Title: Deciphering the Timing and Impact of Life-extending Drugs: A Novel Analytic Approach that Differentiates Early, Midlife, and Senescence Phase Efficacies

Authors: Nisi Jiang[1,2]†, Catherine J. Cheng[1,2]†, Qianqian Liu[3]†, Randy Strong[1,4,5], Jonathan Gelfond[1,3]*, James F. Nelson[1,2]*

Affiliations:

[1]The Sam and Ann Barshop Institute for Longevity and Aging Studies, UT Health San Antonio; San Antonio, TX, U.S.A.

[2]Department of Cellular and Integrative Physiology, UT Health San Antonio; San Antonio, TX, U.S.A.

[3]Department of Population Health Sciences, UT Health San Antonio; San Antonio, TX, U.S.A.

[4]Department of Pharmacology, UT Health San Antonio; San Antonio, TX, U.S.A.

[5]South Texas Veterans Health Care System, San Antonio, TX, U.S.A.

*Corresponding authors. Emails: nelsonj@uthscsa.edu; gelfondjal@uthscsa.edu.

† These authors contributed equally to this work.

Abstract:

Evidence that life-extending interventions are not uniformly effective across the lifespan calls for an analytic tool that can estimate age-specific treatment effects on mortality hazards. Here we report such a tool, applying it to mouse data from 42 agents tested in the NIA Interventions Testing Program. This tool identified agents that either reduced (22) or increased (16) mortality hazards or did both (6), all with marked variation in the duration of efficacy and magnitude of effect size. Only 7 reduced mortality hazards after the 90% mortality, when the burden of senescence is greatest. Sex differences were apparent in all parameters. This new analytic tool complements the commonly used log-rank test. It detects more potential life-extending candidates (22 versus 10) and indicates when during the life course they are effective. It also uncovers adverse effects. Most importantly, it identifies agents that specifically reduce mortality hazards during the senescent phase of life.

Keywords: mortality hazard, life-extending interventions, statistical analysis, mice, longevity

Introduction:

The search for pharmacological interventions that extend the healthy lifespan has increased markedly in recent years, spurred by the discovery of a wide range of compounds, such as rapamycin and acarbose, that lengthen life of model organisms[1-3]. Whether these life-extending agents act broadly by extending survival throughout the lifespan or only affect survival during part of the life course remains unclear, in part due to the inadequacy of statistical tests used in interventional research. The log-rank test[4], widely employed in clinical trials, is the statistical tool most commonly used in aging research to determine whether an intervention, be it pharmacologic, genetic, or nutritional, is life-extending. However, its use as the primary and often only tool for this purpose is questionable for several reasons. First is its requirement for proportional hazards between compared groups, implying that treatment effects on mortality remain constant over time[5]. This assumption does not align with the evidence. Many aging interventions exert varying impacts at different life stages. For example, in an earlier analysis of data from the Interventions Testing Program (ITP), we found that many tested interventions do not adhere to the PH assumption, challenging the applicability of the log-rank test in these contexts[6]. For these interventions, we utilized the Gehan test, more robust to the PH consistency requirement and more sensitive to the effects during early adulthood. Notably, it identified five new life-extending candidates[6]. Despite its strengths, the Gehan test has its own drawback: a diminished sensitivity to effects manifesting at later life stages[7].

To assess the effects of interventions on the final phase of the aging process, methods like the Wang-Allison test have been developed to determine if treatments extend the "maximum" lifespan[8]. However, these approaches predominantly assess cumulative survival rates[9]. They do not evaluate whether an intervention specifically reduces age-specific mortality in the last phase of life when frailty, cognitive impairment, chronic disease, and other burdens of senescence peak. Although the Gompertz model has been used for evaluating age-specific or time-varying effects, it is limited by its strict parametric assumptions about the shape of the hazard function[10]. The limitations of these approaches underscore the need for a flexible tool for evaluating longevity interventions, one that accommodates the variable impacts of treatments across an organism's lifespan. Such methods should pinpoint when, for how long, and to what extent an intervention significantly alters the mortality risk. This capability is particularly crucial for identifying interventions that mitigate mortality toward the end of life when the exponential increase in the burden of senescence is greatest.

Here we introduce a novel analytic tool, a time-varying hazard ratio analysis, that detects age-specific drug effects on the mortality hazard with notable precision, thus addressing the major limitation of the log-rank test. For this investigation, we utilized publicly available data from the ITP up to 2022, comprising 42 drugs evaluated in over 27,000 genetically heterogeneous mice at 3 geographically distinct sites[11]. These agents were tested alone or in combination in 132 trials, examining the effects of sex, dosage, and age of treatment initiation. Ten of these agents have been identified by log-rank testing to significantly extend lifespan in at least one sex[12]. This is the largest compendium of mouse survival data from tests of compounds with lifespan-extension potential, an exemplary resource for testing the efficacy of our analytic tool.

Results

A new analytic tool to determine the timing and impact of life-extending candidates.

Figure 1 illustrates how the application of the time-varying hazard analysis identifies age-specific effects of an intervention on the mortality hazard, using the ITP test of green tea extract

(GTE) in females as an example. Details of the analysis are described in "Online Methods". It should be noted that GTE had no effect on survival by log-rank testing[13]. Figure 1A shows the Kaplan-Meier survival plots for treatment and control groups. These plots indicate that the proportional hazard assumption is likely violated due to the crossing survival curves, which was confirmed by the z-test (Table S1).

Figure 1B is a graphical representation of the mortality hazards of the control and GTE-treated groups throughout the period of testing, using a method described previously[14,15]. The mortality hazard of the GTE-treated group is reduced relative to that of the control group before the median lifespan, but shortly thereafter crosses over, exceeding that of the control group.

Figure 1C shows the application of the time-varying hazard ratio analysis to the GTE data. The log ratio of the mortality hazards of GTE treated and control groups shown in Figure 1B is calculated along with its 95% confidence intervals. Negative values (i.e., log hazard ratios < 0) indicate that GTE has a beneficial effect (i.e., lower mortality hazard than the control group), while positive values (i.e., log hazard ratio > 0) reveal adverse effects of GTE treatment. The 95% confidence intervals of the mortality hazard ratio were estimated using 1,000 bootstrapped replications[15], shown as dashed lines in the figure. Ages when the hazard ratio is < 0 and the upper 95% confidence limit is also < 0 indicate when the treatment is significant for reducing mortality. The duration (age range) of significance is bounded by the ages when the upper 95% confidence interval crosses 0, as illustrated. Conversely, ages when the hazard ratio is > 0 and the lower 95% confidence limit is > 0, correspond to ages when the treatment is significantly increasing mortality, and the duration of significance is bounded in the same way. This analysis reveals that GTE both reduced mortality hazards during midlife and increased mortality hazards toward the end of life.

