \alpha m_i(t)\}, \quad \text{(Eq. 1)}$$

where $h_i(t|\cdot)$ is the mortality rate for the i^{th} individual at time point t, $m_i(t)$ is the "true" (i.e., unobserved) LPC level (see below in the text after Eq. 2) at time t, $h_0(\cdot)$ is the baseline mortality rate, w_i is a vector of baseline covariates, and γ is a vector of respective regression coefficients.

We highlight the main parameter of our interest that we report in respective tables, α (a scalar), which is called the association parameter. In JM, the association parameter quantifies the relationship between longitudinal data (repeated measures of (transformed) LPC levels in our case) and time-to-event data (survival times in our application). This parameter is crucial because it helps us understand how changes in the longitudinal measurements are associated with the risk of the respective event. Positive values of α indicate that higher values of the longitudinal measurements are associated with an increased risk of the event. Respectively, negative values of α imply that higher values of the longitudinal measurements are associated with a decreased risk of the event. A zero association parameter suggests no relationship between the longitudinal measurements and the risk of the event (i.e., changes in the longitudinal measurements do not affect the event's likelihood). We also report hazard ratios (HRs) computed from the association parameter as $HR = exp(\alpha)$. They are interpreted similarly to HRs from the Cox model, e.g., HR = 1.5 means that for each unit increase in the

longitudinal measurement (i.e., (transformed) LPC level), the risk of the event (mortality) increases by 50%, and HR = 0.7 means that for each unit increase in the longitudinal measurement, the risk of the event decreases by 30%.

The longitudinal sub-model of JM in its basic form [4, 5, 33] is a linear mixed effects model that describes changes in the longitudinal outcome ((transformed) LPC levels) as a function of time and other covariates:

$$y_i(t) = m_i(t) + \varepsilon_i(t) = x_i^T(t)\beta + z_i^T(t)b_i + \varepsilon_i(t), \qquad \text{(Eq. 2)}$$

where $y_i(t)$ is the observed LPC level at time point t in the i^{th} individual, $x_i(t)$ and $z_i(t)$ are the corresponding fixed and random effects, β and b_i are the respective vectors of parameters (that model population- and individual-level characteristics of LPC trajectories, respectively), and $\varepsilon_i(t)$ is the error term (independent of b_i), normally distributed with zero mean and variance σ^2. The difference between the observed value $y_i(t)$ and the error term $\varepsilon_i(t)$, $m_i(t) = x_i^T(t)\beta + z_i^T(t)b_i$, represents the "true" LPC level included in Eq. 1.

In addition to the general form of JM as in Eqs. 1–2, we used the JM versions where the association between the longitudinal and time-to-event outcomes is captured by a latent Gaussian process [34–36], as implemented in the R-package *joineR*. These models (also known as the joint models with shared random effects, JM-SRE) allow for different specifications of associations of individual changes of metabolites with the mortality rate. The general formula for the longitudinal part is as in Eq. 2, but the expression for the hazard rate differs from Eq. 1:

$$h_i(t \mid b_i, w_i) = h_0(t)\exp\{\gamma^T w_i + D_i(t)(\alpha^T b_i)\}, \qquad \text{(Eq. 3)}$$

where α is a vector of association parameters corresponding to random effects b_i and $D_i(t)$ is the corresponding design matrix. We used two specifications of the JM: the intercept model ("int") and the intercept and slope model ("intslope"). In the latter case, we used the option "*sepassoc=TRUE*" in function *joint* from the R-package *joineR*. In the "int" model,

$$D_i(t)(\alpha^T b_i) = \alpha_0 b_{0i}, \qquad \text{(Eq. 4)}$$

that is, the individual intercept of LPC (representing individual differences in baseline levels of LPC) enters the hazard rate, and in the "intslope" model,

$$D_i(t)(\alpha^T b_i) = \alpha_0 b_{0i} + \alpha_1 b_{1i} t, \qquad \text{(Eq. 5)}$$

i.e., the individual intercepts and slopes of LPC (representing individual differences in both baseline levels of LPC and rates of change in LPC levels over time) are both tested for their association with the mortality rate. We report the association parameters α_0 and α_1 in respective tables, which are interpreted similarly to the association parameters in the base JM (see above), for respective variables (individual intercepts and slopes of LPC). We also compute HRs per standard deviation (SD) of the corresponding variables, which are interpreted accordingly (e.g., if $\alpha_0 = -0.25$ and SD of b_{0i} is 1.42, then we would have $HR = exp\ (-0.25 \times 1.42) \approx 0.7$, meaning that for each SD increase in the baseline levels of LPC, the mortality risk decreases by about 30%).

Joint models: Specific versions used in applications

In our applications of the basic form of JM [33], the longitudinal trajectories of different LPC species were modeled by a linear mixed effects model (Eq. 2) with linear random effects, i.e., random intercept and random slope. Time since the baseline visit was used as a time variable (as implemented in the R-package *JM*). Additional covariates were included in the fixed effects part of the longitudinal sub-model of JM: sex (1: male, 0: female), age at baseline visit, country (1: Denmark, 0: USA), education (1: below high school, 0: otherwise), smoking (smoked >100 cigarettes in lifetime: yes (1)/no (0)), medication use (anti-diabetic, lipid-lowering, anti-hypertensive, heart disease) (1: used, 0: not used), APOE4 (1: carriers of apolipoprotein E (*APOE*) ε4 allele; 0 – non-carriers of ε4). The medications listed above include all available groups constructed by the LLFS investigators from original medication records using the Anatomical Therapeutic Chemical Classification System codes.

The time-to-event outcome (i.e., the mortality rate) was modeled as in Eq. 1, with the following (time-independent) covariates: sex, age at baseline, country, education, smoking, APOE4, physical activity at baseline (1: no intense physical activity at baseline; 0: intense physical activity), Short Physical Performance Battery (SPPB) total score, body mass index (BMI) at baseline, and prevalence of heart disease, stroke, lung disease, cancer, hypertension, or diabetes at baseline (1: prevalence of any of the diseases, 0: no prevalence). In addition, two genetic principal components (PCs) were included as covariates in the hazard rate (we tested models with different numbers of PCs and the results were similar; data not shown). The baseline mortality rate $h_0(t)$ was modeled by a piecewise constant function. The pseudo-adaptive Gauss-Hermite quadrature rule [66] was used to approximate the required integrals in the estimation procedure. Analyses were performed in

the total sample and separately for females and males (in which case sex was not included as a covariate).

In the specification of JM implemented in the R-package *joineR*, we used the same list of covariates in the longitudinal and survival sub-models. Unlike the *JM* package, the baseline hazard is represented semi-parametrically in *joineR*. Individual values of random intercepts and slopes were used in the expression of the hazard rate as shown in Eqs. 3–5, instead of the "true" level of the metabolite as in Eq. 1.

R version 4.4.1 was used to run the R-packages *JM* (version 1.5-2) and *joineR* (version 1.2.8) estimating respective models.

Stochastic process models: General specifications

For SPM applications, we used a one-dimensional version with time-dependent components [15]. The age trajectory of a repeatedly measured variable $Y(t,c)$, where t is age and c denotes covariates, is represented as a stochastic process with the following equation (in our applications, this equation models age trajectories of (transformed) LPC):

$$dY(t,c) = a(t,c)(Y(t,c) - f_1(t,c))\,dt + b(t,c)dW(t), \quad \text{(Eq. 6)}$$

with initial condition $Y(t_0, c)$ (t_0 denotes age at entering the study). Note that SPM equations represent individual trajectories/rates; we do not use an index to indicate that t, c, and $Y(t, c)$ are individual-based quantities, for simplicity of notation and visualization. Here $W(t)$ is the stochastic (Wiener) process (assumed to be independent of $Y(t_0, c)$) that defines random paths of $Y(t, c)$, $b(t, c)$ is the volatility coefficient controlling the volatility of $Y(t, c)$, $f_1(t, c)$ is the long-term mean of the stochastic process (or the equilibrium trajectory), and $a(t, c)$ is the negative feedback coefficient regulating how fast the trajectory of $Y(t, c)$ returns to $f_1(t, c)$ when it deviates from it. The SPM expresses the hazard rate (i.e., the mortality rate in our case) as a function of age (t), the vector of covariates (c), and the value of the longitudinal variable $Y(t, c)$:

$$\mu\big(t,c,Y(t,c)\big) = \mu_0(t,c) + Q(t,c)\big(Y(t,c) - f_0(t,c)\big)^2. \quad \text{(Eq. 7)}$$

Here $\mu_0(t,c)$ is the baseline hazard (i.e., mortality in our case) rate, $Q(t,c)$ is the multiplier scaling the quadratic term of the hazard at different ages and values of covariates, and $f_0(t,c)$ represents the values of $Y(t,c)$ (i.e., LPC) minimizing the risk (mortality) at age t and covariate values c.

The main characteristic feature of SPM is that Eqs. 6–7 embed several aging-related concepts (see more details in [16, 17]), thus facilitating more detailed analyses and interpretation of results in the context of aging, compared to analyses by JM (see more details in Supplementary Text Stochastic process models: Interpretation and illustration of components, parameters, and related null hypotheses; Supplementary Figures 2–9 provide further illustration of the model setup and mechanics): (a) *homeostatic regulation*, which is a fundamental feature of a living organism; (b) *allostasis and mean allostatic ("equilibrium") levels* ($f_1(t,c)$), featuring the effect of allostatic adaptation [67], i.e., the LPC levels forced by organism's regulatory systems functioning at non-optimal levels; (c) *adaptive capacity* ($a(t,c)$), modeling the rate of adaptive response (associated with *biological resilience* [17, 41, 68]) to any factors causing deviations of $Y(t,c)$ from its dynamic equilibrium levels $f_1(t,c)$; (d) *physiological or biological optima ("sweet spots"* [38–40]) naturally represented by $f_0(t,c)$; (e) *vulnerability component of stress resistance* (associated with *biological robustness* [68–71]) captured by the U-(J-)shape of the hazard and regulated by the multiplier $Q(t,c)$ in the quadratic part of the hazard; (f) *allostatic load (AL)* computed as $AL(t,c) = |f_0(t,c) - f_1(t,c)|$ and representing the practical realization of the theoretical concept of AL suggested in the literature [67, 72–74] (the larger the value of this AL measure, the greater the price or load is for an organism in terms of an increased mortality risk, compared to the best-case scenario where LPC levels follow the optimal trajectory $f_0(t,c)$, i.e., when AL = 0).

Stochastic process models: Specific parameterizations used in applications

For applications, we used the following specification of SPM: (a) the Gompertz baseline hazard (represents a common pattern of mortality rate at adult and old ages): $\ln \mu_0(t,c) = \ln a_{\mu_0} + b_{\mu_0}(t - t_{\min}) + \beta_{\mu_0} c$; (b) constant volatility coefficient (based on our prior simulations showing the best accuracy of parameter estimates for models with constant $b(.)$ [15]): $b(t,c) = \sigma_1 + \beta_W c$; and (c) linear functions of age for other components (to estimate age trends in the respective components): $a(t,c) = a_Y + b_Y(t - t_{\min}) + \beta_Y c$, where $a_Y < 0$, $b_Y \geq 0$ and $t_{\min} = 50$; $f_o(t,c) = a_{f_0} + b_{f_0}(t - t_{\min}) + \beta_{f_0} c$; $f_1(t,c) = a_{f_1} + b_{f_1}(t - t_{\min}) + \beta_{f_1} c$; $Q(t,c) = a_Q + b_Q(t - t_{\min}) + \beta_Q c$; and $Y(t_0,c) \sim N(f_1(t_0,c), \sigma_0^2)$. Based on our prior experience (dictated by technical complexities of the estimation algorithm), we included all covariates used in JM (except age at baseline because age is used as the time

variable in SPM) in $\mu_0(t,c)$, whereas only one covariate (sex) was included in all other components.

In-house MATLAB codes (run in MATLAB version R2024a) implementing estimation algorithms with covariates in discrete-time approximations of SPM [18, 55] were used to estimate SPM parameters. Likelihood ratio tests were used to test several null hypotheses (H0s) about the functional forms of each model's components (i.e., to test whether they depend on age and on respective covariates, or are non-zero). First, an "unrestricted" model (with the parameterization presented above) with no restrictions on parameters was estimated. Then, other models that contain one or more restrictions on parameters were estimated to test respective H0s:

1. H0: $Q(t,c)$ (Qzero; interpretation: no quadratic term in the hazard);
2. H0: $Q(t,c) = Q(c)$, i.e., $b_Q = 0$ (QnoT; the term in the quadratic part of the hazard does not depend on age, i.e., robustness to deviations of LPC from the optimal trajectory does not depend on age);
3. H0: $Q(t,c) = Q(t)$, i.e., $\beta_Q = 0$ (QnoC; the term in the quadratic part of the hazard does not depend on sex, i.e., robustness to deviations of LPC from the optimal trajectory does not depend on sex);
4. H0: $a(t,c) = a(c)$, i.e., $b_Y = 0$ (AnoT; the feedback coefficient does not depend on age, i.e., the adaptive capacity (resilience) is age-independent);
5. H0: $a(t,c) = a(t)$, i.e., $\beta_Y = 0$ (AnoC; the feedback coefficient does not depend on sex, i.e., the adaptive capacity (resilience) is the same in females and males);
6. H0: $b(t,c) = b(t)$, i.e., $\beta_W = 0$ (BnoC; the volatility coefficient does not depend on sex);
7. H0: $f_1(t,c) = f_1(c)$, i.e., $b_{f_1} = 0$ (F1noT; equilibrium LPC levels are the same for all ages);
8. H0: $f_1(t,c) = f_1(t)$, i.e., $\beta_{f_1} = 0$ (F1noC; equilibrium LPC levels do not differ by sex);
9. H0: $f_0(t,c) = f_0(c)$, i.e., $b_{f_0} = 0$ (F0noT; LPC "sweet spots" are the same for all ages);
10. H0: $f_0(t,c) = f_0(t)$, i.e., $\beta_{f_0} = 0$ (F0noC; optimal LPC levels minimizing mortality risk coincide for females and males);
11. H0: $f_1(t,c) = f_0(t,c)$, i.e., $AL(t,c) = 0$ (ALzero, zero AL);
12. H0: $f_1(t,c) = f_1(c)$ and $f_0(t,c) = f_0(c)$, i.e., $\beta_{f_1} = 0$ and $b_{f_0} = 0$, or $AL(t, c) = AL(c)$ (ALnoT, AL does not accumulate with age).

Sensitivity analyses

The LLFS is a family-based study that contains related individuals. Currently available JM tools allowing analyses of related individuals (R-packages *merlin* and *rstanarm*) were not usable for our applications because of technical issues. SPM tools for related samples currently do not exist. Therefore, we used the available tools for unrelated individuals. To test whether this could affect our results, we performed sensitivity analyses implementing the "familial bootstrap" approach [37]. Specifically, we collected estimates of the JM from 100 bootstrap samples constructed from data on the families generated (with replacement) from the original analytic sample (note that, even though the number of families in each generated sample was the same, the numbers of individuals were different). Then, we computed relevant quantities from all 100 samples (e.g., medians of hazard ratios of the association parameter α (Eq. 1) in JM, along with the range of the hazard ratios). The respective estimates are provided in Supplementary Materials and discussed in Results.

Abbreviations

AL: allostatic load; *APOE*: apolipoprotein E; BMI: body mass index; CI: confidence interval(s); H0s: null hypotheses; HR: hazard ratio(s); JM: joint model(s); JM-SRE: joint model(s) with shared random effects; LLFS: Long Life Family Study; LPC: lyso-phosphatidylcholine(s); PCs: principal components; SPM: stochastic process model(s); SPPB: Short Physical Performance Battery.

AUTHOR CONTRIBUTIONS

KGA: conceived and designed the study, contributed to statistical analyses, and wrote the manuscript; OB: prepared phenotypic analytic files, performed statistical analyses, prepared tables and figures, and contributed to writing Materials and Methods section; SVU, AK, ES, YG, AIY, MAP: provided feedback on analyses and results; MAP: provided funding; MSH, GJP: generated and cleaned metabolomics data. All authors provided critical comments on the final version of the manuscript.

CONFLICTS OF INTEREST

The authors declare there are no conflicts of interest related to this study.

ETHICAL STATEMENT AND CONSENT

Written informed consent was obtained from all subjects following protocols approved in the US by the

respective field center's Institutional Review Boards (IRBs) and, in Denmark, by the Regional Committees on Health Research Ethics for Southern Denmark. In this paper, we performed secondary analyses of LLFS data collected at all field centers. This study was approved by the Duke University Health System IRB, protocol number is Pro00103124.

FUNDING

Research reported in this publication was supported by the National Institute on Aging of the National Institutes of Health (NIA/NIH) under Award Number U19AG063893. This content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

1. Therneau T, Grambsch P. Modeling Survival Data: Extending the Cox Model. New York: Springer-Verlag. 2000. https://doi.org/10.1007/978-1-4757-3294-8

2. Prentice RL. Covariate measurement errors and parameter estimation in a failure time regression model. Biometrika. 1982; 69:331–42. https://doi.org/10.1093/biomet/69.2.331

3. Sweeting MJ, Thompson SG. Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. Biom J. 2011; 53:750–63. https://doi.org/10.1002/bimj.201100052 PMID:21834127

4. Rizopoulos D. Joint Models for Longitudinal and Time-to-Event Data With Applications in R. Boca Raton, FL: Chapman and Hall/CRC. 2012.

