

Compression of morbidity by interventions that steepen the survival curve

Corresponding Author: Professor Uri Alon

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

I'm glad someone is giving thought to this idea, which is straight forward in intent -- to find a predictive tool that links prospective gerotherapeutics to the goal of compressing the sickspan. I have some specific suggestions.

1. As currently stated, the authors indicate that healthspan is followed by sickspan. This rarely happens as discrete sequential events. Most people transition into and out of sickspan during the course of life, and the transitions are more frequent and closer together the older we get. It might be a good idea for the authors to acknowledge this reality. The proper goal is to delay this from occurring, expand the healthspan between negative health events, and lessen the frequency of bouts of sickspan.
2. It is stated on line 46 that experimental longevity interventions delay morbidity and reduce the decline in physiological systems. This has not yet been proven to occur, so I suggest this be restated as the goal of geroscience -- not a fact.
3. Line 57. The hypothesis does not predict anything. Your hypothesis is that you may have a way to predict whether an intervention yields the desired effect. I suggest rephrasing this to something like the following: "Here we present a hypothesis that it may be possible to predict which longevity interventions compress sickspan."
4. Lines 76-90. You're suggesting here that altering the slope of the survival curve could mean a deceleration in the rate of aging, with your subsequent hypothesis that by doing so, you'll also be compressing the sickspan. You may also want to consider just looking at changes in the slope of the mortality, which throughout history for humans and most other species, has remained remarkably constant. The key here is whether changing the slope is a demonstration of sickspan compression, or just delayed aging with no knowledge of effect on the healthspan/sickspan distribution. Personally, I'm very skeptical of our ability to measure sickspan in animals that has any crossover meaning to humans given the complexity of what drives morbidity in humans. Keep in mind that just being sick is not everything that's important when it comes to old-age survival in humans. For many people, it's okay to be sick and have some disease in older ages (it would be rare not to), as long as they're happy. Recent research by Carstensen at Stanford indicates clearly that older cohorts are the happiest among all age groups. I'm just sending out a word of caution to not be dogmatic when it comes to measuring or thinking about sickspan.
5. When describing the model for measuring sickspan, unless I'm misinterpreting, they're no coming back from crossing the threshold. This is not the reality that humans experience.
6. Lines 153-156. In the absence of sickspan data in mice, all you can do is speculate on whether a slope change in the survival curve is slowing aging AND compression sickspan. This is a major weakness of this article.
7. I'm not sure I need to go further with this review -- you do not have a measure of your primary outcome variable.

While I like the idea of going along this line of reasoning, there are fundamental problems with the baseline measure of interest; there are fundamental problems crossing over species; and there are fundamental problems with the singular transition to a state of sickspan from healthspan when, in the real world, this rarely happens.

(Remarks to the Author)

Yang et al present a fascinating study of the underlying dynamical processes that determine the relative timing of failure of health and lifespan during aging. Because these dynamical processes cannot yet be directly observed, Yang et al cleverly apply a modelling approach to understand them. Yang et al consider an ODE model and perform simulation study to infer the dynamics of hidden physiologic changes based on the timing of the outcome they produce: death.

The authors make a useful and non-obvious prediction from their model—interventions in aging that produce deviations from temporal scaling will shorten the “sickspan” of individuals. A shortening of sickspan is what people would typically hope for in an intervention in aging, so the authors work helps to direct attention towards clinically important interventions.

The model is useful because it suggests a way that we can use existing lifespan data to understand healthspan—a quantity for which less interventional data is currently available. By analyzing the shape of mouse survival curves in a set of interventions in aging, the authors make predictions about the relative effect these interventions should have on healthspan. The manuscript's strengths thereby lie first in its rigorous quantitative approach and second its ability to generate testable predictions that, if correct, could be very useful to anyone working to understand the basic biology of aging or design new therapies.

There are three limitations in this study. First, the statistical approaches for analyzing lifespan distributions appear to vary in rigor across the manuscript and sometimes do not seem so well justified. Second, the model makes some important assumptions that are not discussed in the text and also may not be so well justified. These first two issues can likely be addressed with some additional analysis and additional discussion in the text. The final issue is the lack of model validation. The paper analyzes a lot of high-quality lifespan data to make predictions of health-span, but does not seem to include any health-span data. The paper would be greatly strengthened by incorporation of some healthspan data.

Is it really the case that no one has previously performed measurements of an intervention on both mouse health and lifespan? If such a data set exists, is it not of sufficient quality to validate the model? Just an exploration of these issues would be helpful for readers to understand the value of the model. Data sets of sufficient quality do exist in invertebrates, for example Bansal et al PNAS 2015 or Oswal et al. PLOS Computational Biology 2022 . But maybe the underlying biology differs in mouse enough that these data cannot be used for validation. Perhaps the authors can explain.

****Major issues****

1. Past work including Hughes and Hekimi Genetics 2016, Swindell Experimental Gerontology 2009, and Stroustrup et al Nature 2016 identify temporal scaling via application of accelerated failure time models. These models explicitly incorporate both censoring and measurement uncertainty into the alignment of curves across timescales. In several places in this study, alignment is performed by dividing each individual's death time by the population's median lifespan. At first glance, this would seem to be a less statistically sound approach. In particular, the applied method doesn't seem to consider how statistical uncertainty in the estimation population median will affect estimates of the extent of scaling. The authors do apply AFT regression models as a statistical test for deviations from temporal scaling in Supplementary Figure 4, but it's unclear why different methods for aligning curves are used in different parts of the manuscript. If the authors see an advantage to switching to different statistical methods, it would be helpful to provide more justification.

2. The manuscript do not appear to include empirical validation of the model's predictions. This leaves the reader wondering whether the author's model describes the only possible underlying dynamics that are consistent with the observed lifespan data, or, alternatively, whether some completely different model could be both internally self-consistent (just like the authors' proposed model) but make opposite predictions (e.g that deviations from temporal scaling suggest an expansion of sick-span).

This reviewer can imagine a few possible sources of data for validation that might reduce such ambiguity, for example Chan, Raj, et al. eLife 2022, or Martin-Montalvo et al, Nature Communications 2013, or Xie et al Nature Communications 2022. Perhaps none of these contain data of sufficient quality—and in this case, would further justify the value of the new model proposed here! Yet it would be reassuring for readers if the authors at least attempted to explore health-span data as a means for validating their model.

3. The authors seem to make a large simplifying assumption—that a single damage X determines both morbidity and mortality. This assumption is stated very clearly in the equations of figure 2—one of the advantages of having a rigorous phenomenological model that so much other work lacks! But the implications of this assumption are not really considered in the text. The mouse has many different organs—might damage of some organs effect health differently from lifespan? In such a case there might be multiple qualitatively distinct forms of damage rather than one single X. There is some (indirect) evidence in other organisms that health and lifespan may have overlapping but distinct determinants (Oswal, PLoS Computational Biology 2022). How would the current model fare if such a separation of molecular determinants of health and outcomes existed in mice?

****Minor issues****

1. The empirical grounding of this paper involves the analysis of mouse lifespan. It would be very helpful to get a mouse aging researcher to comment on the analysis, as there may be species-specific limitations to the data set not apparent to non-specialists.

2. The authors seem to miss some previous work on survival shape in mice: Hughes and Hekimi Genetics 2016, Swindell Experimental Gerontology 2009.

Reviewer #3

(Remarks to the Author)

The manuscript aims to empirically assess the theoretical implications of the saturated removal model of aging. The model suggests that interventions shifting the survival curve proportionally reshape both sickspan and healthspan, while those altering the shape of the curve affect the proportion of life spent in good or bad health. Lacking sickspan data for lab organisms, the study evaluates the interventions' effects on the shape of the survival curve to gauge their potential sickspan reduction.

While the approach is sound given the absence of empirical sickspan data, the manuscript has significant weaknesses. It overlooks the relevant and pertinent literature on the dynamics of morbidity in increased lifespans, such as the expansion, compression, and dynamic equilibrium scenarios. Authors can reference Jagger et al.'s International Handbook of Health Expectancies (in particular Chap. 2) for an entry point to this branch of literature. They should also explore James Carey's work on lifespan extension in laboratory experiments. He has produced both empirical and methodological contributions on the subject, which could provide valuable insights for the framing of the study and the interpretation of the results.

The authors' interpretation of results faces two key issues. The first problem is the assumption that the rectangularization of the survival curve implies reduced frailty and a shift towards a population where the individuals are more similar and less frail, with a consequent reduction of the sickspan compared to the total lifespan. At the level of population level, the rectangularization of the survival curve has widely taken place over the past decades (commonly called reduction of lifespan inequality) and, yet, both expansion and compression of morbidity, in both absolute and relative terms, are routinely observed in empirical studies of population health. The debate on whether increased lifespan causes compression or extension of morbidity remains open, challenging the model's prediction. This is a very important point to consider when interpreting the results. In particular, caution is necessary to assume that this specific prediction of the model properly describes the reality.

The second problem in the interpretation of the results is in the assumption that age-related morbidity is caused by the same or similar class of factors that causes age-related mortality. While this might be the case the molecular level, at the population level is not clear whether this is the case, with one more generally associated with disease prevalence and progression and the other one more associated with biological senescence.

Additional minor comments include improving the presentation of morbidity and mortality thresholds, which are less straightforward to interpret than the other parameters. Clarifying distinctions between interventions affecting damage removal versus morbidity/mortality thresholds is crucial. Additionally, details on how the saturated removal model accounts for various aging features (lines 234-237) would be beneficial for a full understanding of the model.

In summary, while the manuscript holds promise, addressing these weaknesses is important for its viability in a high-impact publication.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Response to comments was appropriate and thorough.

I would encourage the authors to expand on this work in the future by examining population subgroups that exhibit variations on the steepness of the survival curve, and then examine whether morbidity compression/expansion occurred. This could be done using US data based on level of completed education (e.g., compare the extremes). There will be some complicating effects of selective survival, but I suspect this will corroborate your hypothesis.

Reviewer #2

(Remarks to the Author)

The authors include substantial new data and analysis with the potential to greatly strengthen their manuscript. Some concerns remain about the details of statistical analysis of the new data, in particular whether or not some environmental covariates have been appropriately controlled for. Also, at several places in their text, the authors describe their new data as providing stronger evidence for their model than their statistical analyses would seem to suggest.