Figure 1D is the graphical representation of the results of the new analysis. It combines the features of Figure 1C into an annotated horizontal heatmap to facilitate cross-drug comparisons. Colors encode the drug effects on a single horizontal band from birth to the death of the last mouse in either the control group or treated group, whichever is first. The heatmap is blank until treatment begins. During the period of treatment, gray designates ages with no significant treatment effect. Green and red designate ages when the treatment is significantly reducing or increasing the mortality rate, respectively. Color intensity is directly proportional to effect size (log HR). This graphic representation facilitates comparisons across various treatments as shown in Figures 2 and 3. Unlike the log-rank test, this tool can identify at what ages, for how long, and to what extent an intervention significantly alters the age-specific mortality hazard. It also can show whether the intervention decreases or increases the mortality rate.

Greater sensitivity and precision in identifying mortality-modifying interventions

Figure 2 shows all the interventions identified by the new analytic tool that significantly reduced (or increased) the age-specific mortality hazard during treatment, using the annotated horizontal heatmap graphical representation. The hazard ratio plots used to generate these heatmaps, calculated by the time-varying hazard ratio analysis, are shown in Figures S2 and S3, for males and females, respectively. In this Figure, the interventions are ranked from the earliest to the oldest age of cessation of beneficial effect in males. Figure S1 shows the results ranking by the age of cessation of beneficial effect in females. Thirty-two compounds, consisting either of a single agent or a combination of two agents, at one or more doses, initiated at varying ages, significantly modified the mortality hazard in one or both sexes at one or more periods during

the treatment period. This analysis identified 12 new compounds that significantly reduced mortality in at least one sex during treatment but were overlooked by the log-rank test: namely, candesartan cilexetil (CC), caffeic acid phenethyl ester (CAPE), 17-dimethylaminoethylamino-17-demethoxygeldanamycin hydrochloride (DMAG), enalapril, GTE, L-leucine, metformin, oxaloacetic acid (OAA), PB125, simvastatin, syringaresinol (Syr), and ursodeoxycholic acid (UDCA). The new analysis also identified 16 compounds that were detrimental (i.e., increased mortality) in one or both sexes at one or more periods of treatment. The duration of significant benefit or detriment varied markedly, from a few days (e.g., simvastatin, minocycline) to the entire treatment period (e.g., rapamycin + acarbose). Most drugs only reduced mortality or only increased mortality. Two exceptions were UDCA in males and GTE in females. Both were beneficial for several months before the median lifespan and detrimental for several months in very old mice. Effect sizes, indicated by the color intensity, varied markedly during the periods of benefit and detriment. Acarbose had its greatest benefit at the initiation of treatment, waning progressively thereafter. Effect sizes of other compounds, such as butanediol and captopril in males and many of the different rapamycin trials in females peaked during the middle of treatment. A few interventions showed a steady increase in effect with continued treatment (e.g., glycine in males and leucine in females).

Only a fraction of interventions reduced mortality at later ages

One of the benefits of the new analytic tool is its ability to estimate when during the life course and for how long an agent exerts its effect on survival. In males, 17 drugs reduced mortality hazards at some point during the life course (Figure 2). Of these, 9 compounds only reduced mortality risk in early and mid-adulthood (i.e., before reaching the median lifespan): Syr, (R/S)-1,3-butanediol (BD), CC, captopril, enalapril, UDCA, metformin, and DMAG, and nordihydroguaiaretic acid (NDGA) at 800 ppm. The two higher doses of NDGA had a slightly longer period of benefit, but only a few days beyond the median lifespan. By contrast, in females, of the 11 agents that reduced mortality risk at some stage of life, a much smaller fraction only reduced mortality during early- to mid-adulthood: GTE and OAA. In males, five compounds tested in 11 trials demonstrated reduced mortality after attainment of median lifespan, although these effects vanished before mice attained ages reaching the 90% mortality benchmark: 17α-estradiol, aspirin at 21 ppm, Protandim, high doses of NDGA, and 3 of 4 late-onset (20 mo) rapamycin treatments. Notably, only 6 of the 17 compounds that reduced mortality in males did so at ages beyond the 90% mortality threshold: canagliflozin, acarbose, 17α-estradiol, glycine, simvastatin, rapamycin, and cocktails of either acarbose or metformin with rapamycin. In females, in contrast to males, most of the drugs with beneficial effects (i.e., 9 of 11) reduced mortality mainly at ages after attainment of median lifespan. 8 trials involved drugs that reduced mortality at ages after the median lifespan, but half of these lost efficacy before reaching 90% mortality, including CAPE, glycine, CC, PB125, acarbose, and one trial of late-onset rapamycin treatments; 14 trials had mortality-reducing compounds with efficacy beyond the 90% mortality milestone: predominantly rapamycin-related (10 out of 14), and BD, L-leucine, and captopril (Figure S1).

Some compounds have adverse effects on mortality hazards

Unexpectedly, the new analysis identified 16 compounds that adversely affected mortality at specific stages in the life course in at least one sex (Figure 3). Since the inception of the ITP, concerns have lingered about the potential lifespan-shortening and mortality hazard-increasing effects of some compounds. However, no substances were identified with significant adverse effects using the log-rank test. The analytic tool used here revealed 20 trials with 16 compounds that increased the mortality hazard at one or more periods of treatment: 5 in males and 15 in

females. In males, UDCA, 3-(3-hydroxybenzyl)-5-methylbenzo[d]oxazol-2(3H)-one (MIF098), 2-(2-Hydroxyphenyl)benzoxazole (HBX), resveratrol, and INT-767 had significant negative impacts on the mortality hazard ratio. In females, CC, metformin, DMAG, canagliflozin, metformin combined with rapamycin, 17α-estradiol, GTE, minocycline, beta-guanidinopropionic acid (bGPA), geranylgeranyl acetone (GGA), fish oil, nicotinamide riboside (NR), UDCA, and MIF098 exhibited detrimental effects. UDCA and MIF098 affected both sexes adversely. Notably, some compounds displayed dual effects. In males, UDCA treatment showed early protective effects during pre-median lifespan stages but at later ages manifested significant negative impacts. In females, three experiments demonstrated mixed outcomes. GTE, akin to UDCA's pattern in males, had early beneficial effects but turned detrimental nearing the 90% survival mark. CC exhibited significant early adverse effects but became protective post-median lifespan. Most intriguing was the metformin and rapamycin combination (MetRapa), which presented pronounced benefits beyond the 90% mortality threshold but briefly exhibited significant detrimental effects at advanced ages in females. Metformin given alone also had detrimental effects briefly at the end of life in females.

