Average and upper-tail lifespan gains are not interchangeable: a publication-balanced DrugAge analysis

Paired endpoint differences across 990 curated experimental records

Unreviewed AI-assisted computational research | Version 1.0 | 5 September 2026

Abstract

Background: a gain in average or median lifespan need not imply the same proportional gain near the upper end of survival. Hypothesis: among DrugAge records with both endpoints, maximum/upper-tail changes are smaller than average/median changes. Methods: a prospective local plan specified paired differences, equal weighting of publication-source groups, 10,000 publication bootstrap samples and sensitivity checks. Results: after removing seven exact duplicates, 990 paired records from 248 source identifiers and 298 compounds were eligible. The publication-balanced difference was -2.77 percentage points (95% bootstrap CI -4.02 to -1.58; two-paper Holm-adjusted $p = 9.68 \times 10^{-6}$). Median publication difference was -1.95 points; 149 groups had negative differences and 94 positive differences. Excluding a mismatched source identifier yielded -2.77 points. Maximum values were absent in 70.55% of deduplicated records. Conclusion: in this selected database subset, central and upper-tail lifespan changes are not interchangeable. The estimate is a descriptive synthesis of curated experiments, not a pooled causal drug effect or evidence of human life extension.

Background and hypothesis

DrugAge assembles experimental lifespan effects of compounds across model organisms [1,2]. Its original publication already reported a correlation between average/median and maximum lifespan changes [1]. Correlation does not imply agreement: two endpoints can be correlated while differing systematically. We therefore examine their paired difference under publication-balanced weighting and document missingness and source-data limitations. This is an updated robustness analysis of an existing question, not a new drug discovery or a comprehensive systematic review.

We hypothesized a negative within-record difference, defined as maximum/upper-tail percent change minus average/median percent change. The comparison is in percentage points; a difference of -3 points does not mean that lifespan itself fell by 3%. A later methodological appraisal of DrugAge highlighted substantial reporting and selection concerns [3]. Our estimand explicitly concerns records selected into DrugAge and further selected by availability of both endpoints.

Methods

Provenance. We downloaded the official DrugAge ZIP on 5 September 2026 at 08:07:34 UTC. Its SHA-256 is d12f59717de60a207748f53bdbcdb484ed07e30d8608eb8e4a101e2a08d7fa80. The export contains 3,423 records, matching the enclosed build 5 notes dated 29 November 2024 [4]. This is a frozen release, not an assertion of literature coverage through September 2026. HAGR data are reused with attribution under CC BY 3.0 [5]. The manifest preserves source URLs, dates, file sizes and checksums.

Eligibility and cleaning. We removed seven exact duplicate rows before adding source-row identifiers. Species names were trimmed and lowercased to reconcile capitalization variants; raw labels were retained. Eligible paired rows had finite numeric average/median and maximum values and a nonmissing source identifier. Blanks were missing rather than zero. No paired change was ≤ -100%, so all could also

enter a log-response-ratio sensitivity. No eligibility requirement used statistical significance. Dose, strain, sex and treatment-start variants were retained as distinct rows; they were not assumed statistically independent.

Endpoint definitions. DrugAge uses median instead of mean when available, so the central endpoint mixes these summaries [6]. Its maximum field can refer to upper-tail survival, including the 90th percentile in mouse studies [7,8]. We use maximum/upper-tail to acknowledge this heterogeneity. The data do not provide harmonized survival curves or sampling variances. The contrast compares percent changes relative to each endpoint's own control value; it does not estimate years gained, a hazard ratio, or a change in the aging rate.

Primary estimand and uncertainty. First we calculated a paired difference for each eligible record. We then averaged differences within each source key in the pubmed_id field and averaged these source means, giving each group equal weight. We use publication-source group rather than verified publication because one audited key was mismatched. The primary confidence interval resamples the 248 groups with replacement 10,000 times, using seed 20260905 and percentile endpoints. The two-sided primary p value is a one-sample t test on the group means. Holm correction spans the primary tests of both papers [9]. The bootstrap treats source groups as exchangeable and cannot remove dependence between publications from the same laboratory or underlying cohort.