In this response to response to reviewers, the original text colors will be removed by the manuscript handling system. The reviewer responds inline with their text marked by the label ****Reviewer #2 response****

****Major issues****

1. Past work including Hughes and Hekimi Genetics 2016, Swindell Experimental Gerontology 2009, and Stroustrup et al

Nature 2016 identify temporal scaling via application of accelerated failure time models. These models explicitly incorporate both censoring and measurement uncertainty into the alignment of curves across timescales. In several places in this study, alignment is performed by dividing each individual's death time by the population's median lifespan. At first glance, this would seem to be a less statistically sound approach. In particular, the applied method doesn't seem to consider how statistical uncertainty in the estimation population median will affect estimates of the extent of scaling. The authors do apply AFT regression models as a statistical test for deviations from temporal scaling in Supplementary Figure 4, but it's unclear why different methods for aligning curves are used in different parts of the manuscript. If the authors see an advantage to switching to different statistical methods, it would be helpful to provide more justification.

We thank the reviewer for this comment. We now apply the AFT+KS test to all survival curves as suggested. This makes the analysis more rigorous and uniform. We write on page 2 line 73-77:

"Longevity-enhancing interventions shift the survival curve to greater ages, increasing the median survival age t_{50} (Fig. 1a). They often preserve the curve shape, a phenomenon called scaling 26,30,31. Interventions are said to be scaling if their survival curves collapse onto the control survival curve when age is normalized by the median lifespan (Fig. 1b, upper panel; throughout we use accelerated failure time model for rescaling and associated statistical test for scaling 30,38, see Methods)."

****Reviewer #2 response****

These updates address my concerns.

2. The manuscript do not appear to include empirical validation of the model's predictions. This leaves the reader wondering whether the author's model describes the only possible underlying dynamics that are consistent with the observed lifespan data, or, alternatively, whether some completely different model could be both internally self-consistent (just like the authors' proposed model) but make opposite predictions (e.g. that deviations from temporal scaling suggest an expansion of sickspan).

This reviewer can imagine a few possible sources of data for validation that might reduce such ambiguity, for example Chan, Raj, et al. eLife 2022, or Martin-Montalvo et al, Nature Communications 2013, or Xie et al Nature Communications 2022. Perhaps none of these contain data of sufficient quality—and in this case, would further justify the value of the new model proposed here! Yet it would be reassuring for readers if the authors at least attempted to explore health-span data as a means for validating their model.

We thank the reviewer for this comment that helped us to add new data to the revised manuscript that significantly strengthens it, as a new Fig 3. We add data from three model organisms- nematodes, fruit flies and mice- from studies that measured longitudinal health parameters and determined healthspan and sickspan on the level of the individual organism, together with mortality, under various longevity interventions. We find in all three organisms that relative sickspan drops with survival curve steepness. This provides support to the prediction of our model, and enhances one's confidence in the relation of steepness and sickspan.

****Reviewer #2 response****

The authors include new data with the potential to greatly strengthen the generality of their model. However, it's unclear to what extent the authors' new data supports the conclusion that a saturating-removal model describes the aging of mice and *C. elegans*. Looking at the shaded confidence regions in Fig. 3 a, c, and d, it would appear that the confidence intervals of the authors' estimates of the relationship between steepness intervention/control and relative sickspan overlap zero. That is, the evidence for mice and *C. elegans* does not exclude the possibility that there is a very small or no relationship between an interventions' effect on IQD and its effect on sickspan. The authors could clarify their text to better convey the statistical uncertainty here. It might also help to include the actual slope parameter estimates along with confidence intervals, so readers don't have to infer the strength of statistical support from shaded confidence regions alone.

It may also be possible that small refinements to the statistical methods might help here. In particular, the authors might want to confirm that they are optimally controlling for experimental variation when calculating IQDs. For *C. elegans* in particular, the conventional experimental practice is to distribute large populations across plates each housing less than 100 individuals. Populations housed on different plates show consistent differences in health-span and lifespan, presumably due to imperfectly controlled environmental conditions. To minimize the impact of such differences, previous analyses of survival curve shape (e.g. Stroustrup et al 2016) used AFT regression models to estimate and remove such plate-associated differences, fitting a regression model with plate name as the only covariate and then performing survival curve shape analysis (which in this case would be the IQD calculation) on the regression residuals. Such accounting for experimental variation might provide better estimates of IQD which would then improve any downstream analyses that depend on those IQDs.

We also now highlight a case where a different effect not considered in our model makes a big difference- extrinsic mortality in human data causes rectangularization, and thus the model's intrinsic parameter effects do not appear to be a main drive of morbidity expansion/compression.

We now write in the results new section on human data in the results:

"We find that the main driver for rectangularization of the survival curve in human data is the decline in extrinsic mortality. Similar results are found when fitting human hazard to the Gamma-Gompertz-Makham equation 66,67 -- the age independent extrinsic mortality drops with cohort birth year (Fig. 6c), whereas the Gompertz slope and intercept remain nearly constant (Fig. 6 d,e). The dominance of the extrinsic mortality drop in human cohorts is consistent with previous studies 7,9. This is in contrast to model organism data analyzed above in Fig. 3. Unlike human populations, these animal cohorts were studied in controlled environments, and extrinsic mortality is not the main determinant of their survival curve

steepness. “

****Reviewer #2 Response****

The inclusion of this important caveat / clarification to the authors' model strengthens the manuscript.

3. The authors seem to make a large simplifying assumption—that a single damage X determines both morbidity and mortality. This assumption is stated very clearly in the equations of figure 2—one of the advantages of having a rigorous phenomenological model that so much other work lacks! But the implications of this assumption are not really considered in the text. The mouse has many different organs—might damage of some organs effect health differently from lifespan? In such a case there might be multiple qualitatively distinct forms of damage rather than one single X. There is some (indirect) evidence in other organisms that health and lifespan may have overlapping but distinct determinants (Oswal, PLoS Computational Biology 2022). How would the current model fare if such a separation of molecular determinants of health and outcomes existed in mice?

We now clearly discuss the simplifying model assumption and its implications.

One implication is if the death threshold changes separately from the disease threshold, one has different conclusions (steepening survival but morbidity expansion).

We discuss this in the results. On page 5, line 170-173, we write:

“There is one exception to the rule that interventions that steepen survival curves compress relative sickspan. This occurs when the death threshold X_c rises but the disease threshold X_d does not (Fig. 2c,d, green lines). This might apply to human data when medical care extends life in sick individuals, without affecting age of onset of sickspan. We discuss human mortality in the last section. “

And in the discussion, page 14, line 438-447, we write:

“One exception is raising the death threshold without correspondingly raising the disease threshold. This steepens the survival curve, but increases relative sickspan, since individuals spend more time between the disease and death thresholds. As a thought experiment, consider artificial life expansion by intensive care units - such perturbations increase sickspan by delaying death. A steep survival curve in these situations can result by artificially keeping individuals alive to the same age. We assume that the longevity interventions in the controlled experiments in model organisms that we analyzed here do not cause such uncoordinated changes in death and disease thresholds. In humans, sex differences in the death and morbidity thresholds between females and males could explain the 'health survival paradox' - that females have longer lifespans but expanded morbidity. This would occur if X_c is higher in females (and X_d is either the same or lower in females).”

In the discussion we also add a paragraph on limitations of our main assumption, as suggested by the reviewer. on page 14-15, line 456-465:

“The SR model makes a major simplifying assumption—that a single damage variable X determines both morbidity and mortality. This provides quantitative predictions that can be tested against data, as in Fig. 3,6. Future work can explore more complex situations — for example that damage in each organ affects health differently from lifespan. In such a case there might be multiple qualitatively distinct damage variables rather than a single X. There is evidence that health and lifespan may have overlapping but distinct determinants, and that cognitive functions may have somewhat different aging patterns than other physiological functions 44,72–75 . Furthermore, on the population level, genetic and environmental heterogeneity can affect the onset of morbidity and complicate interpretation of the model. Extensions of the SR models can in principle address this complexity 72. “

****Reviewer #2 Response****

The inclusion of this important caveat / clarification to the authors' model strengthens the manuscript.

****Reviewer #2 Response****

All minor issues have been addressed.

Reviewer #3

(Remarks to the Author)

The authors have properly addressed all the comments. The current version of the manuscript has substantially improved and is ready for publication.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Cover letter of the revision

Sunday, September 29, 2024

Dear Editor,

We would like to submit our revision to our previous submission NCOMMS-23-49968, now entitled "**Compression of morbidity by interventions that steepen the survival curve**".

Thank you for your positive consideration and for the reviewer comments. The comments helped us to greatly improve this study. The main step forward is analysis of experimental data on health, which was lacking. We added data from mouse, *C. elegans* and *Drosophila*. The data strongly supports the model prediction. We also add new analysis and robustness tests, and clarify the text. Although the theory and data is focused on lab animals, we also add an analysis of human data. We believe the revised paper is much stronger, more rigorous and clear.

Below we provide point-to-point response to review comments.

Sincerely,
Yifan Yang, Uri Alon

Reviewer #1:

(Reviewer remarks are in purple, our responses are in black, and quotes from the manuscript are in blue)

I'm glad someone is giving thought to this idea, which is straight forward in intent -- to find a predictive tool that links prospective gerotherapeutics to the goal of compressing the sickspan. I have some specific suggestions.

We thank the reviewer for this endorsement.

1. As currently stated, the authors indicate that healthspan is followed by sickspan. This rarely happens as discrete sequential events. Most people transition into and out of sickspan during the course of life, and the transitions are more frequent and closer together the older we get. It might be a good idea for the authors to acknowledge this reality. The proper goal is to delay this from occurring, expand the healthspan between negative health events, and lessen the frequency of bouts of sickspan.

We have corrected this in the revised version. In the intro we now write in page 1, line 27-30:

“The lifespan of an individual typically begins with a period of general health, called the healthspan, followed by a period of intermittent or continuous morbidity and disability called the sickspan¹⁻³. The sickspan burdens human quality of life and is a growing driver of economic expenditure across societies⁴⁻⁶. “

2. It is stated on line 46 that experimental longevity interventions delay morbidity and reduce the decline in physiological systems. This has not yet been proven to occur, so I suggest this be restated as the goal of geroscience -- not a fact.