Sex differences in the effect of pharmacological interventions

Marked sex differences in the responses to life-extending drugs are one of the key outcomes of the ITP[12]. In addition to those already noted, this new analytic tool unveiled even more sex differences in the response to the pharmacological interventions of the ITP. It identified 6 additional compounds that only benefited males: Syr, enalapril, simvastatin, metformin, DMAG, and UDCA, and 5 drugs that only reduced mortality in females: OAA, CAPE, PB125, Leu, and GTE. Notably, 7 interventions, including UDCA, CC, metformin, DMAG, canagliflozin, MetRapa, and 17α-estradiol, exhibited beneficial effects in males but detrimental effects in females (Figure 3). More compounds adversely affected survival in females (13) than in males (4). Moreover, most drugs with negative effects exerted their effect on females almost from the beginning of treatment. The detrimental effects waned during the 2nd year of life but sometimes reappeared in the final stage of life. Six agents only had deleterious effects late in life: fish oil, GTE, metformin, and MetRapa in females, and MIF098, INT-767, and resveratrol in males. These results underscore the necessity of including both sexes when testing longevity interventions.

Discussion

The analytic method presented here promises to be broadly useful and impactful for interventional research on aging. Revealing the age-specific effects of interventions on the mortality hazard opens the door to asking more nuanced and targeted questions about the actions of an intervention. Answers to such questions can ultimately lead to greater life-extending efficacy of interventions and a better understanding of the underlying mechanisms that the interventions target. The analysis does this by providing estimates of when and for how long during the life course an intervention reduces (or in the case of detrimental effects, increases) age-specific mortality. It also provides an estimate of the effect size of an intervention and how the strength of its effect changes over the course of treatment. None of this information is attainable by traditional methods such as the log-rank test, the current standard for evaluating longevity interventions.

This new method can distinguish interventions that specifically reduce mortality during senescence from those that only affect survival during midlife or earlier. This is an important distinction in the search for therapeutic interventions that benefit individuals of advanced age when the burdens of senescence are greatest. This analytic tool is also sensitive to adverse effects—critically important for pre-clinical models that aim to be translatable. Furthermore, the

method is sensitive to sex differences in timing, duration, and efficacy of interventions, as well as adverse outcomes—providing further impetus to probe the mechanisms underlying the growing number of sexually dimorphic traits in aging. Here we discuss some of the ways the new information provided by this analytic tool can assist drug discovery, the search for the underlying mechanisms that drive aging, and other areas of Geroscience. These are only a few examples of how this analytical tool can be utilized. Additional applications will likely emerge as its adoption spreads within the geroscience community.

A major discovery using this tool is that the effect of virtually every intervention analyzed was non-uniform across the life course. This observation is not readily apparent by visual inspection of most Kaplan-Meier plots and is not obtainable from log-rank tests. Very few interventions significantly reduced (or increased) mortality through the entire course of treatment. Most were only effective for less than half of the treatment duration. This calls for explanation, and the answers are likely to lead to better interventions and greater insight into the mechanisms of aging. One possibility is that the aging process may impact drug efficacy. The decrease, increase, or loss of efficacy of an intervention may reflect age-related changes in pharmacokinetics or pharmacodynamics, leading to suboptimal (over or under) dosage. The ITP database provides some insights into this question, because some interventions were tested at several doses. Acarbose, for example, was tested at three doses. Acarbose efficacy in males, measured as the reduced mortality hazard ratio, increased with increasing doses during the initial period of treatment, but paradoxically, its beneficial effect ceased at progressively earlier ages with increasing doses. This finding opens the door to developing age-specific doses to sustain efficacy for longer periods and raises awareness of the importance of understanding the role of aging in pharmacokinetics. An alternative explanation for the complex response to varying doses of acarbose is an age-related change in pharmacodynamics. It is plausible that the aging processes or causes of mortality change with age and the intervention loses efficacy because it no longer targets the underlying pathways. Whatever the reason, this tool has uncovered a critical variable that needs to be considered in interventional geroscience.

Another important outcome of the application of this analytic tool to longevity data is the finding that only a subset of the interventions in the ITP database affected age-specific mortality rates in the last half of the lifespan, and even fewer affected mortality rates at ages after the age when 90% of the control cohort has died (maximum lifespan)[8]. Diet restriction has long been considered an example of an intervention that retards aging processes broadly, because it extends the age of 90% mortality, distinguishing it from many interventions that do not—the latter often only extending the median lifespan[16,17]. Several studies, including the ITP, use the Wang-Allison test as a discriminator for interventions that do or do not extend the maximum lifespan based on the 90% mortality measure. However, this test does not distinguish whether an increase in age of 90% mortality reflects the effects of reduced mortality accumulated during earlier ages from the effects of the age-specific mortality reduction at or near the age of 90% mortality. Also, it provides no information on the effect of the intervention on the last 10% of the population. This distinction is of particular importance to a major goal of Geroscience: namely, to identify compounds and discover the underlying mechanisms that unequivocally extend the maximum lifespan of a species by reducing age-specific mortality during the later stages of life when the burden of senescence is greatest. The analytic tool described here provides such a measure by indicating whether the intervention specifically reduces mortality rates in the final stage of life, whether it be after the age at which 90% of the control population has died or some other late age. Only a subset of the interventions that have been reported by the ITP as "lifespan extending" using log-rank analysis reduced mortality hazard after the median lifespan, and even fewer did so during what is considered the maximum lifespan.

Nevertheless, compounds that only reduce mortality during the first half of adult life should not be discounted. Reducing mortality at any stage of life can be impactful, especially when considering its potential translatability to humans. For example, the male mortality disadvantage, compared to females, is greatest in the first half of adult life in both humans and UM-HET3 mice[15]. It is noteworthy that most of the compounds that are only effective in males are only effective during the first half of the lifespan. Castration of UM-HET3 males before puberty eliminates this mortality disadvantage[14]. If any of the drugs that only eliminate the male mortality disadvantage during this period can do so without interfering with male reproductive function, the societal impact if clinically translatable would be great[12,18].

This method is not only more sensitive to agents that reduce age-specific mortality, it also is more sensitive to those that increase mortality. The ITP never identified adverse effects using the log-rank test. This new tool revealed 20 trials involving 16 compounds that increased mortality hazards at certain life stages in at least one gender. There was a marked sex difference. Only 5 trials showed detrimental effects in males compared to 15 trials in females. MIF098 was the only drug that adversely affected both sexes. Some compounds, including canagliflozin and high doses of 17α-estradiol markedly reduced mortality in males but were harmful in females. These findings underscore the need for sex-specific testing of life-extending candidates.