Sensitivity and subgroup analyses. Planned checks used row-weighted and compound-balanced estimates, the median source difference, leave-one-source-out estimates, and a mean log-response-ratio contrast. Worm, fly and mouse subgroup tests each used publication-source means; Holm correction was applied across these three two-sided tests. These subgroup tests assess a difference from zero within species and do not test differences between species. Compound-level confidence intervals are descriptive because compounds share publications. A post-hoc ITP-labelled mouse analysis probes a more standardized study program; it is not an independent dataset, and ITP versus non-ITP was not randomized.

Source and numerical audits. Independent Python parsing reproduced row counts, the primary mean and its standard error; an independent NumPy random stream gave a similar bootstrap interval (-3.99 to -1.58 points). R t probabilities were checked by numerical integration. NCBI metadata were retrieved for all 248 keys and titles inspected for obvious mismatch. The first, middle and last keys numerically were selected for a closer source check: 101039, 24245565 and 38753230. Key 101039 links to an unrelated radiology article, despite DrugAge rows describing 2-SeCD. A relevant 2-SeCD paper was found at DOI 10.1039/C7RA07210D, but its numeric table was not fully verified. The original key was retained in the prespecified analysis and excluded in a post-hoc sensitivity. The bibliography flags the mismatch rather than presenting the radiology article as a lifespan source.

For source 24245565, the acarbose central changes (male 22%, female 5%) and upper-tail changes (11%, 9%) agree with the primary abstract [7]. For source 38753230, the checked male 16-hydroxyestriol and late-start canagliflozin central changes agree with the primary paper [8]. Its treatment name differs from the export's 16-alpha-hydroxyestradiol label, and source-table values are rounded in DrugAge. These checks are sampled; they do not certify every record, chemical label or experiment. The locally committed plan predates download, but there was no independent registry registration or external data re-extraction.

Results

Of 3,416 deduplicated rows, 990 (28.98%) had both endpoints. These represented 248 source identifiers, 298 normalized compound names and 20 species. Maximum values were missing from 2,410 rows (70.55%) and average/median values from 51 rows (1.49%); no source identifiers were missing. Complete-pair availability was 293/1,579 in C. elegans, 217/981 in D. melanogaster and 298/373 in M. musculus. The source-weighted mean central change was 13.43%, compared with 10.66% for the maximum/upper-tail endpoint. These are properties of the included records and not expected benefits of taking a drug.