We agree and have corrected this on line 46 - specifically mentioning experiments on model organisms. We now write on page 2, line 44-51:

“These longevity interventions are developed to also delay morbidity in model organisms - the treated animals are usually reported to have better health parameters than equally aged untreated animals as measured at a single time point¹¹. However, whether they compress the sickspan in absolute or relative terms has not generally been reported. Data from *Caenorhabditis elegans* suggests that many interventions that extend lifespan extend sickspan proportionally^{26,27}, and thus do not compress morbidity in absolute or relative terms. A similar effect has been seen in medflies²⁸. Notably, however, data from intervention studies that longitudinally follow health at many timepoints is rare, and thus the question of sickspan relative to lifespan is understudied.

”

3. Line 57. The hypothesis does not predict anything. Your hypothesis is that you may have a way to predict whether an intervention yields the desired effect. I suggest

rephrasing this to something like the following: "Here we present a hypothesis that it may be possible to predict which longevity interventions compress sickspan."

We agree and have rephrased, on line 56-57:

"Here we present a theory that can predict which longevity interventions compress sickspan. The theory is based on the survival curve, the fraction of organisms that survive to a given age ²⁹. "

4. Lines 76-90. You're suggesting here that altering the slope of the survival curve could mean a deceleration in the rate of aging, with your subsequent hypothesis that by doing so, you'll also be compressing the sickspan. You may also want to consider just looking at changes in the slope of the mortality, which throughout history for humans and most other species, has remained remarkably constant. The key here is whether changing the slope is a demonstration of sickspan compression, or just delayed aging with no knowledge of effect on the healthspan/sickspan distribution.

We thank the reviewer for this comment which helped us to clarify a subtle point. A key concept in the field is the slope of the mortality (log hazard) curve. In our analysis we consider the steepness of the survival curve (not hazard curve). This difference has important implications. For example, changing production rate η strongly affects the hazard slope, but does not change survival curve steepness. Conversely, changing removal rate β does not affect the hazard slope, but does change survival curve steepness. (new Fig. 1d) Thus we believe the concepts of this paper add new and potentially meaningful parameters (for molecular/cellular production and removal rates) that go beyond the traditional analysis of slope and intercept of mortality (hazard) curves.

We now add in the revised discussion, page 14, line 448-455:

"Note that we consider survival curve steepness rather than the more commonly studied hazard (Gompertz) slope - the latter is not directly predictive of morbidity compression in our model. Hazard curves are the logarithmic derivative of the survival curves. For example, changes in damage production rate affect hazard slope but not survival steepness, and conversely changes in removal rate do not affect hazard slope but do affect survival steepness. Changes in thresholds and noise affect both. Historically, hazard slope has not changed much in humans, whereas survival curves have steepened. Survival curves are also more robustly measured in experiments than hazard curves."

To clarify the definition of steepness, we design a new Fig. 1 as below:

"

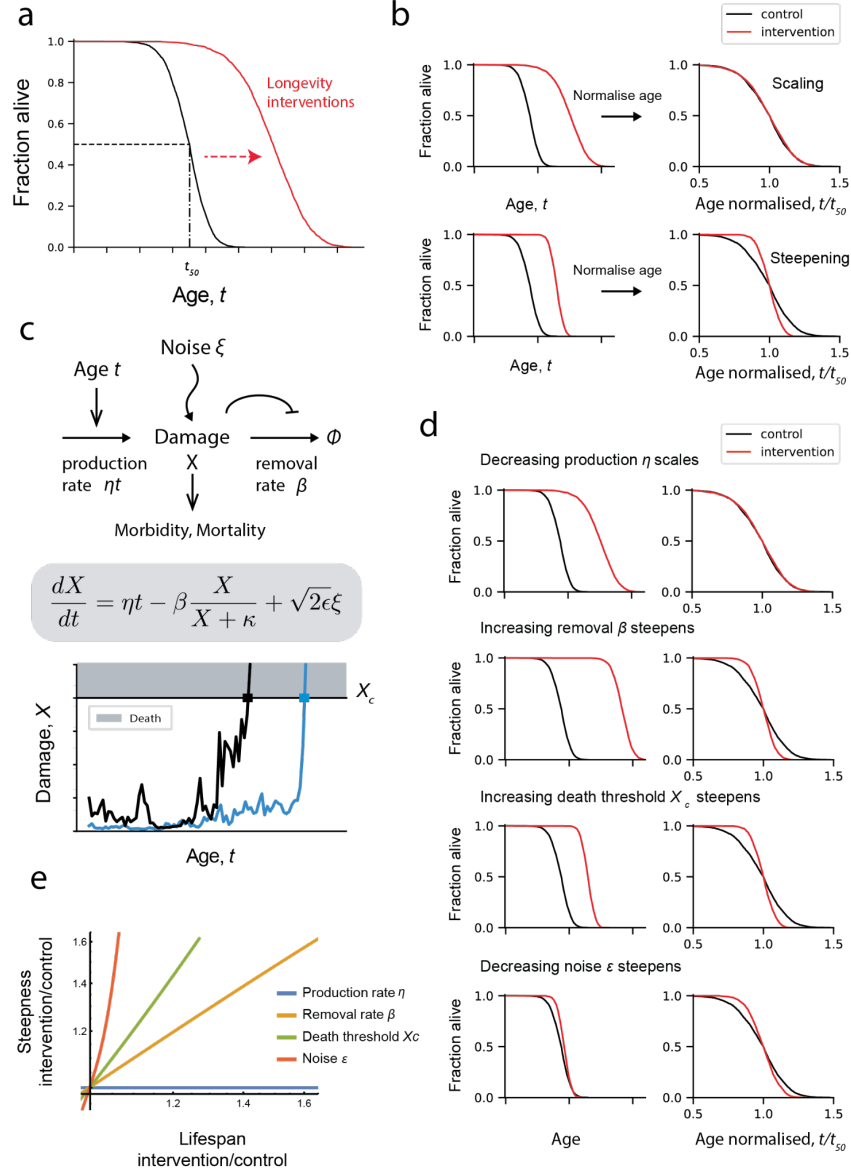


Figure 1. Survival curves in the saturating removal model scale when interventions affect damage production, and steepen when interventions affect other parameters. a, Survival curves describe the fraction of organisms surviving to each age; median age t_{50} is extended by longevity interventions. **b**, Scaling is defined by survival curves collapsing on the same curve when time is re-normalized by their averages³⁰. Steepening occurs when the shape has higher steepness. **c**, The saturating removal model posits a damage X that drives aging whose production rises with age and whose removal saturates, with noise. Simulations (black and blue lines) show stochastic trajectories of 2 individuals that cross the death threshold X_c at different times. **d**, Effect of a 2-fold change in each model parameter on survival curve shape **e**, Effect of parameter changes on t_{50} and steepness defined as median divided by interquartile range. Baseline parameters for mice are $\eta=2.3 \times 10^{-4}/\text{day}^2$, $\beta=0.15/\text{day}$, $\kappa=0.5$, $\epsilon=0.16/\text{day}$, $X_c=17$ ³².

”

We also write at the end of page 2, line 83-85:

“Other interventions increase median lifespan but steepen the survival curve shape (Fig 1b, lower panel) ^{30,37} . A useful metric for steepness is the median age divided by the interquartile range of survival ²⁹ .”

And also write on page 4, line 134-143:

“A key feature of the SR model is its ability to provide scaling of survival curves ³². Changes in the rate of damage production, namely the parameter η , shift median survival but preserve the shape of the curve (Fig. 1d), providing near-perfect scaling.

Changing any of the other model parameters, such as removal rate β , affects not only lifespan but also the shape of the survival curve (Fig. 1d). Increasing damage removal β shifts the survival curve to the right and steepens it. Increasing death threshold X_c mildly extends median lifespan and steepens the curve by preventing early deaths. Decreasing the noise amplitude ϵ causes a small increase in lifespan and a large increase in steepness - less noise leads to more deterministic death times. These effects are shown on a steepness-longevity plot in Fig. 1e, and separately for steepness and longevity in Supp. Fig. 1.”

Personally, I'm very skeptical of our ability to measure sickspan in animals that has any crossover meaning to humans given the complexity of what drives morbidity in humans. Keep in mind that just being sick is not everything that's important when it comes to old-age survival in humans. For many people, it's okay to be sick and have some disease in older ages (it would be rare not to), as long as they're happy. Recent research by Carstensen at Stanford indicates clearly that older cohorts are the happiest among all age groups. I'm just sending out a word of caution to not be dogmatic when it comes to measuring or thinking about sickspan.

This comment helped us better analyze and discuss the complexity of human data and to mention Carstensen's work (reference 94).

Although most of our study considers model organisms in controlled experiments, we now analyze human hazard curves from different countries (New Fig 6). We consider the fact that in human data from the last century, survival curves rectangularize- and note that this is mainly due to a drop in extrinsic mortality rather than a change in intrinsic parameters. The present theory thus does not easily apply to such cohort data due to the complexity of human heterogeneity.

We now added a new section in the results (page 11-12, line 335-370) and a new Figure 6:

”

Steepening of human survival curves is largely due to reduction in extrinsic mortality

The present study primarily aims to understand lab organism data. Considerations in humans are far more complex because as compared to model organisms with defined genetics living in controlled environments, sickspan is affected by additional variations in lifestyle, environment, socioeconomic status and medical care. We therefore provide a tentative account of how the present considerations might apply to human data.

There has been a global trend of rectangularization of human survival curves in the past century ⁸ -- in many countries, the median lifespan has increased and survival curve steepness has increased as well ⁶² (Fig. 6a). There is a tight correlation between life expectancy and lifespan equality (a measure of steepness) ⁶³. There are also exceptions in certain countries in which lifespan variability has increased or remained constant in recent decades ⁶².

Relative sickspan has not been as extensively measured as lifespan, with data accumulating only recently. Available data indicates that relative sickspan since 1980 has compressed in some countries and has grown in others ^{7,9}.