5. Elashoff RM, Li G, Li N. Joint Modeling of Longitudinal and Time-to-Event Data. Boca Raton, FL: CRC Press. 2016.

6. Klimczak-Tomaniak D, de Bakker M, Bouwens E, Akkerhuis KM, Baart S, Rizopoulos D, Mouthaan H, van Ramshorst J, Germans T, Constantinescu A, Manintveld O, Umans V, Boersma E, Kardys I. Dynamic personalized risk prediction in chronic heart failure patients: a longitudinal, clinical investigation of 92 biomarkers (Bio-SHiFT study). Sci Rep. 2022; 12:2795. https://doi.org/10.1038/s41598-022-06698-3 PMID:35181700

7. de Bakker M, Petersen TB, Rueten-Budde AJ, Akkerhuis KM, Umans VA, Brugts JJ, Germans T, Reinders MJT, Katsikis PD, van der Spek PJ, Ostroff R, She R, Lanfear D, et al. Machine learning-based biomarker profile derived from 4210 serially measured proteins predicts clinical outcome of patients with heart failure. Eur Heart J Digit Health. 2023; 4:444–54. https://doi.org/10.1093/ehjdh/ztad056 PMID:38045440

8. de Bakker M, Loncq de Jong M, Petersen T, de Lange I, Akkerhuis KM, Umans VA, Rizopoulos D, Boersma E, Brugts JJ, Kardys I. Sex-specific cardiovascular protein levels and their link with clinical outcome in heart failure. ESC Heart Fail. 2024; 11:594–600. https://doi.org/10.1002/ehf2.14578 PMID:38009274

9. Woodbury MA. A random-walk model of human mortality and aging. Theor Popul Biol. 1977; 11:37–48. https://doi.org/10.1016/0040-5809(77)90005-3 PMID:854860

10. Yashin AI, Manton KG, Vaupel JW. Mortality and aging in a heterogeneous population: a stochastic process model with observed and unobserved variables. Theor Popul Biol. 1985; 27:154–75. https://doi.org/10.1016/0040-5809(85)90008-5 PMID:4023952

11. Yashin AI, Manton KG, Stallard E. The propagation of uncertainty in human mortality processes operating in stochastic environments. Theor Popul Biol. 1989; 35:119–41. https://doi.org/10.1016/0040-5809(89)90013-0 PMID:2727949

12. Yashin AI, Arbeev KG, Akushevich I, Kulminski A, Akushevich L, Ukraintseva SV. Stochastic model for analysis of longitudinal data on aging and mortality. Math Biosci. 2007; 208:538–51. https://doi.org/10.1016/j.mbs.2006.11.006 PMID:17300818

13. Arbeev KG, Ukraintseva SV, Akushevich I, Kulminski AM, Arbeeva LS, Akushevich L, Culminskaya IV, Yashin AI. Age trajectories of physiological indices in relation to healthy life course. Mech Ageing Dev. 2011; 132:93–102. https://doi.org/10.1016/j.mad.2011.01.001 PMID:21262255

14. Arbeev KG, Akushevich I, Kulminski AM, Arbeeva LS, Akushevich L, Ukraintseva SV, Culminskaya IV, Yashin AI. Genetic model for longitudinal studies of aging, health, and longevity and its potential application to incomplete data. J Theor Biol. 2009; 258:103–11. https://doi.org/10.1016/j.jtbi.2009.01.023 PMID:19490866

15. Yashin AI, Arbeev KG, Akushevich I, Kulminski A, Ukraintseva SV, Stallard E, Land KC. The quadratic

hazard model for analyzing longitudinal data on aging, health, and the life span. Phys Life Rev. 2012; 9:177–88.
https://doi.org/10.1016/j.plrev.2012.05.002
PMID:22633776

16. Arbeev KG, Ukraintseva SV, Yashin AI. Dynamics of biomarkers in relation to aging and mortality. Mech Ageing Dev. 2016; 156:42–54.
https://doi.org/10.1016/j.mad.2016.04.010
PMID:27138087

17. Arbeev KG, Bagley O, Yashkin AP, Duan H, Akushevich I, Ukraintseva SV, Yashin AI. Understanding Alzheimer's disease in the context of aging: Findings from applications of stochastic process models to the Health and Retirement Study. Mech Ageing Dev. 2023; 211:111791.
https://doi.org/10.1016/j.mad.2023.111791
PMID:36796730

18. Arbeev KG, Ukraintseva SV, Bagley O, Zhbannikov IY, Cohen AA, Kulminski AM, Yashin AI. "Physiological Dysregulation" as a Promising Measure of Robustness and Resilience in Studies of Aging and a New Indicator of Preclinical Disease. J Gerontol A Biol Sci Med Sci. 2019; 74:462–8.
https://doi.org/10.1093/gerona/gly136
PMID:29939206

19. Yashin AI, Arbeev KG, Akushevich I, Ukraintseva SV, Kulminski A, Arbeeva LS, Culminskaya I. Exceptional survivors have lower age trajectories of blood glucose: lessons from longitudinal data. Biogerontology. 2010; 11:257–65.
https://doi.org/10.1007/s10522-009-9243-1
PMID:19644762

20. Yashin AI, Arbeev KG, Kulminski A, Akushevich I, Akushevich L, Ukraintseva SV. Health decline, aging and mortality: how are they related? Biogerontology. 2007; 8:291–302.
https://doi.org/10.1007/s10522-006-9073-3
PMID:17242962

21. Yashin AI, Arbeev KG, Ukraintseva SV, Akushevich I, Kulminski A. Patterns of aging related changes on the way to 100: An approach to studying aging, mortality, and longevity from longitudinal data. N Amer Actuarial J. 2012; 16:403–33.
https://doi.org/10.1080/10920277.2012.10597640

22. Yashin AI, Arbeev KG, Wu D, Arbeeva L, Kulminski A, Kulminskaya I, Akushevich I, Ukraintseva SV. How Genes Modulate Patterns of Aging-Related Changes on the Way to 100: Biodemographic Models and Methods in Genetic Analyses of Longitudinal Data. N Am Actuar J. 2016; 20:201–32.
https://doi.org/10.1080/10920277.2016.1178588
PMID:27773987

23. Yashin AI, Ukraintseva SV, Arbeev KG, Akushevich I, Arbeeva LS, Kulminski AM. Maintaining physiological state for exceptional survival: What is the normal level of blood glucose and does it change with age? Mech Ageing Dev. 2009; 130:611–8.
https://doi.org/10.1016/j.mad.2009.07.004
PMID:19635493

24. Wojczynski MK, Jiuan Lin S, Sebastiani P, Perls TT, Lee J, Kulminski A, Newman A, Zmuda JM, Christensen K, Province MA. NIA Long Life Family Study: Objectives, Design, and Heritability of Cross-Sectional and Longitudinal Phenotypes. J Gerontol A Biol Sci Med Sci. 2022; 77:717–27.
https://doi.org/10.1093/gerona/glab333
PMID:34739053

25. Law SH, Chan ML, Marathe GK, Parveen F, Chen CH, Ke LY. An Updated Review of Lysophosphatidylcholine Metabolism in Human Diseases. Int J Mol Sci. 2019; 20:1149.
https://doi.org/10.3390/ijms20051149
PMID:30845751

26. Barupal DK, Baillie R, Fan S, Saykin AJ, Meikle PJ, Arnold M, Nho K, Fiehn O, Kaddurah-Daouk R, and Alzheimer Disease Metabolomics Consortium. Sets of coregulated serum lipids are associated with Alzheimer's disease pathophysiology. Alzheimers Dement (Amst). 2019; 11:619–27.
https://doi.org/10.1016/j.dadm.2019.07.002
PMID:31517024

27. Diray-Arce J, Conti MG, Petrova B, Kanarek N, Angelidou A, Levy O. Integrative Metabolomics to Identify Molecular Signatures of Responses to Vaccines and Infections. Metabolites. 2020; 10:492.
https://doi.org/10.3390/metabo10120492
PMID:33266347

28. Semba RD. Perspective: The Potential Role of Circulating Lysophosphatidylcholine in Neuroprotection against Alzheimer Disease. Adv Nutr. 2020; 11:760–72.
https://doi.org/10.1093/advances/nmaa024
PMID:32190891

29. Peña-Bautista C, Álvarez-Sánchez L, Roca M, García-Vallés L, Baquero M, Cháfer-Pericás C. Plasma Lipidomics Approach in Early and Specific Alzheimer's Disease Diagnosis. J Clin Med. 2022; 11:5030.
https://doi.org/10.3390/jcm11175030
PMID:36078960

30. Dorninger F, Moser AB, Kou J, Wiesinger C, Forss-Petter S, Gleiss A, Hinterberger M, Jungwirth S, Fischer P, Berger J. Alterations in the Plasma Levels of Specific Choline Phospholipids in Alzheimer's Disease Mimic Accelerated Aging. J Alzheimers Dis. 2018; 62:841–54.

https://doi.org/10.3233/JAD-171036
PMID:29480199

31. Liu D, Aziz NA, Landstra EN, Breteler MMB. The lipidomic correlates of epigenetic aging across the adult lifespan: A population-based study. Aging Cell. 2023; 22:e13934.
https://doi.org/10.1111/acel.13934
PMID:37496173

32. Knuplez E, Marsche G. An Updated Review of Pro-and Anti-Inflammatory Properties of Plasma Lysophosphatidylcholines in the Vascular System. Int J Mol Sci. 2020; 21:4501.
https://doi.org/10.3390/ijms21124501
PMID:32599910

33. Rizopoulos D. JM: An R Package for the Joint Modelling of Longitudinal and Time-to-Event Data. J Stat Softw. 2010; 35:1–33.
https://doi.org/10.18637/jss.v035.i09

34. Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. Biometrics. 1997; 53:330–9.
PMID:9147598

35. Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. Biostatistics. 2000; 1:465–80.
https://doi.org/10.1093/biostatistics/1.4.465
PMID:12933568

36. Henderson R, Diggle P, Dobson A. Identification and efficacy of longitudinal markers for survival. Biostatistics. 2002; 3:33–50.
https://doi.org/10.1093/biostatistics/3.1.33
PMID:12933622

37. Borecki IB, Province MA. Genetic and genomic discovery using family studies. Circulation. 2008; 118:1057–63.
https://doi.org/10.1161/CIRCULATIONAHA.107.714592
PMID:18765388

38. Johnson JD, Alejandro EU. Control of pancreatic beta-cell fate by insulin signaling: The sweet spot hypothesis. Cell Cycle. 2008; 7:1343–7.
https://doi.org/10.4161/cc.7.10.5865
PMID:18418065

39. Lee JH, Han K, Huh JH. The sweet spot: fasting glucose, cardiovascular disease, and mortality in older adults with diabetes: a nationwide population-based study. Cardiovasc Diabetol. 2020; 19:44.
https://doi.org/10.1186/s12933-020-01021-8
PMID:32238157

40. Vishnyakova O, Song X, Rockwood K, Elliott LT, Brooks-Wilson A. Physiological phenotypes have optimal values relevant to healthy aging: sweet spots deduced from the Canadian Longitudinal Study on Aging. Geroscience. 2024; 46:1589–605.
https://doi.org/10.1007/s11357-023-00895-2
PMID:37688655

41. Ukraintseva S, Arbeev K, Duan M, Akushevich I, Kulminski A, Stallard E, Yashin A. Decline in biological resilience as key manifestation of aging: Potential mechanisms and role in health and longevity. Mech Ageing Dev. 2021; 194:111418.
https://doi.org/10.1016/j.mad.2020.111418
PMID:33340523

42. Li X, Yin Z, Yan W, Wang M, Chang C, Guo C, Xue L, Zhou Q, Sun Y. Association between Changes in Plasma Metabolism and Clinical Outcomes of Sepsis. Emerg Med Int. 2023; 2023:2590115.
https://doi.org/10.1155/2023/2590115
PMID:37346225

43. Trovato FM, Zia R, Napoli S, Wolfer K, Huang X, Morgan PE, Husbyn H, Elgosbi M, Lucangeli M, Miquel R, Wilson I, Heaton ND, Heneghan MA, et al. Dysregulation of the Lysophosphatidylcholine/Autotaxin/Lysophosphatidic Acid Axis in Acute-on-Chronic Liver Failure Is Associated With Mortality and Systemic Inflammation by Lysophosphatidic Acid-Dependent Monocyte Activation. Hepatology. 2021; 74:907–25.
https://doi.org/10.1002/hep.31738
PMID:33908067

44. Banoei MM, Vogel HJ, Weljie AM, Yende S, Angus DC, Winston BW. Plasma lipid profiling for the prognosis of 90-day mortality, in-hospital mortality, ICU admission, and severity in bacterial community-acquired pneumonia (CAP). Crit Care. 2020; 24:461.
https://doi.org/10.1186/s13054-020-03147-3
PMID:32718333

45. Trovato FM, Zia R, Artru F, Mujib S, Jerome E, Cavazza A, Coen M, Wilson I, Holmes E, Morgan P, Singanayagam A, Bernsmeier C, Napoli S, et al. Lysophosphatidylcholines modulate immunoregulatory checkpoints in peripheral monocytes and are associated with mortality in people with acute liver failure. J Hepatol. 2023; 78:558–73.
https://doi.org/10.1016/j.jhep.2022.10.031
PMID:36370949

46. Tian Q, Adam MG, Ozcariz E, Fantoni G, Shehadeh NM, Turek LM, Collingham VL, Kaileh M, Moaddel R, Ferrucci L. Human Metabolome Reference Database in a Biracial Cohort across the Adult Lifespan. Metabolites. 2023; 13:591.
https://doi.org/10.3390/metabo13050591
PMID:37233632

47. Tian Q, Mitchell BA, Zampino M, Ferrucci L. Longitudinal associations between blood lyso-phosphatidylcholines and skeletal muscle mitochondrial function. Geroscience. 2022; 44:2213–21.
https://doi.org/10.1007/s11357-022-00548-w
PMID:35389191

48. Semba RD, Zhang P, Adelnia F, Sun K, Gonzalez-Freire M, Salem N Jr, Brennan N, Spencer RG, Fishbein K, Khadeer M, Shardell M, Moaddel R, Ferrucci L. Low plasma lysophosphatidylcholines are associated with impaired mitochondrial oxidative capacity in adults in the Baltimore Longitudinal Study of Aging. Aging Cell. 2019; 18:e12915.
https://doi.org/10.1111/acel.12915
PMID:30719830

49. Tian Q, Shardell MD, Kuo PL, Tanaka T, Simonsick EM, Moaddel R, Resnick SM, Ferrucci L. Plasma metabolomic signatures of dual decline in memory and gait in older adults. Geroscience. 2023; 45:2659–67.
https://doi.org/10.1007/s11357-023-00792-8
PMID:37052768

50. Hornburg D, Wu S, Moqri M, Zhou X, Contrepois K, Bararpour N, Traber GM, Su B, Metwally AA, Avina M, Zhou W, Ubellacker JM, Mishra T, et al. Dynamic lipidome alterations associated with human health, disease and ageing. Nat Metab. 2023; 5:1578–94.
https://doi.org/10.1038/s42255-023-00880-1
PMID:37697054

51. Mohammadzadeh Honarvar N, Zarezadeh M, Molsberry SA, Ascherio A. Changes in plasma phospholipids and sphingomyelins with aging in men and women: A comprehensive systematic review of longitudinal cohort studies. Ageing Res Rev. 2021; 68:101340.
https://doi.org/10.1016/j.arr.2021.101340
PMID:33839333

52. He L, Zhbannikov I, Arbeev KG, Yashin AI, Kulminski AM. A genetic stochastic process model for genome-wide joint analysis of biomarker dynamics and disease susceptibility with longitudinal data. Genet Epidemiol. 2017; 41:620–35.
https://doi.org/10.1002/gepi.22058
PMID:28636232

53. Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R. Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. BMC Med Res Methodol. 2016; 16:117.
https://doi.org/10.1186/s12874-016-0212-5
PMID:27604810

54. Arbeev KG, Bagley O, Ukraintseva SV, Wu D, Duan H, Kulminski AM, Stallard E, Christensen K, Lee JH, Thyagarajan B, Zmuda JM, Yashin AI. Genetics of physiological dysregulation: findings from the long life family study using joint models. Aging (Albany NY). 2020; 12:5920–47.
https://doi.org/10.18632/aging.102987
PMID:32235003

55. Arbeev KG, Cohen AA, Arbeeva LS, Milot E, Stallard E, Kulminski AM, Akushevich I, Ukraintseva SV, Christensen K, Yashin AI. Optimal Versus Realized Trajectories of Physiological Dysregulation in Aging and Their Relation to Sex-Specific Mortality Risk. Front Public Health. 2016; 4:3.
https://doi.org/10.3389/fpubh.2016.00003
PMID:26835445

56. Sebastiani P, Hadley EC, Province M, Christensen K, Rossi W, Perls TT, Ash AS. A family longevity selection score: ranking sibships by their longevity, size, and availability for study. Am J Epidemiol. 2009; 170:1555–62.
https://doi.org/10.1093/aje/kwp309
PMID:19910380

57. Elo IT, Mykyta L, Sebastiani P, Christensen K, Glynn NW, Perls T. Age validation in the long life family study through a linkage to early-life census records. J Gerontol B Psychol Sci Soc Sci. 2013; 68:580–5.
https://doi.org/10.1093/geronb/gbt033
PMID:23704206

58. Stancliffe E, Schwaiger-Haber M, Sindelar M, Murphy MJ, Soerensen M, Patti GJ. An Untargeted Metabolomics Workflow that Scales to Thousands of Samples for Population-Based Studies. Anal Chem. 2022; 94:17370–8.
https://doi.org/10.1021/acs.analchem.2c01270
PMID:36475608

59. Mahieu NG, Spalding JL, Gelman SJ, Patti GJ. Defining and Detecting Complex Peak Relationships in Mass Spectral Data: The Mz.unity Algorithm. Anal Chem. 2016; 88:9037–46.
https://doi.org/10.1021/acs.analchem.6b01702
PMID:27513885

60. Cho K, Schwaiger-Haber M, Naser FJ, Stancliffe E, Sindelar M, Patti GJ. Targeting unique biological signals on the fly to improve MS/MS coverage and identification efficiency in metabolomics. Anal Chim Acta. 2021; 1149:338210.
https://doi.org/10.1016/j.aca.2021.338210
PMID:33551064

61. Stancliffe E, Schwaiger-Haber M, Sindelar M, Patti GJ. DecoID improves identification rates in metabolomics through database-assisted MS/MS deconvolution. Nat Methods. 2021; 18:779–87.
https://doi.org/10.1038/s41592-021-01195-3
PMID:34239103

62. Jin Z, Kang J, Yu T. Missing value imputation for LC-MS metabolomics data by incorporating metabolic network and adduct ion relations. Bioinformatics. 2018; 34:1555–61. https://doi.org/10.1093/bioinformatics/btx816 PMID:29272352

63. Fan S, Kind T, Cajka T, Hazen SL, Tang WHW, Kaddurah-Daouk R, Irvin MR, Arnett DK, Barupal DK, Fiehn O. Systematic Error Removal Using Random Forest for Normalizing Large-Scale Untargeted Lipidomics Data. Anal Chem. 2019; 91:3590–6. https://doi.org/10.1021/acs.analchem.8b05592 PMID:30758187

64. Fahy E, Subramaniam S. RefMet: a reference nomenclature for metabolomics. Nat Methods. 2020; 17:1173–4. https://doi.org/10.1038/s41592-020-01009-y PMID:33199890

65. Yang J, Loos RJ, Powell JE, Medland SE, Speliotes EK, Chasman DI, Rose LM, Thorleifsson G, Steinthorsdottir V, Mägi R, Waite L, Smith AV, Yerges-Armstrong LM, et al. FTO genotype is associated with phenotypic variability of body mass index. Nature. 2012; 490:267–72. https://doi.org/10.1038/nature11401 PMID:22982992

66. Rizopoulos D. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. Comput Stat Data Anal. 2012; 56:491–501. https://doi.org/10.1016/j.csda.2011.09.007

67. McEwen BS, Wingfield JC. The concept of allostasis in biology and biomedicine. Horm Behav. 2003; 43:2–15. https://doi.org/10.1016/s0018-506x(02)00024-7 PMID:12614627

68. Ukraintseva S, Yashin AI, Arbeev KG. Resilience Versus Robustness in Aging. J Gerontol A Biol Sci Med Sci. 2016; 71:1533–4. https://doi.org/10.1093/gerona/glw083 PMID:27146372

69. Strehler B. Time, Cells, and Aging. London: Academic Press. 1962.

70. Strehler BL, Mildvan AS. General theory of mortality and aging. Science. 1960; 132:14–21. https://doi.org/10.1126/science.132.3418.14 PMID:13835176

71. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. Hallmarks of aging: An expanding universe. Cell. 2023; 186:243–78. https://doi.org/10.1016/j.cell.2022.11.001 PMID:36599349

72. 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; 629–49.