This new analytic tool can detect reversals of the benefit of compounds across the life course. UDCA in males and GTE in females reduced mortality before the median lifespan but increased mortality at later ages—another discriminator not possible using the log-rank test. There is precedence for this reversal. In humans, individuals reporting the lowest intake of dietary protein had reduced mortality from cardiovascular disease and cancer before 65 years of age, but this relationship reversed after 65[19]. Mice with reduced branch chain amino acid intake had extended life when the diet began in early adulthood, but their lifespan was unaffected when the diet was initiated at a later age[20]. Age-related changes in pharmacokinetics and pharmacodynamics may play a role here. For example, blood levels of canagliflozin, whose beneficial effects in males diminish with age and are absent in females are 2-3-fold higher in older males and similarly elevated at all ages in females[21].

Another strength of this tool is its heightened sensitivity to potential life-extending candidates. It identified over twice as many as the log-rank test. This is due in part to its ability to identify age-specific effects on the mortality hazard unimpeded by the requirement of the log-rank test for consistent proportional hazard across the duration of treatment. The newly identified compounds generally have smaller effect sizes and shorter durations of positive effect compared to those identified by the log-rank test. Given their geroprotective potential, they deserve further study. It is important to emphasize that neither this nor any other statistical tool should be used as a final arbiter of any candidate for mortality reduction and lifespan extension (or adverse effect), but rather should be considered a screening tool for identifying potential candidates that deserve follow-up—for example with different doses. We would argue that Type 1 errors (i.e., false positives) during initial screens are more acceptable and preferable to false negatives.

It is important to note the limitations of this method and consider ways to increase its utility. While the flexible estimation of the hazard ratio makes few assumptions about the proportional hazard, the precision of the hazard ratio varies throughout the lifespan. The hazard ratio precision tends to be low during early life due to the lower rate of mortality. The method requires a larger sample size than the log-rank test, and the log-rank test has higher power when the PH assumption is met. The bootstrap confidence intervals mitigate the false positive findings for a

wide range of sample sizes, but the Type I error rate is controlled at each specific time point which does not ensure tight Type I error rates for the full lifespan that a permutation test might achieve. The method currently does not explicitly consider uncertainty in the Time axis so the ages at which the treatment effect becomes nonzero are presented as point-estimates without confidence intervals. However, this limitation did not prevent the consistent findings between similar treatments such as the early effects of ACE inhibitors (Enalapril and Captopril) or early effects of different doses of NDGA. Statistically testing whether two different treatments have the same effect relative to control is more complex (testing whether the ratio of hazard ratios is 1) and may require comparisons across cohorts. While this method allows the estimation of time-varying treatment effects relative to control, future extensions of the method could explicitly test and estimate differences in active treatments in terms of time and magnitude of reduction or increase in the mortality hazard ratio.

353 **Author contributions**
354 Conceptualization: CJC, NJ, JG, QL, JN, RS
355 Methodology: CJC, NJ, JG, QL, JN
356 Funding acquisition: RS, JN
357 Data analysis: QL, NJ, CJC, JG, JN
358 Writing: NJ, JN, JG
359
360 **Acknowledgments**

361
362 Dr. Strong has been honored with the Senior Research Career Scientist award (# IK6 BX006289)
363 from the Department of Veterans Affairs. The funding for this research was provided by the Center
364 for Testing Potential Anti-Aging Interventions (5U01AG022307), the Nathan Shock Center of
365 Excellence in Basic Biology of Aging (5P30AG013319), as well as NIH grant 5T32AG021890-15,
366 alongside a fellowship from the Glenn Foundation awarded to CJC.
367
368 **Conflict of interest**
369 The authors declare no conflict of interest.
370
371 **Funding**
372 National Institute on Aging grant 5U01AG022307 (RS, JN)
373 National Institute on Aging grant 5P30AG013319 (RS, JN)
374

Online Methods:

Data availability, mouse model, and husbandry

The datasets employed in this study are sourced from the Mouse Phenome Database (MPD; phenome.jax.org), encompassing all data from the Interventions Testing Program (ITP) spanning from 2004 to 2022. This dataset incorporates 13 distinct cohorts, integrating data across three research facilities to ensure the robustness and reproducibility of the findings. The ITP employed the UM-HET3 mouse line, a genetically heterogeneous model, chosen for its relevance to the genetically diverse human population. UM-HET3 mice are bred according to a specific crossbreeding protocol: BALB/cByJ females are mated with C57BL/6J males to produce F1 hybrid females, which are then bred with F1 hybrid males derived from mating C3H/HeJ females with DBA/2J males. This breeding strategy is designed to maximize genetic diversity within the model, thereby approximating the genetic variability inherent in human populations and increasing the translational value of the research findings. The mice designated for longevity assays were maintained under controlled environmental conditions, with a constant ambient temperature of 25°C and a regulated photoperiod of 12 hours light/12 hours darkness. Nutritional needs were met with ad libitum access to the Purina 5LG6 diet, alongside specific drugged food formulations as per experimental requirements. Housing protocols were optimized for social enrichment and welfare, accommodating up to three males or five females per standard laboratory enclosure, in accordance with established ethical guidelines. Rigorous daily health assessments were conducted by trained staff to monitor the well-being of the subjects, promptly identify morbidity signs, and implement early intervention strategies as necessary. This proactive health management approach minimized unnecessary suffering and ensured the reliability of longevity data. The specifics of drug administration, including dosage, frequency, and duration, as well as the rationale behind the selection of intervention agents, are detailed in the original published reports, providing a comprehensive overview of the therapeutic strategies explored in this body of research.

Description of the time-varying hazard ratio test

Our investigation into the impact of various treatments on age-specific mortality utilized a piecewise polynomial B-spline hazard model. This model, which assumes a Poisson distribution, was applied using the bshazard package in R, offering a robust tool for analyzing the complex interplay between treatment effects and mortality over time. By integrating this model, we were able to capture the nuanced variations in mortality risk associated with different treatment regimens across the lifespan of the subjects. A key aspect of our analytical strategy was the generation of a nonparametric smoothed estimate of the baseline hazard rate. This was achieved by stratifying the survival data by both treatment and sex within each cohort, thereby allowing for a precise adjustment for the site-specific effects that might otherwise confound the treatment impact assessment. Importantly, this approach facilitated a refined understanding of how baseline mortality risks shift in response to treatment interventions, while accounting for potential biological differences in treatment efficacy between males and females. In our analysis, mortality events occurring prior to the initiation of treatment were excluded to ensure that the hazard ratio estimates accurately reflect the treatment's effect on survival. This exclusion criterion is crucial for eliminating bias arising from pre-treatment mortality, thus enhancing the validity of our findings. The confidence intervals for the treatment hazard ratio were estimated using 1,000 bootstrapped replications. This age-specific analysis is similar to that reported in the estimated age-specific effects of sex[15]. The conventional evaluation of the time-varying hazard (age-specific mortality) was conducted using the test of proportional

hazards (PH) assumption (z-test)[22]. This test assesses whether the hazard ratio (treatment effect) varies across the lifespan. This test was performed for each sex and treatment combination. Test results are shown in Table S1. The distribution of PH violation p-values was 5% with p<.01 (5 times expected), 12% with p<.05 (2.4 times expected), 19% with p<.1 (1.9 times expected). The p-values were combined using Fisher's Method indicating that the PH assumption does not hold (p<.0001) for a subset of interventions.