We asked whether the rectangularization might be due to changes in SR model parameters, to test whether such changes might predict relative sickspan in each country. To do so, we analyzed human cohort mortality data from the Human Mortality Database (HMD) ⁶⁴. Human data shows a sizable extrinsic mortality that dominates mortality until age 30-40. In order to account for this, we fitted human hazard curves to the SR model with an added extrinsic mortality term (Fig. 6b). The fits are excellent (adjusted $R^2=0.9994$, F-test $p<10^{-16}$, see Fig. 6b).

We find that the main driver for rectangularization of the survival curve in human data is the decline in extrinsic mortality. Similar results are found when fitting human hazard to the Gamma-Gompertz-Makham equation ^{65,66} -- the age independent extrinsic mortality drops with cohort birth year (Fig. 6c), whereas the Gompertz slope and intercept remain nearly constant (Fig. 6d,e). The dominance of the extrinsic mortality drop is consistent with previous studies ^{8,10}. This contrasts with sickspan and mortality data of model organisms analyzed above in Fig. 3. Unlike human populations, these animal cohorts live in controlled environments, where extrinsic mortality can be assumed to be constant.

In addition to changes in extrinsic mortality, there have been age-specific improvements in mortality rates. Until recent decades this was primarily a reduction in young deaths (younger than median lifespan), and in the past few decades this shifted to a reduction in old deaths (older than median lifespan) ⁶⁷. These changes are likely to be due to public health and medical interventions, as well as changes in education, socioeconomic status and other factors ⁶⁸. Such age-specific changes confound an attempt to fit changes in SR model parameters to this human data. Thus, why some countries show a reduction in relative sickspan whereas others show an increase ⁷ may not be directly linked to changes in SR model parameters.

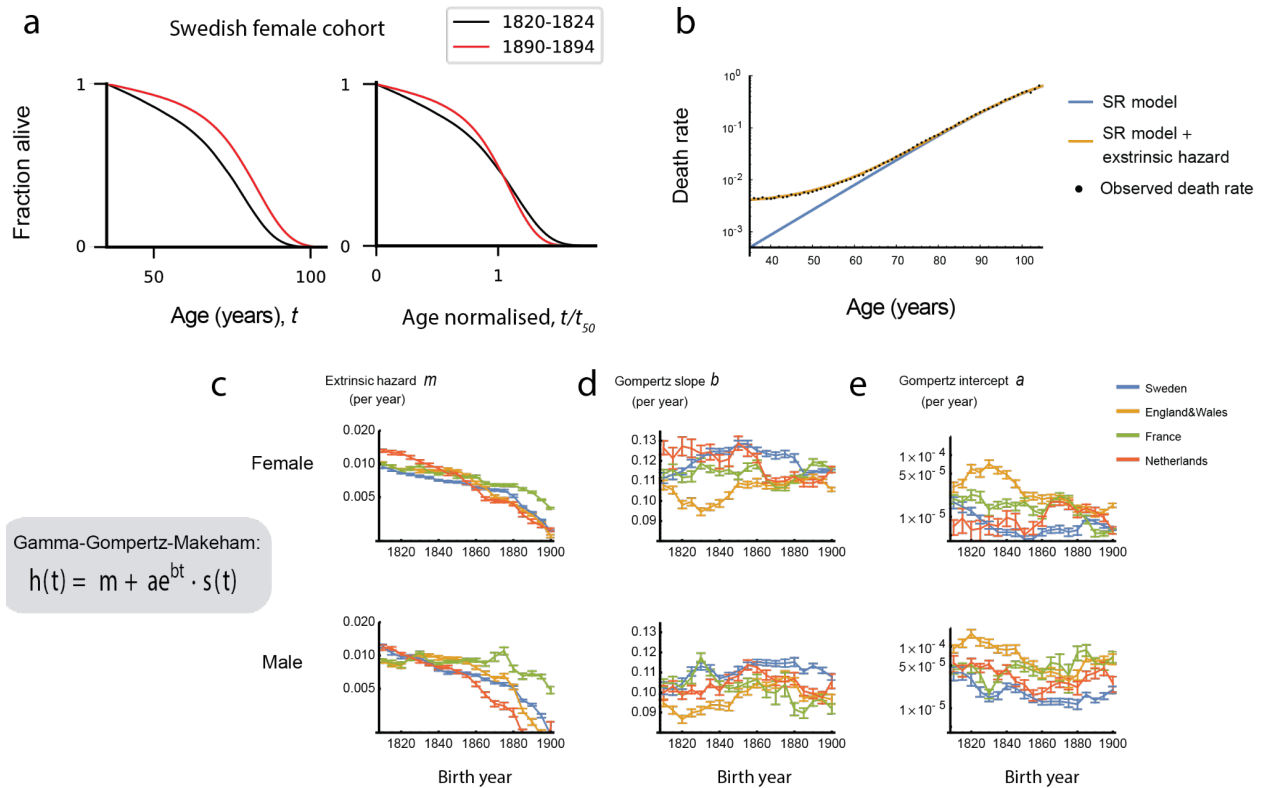


Figure 6. Steepening (rectangularization) of human survival curves across the past century are mainly driven by declining extrinsic mortality. **a**, Survival curves and rescaled survival curves of 2 cohorts of Swedish women born in 1820-1824 and 1890-1894. **b**, Hazard curve (death rates by age) of Swedish women born in 1890-1894, and of SR model fitted to this cohort. The SR model parameters that correspond to the blue and orange hazard curves are $\eta=1.04 \times 10^{-5}/\text{day}^2$, $\beta=0.45/\text{day}$, $\kappa=0.50$, $\varepsilon=0.48/\text{day}$, $X_c=19.9$ ³³. The extrinsic hazard (Makeham term) in the orange curve is $m=0.0036$ per year. **c-e**, Mortality component trends of Swedish, Dutch, French (civilian) and English & Welsh combined (civilian) cohorts born from 1810 to 1910. Extrinsic hazard rates, Gompertz slopes and Gompertz intercepts are estimated using Gamma-Gompertz-Makeham model⁶⁶. In the SR model, Gompertz slope can be approximated by $\eta X_c/\varepsilon$, while the Gompertz intercept is proportional to $e^{\beta X_c/\varepsilon}$ ³³. The Gamma/Gompertz correction term $s(t)$ is $s(t)=e^c/(e^c-1+e^{bt})$, where c is an additional parameter that ranges from 7.5 to 10, such that $s(t)$ only comes into effect at old age.

“

We also address the complexity of human mortality in the discussion, citing Carstensen’s important findings. We write in the revised discussion (page 16, line 525-531):

“Future work can further explore how the SR model relates to human morality and sickspan. Rectangularization of the human survival curve over the past century is primarily due to reduction in extrinsic mortality rather than changes in intrinsic parameters. Old-age survival and morbidity in humans is impacted by heterogeneity of socioeconomic status^{90–92}, lifestyle, genetics⁷², medical care and public health⁹². **An important factor is emotional health - older cohorts appear to have high average well-being⁹³, whereas social isolation in the old has major negative effects⁹⁴.** Such factors can complicate a direct comparison to the SR model.”

5. When describing the model for measuring sickspan, unless I'm misinterpreting, they're not coming back from crossing the threshold. This is not the reality that humans experience.

This comment helped us better explain in the results section the reason for our modeling choice.

The first reason for choosing a continuous sickspan is to relate to the longitudinal animal data available. The longitudinal animal studies that we consider (*Drosophila*, mice and *C. elegans*), all effectively define sickspan as a continuous, end-of-life period. Thus healthspan and sickspan in these studies are continuous and complementary portions of the lifespan. This aligns with our definition. Notably, the deficits considered in these animal studies are severe, and are not expected to be easily reversible.

The second reason to choose a continuous sickspan definition is to avoid complicating model assumptions and unknown parameters. We define sickspan by the simplest definition based on the first crossing of a threshold. Adding more detailed intermittent sickness would require more parameters and model assumptions. For example, for how long does X need to cross the threshold in order to change health status? This would depend on the timescale of the relevant physiological systems. Measures such as human frailty index are composed of both reversible and irreversible deficits, and thus there is a range of timescales. Future work in humans will be required to address this issue.

We now write in results on page 5, line 158-162:

"We thus consider here a definition of sickspan as a continuous end-of-life period that begins upon the first crossing of X_d . This aligns with current definitions in model organisms of an end of life period of morbidity when severe health deficits cross a threshold (refs). An alternative definition would allow for sickspan to be intermittent when X drops below X_d for sufficient time. Such a definition requires additional parameters and assumptions and we leave it for future study."

6. Lines 153-156. In the absence of sickspan data in mice, all you can do is speculate on whether a slope change in the survival curve is slowing aging AND compression sickspan. This is a major weakness of this article.

This comment helped us add crucial new data and greatly improved the paper - see the new Fig 3. We add data in three model organisms- nematodes, fruit flies and mice - from the rare studies that measured longitudinal health parameters together with mortality under various longevity interventions. We find in all three organisms that relative sickspan drops with survival curve steepness. Scaling interventions have approximately scaled sickspan. This analysis provides support to the prediction of our theory, and enhances one's confidence in the relation of steepness and sickspan.