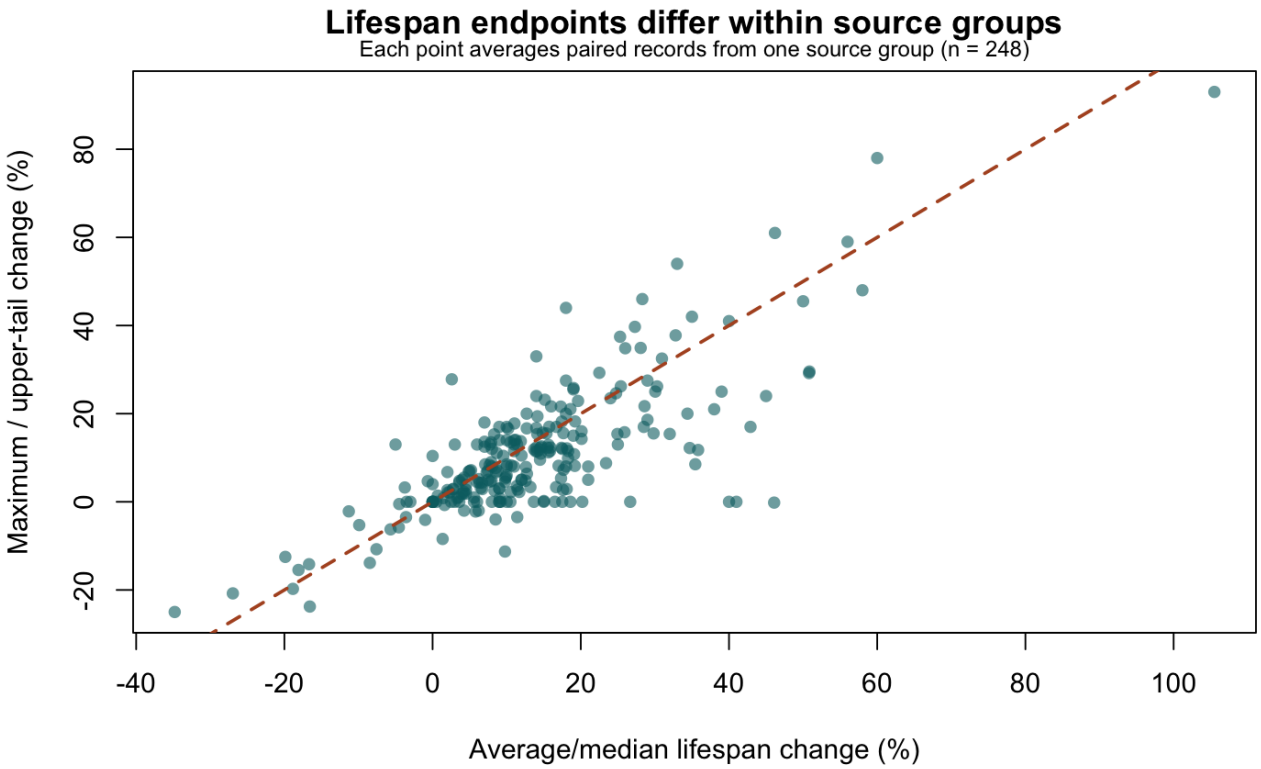

Figure 1. Each point is one source group's mean of paired DrugAge records. The dashed identity line represents equal percent changes at both endpoints. The point cloud describes curated published experiments, not the universe of tested compounds.

The mean paired contrast was -2.7698 percentage points (SE 0.6131; t-based 95% CI -3.9773 to -1.5623; bootstrap 95% CI -4.0247 to -1.5805). The raw and two-paper Holm-adjusted p values were both 9.68×10^{-6}. Of 248 source means, 149 were negative, five were exactly zero and 94 were positive. The median was -1.95 points. Thus the average contrast does not describe every study or compound.

Table 1. Species-specific publication-source contrasts. Holm correction covers the three subgroup t tests; bootstrap confidence intervals are pointwise and unadjusted.

Subset	Rows / sources	Difference, pp (95% bootstrap CI)	Holm p
C. elegans	293 / 95	-2.50 (-4.83, -0.30)	0.0319
D. melanogaster	217 / 51	-3.70 (-5.99, -1.38)	0.0088
M. musculus	298 / 66	-2.02 (-3.61, -0.46)	0.0313