73. McEwen BS, Stellar E. Stress and the individual. Mechanisms leading to disease. Arch Intern Med. 1993; 153:2093–101. PMID:8379800

74. McEwen BS. Stress, adaptation, and disease. Allostasis and allostatic load. Ann N Y Acad Sci. 1998; 840:33–44. https://doi.org/10.1111/j.1749-6632.1998.tb09546.x PMID:9629234

SUPPLEMENTARY MATERIALS

Stochastic process models: Interpretation and illustration of components, parameters, and related null hypotheses

As described in the main text (section Stochastic process models: General specifications), for stochastic process models (SPM) applications, we used a one-dimensional version with time-dependent components [1]. This model has two equations: one for modeling individual age trajectories of LPC species $Y(t,c)$ (where t is age and c denotes covariates; see section Stochastic process models: Specific parameterizations used in applications in the main text) and one representing individual mortality rate $\mu\big(t,c,Y(t,c)\big)$ as a function of age, covariates, and LPC levels:

$$\mathrm{d}Y(t,c) = a(t,c)\big(Y(t,c) - f_1(t,c)\big)\mathrm{d}t + b(t,c)\mathrm{d}W(t), \quad \text{(Eq. 6)}$$

$$\mu\big(t,c,Y(t,c)\big) = \mu_0(t,c) + Q(t,c)\big(Y(t,c) - f_0(t,c)\big)^2. \quad \text{(Eq. 7)}$$

Note that these equations model individual trajectories/rates; we do not use an index to indicate that t, c, and $Y(t,c)$ are individual-based quantities, for simplicity of notation and visualization. Here, $Y(t_0,c)$ is the initial condition (t_0 denotes age at entering the study) and $W(t)$ is the stochastic Wiener process (also known as Brownian motion [2]), which is assumed to be independent of $Y(t_0,c)$. This process defines random paths of $Y(t,c)$. Below, we provide a detailed description of the model's components and parameters, and illustrate the meaning of related null hypotheses. This can assist in the interpretation of SPM results presented in Figure 1 and Supplementary Figure 10A, 10B, and Supplementary Tables 2–4.

Baseline mortality rate $\mu_0(t,c)$ represents the part of the mortality rate in Eq. (7) that is not related to LPC ($Y(t,c)$). As SPM is a parametric model, we need to specify the parametric form of $\mu_0(t,c)$, along with other components. For our applications, we chose the Gompertz baseline hazard, which is commonly used in demography to represent a typical pattern of mortality rate at adult and old ages. In this parameterization, the logarithm of the baseline mortality rate is a linear function of age and covariates: $\ln\mu_0(t,c) = \ln a_{\mu_0} + b_{\mu_0}(t - t_{\min}) + \beta_{\mu_0}c$, where $t_{\min} = 50$ in our applications. Here, a_{μ_0} is the baseline mortality rate corresponding to age $t_{\min}$ and zero values of covariates in a (column) vector c. Parameter b_{μ_0} represents the rate of change in $\ln\mu_0(t,c)$ with age, and β_{μ_0} is a (row) vector of parameters corresponding to covariates in c. The baseline mortality rate is not of direct interest in our applications. Therefore, we do not test any null hypothesis related to $\mu_0(t,c)$.

The quadratic hazard term $Q(t,c)$ (assumed to be non-negative for all t and c) is the multiplier scaling the quadratic component of the hazard in Eq. (7). Such a quadratic shape of the mortality rate used in SPM reflects common epidemiological observations (including our own research [3–6]) that mortality as a function of various biomarkers has a U-shape. This means that there is an optimal value of a biomarker ($f_0(t,c)$) that minimizes the mortality risk at a specific age (and specific values of covariates, if relevant) and that the deviations of the biomarker to smaller or larger values induce an additional mortality risk. This is captured by the quadratic shape of the mortality rate, and $Q(t,c)$ controls the width of the U-shape. We test several null hypotheses (H0) related to this component of SPM.

First, we test H0: $Q(t,c) = 0$ (denoted Qzero). This is the key H0 to test because if we are not able to reject it, it indicates that the respective biomarker (LPC) is not related to the risk of death, making testing any other H0s for that particular LPC irrelevant. Supplementary Figure 2A illustrates the situation when Qzero is rejected, i.e., LPC values are related to the mortality risk (as illustrated by a U-shape of the mortality rate). Supplementary Figure 2B, conversely, shows the case when $Q(t,c) = 0$, so that the mortality rate equals $\mu_0(t,c)$ and it is independent of LPC values.

Second, we test H0 about the age pattern of $Q(t,c)$. We use a parsimonious parameterization of $Q(t,c)$: $Q(t,c) = a_{\mathrm{Q}} + b_{\mathrm{Q}}(t - t_{\min}) + \beta_{\mathrm{Q}}c$. The parameter a_{Q} corresponds to the "baseline" width of the U-shape at age $t_{\min}$ and zero covariate(s) c. We use only one covariate (sex) in $Q(t,c)$, so this corresponds to the width of the U-shape in 50-year-old females. The parameter b_{Q} models how the width of the U-shape changes with age. We test H0: $b_{\mathrm{Q}} = 0$ (QnoT). Supplementary Figure 3A presents three possible age patterns of $Q(t,c)$: age independent ($b_{\mathrm{Q}} = 0$), declining with age ($b_{\mathrm{Q}} < 0$), and increasing with age ($b_{\mathrm{Q}} > 0$). Supplementary Figure 3B–3D display corresponding values of the quadratic part in the hazard (i.e., $Q(t,c)\big(Y(t,c) - f_0(t,c)\big)^2$) for different ages and LPC levels (with c = 0 for the simplicity of illustration). For example, if $b_{\mathrm{Q}} > 0$ (Supplementary Figure 3D), then the U-shape of the mortality rate shrinks with age so that the same deviation of LPC from an optimal (age-specific) level results in a larger additional mortality

risk at older ages compared to younger ages, i.e., the impact of deviations of LPC trajectories from optimal levels aggravates with age. Such observations would correspond to an increase in vulnerability to deviations of LPC from optimal values, which is the manifestation of *age-related decline in biological robustness* [7, 8].

Third, we test H0 about the dependence of $Q(t,c)$ on c (sex-dependence), H0: $\beta_{\mathrm{Q}} = 0$ (QnoC). The parameter β_{Q} specifies how/if the U-shape of the mortality rate as a function of LPC is different in females and males. If $\beta_{\mathrm{Q}} = 0$, then the width of the U-shape is the same in females and males at each age. If $\beta_{\mathrm{Q}} < 0$ ($\beta_{\mathrm{Q}} > 0$), then the U-shape is wider (narrower) in males compared to females at each age. Supplementary Figure 4A, 4B show the patterns of $Q(t,c)$ and the corresponding values of the quadratic part in the hazard ($Q(t,c)\big(Y(t,c) - f_0(t,c)\big)^2$) for females and males when $\beta_{\mathrm{Q}} > 0$ (assuming zero optimal levels $f_0(t,c)$ in females and males, for the simplicity of illustration). In this case, the U-shape is narrower in males at each age, and the same deviation of LPC from the optimal level results in higher additional mortality risk in males compared to females (thus increasing the overall mortality risk in males even if the baseline mortality rates are the same in both sexes).

The (negative) feedback coefficient $a(t,c)$ in Eq. (6) regulates the dynamic behavior of LPC trajectories ($Y(t,c)$). The particular form of the equation used for modeling $Y(t,c)$ in SPM was selected to incorporate *homeostatic regulation* in the model, which is a fundamental feature of living organisms. The trajectory of $Y(t,c)$ modeled by this equation tends to move back to its long-term mean trajectory (or "equilibrium" level) ($f_1(t,c)$) if it deviates from this equilibrium trajectory. For this to happen, the feedback coefficient has to be negative, hence the restriction used in the model: $a(t,c) < 0$, for all values of age t and covariate(s) c observed in the data. The rate of return to the equilibrium level is regulated by this coefficient $a(t,c)$. The larger the absolute value of the feedback coefficient $a(t,c)$, the faster $Y(t,c)$ returns to its equilibrium level. This coefficient is also called the *adaptive capacity* [1] because it represents the rate of adaptive response (associated with *biological resilience* [7, 9, 10]) to any factors causing deviations of $Y(t,c)$ from their dynamic equilibrium levels $f_1(t,c)$.

In our applications, we use a linear function for $a(t,c)$: $a(t,c) = a_{\mathrm{Y}} + b_{\mathrm{Y}}(t - t_{\min}) + \beta_{\mathrm{Y}}c$, where $a_{\mathrm{Y}} < 0$, $b_{\mathrm{Y}} \geq 0$. The parameter a_{Y} represents the "baseline" value of this coefficient corresponding to age $t_{\min}$ and zero covariate(s) c (i.e., 50 years old females in our case). The parameter b_{Y} models the rate of change in $a(t,c)$ with age. We test H0: $b_{\mathrm{Y}} = 0$ (AnoT). Supplementary Figure 5A shows examples of the absolute value of the feedback coefficient ($|a(t,c)|$) for zero and positive b_{Y} (with c = 0 for the simplicity of illustration). Supplementary Figure 5B displays sample trajectories of $Y(t,c)$ in these two cases. As one can see, in case of a positive b_{Y} (when the absolute value of the feedback coefficient becomes smaller with age), it takes more time for a trajectory of $Y(t,c)$ to go back to the equilibrium level $f_1(t,c)$ at older ages compared to younger ages. This illustrates the *aging-related decline in adaptive capacity* or the associated notion of the *decline in biological resilience*, which is a key manifestation of aging [9]. Note that we show the absolute value of $a(t,c)$, $|a(t,c)|$, rather than $a(t,c)$ in Supplementary Figure 5 so that a decline in the displayed quantity would have the interpretation of a decline in adaptive capacity/biological resilience.

We also test H0: $\beta_{\mathrm{Y}} = 0$ (AnoC). The parameter β_{Y} specifies the difference in the baseline levels of the adaptive capacity between males and females. If it is zero, then the baseline level of $a(t,c)$ is the same in females and males. If $\beta_{\mathrm{Y}} > 0$ ($\beta_{\mathrm{Y}} < 0$), then females are more (less) resilient compared to males in terms of a faster (slower) rate of return of $Y(t,c)$ (LPC) to its equilibrium levels. Supplementary Figure 5C, 5D present examples of $|a(t,c)|$ and sample trajectories of $Y(t,c)$ in females and males. This illustrates the situation when $\beta_{\mathrm{Y}} > 0$ corresponding to better adaptive capacity in females (i.e., a faster return of $Y(t,c)$ to the equilibrium level $f_1(t,c)$).

The volatility coefficient $b(t,c)$ controls the volatility of the process $Y(t,c)$. The volatility of $Y(t,c)$ represents the intensity of the random fluctuations (or noise) in the process. It determines how much the process can deviate from its mean due to random impacts. Higher volatility means larger deviations from the mean, while lower volatility indicates smaller deviations. Supplementary Figure 6 shows examples of the process $Y(t,c)$ with higher and lower volatility. Based on our prior simulations showing the best accuracy of parameter estimates for models with a constant volatility coefficient [1], we use the following specification of $b(t,c)$: $b(t,c) = \sigma_1 + \beta_{\mathrm{W}}c$ (with the constraint: $b(t,c) > 0$ for all values of covariates c, i.e., sex in our case). In our applications, σ_1 represents the value of this coefficient in females and $\sigma_1 + \beta_{\mathrm{W}}$ is the volatility coefficient in males, which can be larger ($\beta_{\mathrm{W}} > 0$), smaller ($\beta_{\mathrm{W}} < 0$), or the same ($\beta_{\mathrm{W}} = 0$) as in females. We test the respective H0: $\beta_{\mathrm{W}} = 0$ (BnoC) to determine if the volatility of LPC is sex-specific.

The equilibrium trajectory $f_1(t,c)$ represents the long-term mean of the process $Y(t,c)$ (see the paragraph

describing $a(t,c)$). This SPM component is also known as the "*mean allostatic trajectory*" because it features the effect of allostatic adaptation [11], i.e., the LPC levels forced by an organism's regulatory systems functioning in non-optimal conditions (as postulated by the theory of allostasis [11–13]). In our applications, we assume that the equilibrium LPC levels can depend on age (t) and covariates (c) (i.e., sex). We use the following specification of $f_1(t,c)$: $f_1(t,c) = a_{f_1} + b_{f_1}(t - t_{\min}) + \beta_{f_1} c$. Here, the parameter a_{f_1} is the equilibrium LPC level at age $t_{\min}$ and zero values of covariates c (i.e., in 50-year-old females). The parameter b_{f_1} quantifies the rate of change in the equilibrium LPC level with age, which can increase ($b_{f_1} > 0$), decrease ($b_{f_1} < 0$), or be stable ($b_{f_1} = 0$). The parameter β_{f_1} determines how/if the baseline equilibrium LPC level differs by sex: higher ($\beta_{f_1} > 0$) or lower ($\beta_{f_1} < 0$) in males, or sex-independent ($\beta_{f_1} = 0$). We test two null hypotheses about $f_1(t,c)$ to determine its age pattern and dependence on sex: H0: $b_{f_1} = 0$ (F1noT), i.e., equilibrium LPC levels are the same for all ages, and H0: $\beta_{f_1} = 0$ (F1noC), i.e., the baseline equilibrium LPC levels do not differ by sex. Supplementary Figure 7A, 7B display sample trajectories of $Y(t,c)$ for different age (Supplementary Figure 7A) and sex (Supplementary Figure 7B) patterns of $f_1(t,c)$.

The optimal trajectory $f_0(t,c)$ represents the values of $Y(t,c)$ minimizing the risk at age t and covariate values c. It is interpreted as a *physiological or biological optimum* (also known as *"sweet spots"* [14–16]). In our applications, $f_0(t,c)$ models LPC levels minimizing mortality risks at respective ages and covariate values. We parameterize $f_0(t,c)$ as a linear function: $f_0(t,c) = a_{f_0} + b_{f_0}(t - t_{\min}) + \beta_{f_0} c$. Here, a_{f_0} is the LPC value corresponding to the minimal mortality risk at age $t_{\min}$ for individuals with zero c (i.e., 50 years old females). The parameter b_{f_0} defines the rate of change in the optimal LPC level with age. It can increase ($b_{f_0} > 0$), decrease ($b_{f_0} < 0$), or remain stable ($b_{f_0} = 0$) with age. The parameter β_{f_0} specifies how/if the baseline optimal LPC level differs by sex: higher ($\beta_{f_0} > 0$) or lower ($\beta_{f_0} < 0$) in males, or sex-independent ($\beta_{f_0} = 0$). Similar to $f_1(t,c)$, we test two null hypotheses about $f_0(t,c)$ to determine its age pattern and dependence on sex: H0: $b_{f_0} = 0$ (F0noT), i.e., optimal LPC levels do not change with age, and H0: $\beta_{f_0} = 0$ (F0noC), i.e., the baseline optimal LPC levels are equal in females and males. Supplementary Figure 8A, 8B present the quadratic part in the hazard (i.e., $Q(t,c)\big(Y(t,c) - f_0(t,c)\big)^2$) for different ages and LPC levels (with c = 0 for the simplicity of illustration) in the case of increasing and declining optimal levels. Supplementary Figure 8C, 8D show corresponding mortality rates $\mu\big(t,c,Y(t,c)\big)$ for different ages and LPC levels (with zero covariates c, for the purpose of this illustration). We also assumed in this illustration that $Q(t,c)$ does not depend on age t so that the width of the U-shape of mortality as a function of the biomarker $Y(t,c)$ is the same for all ages. As Supplementary Figure 8A shows, if the optimal trajectory increases with age, then the U-shape of the quadratic part in the hazard shifts to the right (to larger values of $Y(t,c)$) so that smaller values of $Y(t,c)$ result in a larger additional risk compared to the baseline mortality $\mu_0(t,c)$ observed at the optimal level $f_0(t,c)$ (see Supplementary Figure 8C). Conversely, when the optimal trajectory declines with age (Supplementary Figure 8B), the parabola in the hazard rate shifts to the left (to smaller values of $Y(t,c)$) so that larger values of $Y(t,c)$ induce a larger additional risk (Supplementary Figure 8D).