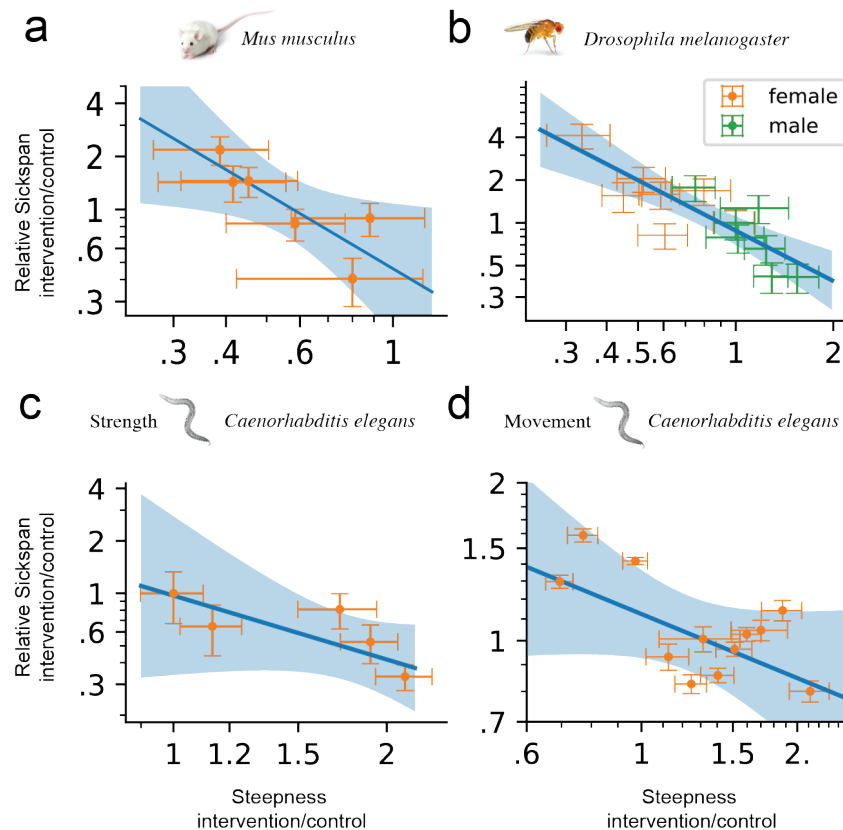


Figure 3. Relative sickspan is compressed by interventions that steepen the survival curves in model organisms, as predicted. Mean absolute sickspan divided by mean lifespan relative to control is plotted versus steepness of survival curve relative to control. **a**, Mice under dietary longevity interventions - caloric restriction and intermittent fasting - in two strains from Luciano et al ³⁹. Sickspan was defined using mouse frailty index. See Supp. Table 1 for intervention details. **b**, Nutritional longevity interventions in two strains of *D. melanogaster*, where sickspan was defined by deficits in climbing, jumping and flying in response to a mechanical challenge. Data from Gaitanidis et al ⁴³. See Supp. Table 2 for intervention details. **c**, Genetic longevity mutants in *C. elegans*, where sickspan was defined by loss of muscle power measured by piezoelectric challenge. Data from Statzer et al ²³. See Supp. Table 3 for strain details. **d**, Genetic, dietary and environmental factors in *C. elegans*, where sickspan was measured by secession of spontaneous movements. Data from Stazer et al ²³ and Oswal et al ⁴⁴. Confidence intervals are calculated using bootstrap resampling of individuals in each cohort. See Supp. Table 4 for strain and intervention details.

“

7. I'm not sure I need to go further with this review -- you do not have a measure of your primary outcome variable.

While I like the idea of going along this line of reasoning, there are fundamental problems with the baseline measure of interest; there are fundamental problems crossing over species; and there are fundamental problems with the singular transition to a state of sickspan from healthspan when, in the real world, this rarely happens.

We benefited from these comments and believe that the revised manuscript addresses these points. Most importantly- **we have added the primary outcome measure of sickspan in three organisms, including mice. This suggests that the predictions may hold across at least these three species.**

Reviewer #2:

(Reviewer remarks are in purple, our responses are in black, and quotes from the manuscript are in blue)

Yang et al present a fascinating study of the underlying dynamical processes that determine the relative timing of failure of health and lifespan during aging. Because these dynamical processes cannot yet be directly observed, Yang et al cleverly apply a modeling approach to understand them. Yang et al consider an ODE model and perform simulation study to infer the dynamics of hidden physiologic changes based on the timing of the outcome they produce: death.

The authors make a useful and non-obvious prediction from their model—interventions in aging that produce deviations from temporal scaling will shorten the “sickspan” of individuals. A shortening of sickspan is what people would typically hope for in an intervention in aging, so the authors work helps to direct attention towards clinically important interventions.

The model is useful because it suggests a way that we can use existing lifespan data to understand healthspan—a quantity for which less interventional data is currently available. By analyzing the shape of mouse survival curves in a set of interventions in aging, the authors make predictions about the relative effect these interventions should have on healthspan. The manuscript's strengths thereby lie first in its rigorous quantitative approach and second its ability to generate testable predictions that, if correct, could be very useful to anyone working to understand the basic biology of aging or design new therapies.

We thank the reviewer for this endorsement

There are three limitations in this study. First, the statistical approaches for analyzing lifespan distributions appear to vary in rigor across the manuscript and sometimes do not seem so well justified. Second, the model makes some important assumptions that are not discussed in the text and also may not be so well justified. These first two issues can likely be addressed with some additional analysis and additional discussion in the text. The final issue is the lack of model validation. The paper analyzes a lot of high-quality lifespan data to make predictions of health-span, but does not seem to include any health-span data. The paper would be greatly strengthened by incorporation of some healthspan data.

Is it really the case that no one has previously performed measurements of an intervention on both mouse health and lifespan? If such a data set exists, is it not of sufficient quality to validate the model? Just an exploration of these issues would be helpful for readers to understand the value of the model. Data sets of sufficient quality do exist in invertebrates, for example Bansal et al PNAS 2015 or Oswal et al. PLOS

Computational Biology 2022 . But maybe the underlying biology differs in mouse enough that these data cannot be used for validation. Perhaps the authors can explain.

We thank the reviewer for this comment that helped us add crucial new data to the paper, to test the theory predictions.

We analyzed longitudinal health data from mice, worms and flies.

This is described in a new section in the results (page 6-7, line 194-241), and a new Figure 3:

“

Steepening interventions compress relative sickspan in invertebrates and mice

To test whether steepening of the survival curve indeed implies a compressed sickspan, we analyzed data from experiments which followed health longitudinally as organisms aged under different interventions. These experiments were performed in mice, *Drosophila melanogaster* and *Caenorhabditis elegans*. In all cases we used the authors' definition of sickspan.

Data on longitudinal health in mice is currently limited, although it is expected to be available from several ongoing experiments ^{40–42}. We analyzed preliminary data from Luciano et al ⁴³ which followed mouse frailty index longitudinally in different longevity interventions. The experiment included dietary interventions (20% and 40% calorie restriction and two intermittent fasting schedules) in two strains (Diversity Outbred and C57BL/6). Sickspan was defined by the authors as periods of life with equal or more than 4 age-related deficits in the mouse frailty index. We evaluated the survival curve steepness relative to control for these experiments and compared it to the relative sickspan (mean absolute sickspan divided by mean lifespan, relative to control). Steepness was defined by the median lifespan divided by the interquartile range (IQR) of lifespan ²⁹. We calculate confidence intervals of both relative sickspan and steepness by bootstrap resampling of individuals.

We find that relative sickspan decreased with steepness as predicted (Fig. 3a, Supp. Table 1). Scaling interventions (steepness intervention/control of 1) showed approximately scaled sickspan (Fig. 3a). These findings agree with the SR model predictions. The relationship is robust to the definition of sickspan (Supp. Fig. 2).

We next considered longitudinal experiments in flies. Carey and colleagues pioneered the use of longitudinal life-history data such as fecundity and mortality to test evolutionary theories of aging ^{44–46} using medflies (*Ceratitis capitata*). More recently, Gaitanidis et al ⁴⁷ measured longitudinal healthspan data in *Drosophila* by quantifying age-related declines of motor functions. The authors assayed escape performance in response to gently banging food vials on the counter, i.e. “startle assays”. Each fly was scored for its jumping, climbing and flying responses to generate a health score for each individual by summing the number of functional deficits. Sickspan was defined as the period where an individual has a score below a threshold value. The experiment was repeated for 5 dietary interventions (including protein restriction, superfood and curcumin) in males and females in two fly strains with different lifespans (Lausanne, Oregon).

We evaluated the survival curve steepness relative to control for these experiments, and compared it to the relative sickspan. Relative sickspan decreased with steepness, as predicted. Scaling interventions (steepness/control of 1) showed approximately scaled sickspan (Fig. 3b, Supp. Table 2). These findings agree with the SR model predictions.

We also considered two *C. elegans* studies. Statzer et al used a piezoelectric system to measure muscle power cross-sectionally throughout the lifespan ²⁶, for five different longevity-affecting mutants: *daf-2(e1368)*, *daf-2(e1370)*, *glp-1* and *eat-2*, and define sickspan by the period of life when muscle power is below 50% of wildtype. Analysis of this data shows that relative sickspan decreases with steepness (Fig. 3c, Supp. Table 3).

Another method to define healthspan used the period of spontaneous movements to define the healthspan⁴⁸. This method was used by both Oswal et al⁴⁸ and Statzer et al²⁶, for several longevity interventions including temperature, food deprivation, food inactivation by UV, *daf-2* (e1368 & e1370), *eat-2*, *glp-1* (both studies), *nuo-6* (see Supp. Table 4 for details of all the intervention and control cohorts). The analysis showed that relative sickspan defined by voluntary movement also decreases with survival curve steepness, as predicted (Fig. 3d).

We conclude that in invertebrates and mice, steepening interventions compress the relative sickspan, whereas scaling interventions approximately scale the sickspan.

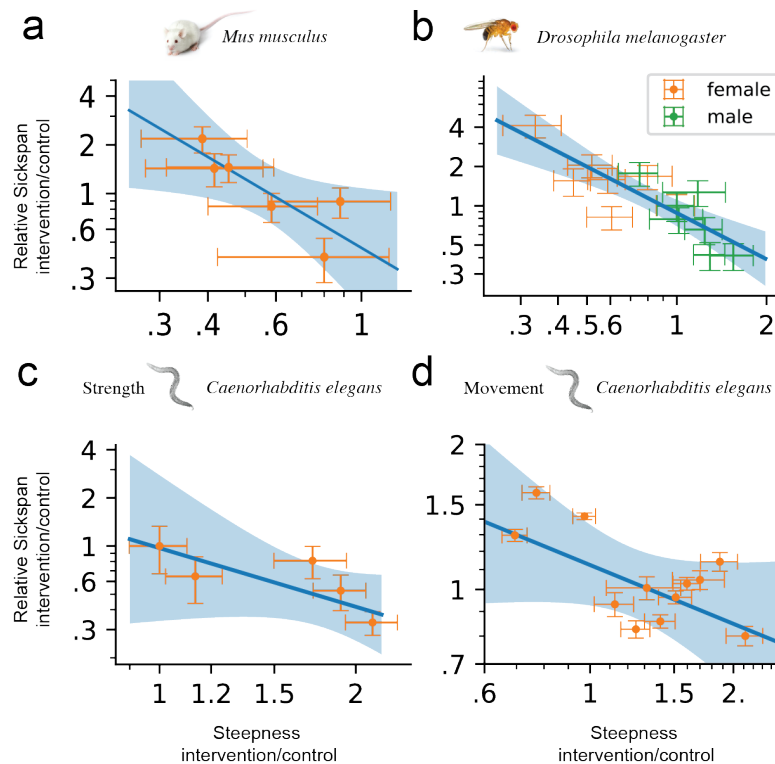


Figure 3. Relative sickspan is compressed by interventions that steepen the survival curves in model organisms, as predicted. Mean absolute sickspan divided by mean lifespan relative to control is plotted versus steepness of survival curve relative to control. **a**, Mice under dietary longevity interventions - caloric restriction and intermittent fasting - in two strains from Luciano et al³⁹. Sickspan was defined using mouse frailty index. See Supp. Table 1 for intervention details. **b**, Nutritional longevity interventions in two strains of *D. melanogaster*, where sickspan was defined by deficits in climbing, jumping and flying in response to a mechanical challenge. Data from Gaitanidis et al⁴³. See Supp. Table 2 for intervention details. **c**, Genetic longevity mutants in *C. elegans*, where sickspan was defined by loss of muscle power measured by piezoelectric challenge. Data from Statzer et al²³. See Supp. Table 3 for strain details. **d**, Genetic, dietary and environmental factors in *C. elegans*, where sickspan was measured by secession of spontaneous movements. Data from Statzer et al²³ and Oswal et al⁴⁴. Confidence intervals are calculated using bootstrap resampling of individuals in each cohort. See Supp. Table 4 for strain and intervention details.