Data Visualization

The visualization method uses a color-coded band to depict treatment effects on hazard ratios, with the pre-treatment phase shown as a blank band. Upon treatment initiation, a gray color indicates no detectable effect, while significant effects are represented by changes in color intensity: beneficial effects cause the band to turn green, with the intensity reflecting the magnitude of negative log hazard ratios, and detrimental effects are shown in red, with intensity corresponding to positive log hazard ratios. The transition points where significant effects begin, or end are marked by dashed lines. Additionally, key lifespan metrics for the control group, such as median and maximum lifespan (when 90% have died), are highlighted to facilitate interpretation. All computational analyses were conducted in R (version 4.3, Vienna, Austria).

References:

1 Harrison, D. E. *et al.* Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. *Nature* **460**, 392-395, doi:10.1038/nature08221 (2009).

2 Harrison, D. E. *et al.* Acarbose, 17-alpha-estradiol, and nordihydroguaiaretic acid extend mouse lifespan preferentially in males. *Aging Cell* **13**, 273-282, doi:10.1111/acel.12170 (2014).

3 Knufinke, M., MacArthur, M. R., Ewald, C. Y. & Mitchell, S. J. Sex differences in pharmacological interventions and their effects on lifespan and healthspan outcomes: a systematic review. *Front Aging* **4**, 1172789, doi:10.3389/fragi.2023.1172789 (2023).

4 Mantel, N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep* **50**, 163-170 (1966).

5 Bouliotis, G. & Billingham, L. Crossing survival curves: alternatives to the log-rank test. *Trials* **12**, A137, doi:10.1186/1745-6215-12-S1-A137 (2011).

6 Jiang, N., Gelfond, J., Liu, Q., Strong, R. & Nelson, J. F. The Gehan test identifies life-extending compounds overlooked by the log-rank test in the NIA Interventions Testing Program: Metformin, Enalapril, caffeic acid phenethyl ester, green tea extract, and 17-dimethylaminoethylamino-17-demethoxygeldanamycin hydrochloride. *bioRxiv*, 2024.2002.2017.579363, doi:10.1101/2024.02.17.579363 (2024).

7 Harrington, D. P. & Fleming, T. R. A class of rank test procedures for censored survival data. *Biometrika* **69**, 553-566 (1982).

8 Wang, C., Li, Q., Redden, D. T., Weindruch, R. & Allison, D. B. Statistical methods for testing effects on "maximum lifespan". *Mech Ageing Dev* **125**, 629-632, doi:10.1016/j.mad.2004.07.003 (2004).

9 Gao, G., Wan, W., Zhang, S., Redden, D. T. & Allison, D. B. Testing for differences in distribution tails to test for differences in 'maximum' lifespan. *BMC Med Res Methodol* **8**, 49, doi:10.1186/1471-2288-8-49 (2008).

10 Pletcher, S. D., Khazaeli, A. A. & Curtsinger, J. W. Why do life spans differ? Partitioning mean longevity differences in terms of age-specific mortality parameters. *J Gerontol A Biol Sci Med Sci* **55**, B381-389, doi:10.1093/gerona/55.8.b381 (2000).

11 Miller, R. A. *et al.* An Aging Interventions Testing Program: study design and interim report. *Aging Cell* **6**, 565-575, doi:10.1111/j.1474-9726.2007.00311.x (2007).

12 Jiang, N. & Nelson, J. F. Sex Differences in Mouse Longevity and Responses to Geroprotective Drugs: Implications for Human Intervention. *Public Policy Aging Rep* **33**, 120-124, doi:10.1093/ppar/prad026 (2023).

13 Strong, R. *et al.* Evaluation of resveratrol, green tea extract, curcumin, oxaloacetic acid, and medium-chain triglyceride oil on life span of genetically heterogeneous mice. *J Gerontol A Biol Sci Med Sci* **68**, 6-16, doi:10.1093/gerona/gls070 (2013).

14 Jiang, N. *et al.* Prepubertal castration eliminates sex differences in lifespan and growth trajectories in genetically heterogeneous mice. *Aging Cell* **22**, e13891, doi:10.1111/acel.13891 (2023).

15 Cheng, C. J., Gelfond, J. A. L., Strong, R. & Nelson, J. F. Genetically heterogeneous mice exhibit a female survival advantage that is age- and site-specific: Results from a large multi-site study. *Aging Cell* **18**, e12905, doi:10.1111/acel.12905 (2019).

16 Yu, B. P., Masoro, E. J. & McMahan, C. A. Nutritional influences on aging of Fischer 344 rats: I. Physical, metabolic, and longevity characteristics. *J Gerontol* **40**, 657-670, doi:10.1093/geronj/40.6.657 (1985).

17 Masoro, E. J. Biology of Aging: Current State of Knowledge. *Archives of Internal Medicine* **147**, 166-169, doi:10.1001/archinte.1987.00370010164033 (1987).

18 Jiang, N., Cheng, C. J., Strong, R. & Nelson, J. F. Castration reduces mortality and increases resilience in male mice: what is next? *Geroscience* **46**, 2787-2790, doi:10.1007/s11357-023-00973-5 (2024).

19 Levine, M. E. *et al.* Low protein intake is associated with a major reduction in IGF-1, cancer, and overall mortality in the 65 and younger but not older population. *Cell Metab* **19**, 407-417, doi:10.1016/j.cmet.2014.02.006 (2014).

20 Richardson, N. E. *et al.* Lifelong restriction of dietary branched-chain amino acids has sex-specific benefits for frailty and lifespan in mice. *Nat Aging* **1**, 73-86, doi:10.1038/s43587-020-00006-2 (2021).

21 Miller, R. A. *et al.* Canagliflozin extends life span in genetically heterogeneous male but not female mice. *JCI Insight* **5**, doi:10.1172/jci.insight.140019 (2020).

22 Grambsch, P. M. & Therneau, T. M. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* **81**, 515-526 (1994).

23 Strong, R. *et al.* Lifespan benefits for the combination of rapamycin plus acarbose and for captopril in genetically heterogeneous mice. *Aging Cell* **21**, e13724, doi:10.1111/acel.13724 (2022).