Note that the equilibrium and optimal trajectories can be different, and the absolute value of this difference, $AL(t,c) = |f_0(t,c) - f_1(t,c)|$, is related to the practical realization of the theoretical concept of the *allostatic load (AL)* suggested in the literature [11–13, 17]. If the optimal and equilibrium trajectories coincide (i.e., $AL(t,c) = 0$), then LPC trajectories ($Y(t,c)$) tend to converge to $f_0(t,c)$ so that the mortality rate $\mu\big(t,c,Y(t,c)\big)$ gets closer to the baseline level $\mu_0(t,c)$ as the quadratic part in Eq. (7) (i.e., $Q(t,c)\big(Y(t,c) - f_0(t,c)\big)^2$) gets close to zero. However, if the equilibrium trajectory differs from the optimal one, then LPC values tend to a trajectory which is different from that minimizing the mortality rate. As a result, the mortality rate fluctuates around the level $\mu_0(t,c) + Q(t,c)\big(f_1(t,c) - f_0(t,c)\big)^2$. As $Q(t,c) \geq 0$ by the assumption of SPM, this means that this level is higher than the baseline mortality $\mu_0(t,c)$. The larger the value of this measure $AL(t,c)$, the larger this additional mortality risk ("load") $Q(t,c)\big(f_1(t,c) - f_0(t,c)\big)^2$. We test two H0s related to $AL(t,c)$. First, we test H0: $f_1(t,c) = f_0(t,c)$, i.e., $AL(t,c) = 0$ (ALzero), that is, the optimal and equilibrium trajectories are the same. Second, we test H0: $b_{f_1} = 0$ and $b_{f_0} = 0$ (ALnoT), i.e., that the difference between the optimal and equilibrium trajectories does not change with age. Supplementary Figure 9A, 9B illustrate the quadratic part in the hazard ($Q(t,c)\big(f_1(t,c) - f_0(t,c)\big)^2$) and the mortality rate ($\mu\big(t,c,Y(t,c)\big)$) (Eq. 7) evaluated at the equilibrium $f_1(t,c)$ and optimal ($f_0(t,c)$) levels for different ages t (with c = 0 for the simplicity of illustration). This shows the case when the optimal and equilibrium trajectories diverge at older ages, i.e., when $AL(t,c)$ increases with

age. As this illustrative example shows, if the equilibrium and optimal levels coincide at age 50, then there is no additional mortality risk when LPC is at the equilibrium level. Therefore, the mortality rate for an "average" individual with LPC at the equilibrium level equals the optimal (baseline) level $\mu_0(50, c)$. However, if the equilibrium and optimal trajectories diverge with age, then the mortality rate of a centenarian whose LPC level follows the equilibrium trajectory will be about 0.32 higher than the mortality rate of a centenarian with the optimal LPC level for that age.

Supplementary References

1. Yashin AI, Arbeev KG, Akushevich I, Kulminski A, Ukraintseva SV, Stallard E, Land KC. The quadratic hazard model for analyzing longitudinal data on aging, health, and the life span. Phys Life Rev. 2012; 9:177–88. https://doi.org/10.1016/j.plrev.2012.05.002 PMID:22633776

2. Durrett R. Brownian Motion: Cambridge University Press). 2019.

3. Arbeev KG, Cohen AA, Arbeeva LS, Milot E, Stallard E, Kulminski AM, Akushevich I, Ukraintseva SV, Christensen K, Yashin AI. Optimal Versus Realized Trajectories of Physiological Dysregulation in Aging and Their Relation to Sex-Specific Mortality Risk. Front Public Health. 2016; 4:3. https://doi.org/10.3389/fpubh.2016.00003 PMID:26835445

4. Yashin AI, Arbeev KG, Akushevich I, Ukraintseva SV, Kulminski A, Arbeeva LS, Culminskaya I. Exceptional survivors have lower age trajectories of blood glucose: lessons from longitudinal data. Biogerontology. 2010; 11:257–65. https://doi.org/10.1007/s10522-009-9243-1 PMID:19644762

5. Yashin AI, Arbeev KG, Ukraintseva SV, Akushevich I and Kulminski A. Patterns of aging related changes on the way to 100: An approach to studying aging, mortality, and longevity from longitudinal data. N Amer Actuarial J. 2012; 16:403–33.

6. Yashin AI, Ukraintseva SV, Arbeev KG, Akushevich I, Arbeeva LS, Kulminski AM. Maintaining physiological state for exceptional survival: What is the normal level of blood glucose and does it change with age? Mech Ageing Dev. 2009; 130:611–8. https://doi.org/10.1016/j.mad.2009.07.004 PMID:19635493

7. Ukraintseva S, Yashin AI, Arbeev KG. Resilience Versus Robustness in Aging. J Gerontol A Biol Sci Med Sci. 2016; 71:1533–4. https://doi.org/10.1093/gerona/glw083 PMID:27146372

8. Arbeev KG, Ukraintseva SV, Bagley O, Zhbannikov IY, Cohen AA, Kulminski AM, Yashin AI. "Physiological Dysregulation" as a Promising Measure of Robustness and Resilience in Studies of Aging and a New Indicator of Preclinical Disease. J Gerontol A Biol Sci Med Sci. 2019; 74:462–8. https://doi.org/10.1093/gerona/gly136 PMID:29939206

9. Ukraintseva S, Arbeev K, Duan M, Akushevich I, Kulminski A, Stallard E, Yashin A. Decline in biological resilience as key manifestation of aging: Potential mechanisms and role in health and longevity. Mech Ageing Dev. 2021; 194:111418. https://doi.org/10.1016/j.mad.2020.111418 PMID:33340523

10. Arbeev KG, Bagley O, Yashkin AP, Duan H, Akushevich I, Ukraintseva SV, Yashin AI. Understanding Alzheimer's disease in the context of aging: Findings from applications of stochastic process models to the Health and Retirement Study. Mech Ageing Dev. 2023; 211:111791. https://doi.org/10.1016/j.mad.2023.111791 PMID:36796730

11. McEwen BS, Wingfield JC. The concept of allostasis in biology and biomedicine. Horm Behav. 2003; 43:2–15. https://doi.org/10.1016/s0018-506x(02)00024-7 PMID:12614627

12. Sterling P and Eyer J. Allostasis: A New Paradigm to Explain Arousal Pathology. In: Fisher S and Reason J, eds. Handbook of Life Stress, Cognition and Health. (New York: John Wiley & Sons). 1988. 629–49.

13. McEwen BS. Stress, adaptation, and disease. Allostasis and allostatic load. Ann N Y Acad Sci. 1998; 840:33–44. https://doi.org/10.1111/j.1749-6632.1998.tb09546.x PMID:9629234

14. Johnson JD, Alejandro EU. Control of pancreatic beta-cell fate by insulin signaling: The sweet spot hypothesis. Cell Cycle. 2008; 7:1343–7. https://doi.org/10.4161/cc.7.10.5865 PMID:18418065

15. Lee JH, Han K, Huh JH. The sweet spot: fasting glucose, cardiovascular disease, and mortality in older adults with diabetes: a nationwide population-based study. Cardiovasc Diabetol. 2020; 19:44. https://doi.org/10.1186/s12933-020-01021-8 PMID:32238157

16. Vishnyakova O, Song X, Rockwood K, Elliott LT, Brooks-Wilson A. Physiological phenotypes have

optimal values relevant to healthy aging: sweet spots deduced from the Canadian Longitudinal Study on Aging. Geroscience. 2024; 46:1589–605. https://doi.org/10.1007/s11357-023-00895-2 PMID:37688655

17. McEwen BS, Stellar E. Stress and the individual. Mechanisms leading to disease. Arch Intern Med. 1993; 153:2093–101. PMID:8379800

18. Borecki IB, Province MA. Genetic and genomic discovery using family studies. Circulation. 2008; 118:1057–63. https://doi.org/10.1161/CIRCULATIONAHA.107.714592 PMID:18765388

Code implementing likelihood estimation procedure of SPM

This section describes the code that can be used to estimate the likelihood function of the Stochastic Process Model (SPM) used in the paper. It provides the code for "unrestricted" model, which can be modified to specify one or more restrictions on parameters to perform hypothesis testing presented in the text.

The function estimates the likelihood for SPM represented by equations:

$$dY(t,c) = a(t,c)(Y(t,c) - f_1(t,c))dt + b(t,c)dW(t)$$

$$\mu(t,c,Y(t,c)) = \mu_0(t,c) + Q(t,c)(Y(t,c) - f_0(t,c))^2$$

with the following specification of components:

$$a(t,c) = a_Y + b_Y(t - t_{min}) + \beta_Y c, a_Y < 0, b_Y \geq 0,$$

$$f_1(t,c) = a_{f_1} + b_{f_1}(t - t_{min}) + \beta_{f_1} c,$$

$$b(t,c) = \sigma_1 + \beta_W c,$$

$$\ln \mu_0(t,c) = \ln a_{\mu_0} + b_{\mu_0}(t - t_{min}) + \beta_{\mu_0} c,$$

$$Q(t,c) = a_Q + b_Q(t - t_{min}) + \beta_Q c,$$

$$f_0(t,c) = a_{f_0} + b_{f_0}(t - t_{min}) + \beta_{f_0} c,$$

$$Y(t_0,c) \sim N(f_1(t_0,c), \sigma_0).$$

Syntax:
function lnLik = LogLik(Param, DataSPM, t_min, NamesCovar)

Parameters:

Param - a column vector of model parameters in the following order:

$$\ln a_{\mu_0}, b_{\mu_0}, \beta_{\mu_0}, a_Q, b_Q, \beta_Q, a_Y, b_Y, \beta_Y, \sigma_0, \sigma_1, \beta_W, a_{f_1}, b_{f_1}, \beta_{f_1}, a_{f_0}, b_{f_0}, \beta_{f_0}$$

DataSPM - a table with the following variables (in any order; can have additional variables which will be ignored): *Age* (start of age interval), *AgeNext* (end of age interval), *IndicatorEvent* (a binary variable indicating an event (1) or no event (0) within the age interval (*Age* , *AgeNext*)), *Yt* (longitudinal variable modeled by Y(t,c)), *IsFirstRow* (a binary variable indicating the first record for an individual: 1 - first record; 0 - otherwise), *IsLastRow* (a binary variable indicating the last record for an individual: 1 - last record; 0 - otherwise), and variables to be included as additional covariates (c), see NamesCovar

t_min - minimal age used in formulas, see above

NamesCovar - cell array with names of variables in DataSPM to be used as additional covariates (c): the first cell contains names of variables to be used as covariates in $\mu_0(t, c)$ and the second cell contains names of variables to be used as covariates in other components

Output:

lnLik - minus logarithm of the likelihood function

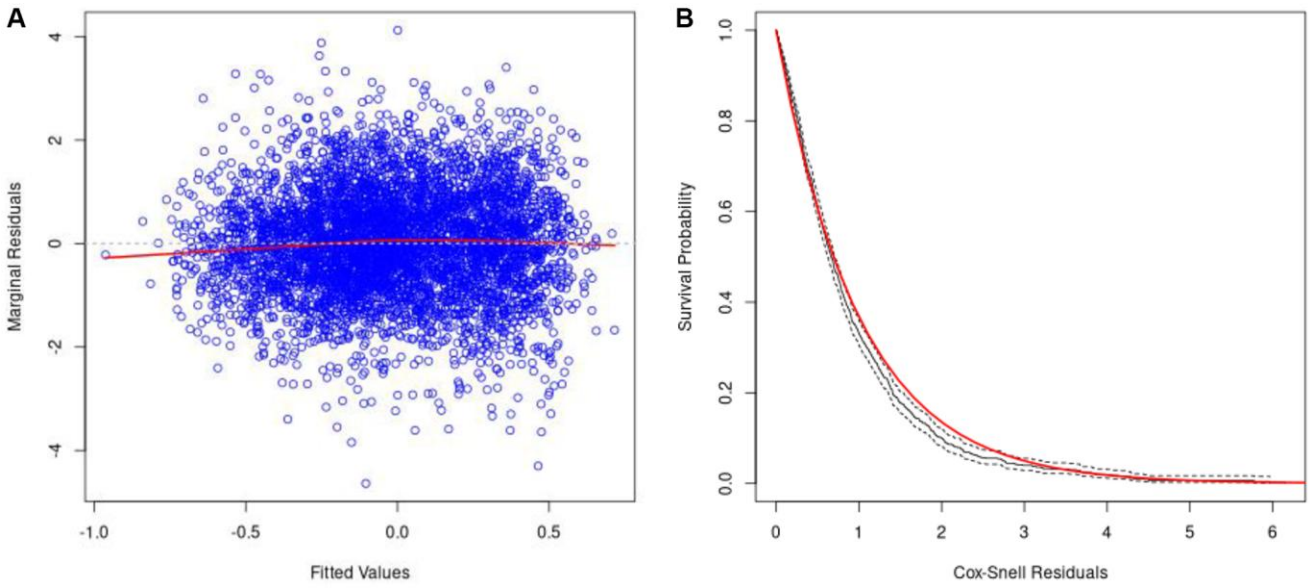
```
function lnLik = LogLikSPM(Param, DataSPM, t_min, NamesCovar)

NumRows = height(DataSPM);

NamesCovarMu0 = NamesCovar{1};
NamesCovarOther = NamesCovar{2};
NumCovarMu0 = length(NamesCovarMu0);
NumCovarOther = length(NamesCovarOther);

ln_a_mu0 = Param(1);
b_mu0 = Param(2);
b_covar_mu0 = Param(3:(3 + NumCovarMu0 - 1));
a_Q = Param(3 + NumCovarMu0);
b_Q = Param(4 + NumCovarMu0);
b_covar_Q = Param((5 + NumCovarMu0):(5 + NumCovarMu0 + NumCovarOther - 1));
a_Y = Param(5 + NumCovarMu0 + NumCovarOther);
b_Y = Param(6 + NumCovarMu0 + NumCovarOther);
b_covar_Y = Param((7 + NumCovarMu0 + NumCovarOther):(7 + NumCovarMu0 + 2*NumCovarOther - 1));
sigma0 = Param(7 + NumCovarMu0 + 2*NumCovarOther);
sigma1 = Param(8 + NumCovarMu0 + 2*NumCovarOther);
b_covar_W = Param((9 + NumCovarMu0 + 2*NumCovarOther):(9 + NumCovarMu0 + 3*NumCovarOther - 1));
a_f1 = Param(9 + NumCovarMu0 + 3*NumCovarOther);
b_f1 = Param(10 + NumCovarMu0 + 3*NumCovarOther);
b_covar_f1 = Param((11 + NumCovarMu0 + 3*NumCovarOther):(11 + NumCovarMu0 + 4*NumCovarOther - 1));
a_f0 = Param(11 + NumCovarMu0 + 4*NumCovarOther);
b_f0 = Param(12 + NumCovarMu0 + 4*NumCovarOther);
b_covar_f0 = Param((13 + NumCovarMu0 + 4*NumCovarOther):(13 + NumCovarMu0 + 5*NumCovarOther - 1));

delta_i_all = DataSPM.IndicatorDeath;
t = DataSPM.Age;
t_next = DataSPM.AgeNext;
Yt = DataSPM.Yt;
IsFirstRow = DataSPM.IsFirstRow;
IsLastRow = DataSPM.IsLastRow;

if NumCovarOther == 1
    Xt_other = DataSPM.(NamesCovarOther{1});
else
    Xt_other = NaN*ones(NumRows, NumCovarOther);
    for i = 1:NumCovarOther
        Xt_other(:, i) = DataSPM.(NamesCovarOther{i});
    end
end

if NumCovarMu0 == 1
    Xt_mu0 = DataSPM.(NamesCovarMu0{1});
```

```
else
    Xt_mu0 = NaN*ones(NumRows, NumCovarMu0);
    for i = 1:NumCovarMu0
        Xt_mu0(:, i) = DataSPM.(NamesCovarMu0{i});
    end
end

lnLik = 0;
for i = 1:NumRows
    delta_i = delta_i_all(i);
    tk = t(i);
    tk_next = t_next(i);
    Ytk = Yt(i);
    Xtk_other = Xt_other(i, :);
    Xtk_mu0 = Xt_mu0(i, :);

    mu0_tk = exp(ln_a_mu0 + b_mu0*(tk - t_min) + Xtk_mu0*b_covar_mu0);
    f0_tk = a_f0 + b_f0*(tk - t_min) + Xtk_other*b_covar_f0;
    Q_tk = a_Q + b_Q*(tk - t_min) + Xtk_other*b_covar_Q;
    mu_tk = mu0_tk + Q_tk*(Ytk - f0_tk)^2;

    if IsFirstRow(i) == 1
        lnLY = 0;
        lnLQ = 0;

        t0 = t(i);
        Yt0 = Yt(i);

        Ybar_tk_prev = a_f1 + b_f1*(t0 - t_min) + Xtk_other*b_covar_f1;
        if (sigma0 > 0)
            lnLY = lnLY - log(sqrt(2*pi)*sigma0) - ((Yt0 - Ybar_tk_prev)^2)/(2*sigma0^2);
        end
    else
        tk_prev = t(i-1);
        Ytk_prev = Yt(i-1);

        a_tk_prev = a_Y + b_Y*(tk_prev - t_min) + Xtk_other*b_covar_Y;
        f1_tk_prev = a_f1 + b_f1*(tk_prev - t_min) + Xtk_other*b_covar_f1;

        Ybar_tk_prev = Ytk_prev + a_tk_prev*(Ytk_prev - f1_tk_prev)*(tk - tk_prev);
        sigma1_tk = sigma1 + Xtk_other*b_covar_W;
        if (sigma1_tk > 0) && ((tk - tk_prev) > 0)
            lnLY = lnLY - log(sqrt(2*pi*(tk - tk_prev))*sigma1_tk) - ((Ytk - Ybar_tk_prev)^2)/(2*(tk -
tk_prev)*sigma1_tk^2);
        end
    end

    if IsLastRow(i) == 1
        if delta_i == 0
            lnLQ = lnLQ - mu_tk*(tk_next - tk);
        elseif delta_i == 1
            lnLQ = lnLQ + log(1 - exp(-mu_tk*(tk_next - tk)));
        end
    else
        lnLQ = lnLQ - mu_tk*(tk_next - tk);
    end
```