****Major issues****

1. Past work including Hughes and Hekimi Genetics 2016, Swindell Experimental Gerontology 2009, and Stroustrup et al Nature 2016 identify temporal scaling via application of accelerated failure time models. These models explicitly incorporate both

censoring and measurement uncertainty into the alignment of curves across timescales. In several places in this study, alignment is performed by dividing each individual's death time by the population's median lifespan. At first glance, this would seem to be a less statistically sound approach. In particular, the applied method doesn't seem to consider how statistical uncertainty in the estimation population median will affect estimates of the extent of scaling. The authors do apply AFT regression models as a statistical test for deviations from temporal scaling in Supplementary Figure 4, but it's unclear why different methods for aligning curves are used in different parts of the manuscript. If the authors see an advantage to switching to different statistical methods, it would be helpful to provide more justification.

We thank the reviewer for this comment. We now apply the AFT+KS test to all survival curves as suggested. This makes the analysis more rigorous and uniform. We write on page 2 line 73-77:

“Longevity-enhancing interventions shift the survival curve to greater ages, increasing the median survival age t_{50} (Fig. 1a). They often preserve the curve shape, a phenomenon called scaling^{26,30,31}. Interventions are said to be scaling if their survival curves collapse onto the control survival curve when age is normalized by the median lifespan (Fig. 1b, upper panel; throughout we use accelerated failure time model for rescaling and associated statistical test for scaling^{30,38}, see Methods).”

2. The manuscript do not appear to include empirical validation of the model's predictions. This leaves the reader wondering whether the author's model describes the only possible underlying dynamics that are consistent with the observed lifespan data, or, alternatively, whether some completely different model could be both internally self-consistent (just like the authors' proposed model) but make opposite predictions (e.g that deviations from temporal scaling suggest an expansion of sick-span).

This reviewer can imagine a few possible sources of data for validation that might reduce such ambiguity, for example Chan, Raj, et al. eLife 2022, or Martin-Montalvo et al, Nature Communications 2013, or Xie et al Nature Communications 2022. Perhaps none of these contain data of sufficient quality—and in this case, would further justify the value of the new model proposed here! Yet it would be reassuring for readers if the authors at least attempted to explore health-span data as a means for validating their model.

We thank the reviewer for this comment that helped us to add new data to the revised manuscript that significantly strengthens it, as a new Fig 3. We add data from three model organisms- nematodes, fruit flies and mice- from studies that measured longitudinal health parameters and determined healthspan and sickspan on the level of the individual organism, together with mortality, under various longevity interventions. We find in all three organisms that relative sickspan drops with survival curve steepness. This provides support to the prediction of our model, and enhances one's confidence in the relation of steepness and sickspan.

“

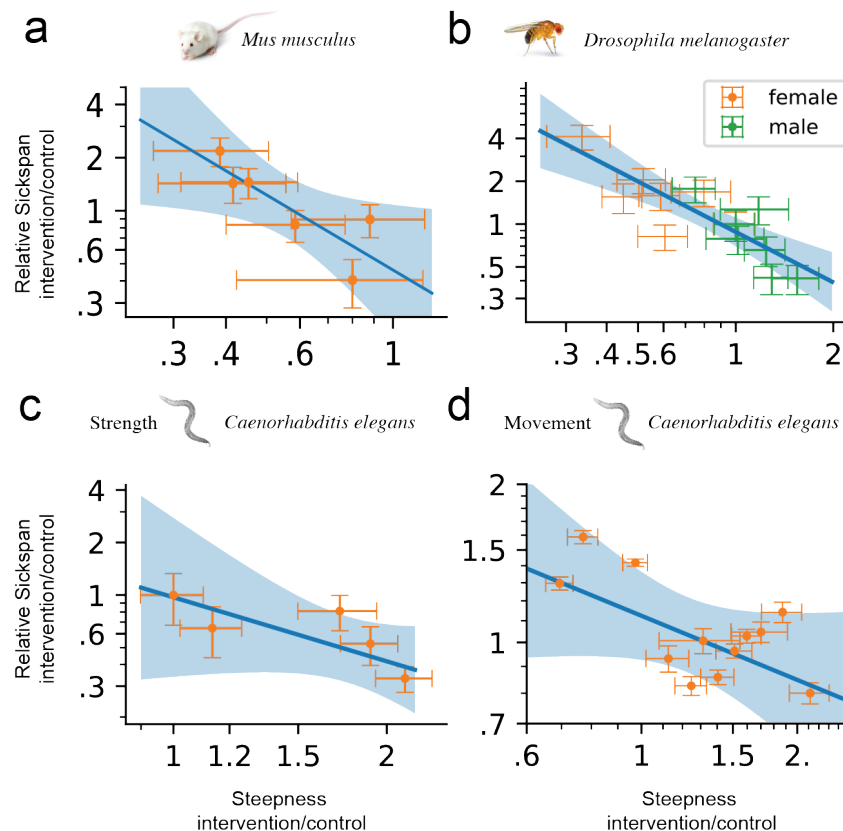


Figure 3. Relative sickspan is compressed by interventions that steepen the survival curves in model organisms, as predicted. Mean absolute sickspan divided by mean lifespan relative to control is plotted versus steepness of survival curve relative to control. **a**, Mice under dietary longevity interventions - caloric restriction and intermittent fasting - in two strains from Luciano et al ³⁹. Sickspan was defined using mouse frailty index. See Supp. Table 1 for intervention details. **b**, Nutritional longevity interventions in two strains of *D. melanogaster*, where sickspan was defined by deficits in climbing, jumping and flying in response to a mechanical challenge. Data from Gaitanidis et al ⁴³. See Supp. Table 2 for intervention details. **c**, Genetic longevity mutants in *C. elegans*, where sickspan was defined by loss of muscle power measured by piezoelectric challenge. Data from Statzer et al ²³. See Supp. Table 3 for strain details. **d**, Genetic, dietary and environmental factors in *C. elegans*, where sickspan was measured by secession of spontaneous movements. Data from Stazer et al ²³ and Oswal et al ⁴⁴. Confidence intervals are calculated using bootstrap resampling of individuals in each cohort. See Supp. Table 4 for strain and intervention details.

“

We also now highlight a case where a different effect not considered in our model makes a big difference- extrinsic mortality in human data causes rectangularization, and thus the model's intrinsic parameter effects do not appear to be a main drive of morbidity expansion/compression.

We now write in the results new section on human data in the results:

“We find that the main driver for rectangularization of the survival curve in human data is the decline in extrinsic mortality. Similar results are found when fitting human hazard to the

Gamma-Gompertz-Makham equation^{66,67} -- the age independent extrinsic mortality drops with cohort birth year (Fig. 6c), whereas the Gompertz slope and intercept remain nearly constant (Fig. 6 d,e). The dominance of the extrinsic mortality drop in human cohorts is consistent with previous studies^{7,9}. This is in contrast to model organism data analyzed above in Fig. 3. Unlike human populations, these animal cohorts were studied in controlled environments, and extrinsic mortality is not the main determinant of their survival curve steepness. “

3. The authors seem to make a large simplifying assumption—that a single damage X determines both morbidity and mortality. This assumption is stated very clearly in the equations of figure 2—one of the advantages of having a rigorous phenomenological model that so much other work lacks! But the implications of this assumption are not really considered in the text. The mouse has many different organs—might damage of some organs effect health differently from lifespan? In such a case there might be multiple qualitatively distinct forms of damage rather than one single X . There is some (indirect) evidence in other organisms that health and lifespan may have overlapping but distinct determinants (Oswal, PLoS Computational Biology 2022). How would the current model fare if such a separation of molecular determinants of health and outcomes existed in mice?

We now clearly discuss the simplifying model assumption and its implications. One implication is if the death threshold changes separately from the disease threshold, one has different conclusions (steepening survival but morbidity expansion).

We discuss this in the results. On page 5, line 170-173, we write:

“There is one exception to the rule that interventions that steepen survival curves compress relative sickspan. This occurs when the death threshold X_c rises but the disease threshold X_d does not (Fig. 2c,d, green lines). This might apply to human data when medical care extends life in sick individuals, without affecting age of onset of sickspan. We discuss human mortality in the last section. “

And in the discussion, page 14, line 438-447, we write:

“One exception is raising the death threshold without correspondingly raising the disease threshold. This steepens the survival curve, but increases relative sickspan, since individuals spend more time between the disease and death thresholds. As a thought experiment, consider artificial life expansion by intensive care units - such perturbations increase sickspan by delaying death. A steep survival curve in these situations can result by artificially keeping individuals alive to the same age. We assume that the longevity interventions in the controlled experiments in model organisms that we analyzed here do not cause such uncoordinated changes in death and disease thresholds. In humans, sex differences in the death and morbidity thresholds between females and males could explain the 'health survival paradox' - that females have longer lifespans but expanded morbidity. This would occur if X_c is higher in females (and X_d is either the same or lower in females).”