Figures:

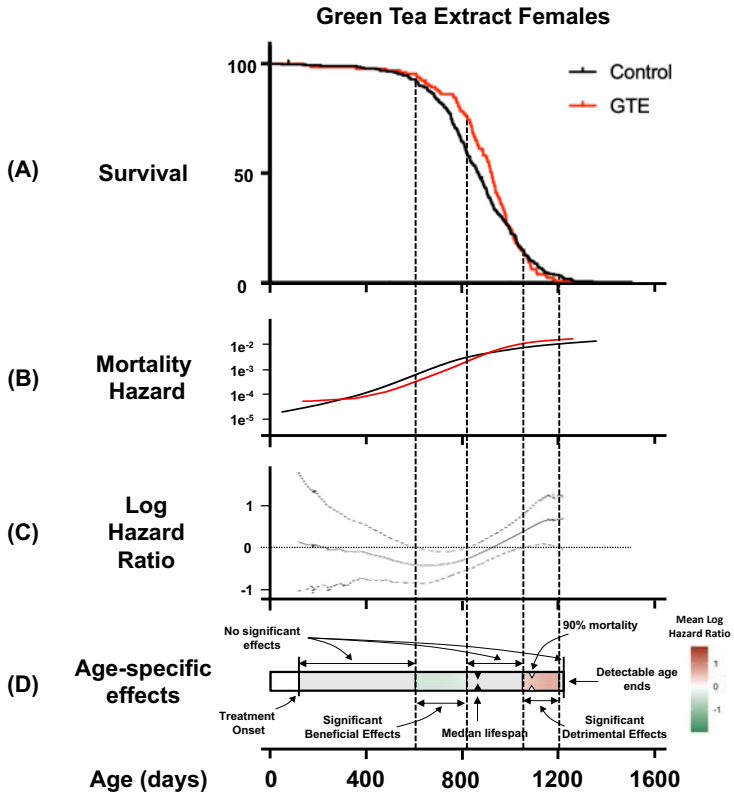

Figure 1. Graphical representation of the analytic tool for determining the timing and impact of life-extending candidates. Survival data are from the test of Green Tea Extract in females[13]**. A)** Kaplan-Meier survival curves of the GTE-treated female mice (Red) and control female mice (Black); **B)** Age-specific mortality hazards of GTE-treated and control mice groups with 95% confidence intervals; **C)** Mortality hazard ratio between GTE-treated and control mice groups and 95% confidence intervals shown as dashed lines; **D)** Life course heat map visualization of the age-specific effects of GTE on the mortality hazard ratio. Vertical dashed lines mark the boundaries of significant effects on the mortality hazard ratio based on the ages when the 95% confidence intervals in Figure C cross 0.

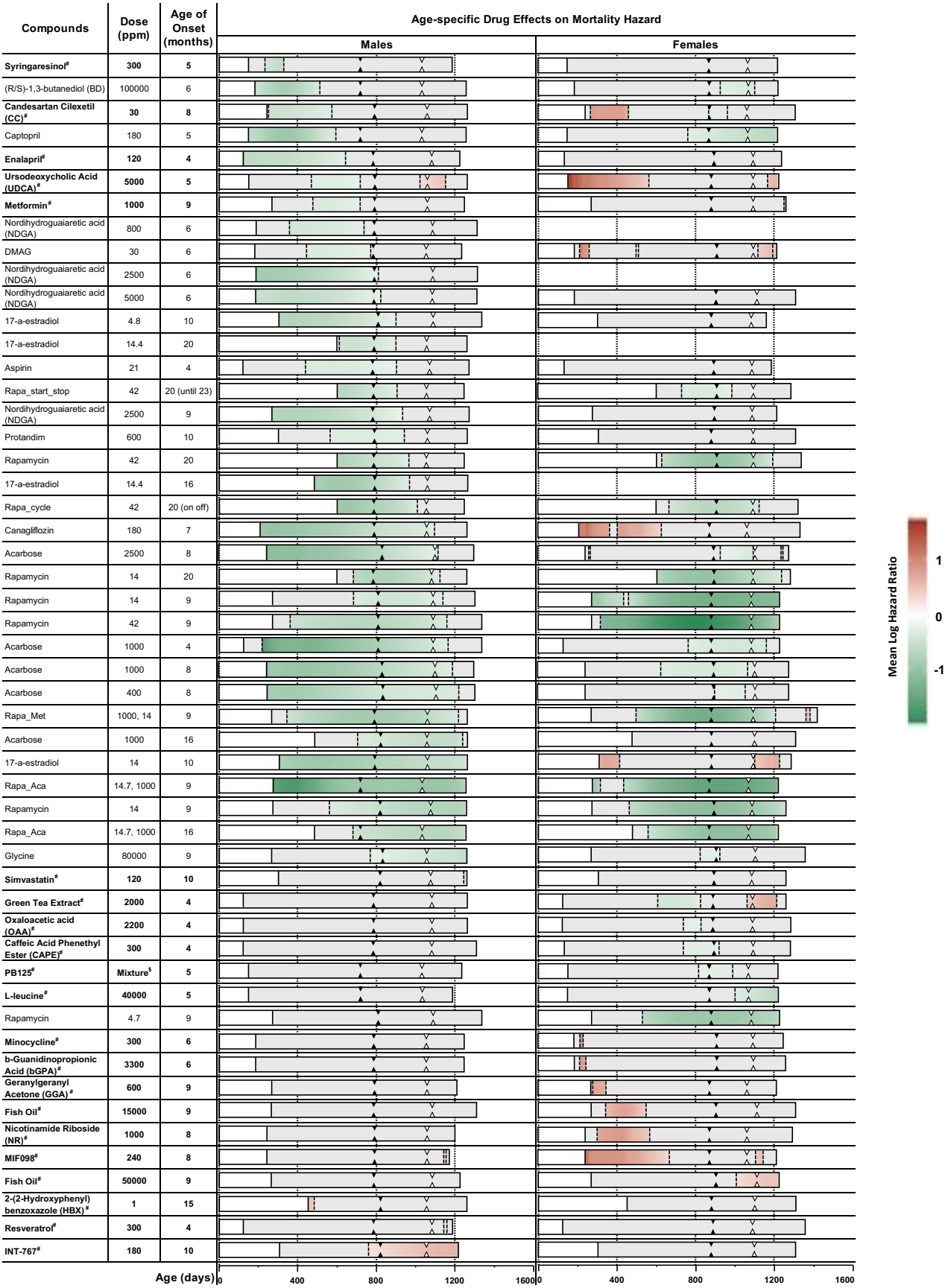
Compounds
Dose (ppm)
Age of Onset (months)
Age-specific Drug Effects on Mortality Hazard
Males
Females
Syringaresinol#
300
5
(R/S)-1,3-butanediol (BD)
100000
6
Candesartan Cilexetil (CC)#
30
8
Captopril
180
5
Enalapril#
120
4
Ursodeoxycholic Acid (UDCA)#
5000
5
Metformin#
1000
9
Nordihydroguaiaretic acid (NDGA)
800
6
DMAG
30
6
Nordihydroguaiaretic acid (NDGA)
2500
6
Nordihydroguaiaretic acid (NDGA)
5000
6
17-a-estradiol
4.8
10
17-a-estradiol
14.4
20
Aspirin
21
4
Rapa_start_stop
42
20 (until 23)
Nordihydroguaiaretic acid (NDGA)
2500
9
Protandim
600
10
Rapamycin
42
20
17-a-estradiol
14.4
16
Rapa_cycle
42
20 (on off)
Canagliflozin
180
7
Acarbose
2500
8
Rapamycin
14
20
Rapamycin
14
9
Rapamycin
42
9
Acarbose
1000
4
Acarbose
1000
8
Acarbose
400
8
Rapa_Met
1000, 14
9
Acarbose
1000
16
17-a-estradiol
14
10
Rapa_Aca
14.7, 1000
9
Rapamycin
14
9
Rapa_Aca
14.7, 1000
16
Glycine
80000
9
Simvastatin#
120
10
Green Tea Extract#
2000
4
Oxaloacetic acid (OAA)#
2200
4
Caffeic Acid Phenethyl Ester (CAPE)#
300
4
PB125#
Mixture$
5
L-leucine#
40000
5
Rapamycin
4.7
9
Minocycline#
300
6
b-Guanidinopropionic Acid (bGPA)#
3300
6
Geranylgeranyl Acetone (GGA)#
600
9
Fish Oil#
15000
9
Nicotinamide Riboside (NR)#
1000
8
MIF098#
240
8
Fish Oil#
50000
9
2-(2-Hydroxyphenyl) benzoxazole (HBX)#
1
15
Resveratrol#
300
4
INT-767#
180
10
Age (days)
0
400
800
1200
1600
Mean Log Hazard Ratio
1
0
-1