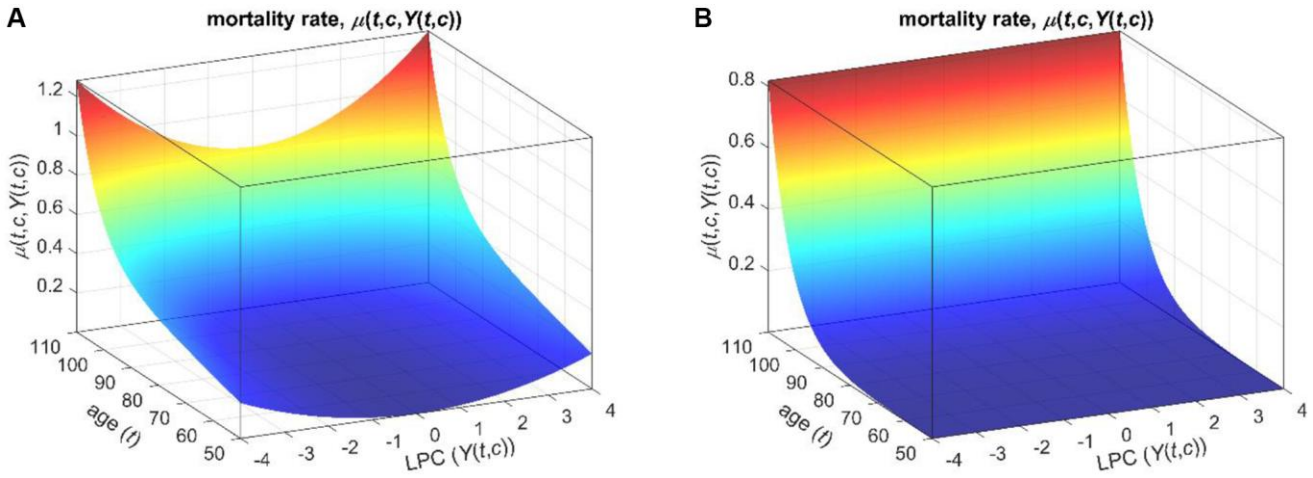
```
    if IsLastRow(i) == 1
        lnLik = lnLik + lnLY + lnLQ;
    end
end
lnLik = -lnLik;
```

Published with MATLAB® R2024b

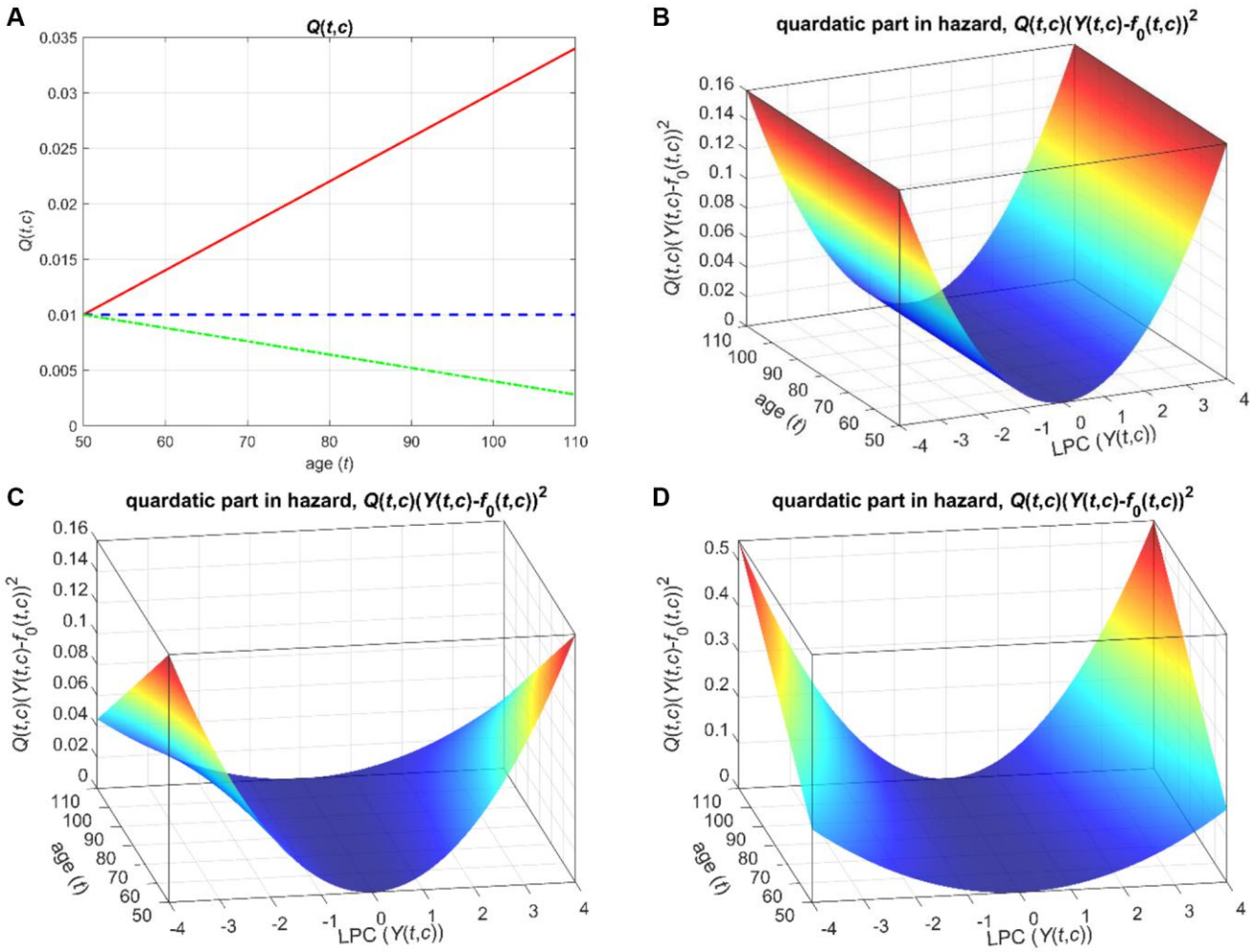
Supplementary Figures



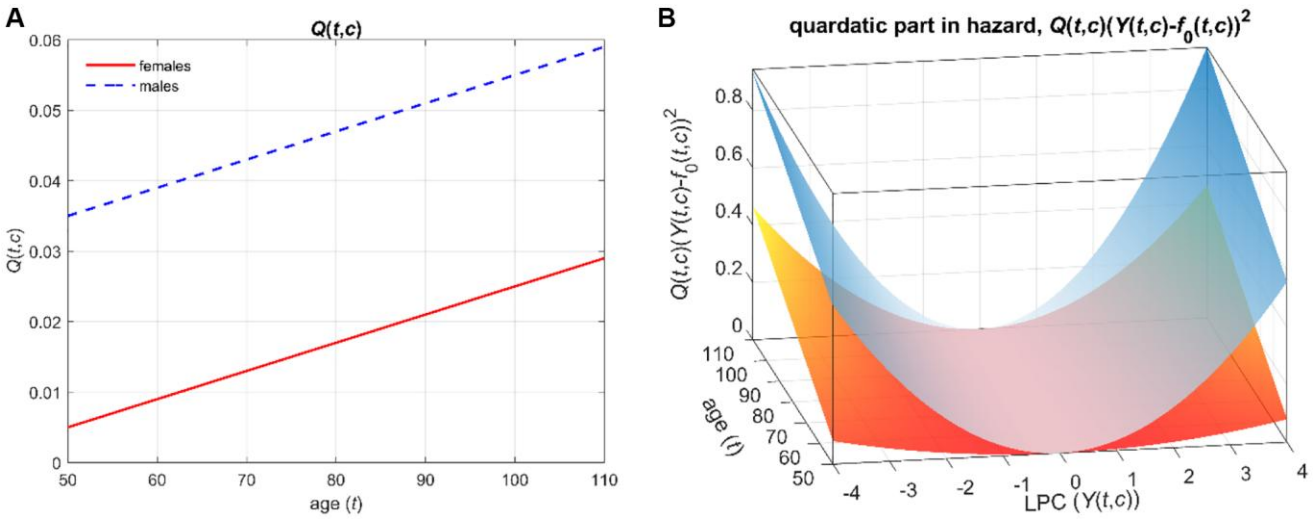
Supplementary Figure 1. Diagnostic plots assessing the goodness-of-fit and assumptions of joint models in applications to LPC 15:0/0:0. (**A**) Plot of the standardized marginal residuals versus the corresponding fitted values for the longitudinal outcome (LPC 15:0/0:0). The red solid line denotes the fit of the loess smoother. (**B**) Residuals analysis for the survival outcome by assessing the overall fit of the survival sub-model using the Cox-Snell residuals. The black solid line denotes the Kaplan-Meier estimate of the survival function of the residuals (with the dashed lines corresponding to the 95% pointwise confidence intervals). The red solid line is the survival function of the unit exponential distribution (this is the distribution if the survival sub-model is correct).



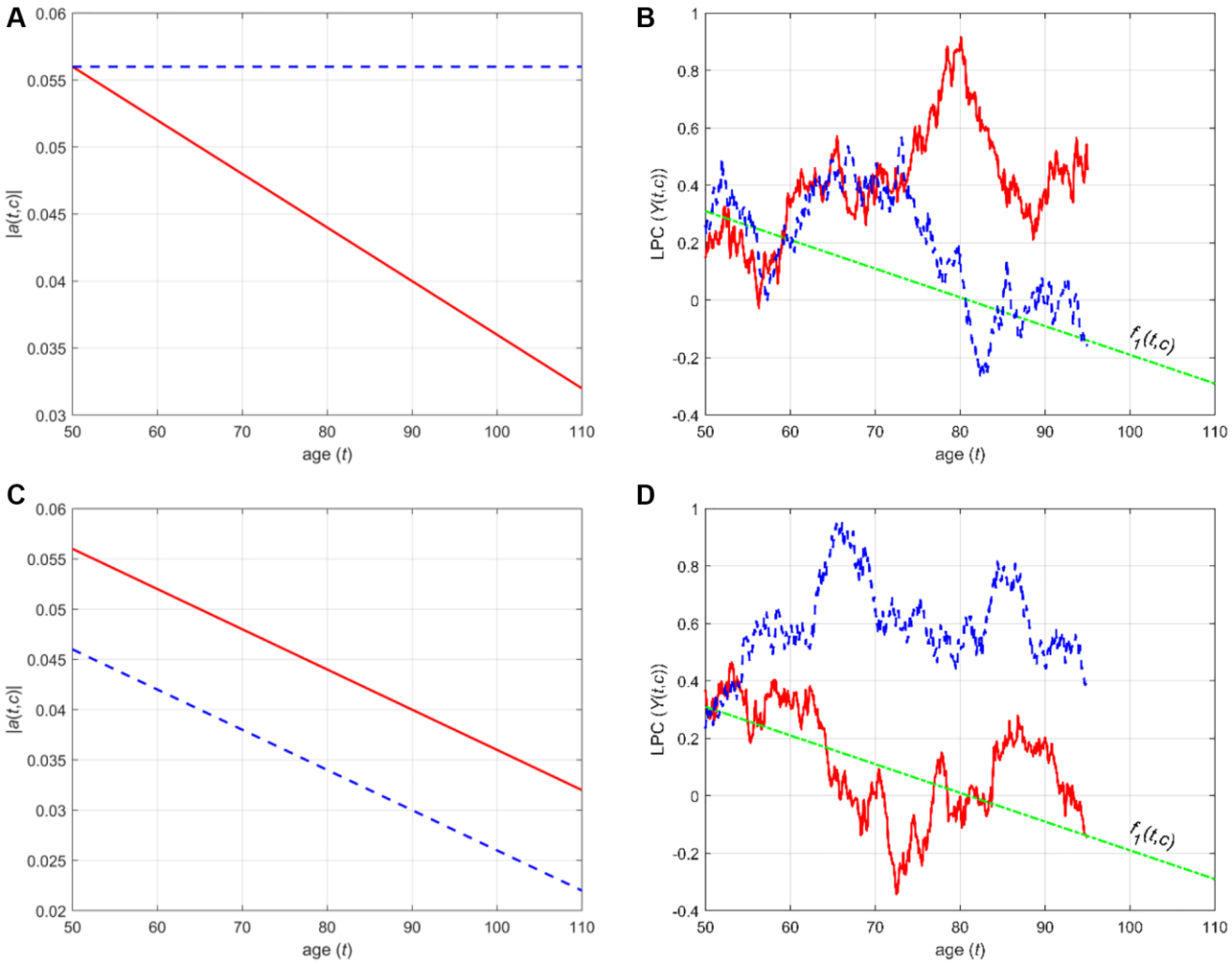
Supplementary Figure 2. Stochastic process models: 3D plots illustrating mortality rates corresponding to different quadratic hazard terms $\boldsymbol{Q}(\boldsymbol{t},\boldsymbol{c})$. (**A**) Mortality rate in the case of non-zero $\boldsymbol{Q}(\boldsymbol{t},\boldsymbol{c})$; (**B**) Mortality rate when $\boldsymbol{Q}(\boldsymbol{t},\boldsymbol{c}) = \boldsymbol{0}$. For this illustration, we modeled Eqs. (6–7) with a single binary covariate *c* and parameters corresponding to a hypothetical biomarker $\boldsymbol{Y}(\boldsymbol{t},\boldsymbol{c})$ (e.g., a transformed LPC) having the standard normal distribution and hazard rates resembling human mortality rates at old ages. The figures display the mortality rate $\boldsymbol{\mu}\big(\boldsymbol{t},\boldsymbol{c},\boldsymbol{Y}(\boldsymbol{t},\boldsymbol{c})\big)$ (Eq. (7)) as a function of age *t*, covariate *c* (shown for *c* = 0), and the hypothetical biomarker $\boldsymbol{Y}(\boldsymbol{t},\boldsymbol{c})$.



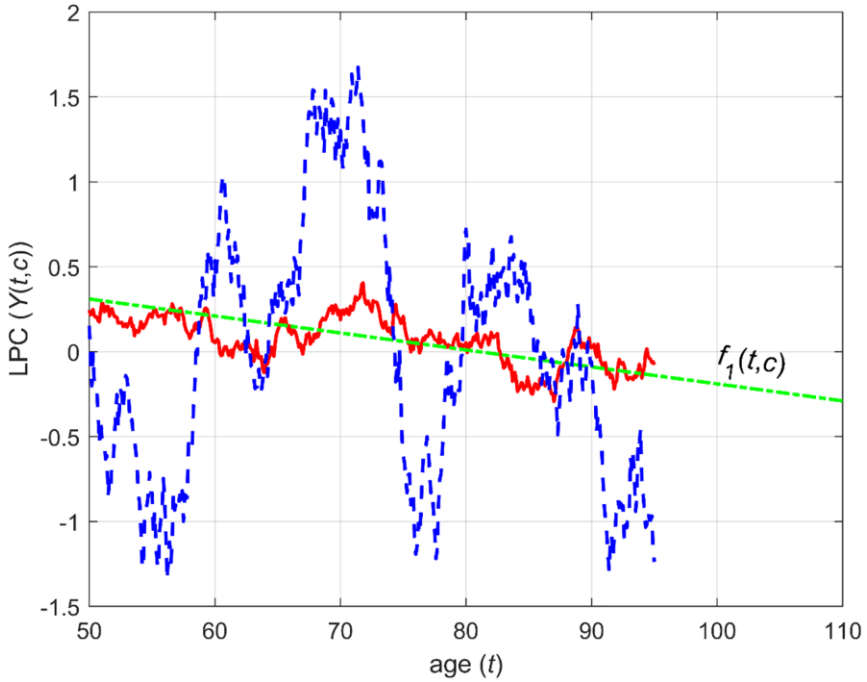
Supplementary Figure 3. Stochastic process models: Illustration of the quadratic part in the hazard corresponding to different age patterns of $\boldsymbol{Q(t,c)}$. (**A**) Age patterns of $\boldsymbol{Q(t,c)}$: age-independent (dashed blue line), declining with age (dash-dotted green line), and increasing with age (solid red line); (**B**) The quadratic part in the hazard in the case of age-independent $\boldsymbol{Q(t,c)}$; (**C**) The quadratic part in the hazard corresponding to $\boldsymbol{Q(t,c)}$ declining with age; (**D**) The quadratic part in the hazard when $\boldsymbol{Q(t,c)}$ is increasing with age. For this illustration, we modeled Eqs. (6–7) with a single binary covariate *c* and parameters corresponding to a hypothetical biomarker $\boldsymbol{Y(t,c)}$ (e.g., a transformed LPC) having the standard normal distribution and hazard rates resembling human mortality rates at old ages. The figures (**B**–**D**) display the quadratic part in the hazard in Eq. (7) (i.e., $\boldsymbol{Q(t,c)\left(Y(t,c)-f_0(t,c)\right)^2}$) as a function of age *t*, covariate *c* (shown for *c* = 0), and the hypothetical biomarker $\boldsymbol{Y(t,c)}$ for respective $\boldsymbol{Q(t,c)}$ modeled as a linear function of age *t* and covariate *c*. We assumed $\boldsymbol{f_0(t,c)=0}$ for simplicity of illustration.



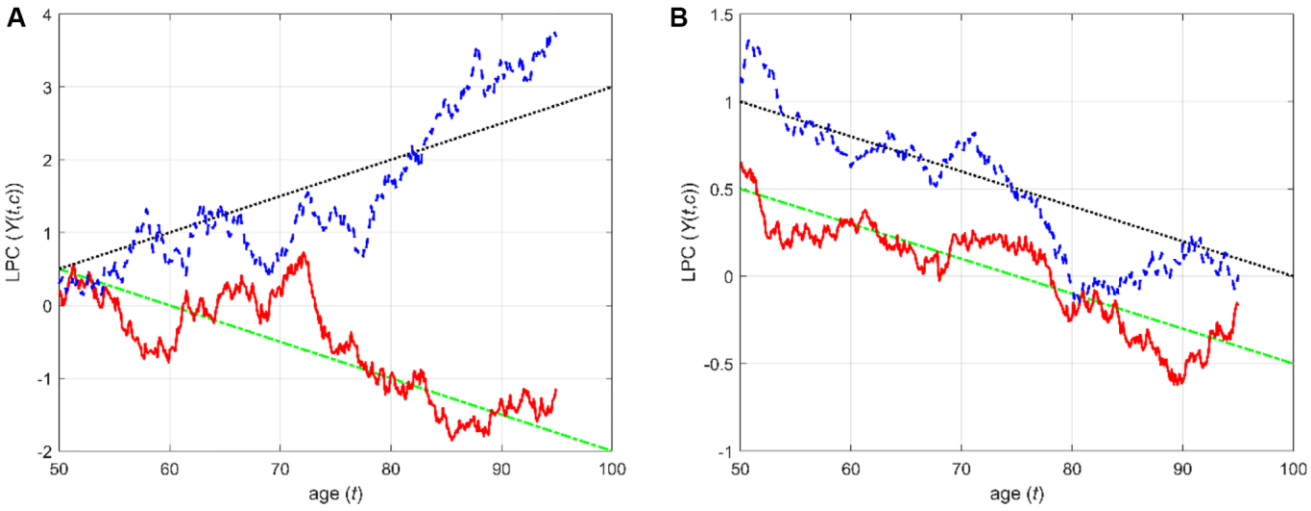
Supplementary Figure 4. Stochastic process models: Illustration of the quadratic part in the hazard corresponding to different patterns of $\boldsymbol{Q(t,c)}$ in females and males. (**A**) Age patterns of $\boldsymbol{Q(t,c)}$ in females (solid red line) and males (dashed blue line); b) The quadratic part in the hazard corresponding to female (surface at the bottom) and male (surface at the top) $\boldsymbol{Q(t,c)}$. For this illustration, we modeled Eqs. (6–7) with a single binary covariate *c* (sex) and parameters corresponding to a hypothetical biomarker $\boldsymbol{Y(t,c)}$ (e.g., a transformed LPC) having the standard normal distribution and hazard rates resembling human mortality rates at old ages. The surfaces in (**B**) show the quadratic part in the hazard in Eq. (7) (i.e., $\boldsymbol{Q(t,c)\left(Y(t,c)-f_0(t,c)\right)^2}$) as a function of age *t*, covariate *c* (*c* = 0 for females, *c* = 1 for males), and the hypothetical biomarker $\boldsymbol{Y(t,c)}$ for respective $\boldsymbol{Q(t,c)}$ modeled as a linear function of age *t* and covariate *c* as shown in Figure A). We assumed $\boldsymbol{f_0(t,c)=0}$ for simplicity of illustration.