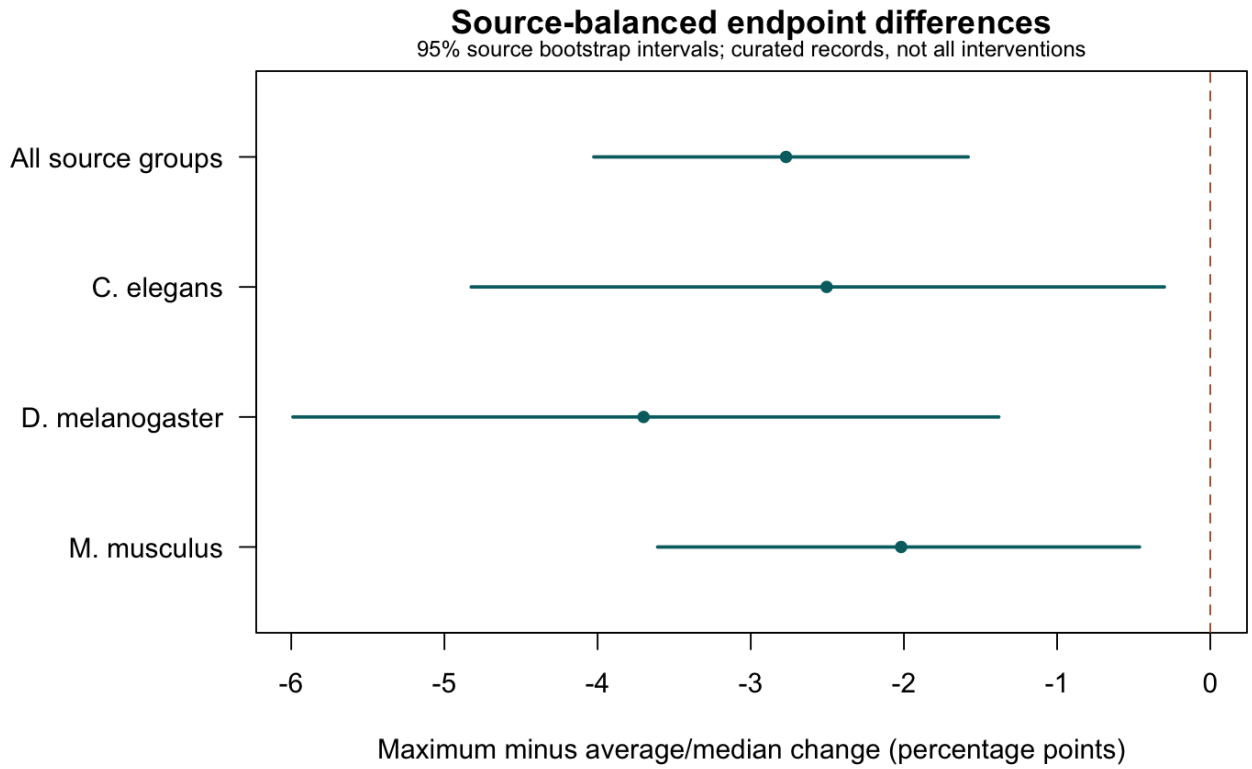

Figure 2. The overall contrast and three prespecified species contrasts. Intervals resample publication-source groups. Negative differences indicate smaller changes in the maximum/upper-tail field, not shorter treated lifespans than controls.

The row-weighted contrast was -2.44 points and the compound-balanced contrast -1.53 points (descriptive bootstrap interval -2.42 to -0.69). Leave-one-source-out estimates ranged from -2.886 to -2.594 points. The publication-balanced log-response-ratio contrast was -0.02378 (95% bootstrap interval -0.03441 to -0.01381). These estimates use different weighting or scales; their agreement in direction is a robustness observation, not multiple independent confirmations.

Excluding key 101039 retained 247 source groups and yielded -2.7681 points (95% bootstrap CI -3.9908 to -1.5848). The post-hoc mouse subset labelled ITP contained 15 sources and gave -1.82 points (-3.06 to -0.82); the other 51 mouse sources gave -2.08 points (-4.08 to -0.07). For the latter, the t-based interval crosses zero (-4.16 to 0.01); this method sensitivity should discourage binary claims about subgroup significance. These subsets share the same curated source and are not external validation.

Discussion and limitations

The study addresses agreement rather than correlation. A negative central-versus-upper-tail contrast is compatible with earlier DrugAge findings that the endpoints are correlated [1]. Publication balancing prevents a paper with many dose, sex or strain records from dominating the overall mean, but it gives a small experiment and a large experiment equal weight. Without sampling variances, this is not an inverse-variance meta-analysis and should not be represented as one.

The main limitation is selection. DrugAge historically favored compounds with at least one significant lifespan extension and later incorporated conflicting results [1]. Within that selected resource, fewer than one third of deduplicated records supply both endpoints. Neither the bootstrap nor a small p value corrects publication bias, database selection, endpoint missingness, correlated studies, or extraction errors. The missingness table shows different coverage across species and describes central effects in paired and unpaired records; it does not establish that missingness is random.

The meaning of maximum varies across source papers, and central summaries mix mean and median. Dose, treatment initiation, sex, strain, shared controls and measurement conventions vary. Within-record

pairing holds those conditions fixed for the endpoint contrast, but comparisons across records still reflect their mix. Rounded percentages may blur small differences. Compound names are normalized only for capitalization and whitespace, not chemically reconciled. The source-ID and chemical-label audit findings demonstrate that manual verification remains necessary. The analysis uses no independent external cohort and makes no causal inference about altering aging.