In the discussion we also add a paragraph on limitations of our main assumption, as suggested by the reviewer. on page 14-15, line 456-465:

“The SR model makes a major simplifying assumption—that a single damage variable X determines both morbidity and mortality. This provides quantitative predictions that can be

tested against data, as in Fig. 3,6. Future work can explore more complex situations — for example that damage in each organ affects health differently from lifespan. In such a case there might be multiple qualitatively distinct damage variables rather than a single X. There is evidence that health and lifespan may have overlapping but distinct determinants, and that cognitive functions may have somewhat different aging patterns than other physiological functions^{44,72–75}. Furthermore, on the population level, genetic and environmental heterogeneity can affect the onset of morbidity and complicate interpretation of the model. Extensions of the SR models can in principle address this complexity⁷². “

****Minor issues****

1. The empirical grounding of this paper involves the analysis of mouse lifespan. It would be very helpful to get a mouse aging researcher to comment on the analysis, as there may be species-specific limitations to the data set not apparent to non-specialists.

We have consulted with Richard Miller (head of the ITP) and incorporated his feedback in the revised manuscript, including specifically confidence intervals for steepness (Supp. Fig. 4). We also consulted Valery Krizhanovksy who studies senescence in mice models.

2. The authors seem to miss some previous work on survival shape in mice: Hughes and Hekimi Genetics 2016, Swindell Experimental Gerontology 2009.

We thank the reviewer for these references, and now cite them in 2 places: on page 2 as examples of interventions that scale and steepen survival curves; and on page 8 as previous work that analyzed the shape of mice survival curves.

Reviewer #3:

(Reviewer remarks are in purple, our responses are in black, and quotes from the manuscript are in blue)

The manuscript aims to empirically assess the theoretical implications of the saturated removal model of aging. The model suggests that interventions shifting the survival curve proportionally reshape both sickspan and healthspan, while those altering the shape of the curve affect the proportion of life spent in good or bad health. Lacking sickspan data for lab organisms, the study evaluates the interventions' effects on the shape of the survival curve to gauge their potential sickspan reduction.

We have now added sickspan data for lab animals (new Fig 3). We find that our prediction holds in three different species (nematodes, flies and mice) under various longevity interventions- the steeper the survival curve, the more compressed the relative sickspan.

We believe this makes the paper much stronger.

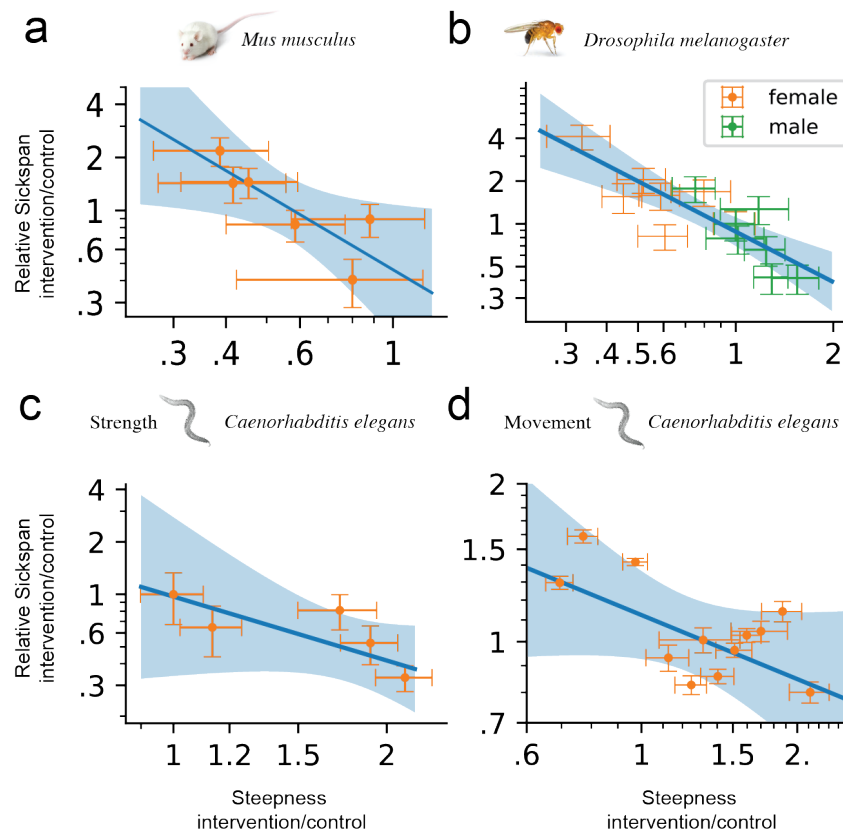


Figure 3. Relative sickspan is compressed by interventions that steepen the survival curves in model organisms, as predicted. Mean absolute sickspan divided by mean lifespan relative to control is

plotted versus steepness of survival curve relative to control. **a**, Mice under dietary longevity interventions - caloric restriction and intermittent fasting - in two strains from Luciano et al ³⁹. Sickspan was defined using mouse frailty index. See Supp. Table 1 for intervention details. **b**, Nutritional longevity interventions in two strains of *D. melanogaster*, where sickspan was defined by deficits in climbing, jumping and flying in response to a mechanical challenge. Data from Gaitanidis et al ⁴³. See Supp. Table 2 for intervention details. **c**, Genetic longevity mutants in *C. elegans*, where sickspan was defined by loss of muscle power measured by piezoelectric challenge. Data from Statzer et al ²³. See Supp. Table 3 for strain details. **d**, Genetic, dietary and environmental factors in *C. elegans*, where sickspan was measured by secession of spontaneous movements. Data from Stazer et al ²³ and Oswal et al ⁴⁴. Confidence intervals are calculated using bootstrap resampling of individuals in each cohort. See Supp. Table 4 for strain and intervention details.

While the approach is sound given the absence of empirical sickspan data, the manuscript has significant weaknesses. It overlooks the relevant and pertinent literature on the dynamics of morbidity in increased lifespans, such as the expansion, compression, and dynamic equilibrium scenarios. Authors can reference Jagger et al.'s International Handbook of Health Expectancies (in particular Chap. 2) for an entry point to this branch of literature.

We thank the reviewer for pointing out these concepts and references. We now cite them in the introduction. We specifically discuss Manton's concept of dynamic equilibrium, cite Jagger et al,

We now write in the introduction on page 1 (line 31-39):

“Understanding what sets lifespan, and how to extend it, is a central goal of longevity research. There are several possible scenarios of lifespan extension regarding sickspan ⁷. In principle, a longevity intervention could expand morbidity, and in the extreme case extend the sickspan exclusively, which is clearly undesirable. Alternatively, both healthspan and sickspan could be simply ‘stretched’ in proportion to lifespan. Finally, as proposed in the field of gerontology ^{1–3}, one could ideally aim to extend the healthspan and compress morbidity. Sickspan can be compressed either in absolute terms, as advanced by Fries ⁸ or relative to lifespan ⁹, as in the dynamic equilibrium concept by Manton in which both healthspan and lifespan are extended by similar absolute amounts ¹⁰. It is generally thought that optimal aging trajectories should restrict morbidity to a brief period before death ^{3,7}.”

While the main body of this work focuses on model organisms rather than humans, we added a new results section entitled “Steepening of human survival curves is primarily due to reduction in extrinsic mortality “ and the new Fig 6, discussing historical steepening of human survival curves. Specifically In relation to the dynamics of morbidity in different countries we write in this section, on page 12 (line 363-370):

“In addition to changes in extrinsic mortality, there have been age-specific improvements in mortality rates. Until recent decades this was primarily a reduction in young deaths (younger than median lifespan), and in the past few decades this shifted to a reduction in old deaths (older than median lifespan) ⁶⁷. These changes are likely to be due to public

health and medical interventions, as well as changes in education, socioeconomic status and other factors ⁶⁸. Such age-specific changes confound an attempt to fit changes in SR model parameters to this human data. Thus, why some countries show a reduction in relative sickspan whereas others show an increase ⁷ may not be directly linked to changes in SR model parameters.”

They should also explore James Carey's work on lifespan extension in laboratory experiments. He has produced both empirical and methodological contributions on the subject, which could provide valuable insights for the framing of the study and the interpretation of the results.

We now cite Carey's work (see quoted below, page 7, line 199-201, line 214-223). Reading Carey's work helped us to find *Drosophila* data that we can analyze for longitudinal healthspan in the new Figure 3.

“Data on longitudinal health in mice is currently limited, although it is expected to be available from several ongoing experiments ^{38–40}. We analyzed preliminary data from Luciano et al ⁴¹ which followed mouse frailty index longitudinally in different longevity interventions.

....

We next considered longitudinal experiments in flies. James Carey and colleagues pioneered the use of longitudinal life-history data such as fecundity and mortality to test evolutionary theories of aging ^{42–44} using medflies (*Ceratitis capitata*). More recently, Gaitanidis et al measured longitudinal healthspan data in *Drosophila* ⁴⁵ by quantifying age-related declines of motor functions. The authors assayed escape performance in response to gently banging food vials on the counter, i.e. “startle assays”. Each fly was scored for its jumping, climbing and flying responses to generate a health score for each individual by summing the number of functional deficits. Sickspan was defined as the period where an individual has a score below a threshold value. The experiment was repeated for 5 dietary interventions (including protein restriction, superfood and curcumin) in males and females in two fly strains with different lifespans (Lausanne, Oregon). “

The authors' interpretation of results faces two key issues. The first problem is the assumption that the rectangularization of the survival curve implies reduced frailty and a shift towards a population where the individuals are more similar and less frail, with a consequent reduction of the sickspan compared to the total lifespan. At the level of population level, the rectangularization of the survival curve has widely taken place over the past decades (commonly called reduction of lifespan inequality) and, yet, both expansion and compression of morbidity, in both absolute and relative terms, are routinely observed in empirical studies of population health. The debate on whether increased lifespan causes compression or extension of morbidity remains open, challenging the model's prediction. This is a very important point to consider when interpreting the results. In particular, caution is necessary to assume that this specific prediction of the model properly describes reality.