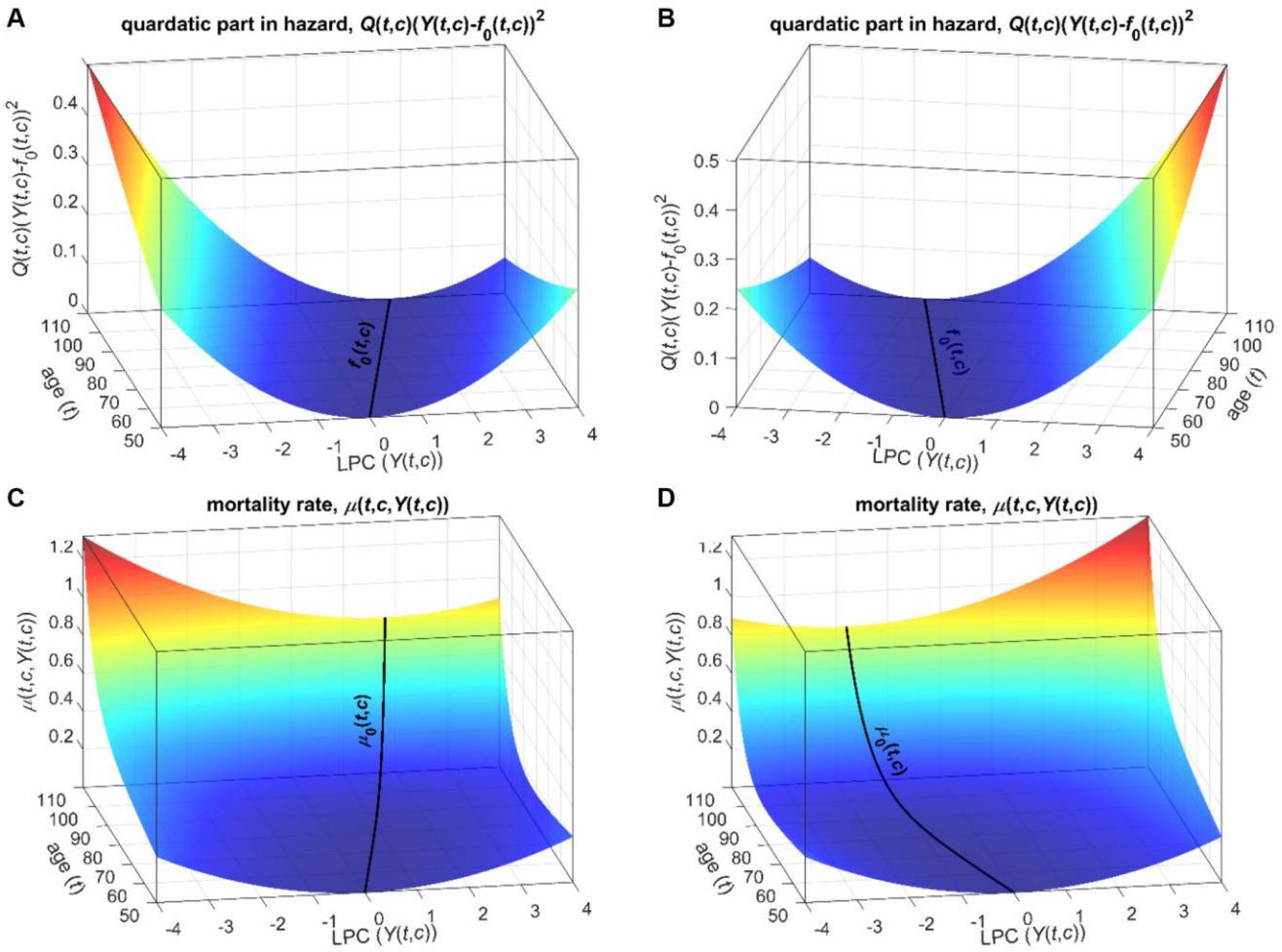
Supplementary Figure 5. Stochastic process models: Illustration of the patterns of the feedback coefficient and sample age trajectories of the process $Y(t, c)$. (**A**) Age-independent (dashed blue line) and declining (solid red line) patterns of the absolute value of the feedback coefficient ($|a(t, c)|$); (**B**) Sample trajectories of $Y(t, c)$ corresponding to respective patterns of $|a(t, c)|$ shown in (A); (**C**) Age patterns of $|a(t, c)|$ for females (solid red line) and males (dashed blue line); (**D**) Sample trajectories of $Y(t, c)$ in females (solid red line) and males (dashed blue line) corresponding to respective patterns of $|a(t, c)|$ shown in (**C**). For this illustration, we modeled Eq. (6) for a hypothetical biomarker $Y(t, c)$ (e.g., a transformed LPC) with respective patterns of the feedback coefficient $a(t, c)$ and a single binary covariate *c* (sex). The sample trajectories of $Y(t, c)$ shown in (**B**) and (**D**) were generated from respective distributions with feedback coefficients shown in (**A**) and (**C**). We assumed that the equilibrium levels $f_1(t, c)$ (shown as dash-dotted green lines in (**B**) and (**D**)) are similar in both cases, for simplicity of illustration.



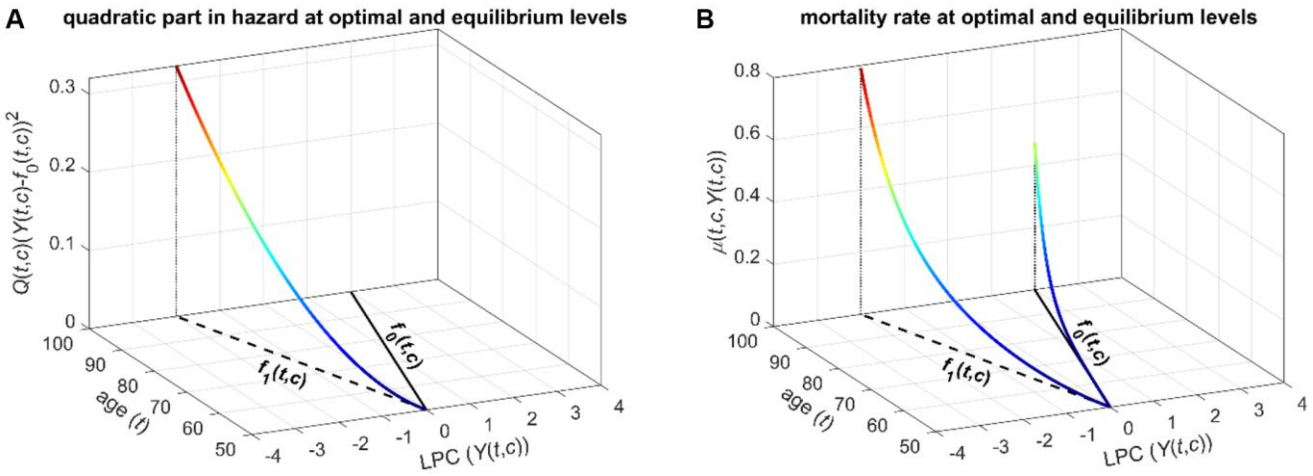
Supplementary Figure 6. Stochastic process models: Examples of trajectories of the stochastic process $\boldsymbol{Y(t,c)}$ with different volatility coefficients. The figure shows sample trajectories of $\boldsymbol{Y(t,c)}$ in the stochastic process models corresponding to higher (dashed blue line) and lower (solid red line) volatility coefficients $\boldsymbol{b(t,c)}$. For this illustration, we modeled Eq. (6) for a hypothetical biomarker $\boldsymbol{Y(t,c)}$ (e.g., a transformed LPC). The sample trajectories of $\boldsymbol{Y(t,c)}$ were generated from respective distributions with different values of the volatility coefficients $\boldsymbol{b(t,c)}$ ($\boldsymbol{b(t,c)=0.1}$ and $\boldsymbol{b(t,c)=0.4}$). We assumed that the equilibrium levels $\boldsymbol{f_1(t,c)}$ (shown as the dash-dotted green line) are similar in both cases, for simplicity of illustration.



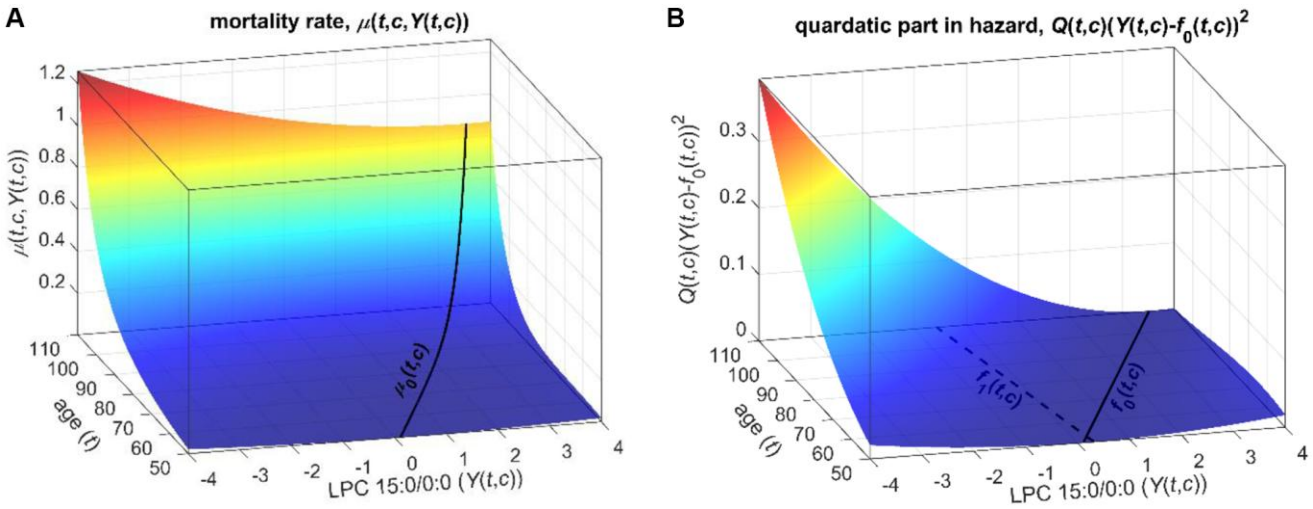
Supplementary Figure 7. Stochastic process models: Examples of trajectories of the stochastic process $\boldsymbol{Y(t,c)}$ with different equilibrium patterns. (**A**) This figure shows two equilibrium trajectories with increasing (black dotted line) and decreasing (green dash-dotted line) age patterns and corresponding sample trajectories of $\boldsymbol{Y(t,c)}$ (blue dashed and red solid lines, respectively); (**B**) This figure shows two equilibrium trajectories shifted by a constant (e.g., representing sex-specific equilibrium levels in the model) (black dotted and green dash-dotted lines) and corresponding sample trajectories of $\boldsymbol{Y(t,c)}$ (blue dashed and red solid lines, respectively). For this illustration, we modeled Eq. (6) for a hypothetical biomarker $\boldsymbol{Y(t,c)}$ (e.g., a transformed LPC) with $\boldsymbol{f_1(t,c)=a_{f_1}+b_{f_1}(t-t_{min})+\beta_{f_1}c}$ and a single binary covariate *c* (sex). The sample trajectories of $\boldsymbol{Y(t,c)}$ were generated from respective distributions with different parameters in $\boldsymbol{f_1(t,c)}$ (positive and negative $\boldsymbol{b_{f_1}}$ and zero $\boldsymbol{\beta_{f_1}}$ in (A), negative $\boldsymbol{b_{f_1}}$ and positive $\boldsymbol{\beta_{f_1}}$ in (B)) so that the black dotted and blue dashed lines correspond to the equilibrium levels and a sample trajectory of $\boldsymbol{Y(t,c)}$ for males).



Supplementary Figure 8. Stochastic process models: Illustration of the mortality rate and the quadratic part in the hazard corresponding to different patterns of $f_0(t,c)$. (**A**) The quadratic part in the hazard in the case of the optimal trajectory increasing with age; (**B**) The quadratic part in the hazard corresponding to the optimal trajectory declining with age; (**C**) The mortality rate when the optimal trajectory increases with age; (**D**) The mortality rate when the optimal trajectory declines with age. For this illustration, we modeled Eqs. (6–7) with a single binary covariate *c* and parameters corresponding to a hypothetical biomarker $Y(t,c)$ (e.g., a transformed LPC) having the standard normal distribution and hazard rates resembling human mortality rates at old ages. Figure A, B show the quadratic parts in the hazard from Eq. (7) (i.e., $Q(t,c)\left(Y(t,c)-f_0(t,c)\right)^2$) as a function of age *t*, covariate *c* (shown for *c* = 0), and the hypothetical biomarker $Y(t,c)$ for increasing (**A**) and declining (**B**) $f_0(t,c)$. The figures also display the optimal trajectories $f_0(t,c)$ (black lines) where the quadratic part equals zero so that the mortality rate is $\mu_0(t,c)$ (Eq. (7)). Respective mortality rates $\mu\left(t,c,Y(t,c)\right)$ (Eq. (7)) are presented in (**C**, **D**). These figures also show the baseline mortality $\mu_0(t,c)$ (black lines) observed at the optimal levels $f_0(t,c)$. For simplicity of illustration, we assumed here that $Q(t,c)$ does not depend on age *t* so that the width of the U-shape of mortality as a function of the biomarker $Y(t,c)$ is the same for all ages.



Supplementary Figure 9. Stochastic process models: Illustration of the quadratic part in the hazard and the mortality rate corresponding to equilibrium and optimal trajectories of the stochastic process $\boldsymbol{Y(t,c)}$. The figure shows the quadratic part in the hazard $\boldsymbol{Q(t,c)\big(Y(t,c)-f_0(t,c)\big)^2}$ (**A**) and the mortality rate $\boldsymbol{\mu\big(t,c,Y(t,c)\big)}$ (**B**) from Eq. (7) for different values of age *t*, covariate *c* = 0, and the hypothetical biomarker $\boldsymbol{Y(t,c)}$ evaluated at the optimal ($\boldsymbol{f_0(t,c)}$, solid black lines) and equilibrium ($\boldsymbol{f_1(t,c)}$, dashed black lines) levels for respective ages *t* and *c* = 0. For this illustration, we modeled Eqs. (6)-(7) with a single binary covariate *c* and parameters corresponding to a hypothetical biomarker $\boldsymbol{Y(t,c)}$ (e.g., a transformed LPC) having the standard normal distribution and hazard rates resembling human mortality rates at old ages.



Supplementary Figure 10: Stochastic process models: 3D plots illustrating results of applications to measurements of LPC 15:0/0:0 and mortality data in the LLFS. (**A**) Estimates of the mortality rate ($\boldsymbol{\mu\big(t,c,Y(t,c)\big)}$) as a function of age (*t*), covariates (*c*), and LPC 15:0/0:0 ($\boldsymbol{Y(t,c)}$); (**B**) Estimates of the quadratic part in the equation for the hazard (mortality) rate ($\boldsymbol{\mu\big(t,c,Y(t,c)\big)-\mu_0(t,c)}$) for different ages (*t*) and values of LPC 15:0/0:0 ($\boldsymbol{Y(t,c)}$). The estimates are shown for zero (mean) values of binary (continuous) covariates *c*. The solid black lines denote the baseline mortality rate $\boldsymbol{\mu_0(t,c)}$ (Supplementary Figure 10A) and the optimal trajectory $\boldsymbol{f_0(t,c)}$ (Supplementary Figure 10B). The dashed black line corresponds to the equilibrium trajectory $\boldsymbol{f_1(t,c)}$. LPC values were transformed (see Data).

Supplementary Tables

Supplementary Table 1. Results of applications of joint models with shared random effects to measurements of LPC species and mortality data in the LLFS: Estimates of the association parameters for the random intercepts and random slopes of the metabolite in the survival sub-model.