The defensible implication is methodological: report central and upper-tail survival outcomes separately and retain their definitions. A stronger study would re-extract survival curves, exact endpoint definitions and uncertainties from primary experiments under a registered systematic-review protocol. This analysis neither ranks treatments for human use nor establishes safety, efficacy, or human lifespan extension.

Data, code and research transparency

The reproduction archive includes the frozen official export, hashes and CC BY 3.0 attribution; the prospective plan and all deviations; screened and paired rows; source means; subgroup, missingness and leave-one-out tables; plots; metadata bibliography with the mismatched key flagged; base-R analysis scripts; and the independent NumPy verifier. Reproduce offline from the archive root with Rscript scripts/robustness.R, then python3 scripts/verify.py using NumPy 2.3.5. Reference fetching is optional and separate from numerical reproduction. No participant-level private data or new animal experiments were used. No ethics approval is asserted.

Prepared using OpenAI Codex for Australis Biogenetics. AI assistance covered question selection, literature retrieval, programming, numerical checks, drafting and formatting. No human scientific author, institutional affiliation, ethics approval, external peer review or expert endorsement is asserted. Numerical cross-checks were performed by a second implementation, not by an independent human reviewer. Publication on a company website is not scientific peer review. The website commissioned the work; no funding award, outside sponsor, or human conflict-of-interest declaration has been supplied. Dataset curators and cited investigators did not endorse this synthesis.

References

[1] Barardo D, Thornton D, Thoppil H, Walsh M, Sharifi S, Ferreira S, Anzic A, Fernandes M, Monteiro P, Grum T, Cordeiro R, De-Souza EA, Budovsky A, Araujo N, Gruber J, Petrascheck M, Fraifeld VE, Zhavoronkov A, Moskalev A, de Magalhães JP. The DrugAge database of aging-related drugs. Aging Cell. 2017;16(3):594-597. https://doi.org/10.1111/acel.12585

[2] de Magalhães JP, Abidi Z, Dos Santos GA, Avelar RA, Barardo D, Chatsirisupachai K, Clark P, De-Souza EA, Johnson EJ, Lopes I, Novoa G, Senez L, Talay A, Thornton D, To PKP. Human Ageing Genomic Resources: updates on key databases in ageing research. Nucleic Acids Research. 2024;52(D1):D900-D908. https://doi.org/10.1093/nar/gkad927

[3] Parish A, Ioannidis JPA, Zhang K, Barardo D, Swindell WR, de Magalhães JP. Reporting quality, effect sizes, and biases for aging interventions: a methodological appraisal of the DrugAge database. npj Aging. 2025;11:96. https://doi.org/10.1038/s41514-025-00287-0

[4] Human Ageing Genomic Resources. DrugAge build 5 release notes. 29 November 2024; accessed 5 September 2026. https://genomics.senescence.info/drugs/release.html

[5] Human Ageing Genomic Resources. Disclaimers, credits, and copyright: CC BY 3.0. Accessed 5 September 2026. https://genomics.senescence.info/legal.html

[6] Human Ageing Genomic Resources. Using HAGR: DrugAge documentation. Accessed 5 September 2026. https://genomics.senescence.info/help.html

[7] Harrison DE, Strong R, Allison DB, et al. Acarbose, 17-α-estradiol, and nordihydroguaiaretic acid extend mouse lifespan preferentially in males. Aging Cell. 2014;13(2):273-282. https://doi.org/10.1111/acel.12170

[8] Miller RA, Harrison DE, Cortopassi GA, et al. Lifespan effects in male UM-HET3 mice treated with sodium thiosulfate, 16-hydroxyestriol, and late-start canagliflozin. GeroScience. 2024;46:4657-4670. https://doi.org/10.1007/s11357-024-01176-2

[9] Holm S. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics. 1979;6(2):65-70. https://www.jstor.org/stable/4615733