Figure 2. Interventions that significantly modified mortality hazard using the time-varying hazard analytic tool. Each row represents an individual trial of one intervention in a single cohort. Each intervention involved one compound or a combination of two, with dosage and starting age of treatment listed. Trials are ranked by the cessation age of beneficial effects in males, from earliest to latest age. The color-coded bands denote the temporal significance of drug effects: white indicates the period before treatment onset, gray marks periods with no significant effects, green indicates periods of significant beneficial effects, and red denotes intervals of significant detrimental effects. The solid black triangle indicates the median lifespan of the control group for each trial, and the open triangle marks the age of 90% mortality of the control group.

Footnotes:

New Compounds that significantly affect mortality (i.e., not identified by the log-rank test) are also noted in **bold** font.

$ PB125 is a mixture of luteolin, withaferin A, and carnosol, dosages refer to publication[23].

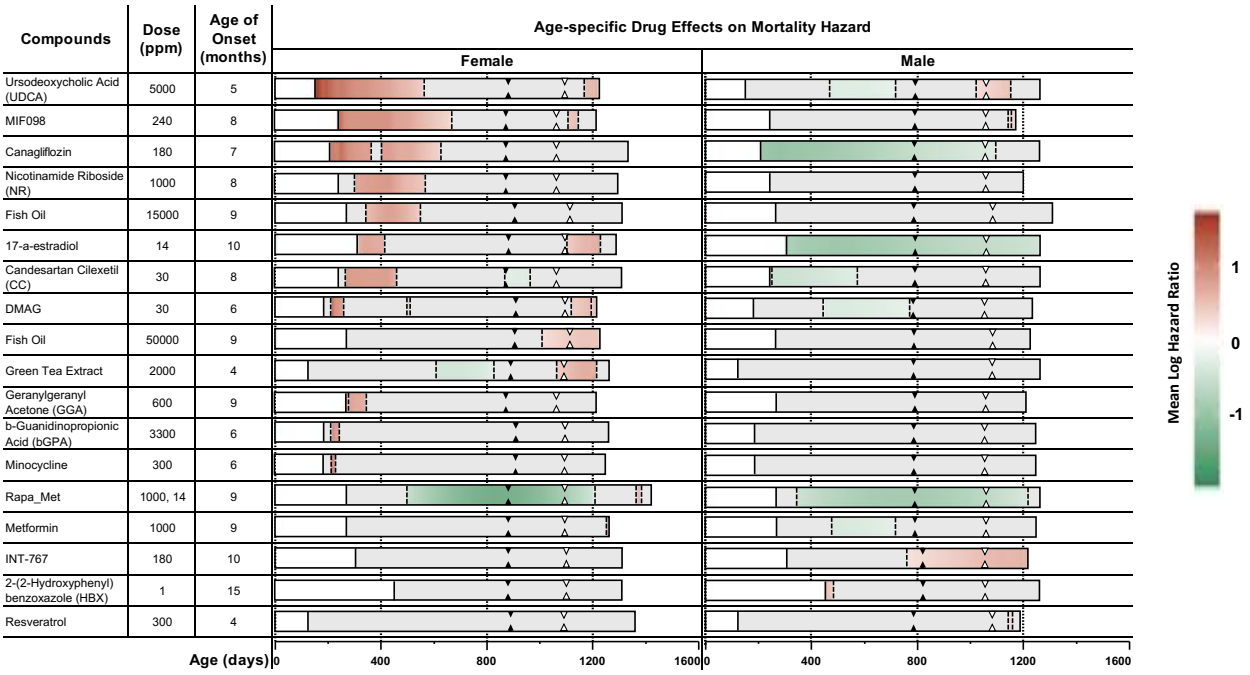

Figure 3. Trials with Drug-Induced Detrimental Effects on Mortality Hazard. These trials are shown ranked by the length of time (longest to shortest) during which significant detrimental effects were observed in females. Each color-coded band across the timeline represents the drug effect phase in relation to the treatment period: white for the pre-treatment phase, gray for phases without significant impact, green for phases with significant beneficial effects, and red for phases with significant adverse effects. Key lifespan indicators are marked by triangles; a solid black triangle denotes the median lifespan in the control group, and an open triangle indicates the point at which 10% of the control group remains alive.