Model	Metabolite	Variable	α_0 (α_1)	HR	95% CI for HR	SD of Variable
int	LPC 0:0/16:0	b_{0i}	**−0.229**	**0.900**	**(0.845, 0.963)**	**0.461**
intslope	LPC 0:0/16:0	b_{0i}	**−0.403**	**0.828**	**(0.808, 0.948)**	**0.470**
intslope	LPC 0:0/16:0	b_{1i}	−6.115	0.990	(0.969, 1.007)	0.002
int	LPC 0:0/16:1	b_{0i}	−0.010	0.996	(0.931, 1.063)	0.437
intslope	LPC 0:0/16:1	b_{0i}	**−0.080**	**0.965**	**(0.817, 0.951)**	**0.445**
intslope	LPC 0:0/16:1	b_{1i}	−2.120	0.996	(0.967, 1.008)	0.002
int	LPC 0:0/18:0	b_{0i}	−0.025	0.986	(0.920, 1.053)	0.572
intslope	LPC 0:0/18:0	b_{0i}	−0.203	0.888	(0.880, 1.036)	0.585
intslope	LPC 0:0/18:0	b_{1i}	−3.645	0.988	(0.988, 1.010)	0.003
int	LPC 0:0/18:1	b_{0i}	**−0.148**	**0.931**	**(0.875, 0.996)**	**0.486**
intslope	LPC 0:0/18:1	b_{0i}	−0.333	0.849		0.490
intslope	LPC 0:0/18:1	b_{1i}	−16.596	0.989		0.001
int	LPC 0:0/18:2	b_{0i}	**−0.213**	**0.909**	**(0.849, 0.976)**	**0.447**
intslope	LPC 0:0/18:2	b_{0i}	−0.207	0.911		0.450
intslope	LPC 0:0/18:2	b_{1i}	−1.224	0.999		0.001
int	LPC 0:0/20:3	b_{0i}	−0.114	0.950	(0.894, 1.010)	0.455
intslope	LPC 0:0/20:3	b_{0i}	−0.16	0.916		0.547
intslope	LPC 0:0/20:3	b_{1i}	−0.55	0.989		0.020
int	LPC 0:0/20:4	b_{0i}	**−0.158**	**0.922**	**(0.865, 0.984)**	**0.515**
intslope	LPC 0:0/20:4	b_{0i}	**−0.236**	**0.885**	**(0.847, 0.980)**	**0.519**
intslope	LPC 0:0/20:4	b_{1i}	−5.689	0.995	(0.989, 1.001)	0.001
int	LPC 0:0/22:6	b_{0i}	**−0.179**	**0.91**	**(0.847, 0.969)**	**0.529**
intslope	LPC 0:0/22:6	b_{0i}	**−0.17**	**0.901**	**(0.822, 0.965)**	**0.610**
intslope	LPC 0:0/22:6	b_{1i}	−0.155	0.997	(0.893, 1.076)	0.022
int	LPC 14:0/0:0	b_{0i}	**−0.225**	**0.922**	**(0.864, 0.979)**	**0.360**
intslope	LPC 14:0/0:0	b_{0i}	**−0.240**	**0.905**	**(0.846, 0.973)**	**0.416**
intslope	LPC 14:0/0:0	b_{1i}	−0.376	0.996	(0.521, 1.173)	0.012
int	LPC 15:0/0:0	b_{0i}	**−0.436**	**0.786**	**(0.728, 0.840)**	**0.553**
intslope	LPC 15:0/0:0	b_{0i}	**−0.452**	**0.775**	**(0.726, 0.846)**	**0.563**
intslope	LPC 15:0/0:0	b_{1i}	−0.765	0.998	(0.994, 1.004)	0.002
int	LPC 16:0/0:0	b_{0i}	**−0.230**	**0.889**	**(0.835, 0.947)**	**0.511**
intslope	LPC 16:0/0:0	b_{0i}	−0.369	0.825		0.522
intslope	LPC 16:0/0:0	b_{1i}	−3.785	0.990		0.003
int	LPC 16:1/0:0	b_{0i}	−0.034	0.985	(0.920, 1.053)	0.457
intslope	LPC 16:1/0:0	b_{0i}	−0.249	0.890	(0.891, 1.038)	0.468
intslope	LPC 16:1/0:0	b_{1i}	−4.560	0.988	(0.982, 1.030)	0.003
int	LPC 17:0/0:0	b_{0i}	**−0.229**	**0.888**	**(0.832, 0.949)**	**0.518**
intslope	LPC 17:0/0:0	b_{0i}	−0.318	0.846		0.524
intslope	LPC 17:0/0:0	b_{1i}	−4.138	0.994		0.002
int	LPC 18:0/0:0	b_{0i}	0.022	1.014	(0.945, 1.080)	0.620
intslope	LPC 18:0/0:0	b_{0i}	0.014	1.009	(0.927, 1.070)	0.625
intslope	LPC 18:0/0:0	b_{1i}	0.217	1.001	(0.994, 1.013)	0.005

int	LPC 18:1/0:0	b_{0i}	**−0.238**	**0.886**	**(0.831, 0.949)**	**0.506**
intslope	LPC 18:1/0:0	b_{0i}	**−0.430**	**0.801**	**(0.804, 0.944)**	**0.515**
intslope	LPC 18:1/0:0	b_{1i}	−5.637	0.988	(0.985, 1.008)	0.002
int	LPC 18:2/0:0	b_{0i}	**−0.277**	**0.883**	**(0.825, 0.943)**	**0.450**
intslope	LPC 18:2/0:0	b_{0i}	**−0.276**	**0.883**	**(0.817, 0.942)**	**0.452**
intslope	LPC 18:2/0:0	b_{1i}	−1.238	0.999	(0.995, 1.002)	0.001
int	LPC 18:3/0:0	b_{0i}	−0.144	0.953	(0.886, 1.020)	0.336
intslope	LPC 18:3/0:0	b_{0i}	−0.148	0.950		0.346
intslope	LPC 18:3/0:0	b_{1i}	−0.640	0.998		0.003
int	LPC 20:2/0:0	b_{0i}	**−0.221**	**0.904**	**(0.849, 0.969)**	**0.458**
intslope	LPC 20:2/0:0	b_{0i}	**−0.227**	**0.900**	**(0.815, 0.960)**	**0.462**
intslope	LPC 20:2/0:0	b_{1i}	−1.342	0.999	(0.988, 1.002)	0.001
int	LPC 20:3/0:0	b_{0i}	**−0.235**	**0.898**	**(0.842, 0.953)**	**0.458**
intslope	LPC 20:3/0:0	b_{0i}	**−0.248**	**0.891**	**(0.807, 0.947)**	**0.463**
intslope	LPC 20:3/0:0		−1.616	0.998	(0.984, 1.000)	0.001
int	LPC 20:4/0:0	b_{0i}	**−0.190**	**0.912**	**(0.855, 0.974)**	**0.487**
intslope	LPC 20:4/0:0	b_{0i}	**−0.183**	**0.914**	**(0.822, 0.976)**	**0.488**
intslope	LPC 20:4/0:0	b_{1i}	−1.252	0.999	(0.994, 1.005)	0.001
int	LPC 20:5/0:0	b_{0i}	**−0.230**	**0.896**	**(0.841, 0.957)**	**0.478**
intslope	LPC 20:5/0:0	b_{0i}	**−0.254**	**0.885**	**(0.830, 0.959)**	**0.483**
intslope	LPC 20:5/0:0	b_{1i}	−1.573	0.998	(0.994, 1.004)	0.001
int	LPC 22:5/0:0	b_{0i}	**−0.215**	**0.892**	**(0.838, 0.955)**	**0.530**
intslope	LPC 22:5/0:0	b_{0i}	−0.291	0.856		0.536
intslope	LPC 22:5/0:0	b_{1i}	−4.164	0.994		0.001
int	LPC 22:6/0:0	b_{0i}	**−0.210**	**0.892**	**(0.830, 0.951)**	**0.546**
intslope	LPC 22:6/0:0	b_{0i}	**−0.210**	**0.892**	**(0.826, 0.951)**	**0.546**
intslope	LPC 22:6/0:0	b_{1i}	0.369	1.000	(0.999, 1.002)	0.0004

Model – type of joint model (int – random intercept of LPC in survival sub-model, intslope – random intercept and slope of LPC in survival sub-model), see section Joint models: General specifications; Variable – b_{0i}: random intercept of the metabolite, b_{1i}: random slope of the metabolite; α_0 (α_1) – estimates of the regression parameters for b_{0i} (b_{1i}) in respective models; HR – hazard ratios for an increase by a standard deviation of Variable; 95% CI for HR – respective 95% confidence intervals for HRs; SD of Variable – standard deviation of Variable. Highlighted in bold are cases where confidence intervals do not contain one. The JM were estimated using the R-package *joineR*. LPC values were transformed (see section Data). Note that for some LPCs, CIs are not available due to technical issues encountered in applications of the R-package *joineR*.

Supplementary Table 2. Results of applications of the stochastic process model to measurements of LPC species and mortality data in the LLFS: Results of testing different null hypotheses on age patterns of the model's components.

LPC	H0						
	Qzero	QnoT	AnoT	F1noT	F0noT	ALzero	ALnoT
LPC 0:0/16:0	0.001	0.64	0.52	<0.0001	0.87	0.88	<0.0001
LPC 0:0/16:1	0.010	1	0.83	<0.0001	0.79	0.97	<0.0001
LPC 0:0/18:0	0.011	1	0.46	0.0002	0.22	0.59	0.0005
LPC 0:0/18:1	0.049	1	1	<0.0001	0.15	0.54	<0.0001
LPC 0:0/18:2	0.031	1	1	<0.0001	0.13	0.055	<0.0001
LPC 0:0/20:3	0.005	0.48	1	<0.0001	0.015	0.074	<0.0001
LPC 0:0/20:4	0.007	0.20	1	<0.0001	0.013	0.018	<0.0001
LPC 0:0/22:6	0.005	0.34	0.37	<0.0001	0.44	0.009	<0.0001

LPC 14:0/0:0	<0.0001	0.25	0.08	<0.0001	0.31	0.31	<0.0001
LPC 15:0/0:0	<0.0001	0.0002	1	<0.0001	0.0002	<0.0001	<0.0001
LPC 16:0/0:0	<0.0001	0.48	0.78	<0.0001	0.45	0.53	<0.0001
LPC 16:1/0:0	0.003	1	0.48	<0.0001	0.81	0.96	<0.0001
LPC 17:0/0:0	0.005	0.43	1	<0.0001	0.63	0.37	<0.0001
LPC 18:0/0:0	0.033	1	1	0.040	0.14	0.37	0.040
LPC 18:1/0:0	0.007	0.19	1	<0.0001	0.23	0.44	<0.0001
LPC 18:2/0:0	0.008	0.11	1	<0.0001	0.019	0.091	<0.0001
LPC 18:3/0:0	0.044	1	1	<0.0001	0.16	0.082	<0.0001
LPC 20:2/0:0	0.002	1	1	<0.0001	1	0.85	<0.0001
LPC 20:3/0:0	0.002	0.10	0.94	<0.0001	0.09	0.34	<0.0001
LPC 20:4/0:0	0.004	0.20	1	<0.0001	0.011	0.014	<0.0001
LPC 20:5/0:0	0.009	0.27	1	<0.0001	0.39	0.012	<0.0001
LPC 22:5/0:0	0.0008	0.047	1	<0.0001	0.005	0.014	<0.0001
LPC 22:6/0:0	0.007	1	1	<0.0001	0.69	0.017	<0.0001

Notes: The table shows results (p-values) of testing the following null hypotheses (H0) corresponding to one or more restrictions on parameters of SPM (see Stochastic process models: Specific parameterizations used in applications): H0: $Q(t,c) = 0$ (Qzero); H0: $Q(t,c) = Q(c)$ (QnoT); H0: $a(t,c) = a(c)$ (AnoT); H0: $f_1(t,c) = f_1(c)$ (F1noT); H0: $f_0(t,c) = f_0(c)$ (F0noT); H0: $f_1(t,c) = f_0(t,c)$, i.e., $AL(t,c) = 0$ (ALzero); H0: $f_1(t,c) = f_1(c)$ and $f_0(t,c) = f_0(c)$, i.e., $AL(t,c) = AL(c)$ (ALnoT).

Supplementary Table 3. Results of applications of the stochastic process model to measurements of LPC species and mortality data in the LLFS: Results of testing different null hypotheses on sex-dependence of the model's components.

LPC	H0						
	MU0noC	QnoC	AnoC	BnoC	F1noC	F0noC	ALLnoC
LPC 0:0/16:0	<0.0001	0.09	0.57	0.40	0.56	0.57	<0.0001
LPC 0:0/16:1	<0.0001	0.037	0.18	0.75	<0.0001	1	<0.0001
LPC 0:0/18:0	<0.0001	0.10	0.59	0.58	<0.0001	1	<0.0001
LPC 0:0/18:1	<0.0001	0.53	0.61	0.63	0.63	1	<0.0001
LPC 0:0/18:2	<0.0001	0.35	0.40	0.14	<0.0001	1	<0.0001
LPC 0:0/20:3	<0.0001	0.05	0.71	0.83	0.16	1	<0.0001
LPC 0:0/20:4	<0.0001	0.06	0.35	0.87	0.007	1	<0.0001
LPC 0:0/22:6	<0.0001	0.94	0.77	0.51	0.10	0.37	<0.0001
LPC 14:0/0:0	<0.0001	0.008	0.16	0.74	<0.0001	0.29	<0.0001
LPC 15:0/0:0	<0.0001	0.13	0.12	0.40	0.001	0.16	<0.0001
LPC 16:0/0:0	<0.0001	0.07	0.82	0.40	0.07	0.50	<0.0001
LPC 16:1/0:0	<0.0001	0.011	0.18	0.33	<0.0001	1	<0.0001
LPC 17:0/0:0	<0.0001	0.29	0.39	0.66	0.0004	0.54	<0.0001
LPC 18:0/0:0	<0.0001	0.11	0.85	0.32	<0.0001	1	<0.0001
LPC 18:1/0:0	<0.0001	0.08	0.85	0.96	0.83	0.65	<0.0001
LPC 18:2/0:0	<0.0001	0.11	0.36	0.16	<0.0001	0.50	<0.0001
LPC 18:3/0:0	<0.0001	0.38	0.84	0.77	1	1	<0.0001
LPC 20:2/0:0	<0.0001	0.025	0.37	0.66	0.46	1	<0.0001
LPC 20:3/0:0	<0.0001	0.05	0.53	0.14	0.018	0.68	<0.0001
LPC 20:4/0:0	<0.0001	0.041	0.72	0.46	<0.0001	1	<0.0001
LPC 20:5/0:0	<0.0001	0.39	0.31	0.37	0.018	0.36	<0.0001
LPC 22:5/0:0	<0.0001	0.043	0.63	0.26	<0.0001	0.81	<0.0001
LPC 22:6/0:0	<0.0001	0.76	0.42	0.71	0.14	1	<0.0001

The table shows results (p-values) of testing the following null hypotheses (H0) corresponding to one or more restrictions on parameters of SPM (see Stochastic process models: Specific parameterizations used in applications): H0: $\mu_0(t,c) = \mu_0(t)$ (MU0noC); H0: $Q(t,c) = Q(t)$ (QnoC); H0: $a(t,c) = a(t)$ (AnoC); H0: $b(t,c) = b(t)$ (BnoC); H0: $f_1(t,c) = f_1(t)$ (F1noC); H0: $f_0(t,c) = f_0(t)$ (F0noC); H0: $\mu_0(t,c) = \mu_0(t)$ and $Q(t,c) = Q(t)$ and $a(t,c) = a(t)$ and $b(t,c) = b(t)$ and $f_1(t,c) = f_1(c)$ and $f_0(t,c) = f_0(c)$ (ALLnoC). Here, c denotes variable SexM in respective components. Note that there are other covariates in $\mu_0(t,c)$ (see Stochastic process models: Specific parameterizations used in applications), which still remain in the restricted models in MU0noC and ALLnoC; for brevity of notation, we do not show them in the formulae above.

Supplementary Table 4. Results of applications of the stochastic process model to measurements of LPC species and mortality data in the LLFS: Estimates of parameters in different components of the model.

LPC	$\mu_0(t,c)$								
	$\ln a_{\mu_0}$	b_{μ_0}	β_{μ_0} (SexM)	β_{μ_0} (IsDK)	β_{μ_0} (LowEduc)	β_{μ_0} (Smoke100)	β_{μ_0} (MedsLipid)	β_{μ_0} (MedsHtn)	β_{μ_0} (MedsNitro)
LPC 0:0/16:0	−4.681	0.119	0.491	0.077	−0.069	0.204	−0.235	−0.184	0.474
LPC 0:0/16:1	−4.721	0.118	0.499	0.082	−0.063	0.202	−0.225	−0.184	0.468
LPC 0:0/18:0	−4.815	0.117	0.511	0.076	−0.063	0.199	−0.225	−0.181	0.470
LPC 0:0/18:1	−4.928	0.117	0.476	0.070	−0.062	0.205	−0.221	−0.182	0.459
LPC 0:0/18:2	−4.920	0.117	0.481	0.073	−0.063	0.204	−0.221	−0.183	0.461
LPC 0:0/20:3	−4.688	0.117	0.498	0.081	−0.064	0.209	−0.228	−0.187	0.469
LPC 0:0/20:4	−4.631	0.116	0.512	0.074	−0.068	0.209	−0.227	−0.191	0.478
LPC 0:0/22:6	−4.833	0.119	0.470	0.071	−0.062	0.205	−0.217	−0.178	0.462
LPC 14:0/0:0	−4.576	0.120	0.495	0.079	−0.074	0.212	−0.249	−0.191	0.488
LPC 15:0/0:0	−3.580	0.127	0.560	−0.020	−0.117	0.225	−0.384	−0.244	0.629
LPC 16:0/0:0	−4.617	0.120	0.502	0.079	−0.074	0.205	−0.245	−0.184	0.481
LPC 16:1/0:0	−4.707	0.118	0.502	0.082	−0.066	0.201	−0.227	−0.184	0.470
LPC 17:0/0:0	−4.864	0.119	0.477	0.057	−0.064	0.205	−0.232	−0.184	0.469
LPC 18:0/0:0	−4.825	0.117	0.513	0.072	−0.061	0.200	−0.226	−0.182	0.470
LPC 18:1/0:0	−4.314	0.116	0.509	0.098	−0.074	0.219	−0.251	−0.197	0.498
LPC 18:2/0:0	−4.282	0.116	0.509	0.064	−0.071	0.215	−0.272	−0.203	0.501
LPC 18:3/0:0	−4.896	0.117	0.477	0.072	−0.062	0.205	−0.219	−0.182	0.459
LPC 20:2/0:0	−4.758	0.118	0.503	0.071	−0.065	0.203	−0.233	−0.184	0.470
LPC 20:3/0:0	−4.422	0.115	0.503	0.078	−0.073	0.214	−0.252	−0.196	0.490
LPC 20:4/0:0	−4.559	0.117	0.514	0.073	−0.070	0.210	−0.225	−0.188	0.477
LPC 20:5/0:0	−4.396	0.115	0.576	0.133	−0.067	0.219	−0.228	−0.210	0.515
LPC 22:5/0:0	−4.404	0.116	0.542	0.086	−0.070	0.209	−0.242	−0.205	0.503
LPC 22:6/0:0	−4.819	0.118	0.491	0.084	−0.065	0.204	−0.215	−0.182	0.463