This comment was eye opening, and helped us to add new data and analysis to the paper. As the referee notes, there has been widespread rectangularization (increased

steepness) of human survival curves in recent decades, but relative sickspan has either grown or shrunk in different countries (e.g. compressed in Austria and expanded in England). We interpreted the historical rectangularization of human survival curves in the context of this study and found that it is due primarily to reduction in extrinsic mortality (age-independent), rather than the intrinsic (aging-related) parameters considered in the SR model. We now add a new section and figure (New Fig 6).

The new section is as follow (page 11-12, line 335-370): “

Steepening of human survival curves is largely due to reduction in extrinsic mortality

The present study primarily aims to understand lab organism data. Considerations in humans are far more complex because as compared to model organisms with defined genetics living in controlled environments, sickspan is affected by additional variations in lifestyle, environment, socioeconomic status and medical care. We therefore provide a tentative account of how the present considerations might apply to human data.

There has been a global trend of rectangularization of human survival curves in the past century ⁸ -- in many countries, the median lifespan has increased and survival curve steepness has increased as well ⁶² (Fig. 6a). There is a tight correlation between life expectancy and lifespan equality (a measure of steepness) ⁶³. There are also exceptions in certain countries in which lifespan variability has increased or remained constant in recent decades ⁶².

Relative sickspan has not been as extensively measured as lifespan, with data accumulating only recently. Available data indicates that relative sickspan since 1980 has compressed in some countries and has grown in others ^{7,9}.

We asked whether the rectangularization might be due to changes in SR model parameters, to test whether such changes might predict relative sickspan in each country. To do so, we analyzed human cohort mortality data from the Human Mortality Database (HMD) ⁶⁴. Human data shows a sizable extrinsic mortality that dominates mortality until age 30-40. In order to account for this, we fitted human hazard curves to the SR model with an added extrinsic mortality term (Fig. 6b). The fits are excellent (adjusted $R^2=0.9994$, F-test $p<10^{-16}$, see Fig. 6b).

We find that the main driver for rectangularization of the survival curve in human data is the decline in extrinsic mortality. Similar results are found when fitting human hazard to the Gamma-Gompertz-Makham equation ^{65,66} -- the age independent extrinsic mortality drops with cohort birth year (Fig. 6c), whereas the Gompertz slope and intercept remain nearly constant (Fig. 6d,e). The dominance of the extrinsic mortality drop is consistent with previous studies ^{8,10}. This contrasts with sickspan and mortality data of model organisms analyzed above in Fig. 3. Unlike human populations, these animal cohorts live in controlled environments, where extrinsic mortality can be assumed to be constant.

In addition to changes in extrinsic mortality, there have been age-specific improvements in mortality rates. Until recent decades this was primarily a reduction in young deaths (younger than median lifespan), and in the past few decades this shifted to a reduction in old deaths (older than median lifespan) ⁶⁷. These changes are likely to be due to public health and medical interventions, as well as changes in education, socioeconomic status and other factors ⁶⁸. Such age-specific changes confound an attempt to fit changes in SR model parameters to this human data. Thus, why some countries show a reduction in relative sickspan whereas others show an increase ⁷ may not be directly linked to changes in SR model parameters.

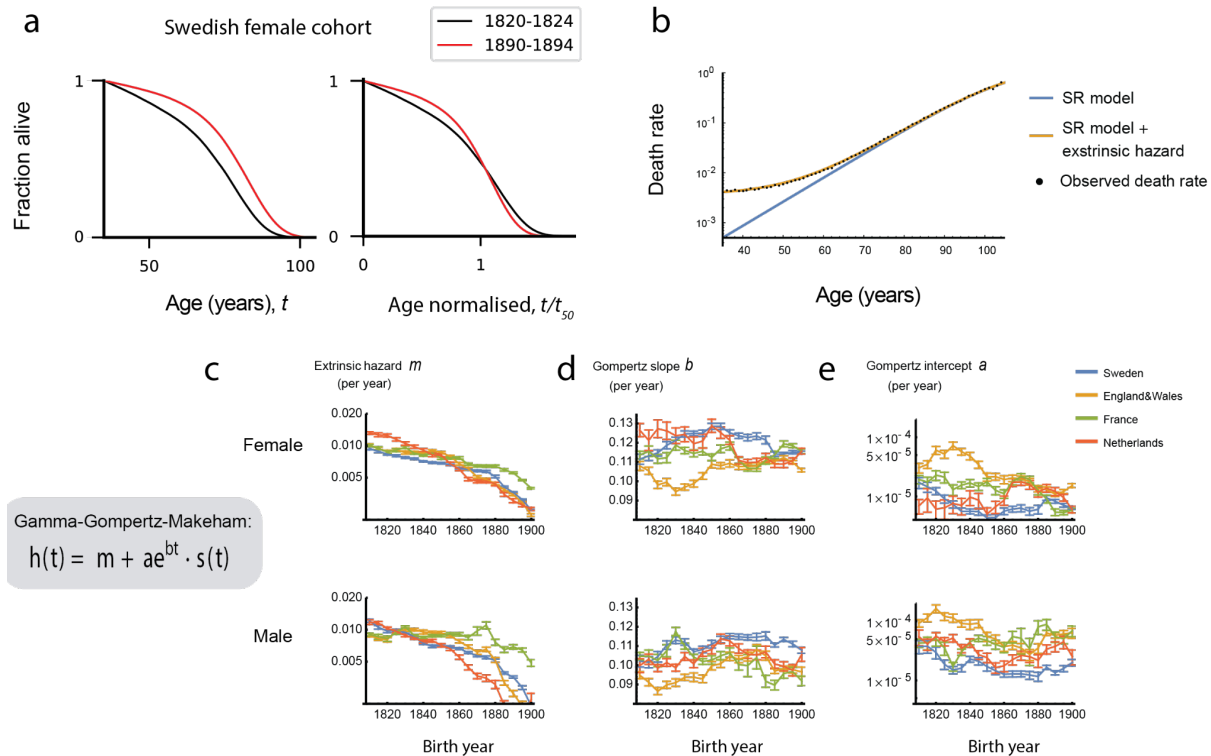


Figure 6. Steepening (rectangularization) of human survival curves in the past century is mainly driven by declining extrinsic mortality. **a**, Survival curves and rescaled survival curves of 2 cohorts of Swedish women born in 1820-1824 and 1890-1894. **b**, Hazard curve (death rates by age) of Swedish women born in 1890-1894, and of SR model fitted to this cohort. The SR model parameters that correspond to the blue and orange hazard curves are $\eta=1.04 \times 10^{-5}/\text{day}^2$, $\beta=0.45/\text{day}$, $\kappa=0.50$, $\varepsilon=0.48/\text{day}$, $X_c=19.9$ ³³. The extrinsic hazard (Makeham term) in the orange curve is $m=0.0036$ per year. **c-e**, Mortality component trends of Swedish, Dutch, French (civilian) and English & Welsh combined (civilian) cohorts born 1810 to 1910. Extrinsic hazard rates, Gompertz slopes and Gompertz intercepts are estimated using the Gamma-Gompertz-Makeham model⁶⁶. In the SR model, Gompertz slope can be approximated by $\eta X_c/\varepsilon$ and intercept is proportional to $e^{\beta X_c/\varepsilon}$ ³³. The Gamma-Gompertz correction term $s(t)$ is $s(t)=e^c/(e^c-1+e^{bt})$, where c is an additional parameter that ranges from 7.5 to 10, such that $s(t)$ only comes into effect at old age.

“

The second problem in the interpretation of the results is in the assumption that age-related morbidity is caused by the same or similar class of factors that causes age-related mortality. While this might be the case the molecular level, at the population level is not clear whether this is the case, with one more generally associated with disease prevalence and progression and the other one more associated with biological senescence.

We agree, and now address this point in the discussion section, page 14-15, line 456-465:

“The SR model makes a major simplifying assumption—that a single damage variable X determines both morbidity and mortality. This provides quantitative predictions that can be tested against data, as in Fig. 3,6. Future work can explore more complex situations — for

example that damage in each organ affects health differently from lifespan. In such a case there might be multiple qualitatively distinct damage variables rather than a single X. There is evidence that health and lifespan may have overlapping but distinct determinants, and that cognitive functions may have somewhat different aging patterns than other physiological functions^{34,41,73–75}. Furthermore, on the population level, genetic and environmental heterogeneity can affect the onset of morbidity and complicate interpretation of the model. Extensions of the SR models can in principle address this complexity³⁴.

”

Additional minor comments include improving the presentation of morbidity and mortality thresholds, which are less straightforward to interpret than the other parameters. Clarifying distinctions between interventions affecting damage removal versus morbidity/mortality thresholds is crucial.

We now clarify this in the new Fig. 2.

Additionally, details on how the saturated removal model accounts for various aging features (lines 234-237) would be beneficial for a full understanding of the model.

We now provide more detail on the aging patterns that the SR model captures quantitatively. We write in the discussion, on page 14, line 419-431:

“

The SR model was previously validated against patterns of aging from diverse studies. The model displays the Gompertz law of mortality in which hazard rises exponentially with age, as well as the late age slowdown of hazard, and explains the Weibull-like shape of mice survival curves³². The model quantitatively predicts the effect of mice parabiosis on senescent cell parameters³³ in agreement with data. Parabiosis in the model is explained by the young mouse sharing removal capacity with the old mouse (e.g. immune cells). The SR model predicts the slowdown in senescent cell removal rate with age in agreement with mice experiments³². It also explains the rapid shifts between survival curves upon dietary midlife interventions in flies, and the change in hazard slope upon temperature changes³². The SR model was extended by Katzir et al to explain in quantitative detail the exponential rise of human age-related disease incidence, as well as the incidence decline at very old ages. The extended model assumed two parameters for each disease - a disease threshold and a fraction of susceptible individuals - and provides excellent fits to the incidence curve of hundreds of age-related diseases from large medical record datasets³⁴.

“

In summary, while the manuscript holds promise, addressing these weaknesses is important for its viability in a high-impact publication.

Thanks to the reviewer comments we have now addressed the major weaknesses- we added data from lab animals on steepness and sickspan, and addressed the applicability to human data.

In summary the reviewer comments helped us add new data on our primary outcome, address human data, make the analysis more rigorous and clear. We

believe that the revised paper is much stronger, more rigorous and comprehensive.