Compounds	Dose (ppm)	Age of Onset (months)
Green Tea Extract[#]	**2000**	**4**
Oxaloacetic acid (OAA)[#]	**2200**	**4**
Caffeic Acid Phenethyl Ester (CAPE)[#]	**300**	**4**
Glycine	80000	9
Candesartan Cilexetil (CC)[#]	**30**	**8**
Rapa_start_stop	42	20 (until 23)
PB125[#]	**See Pub[5]**	**5**
Acarbose	400	8
Acarbose	1000	8
Acarbose	2500	8
(R/S)-1,3-butanediol (BD)	100000	6
Rapa_cycle	42	20 (on off)
Acarbose	1000	4
Rapamycin	42	20
Rapa_Met	1000, 14	9
Rapamycin	14	20
Rapamycin	14	9
Rapa_Aca	14.7, 1000	9
Rapamycin	42	9
Rapamycin	14	9
Rapamycin	4.7	9
Rapa_Aca	14.7, 1000	16
Captopril	180	5
L-leucine[#]	**40000**	**5**
Minocycline[#]	**300**	**6**
b-Guanidinopropionic Acid (bGPA)[#]	**3300**	**6**
Geranylgeranyl Acetone (GGA)[#]	**600**	**9**
Fish Oil[#]	**15000**	**9**
Nicotinamide Riboside (NR)[#]	**1000**	**8**
Canagliflozin	180	7
MIF098[#]	**240**	**8**
Ursodeoxycholic Acid (UDCA)[#]	**5000**	**5**
DMAG	30	6
17-a-estradiol	14	10
Fish Oil[#]	**50000**	**9**
Metformin[#]	**1000**	**9**
Syringaresinol[#]	**300**	**5**
Enalapril[#]	**120**	**4**
Nordihydroguaiaretic acid (NDGA)	800	6
Nordihydroguaiaretic acid (NDGA)	2500	6
Nordihydroguaiaretic acid (NDGA)	5000	6
17-a-estradiol	14.4	20
17-a-estradiol	4.8	10
Aspirin	21	4
Nordihydroguaiaretic acid (NDGA)	2500	9
Protandim	600	10
17-a-estradiol	14.4	16
Acarbose	1000	16
Simvastatin[#]	**120**	**10**
2-(2-Hydroxyphenyl) benzoxazole (HBX)[#]	**1**	**15**
Resveratrol[#]	**300**	**4**
INT-767[#]	**180**	**10**

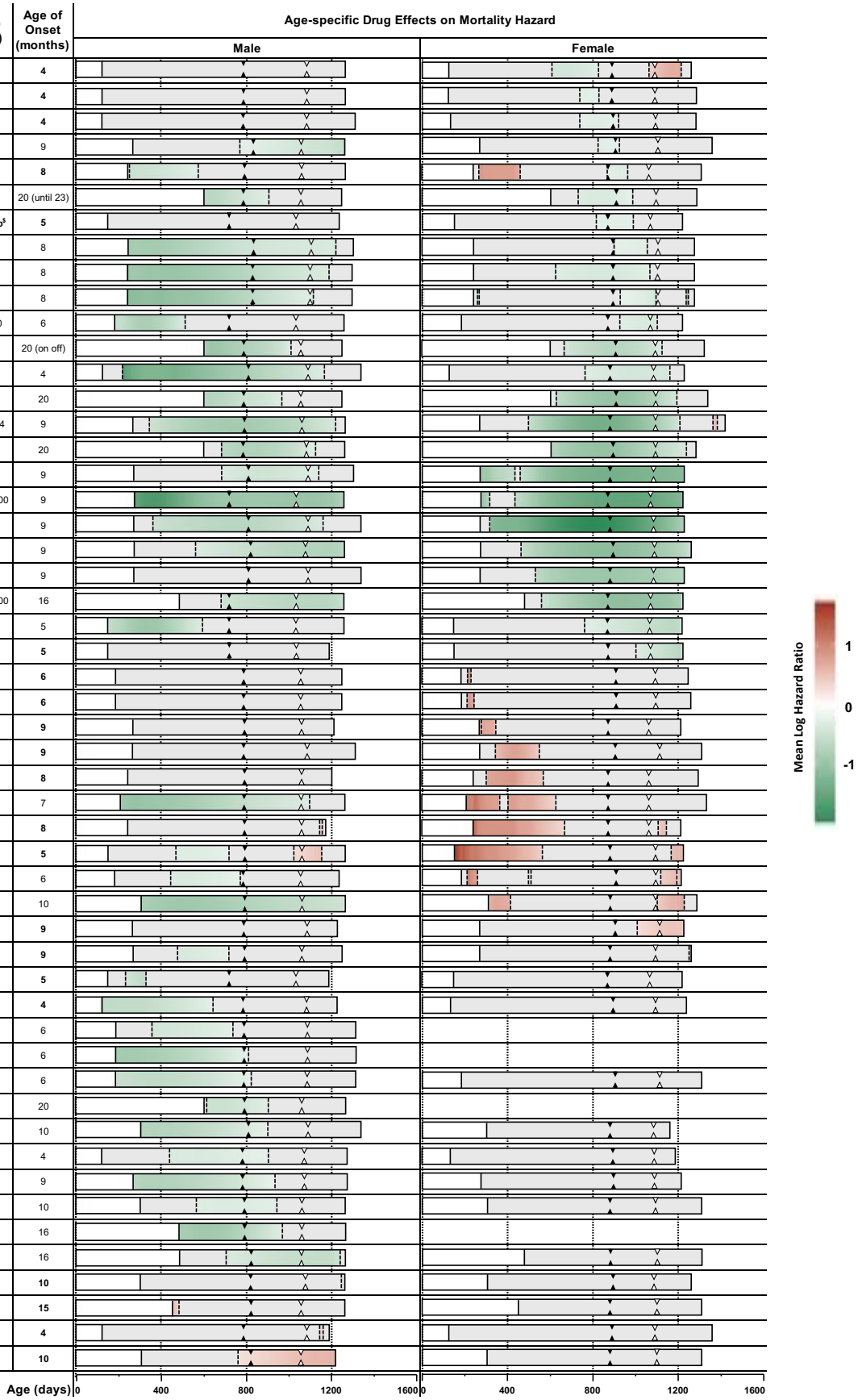

Figure S1. Interventions that significantly modified mortality hazard using the time-varying hazard analytic tool, which is ranked by the cessation age of beneficial effects in females. Each row represents an individual trial of one intervention in a single cohort. Each intervention involved one compound or a cocktail of two, with dosage and starting age of treatment listed. The color-coded bands denote the temporal significance of drug effects: white indicates the period before treatment onset, gray marks periods with no significant effects, green indicates periods of significant beneficial effects, and red denotes intervals of significant detrimental effects. The solid black triangle indicates the median lifespan of the control group for each trial, and the open triangle marks the age of 90% mortality of the control group.

Footnotes:
New Compounds that significantly affect mortality (i.e., not identified by the log-rank test) are also noted in **bold** font.
$ PB125 is a mixture of luteolin, withaferin A, and carnosol, dosages refer to publication[23].