Supplementary Table 4. (Continued)

LPC	$\mu_0(t,c)$								$Q(t,c)$		
	β_{μ_0} (MedsDiab)	β_{μ_0} (APOE4)	β_{μ_0} (NoIntPA)	β_{μ_0} (SPPB)	β_{μ_0} (PREV6)	β_{μ_0} (BMI)	β_{μ_0} (PC1)	β_{μ_0} (PC2)	a_Q	b_Q	β_Q
LPC 0:0/16:0	0.353	0.204	0.032	−0.140	0.191	−0.024	−0.304	−0.125	1.1E-03	4.8E-06	−1.1E-03
LPC 0:0/16:1	0.349	0.203	0.034	−0.140	0.183	−0.023	−0.239	−0.093	1.1E-03	3.3E-11	−1.1E-03
LPC 0:0/18:0	0.354	0.206	0.031	−0.141	0.191	−0.024	−0.158	−0.044	7.9E-04	6.2E-12	−7.9E-04
LPC 0:0/18:1	0.340	0.204	0.034	−0.140	0.199	−0.024	−0.077	0.131	8.8E-05	2.4E-12	−8.8E-05
LPC 0:0/18:2	0.340	0.205	0.032	−0.139	0.200	−0.024	−0.096	0.103	1.1E-04	6.2E-15	−1.1E-04
LPC 0:0/20:3	0.353	0.205	0.035	−0.141	0.199	−0.024	−0.228	−0.150	4.2E-04	1.2E-05	−4.2E-04

LPC 0:0/20:4	0.357	0.211	0.040	−0.145	0.215	−0.025	−0.241	−0.123	5.7E-04	2.4E-05	−5.7E-04
LPC 0:0/22:6	0.354	0.214	0.031	−0.139	0.193	−0.024	−0.201	−0.050	2.0E-04	-3.4E-06	2.0E-05
LPC 14:0/0:0	0.364	0.219	0.038	−0.142	0.210	−0.027	−0.340	−0.319	1.3E-03	1.2E-05	−1.3E-03
LPC 15:0/0:0	0.424	0.286	0.013	−0.168	0.181	−0.046	−0.799	−0.869	1.2E-03	9.6E-05	−1.1E-03
LPC 16:0/0:0	0.356	0.205	0.031	−0.141	0.185	−0.025	−0.372	−0.160	1.3E-03	1.7E-05	−1.3E-03
LPC 16:1/0:0	0.348	0.200	0.033	−0.140	0.185	−0.023	−0.245	−0.131	1.2E-03	1.1E-11	−1.2E-03
LPC 17:0/0:0	0.347	0.207	0.036	−0.140	0.190	−0.024	−0.178	0.058	4.0E-04	6.0E-06	−4.0E-04
LPC 18:0/0:0	0.353	0.206	0.031	−0.142	0.195	−0.024	−0.124	−0.007	5.8E-04	4.7E-12	−5.8E-04
LPC 18:1/0:0	0.349	0.205	0.039	−0.148	0.201	−0.030	−0.389	−0.237	6.9E-04	2.7E-05	−6.9E-04
LPC 18:2/0:0	0.348	0.210	0.044	−0.149	0.219	−0.030	−0.377	−0.311	6.6E-04	3.6E-05	−6.6E-04
LPC 18:3/0:0	0.343	0.204	0.031	−0.139	0.198	−0.024	−0.102	0.087	9.9E-05	1.0E-12	−9.9E-05
LPC 20:2/0:0	0.342	0.201	0.030	−0.140	0.193	−0.024	−0.219	−0.057	1.0E-03	5.0E-12	−1.0E-03
LPC 20:3/0:0	0.362	0.202	0.047	−0.149	0.215	−0.028	−0.313	−0.216	5.6E-04	3.8E-05	−5.6E-04
LPC 20:4/0:0	0.363	0.208	0.035	−0.145	0.210	−0.025	−0.304	−0.198	5.4E-04	2.4E-05	−5.4E-04
LPC 20:5/0:0	0.348	0.207	0.041	−0.153	0.203	−0.032	−0.318	−0.105	3.6E-04	3.9E-05	−3.6E-04
LPC 22:5/0:0	0.351	0.209	0.043	−0.150	0.219	−0.029	−0.286	−0.293	7.0E-04	4.0E-05	−7.0E-04
LPC 22:6/0:0	0.347	0.205	0.030	−0.139	0.195	−0.024	−0.185	−0.016	2.2E-04	1.5E-11	−2.2E-04

Supplementary Table 4. (Continued)

LPC	$a(t,c)$		$Y(t_0,c)$	$b(t,c)$		$f_1(t,c)$			$f_0(t,c)$		
	a_Y	b_Y	β_Y	σ_0	σ_1	β_W	a_{f_1}	b_{f_1}	β_{f_1}	a_{f_0}	b_{f_0}	β_{f_0}
LPC 0:0/16:0	−0.075	1.6E-04	2.0E-04	1.002	0.305	9.3E-03	0.237	−0.012	−0.018	0.200	0.005	2.4E+00
LPC 0:0/16:1	−0.077	5.4E-05	7.9E-03	0.982	0.326	3.7E-03	0.252	−0.007	−0.256	0.154	−0.011	−2.1E-01
LPC 0:0/18:0	−0.063	1.7E-05	2.9E-03	1.006	0.302	−5.9E-03	0.170	−0.004	−0.146	0.298	−0.050	4.9E-01
LPC 0:0/18:1	−0.063	4.9E-10	−3.0E-03	0.987	0.314	5.5E-03	0.198	−0.010	−0.015	4.000	−0.133	−2.6E+00
LPC 0:0/18:2	−0.064	5.9E-13	−5.1E-03	0.952	0.304	1.6E-02	0.233	−0.017	0.188	4.000	−0.133	−2.4E+00
LPC 0:0/20:3	−0.075	7.6E-09	2.1E-03	0.991	0.317	−2.4E-03	0.223	−0.011	−0.044	−1.764	0.096	−7.1E-06
LPC 0:0/20:4	−0.064	2.7E-10	5.2E-03	0.992	0.297	−1.8E-03	0.181	−0.012	0.086	−1.240	0.087	4.5E-06
LPC 0:0/22:6	−0.052	7.5E-10	−1.5E-03	0.989	0.284	6.8E-03	0.139	−0.005	−0.054	2.912	−0.068	−3.4E+00
LPC 14:0/0:0	−0.089	4.7E-04	−4.9E-03	0.980	0.326	−3.8E-03	0.343	−0.013	−0.184	0.242	0.026	1.8E+00
LPC 15:0/0:0	−0.057	4.8E-10	−8.1E-03	0.986	0.276	8.3E-03	0.314	−0.015	−0.104	0.111	0.057	4.8E-01
LPC 16:0/0:0	−0.069	6.7E-05	1.3E-03	1.005	0.298	9.2E-03	0.246	−0.011	−0.057	0.088	0.020	1.1E+00
LPC 16:1/0:0	−0.077	1.8E-04	7.9E-03	0.984	0.316	1.3E-02	0.240	−0.006	−0.278	−0.005	0.009	−3.0E-01
LPC 17:0/0:0	−0.066	1.1E-10	1.8E-03	1.016	0.298	−4.7E-03	0.163	−0.005	−0.115	1.579	−0.033	1.5E+00
LPC 18:0/0:0	−0.049	2.9E-09	−1.0E-03	0.996	0.299	−1.1E-02	0.141	−0.002	−0.145	0.441	−0.074	5.5E-01
LPC 18:1/0:0	−0.064	1.4E-10	1.1E-03	0.987	0.308	5.6E-04	0.232	−0.012	−0.007	−0.532	0.068	4.6E-01
LPC 18:2/0:0	−0.065	1.4E-10	−5.4E-03	0.953	0.301	1.5E-02	0.222	−0.017	0.222	−0.848	0.070	6.2E-01
LPC 18:3/0:0	−0.082	3.4E-10	1.2E-03	0.967	0.331	−3.4E-03	0.250	−0.013	0.000	4.000	−0.133	−2.2E+00
LPC 20:2/0:0	−0.071	1.8E-10	5.2E-03	1.011	0.314	5.0E-03	0.160	−0.008	−0.024	0.323	0.001	−7.0E-02
LPC 20:3/0:0	−0.074	2.0E-05	3.7E-03	0.980	0.307	1.7E-02	0.242	−0.015	0.074	−0.970	0.076	3.8E-01
LPC 20:4/0:0	−0.060	3.7E-11	1.9E-03	0.994	0.288	7.6E-03	0.121	−0.011	0.175	−1.516	0.092	2.4E-06
LPC 20:5/0:0	−0.048	4.0E-11	−5.0E-03	0.975	0.271	8.7E-03	0.241	−0.013	0.075	0.420	0.060	−8.0E-01
LPC 22:5/0:0	−0.065	5.4E-12	2.8E-03	0.983	0.301	1.2E-02	0.108	−0.008	0.142	−0.913	0.082	−1.7E-01
LPC 22:6/0:0	−0.055	3.5E-11	4.3E-03	0.985	0.286	3.9E-03	0.120	−0.006	0.049	2.632	−0.058	−1.8E+00

Notes: Columns $\beta_{\mu_0}(.)$ show coefficients for the variables in the baseline hazard rate ($\mu_0(t,c)$): SexM: sex (1 – male, 0 – female); IsDK: country (1 – Denmark, 0 – USA); LowEduc: low education (1 – below high school, 0 – otherwise); Smoke100: smoking (1 – smoked 100 cigarettes in lifetime, 0 - otherwise); MedsLipid: lipid-lowering medications (1 – taking, 0 – not taking); MedsHtn: hypertension medications (1 – taking, 0 – not taking); MedsNitro: angina medications (1 – taking, 0 – not taking); MedsDiab: diabetes mellitus medications (1 – taking,

0 – not taking); APOE4: *APOE* ε4 carrier status (1 – carrier, 0 – non-carrier); NoIntPA: no intense physical activity (PA) at baseline (1 – no intense PA, 0 – intense PA); SPPB: Short Physical Performance Battery (SPPB) total score at baseline; PREV6: prevalence of major diseases (heart disease, stroke, lung disease, cancer, hypertension, diabetes) at baseline (1: any of the diseases, 0 – none of the diseases); BMI: body mass index (BMI) at baseline; PC1, PC2: first two principal components computed from LLFS whole-genome sequencing data. In other components, respective β's show coefficients for variable SexM. LPC values were transformed (see Data).

Supplementary Table 5. Applications of joint models to measurements of LPC variants and mortality data in the LLFS: Estimates from familial bootstrap.

Metabolite	Median HR (Range)		
	Total	Females	Males
LPC 0:0/16:0	**0.830 (0.743, 0.934)**	**0.843 (0.723, 0.971)**	**0.813 (0.694, 0.968)**
LPC 0:0/16:1	0.961 (0.842, 1.079)	0.989 (0.844, 1.134)	0.935 (0.775, 1.074)
LPC 0:0/18:0	0.977 (0.896, 1.075)	1.003 (0.868, 1.142)	0.942 (0.817, 1.070)
LPC 0:0/18:1	**0.884 (0.796, 0.994)**	0.923 (0.790, 1.083)	**0.831 (0.685, 0.995)**
LPC 0:0/18:2	**0.824 (0.718, 0.942)**	0.910 (0.749, 1.090)	**0.735 (0.611, 0.897)**
LPC 0:0/20:3	**0.852 (0.764, 0.973)**	**0.858 (0.735, 0.976)**	**0.836 (0.691, 0.993)**
LPC 0:0/20:4	**0.860 (0.778, 0.962)**	**0.852 (0.747, 0.986)**	**0.867 (0.736, 0.981)**
LPC 0:0/22:6	**0.842 (0.757, 0.921)**	**0.799 (0.677, 0.900)**	**0.869 (0.735, 0.986)**
LPC 14:0/0:0	**0.828 (0.738, 0.947)**	**0.858 (0.723, 0.995)**	**0.801 (0.670, 0.982)**
LPC 15:0/0:0	**0.713 (0.643, 0.778)**	**0.732 (0.638, 0.850)**	**0.690 (0.591, 0.772)**
LPC 16:0/0:0	**0.831 (0.768, 0.915)**	**0.848 (0.732, 0.998)**	**0.815 (0.739, 0.934)**
LPC 16:1/0:0	0.951 (0.836, 1.056)	0.981 (0.828, 1.135)	0.924 (0.794, 1.044)
LPC 17:0/0:0	**0.821 (0.737, 0.913]**	**0.828 (0.712, 0.955)**	**0.817 (0.705, 0.973)**
LPC 18:0/0:0	1.014 (0.913, 1.115)	1.054 (0.931, 1.214)	0.976 (0.831, 1.142)
LPC 18:1/0:0	**0.820 (0.742, 0.924)**	0.863 (0.731, 1.043)	**0.794 (0.666, 0.932)**
LPC 18:2/0:0	**0.785 (0.678, 0.871)**	0.870 (0.711, 1.049)	**0.707 (0.563, 0.838)**
LPC 18:3/0:0	0.888 (0.756, 1.025)	0.987 (0.855, 1.240)	**0.780 (0.625, 0.985)**
LPC 20:2/0:0	**0.820 (0.715, 0.910)**	0.856 (0.765, 1.048)	**0.774 (0.659, 0.869)**
LPC 20:3/0:0	**0.767 (0.680, 0.873)**	**0.798 (0.657, 0.950)**	**0.746 (0.603, 0.882)**
LPC 20:4/0:0	**0.861 (0.764, 0.949)**	0.897 (0.785, 1.049)	0.808 (0.684, 1.004)
LPC 20:5/0:0	**0.776 (0.696, 0.870)**	**0.742 (0.619, 0.864)**	**0.810 (0.698, 0.989)**
LPC 22:5/0:0	**0.793 (0.714, 0.857)**	**0.781 (0.670, 0.883)**	**0.796 (0.680, 0.898)**
LPC 22:6/0:0	**0.805 (0.733, 0.892)**	**0.759 (0.674, 0.912)**	**0.829 (0.731, 0.936)**

Notes: The table reports medians of hazard ratios (HR) for a unit increase in the transformed metabolite values (ranges in parentheses) for the association parameters for the metabolites in the survival sub-model computed using the familial bootstrap method [18] in 100 bootstrap samples generated from the original sample. See main text (section Materials and Methods: Sensitivity analyses) for details. The joint models were estimated using the R-package *JM*. The cases where the range of HR does not contain one are highlighted in bold. The case where 95% confidence intervals (CI) for HR in the main calculations (Table 1) did not contain one but the HR range in the familial bootstrap included one is highlighted with a yellow background. The cases where 95% CI for HR in the main calculations (Table 1) contained one but the HR range in the familial bootstrap did not include one are highlighted with a grey background.

Supplementary Table 6. Characteristics of the Long Life Family Study metabolomics sample (batch 6, released on October 25, 2023)

Characteristics	Field Center				Total Sample
	BU	NY	PT	DK	
Number of families	244	263	222	77	582
Number of participants at any visit	1,282	875	1,185	1,239	4,581
Number of participants at visit 1	1,176	718	1,131	1,196	4,221

Number of participants at visit 2	682	490	585	798	2,555
Number of participants with genetic PCs	1,270	847	1,174	1,229	4,520
Number (%) of deceased participants	435 (33.9%)	358 (40.9%)	477 (40.3%)	388 (31.3%)	1,658 (36.2%)
Follow-up period (years) (mean ± SD (range))	10.5 ± 4.8 (0.00, 18.22)	10.1 ± 4.5 (0.00, 18.41)	10.8 ± 4.8 (0.29, 18.50)	11.7 ± 5.5 (0.00, 17.93)	10.8 ± 5.0 (0.00, 18.50)
Age at baseline (mean ± SD (range))	69.5 ± 16.0 (32, 110)	73.5 ± 16.1 (24, 108)	71.2 ± 15.9 (36, 104)	67.3 ± 14.3 (36, 104)	70.0 ± 15.7 (24, 110)
Whites (%)	99.2%	98.3%	99.6%	99.0%	99.1%
Females (%)	55.5%	54.6%	55.9%	54.3%	55.1%
Low educated participants (below high school) (%)	5.9%	7.4%	7.2%	27.7%	12.4%
Smokers (smoked >100 cigarettes in lifetime) (%)	41.5%	45.5%	35.6%	49.2%	42.8%
APOE ε4 allele carriers (%)	13.7%	17.3%	17.2%	25.3%	18.4%
Medication use: angina (%)	32.1%	30.4%	32.8%	23.5%	29.6%
Medication use: anti-diabetic (%)	6.8%	7.5%	9.1%	5.5%	7.2%
Medication use: anti-hypertensive (%)	50.4%	52.5%	55.1%	42.0%	49.7%
Medication use: lipid-lowering (%)	34.9%	43.3%	39.2%	21.1%	33.9%
No intense physical activity at baseline (%)	64.2%	56.9%	62.3%	74.0%	65.0%
SPPB total score at baseline (mean ± SD (range))	10.0 ± 3.0 (0, 12)	9.6 ± 3.0 (0, 12)	9.8 ± 3.0 (1, 12)	10.4 ± 2.8 (1, 12)	10.0 ± 3.0 (0, 12)
Prevalence of major diseases at baseline (%)	68.9%	70.4%	69.8%	66.5%	68.8%
BMI at baseline (mean ± SD (range))	27.5 ± 5.1 (16, 57)	26.6 ± 4.2 (17, 41)	27.7 ± 5.2 (17, 52)	26.4 ± 4.2 (13, 54)	27.1 ± 4.8 (13, 57)

Notes: (a) Genetic PCs were computed from LLFS whole-genome sequencing data; (b) Number of missing data: race – 20, education – 10, smoking – 18, *APOE* – 257, angina medications – 273, anti-diabetic drugs – 273, anti-hypertensive drugs – 273, lipid-lowering drugs – 273, no intense physical activity at baseline – 449, SPPB total score at baseline – 179, prevalence of major diseases at baseline – 2, other variables listed in the table have no missing values; (c) The numbers shown in "Number of deceased participants" and "Follow-up period" correspond to the LLFS data release used in this paper (see Data). Abbreviations: BMI: body mass index; BU: Boston; DK: Denmark; NY: New York; PT: Pittsburgh; SD: standard deviation; SPPB: Short Physical Performance Battery.