

# Can Rodent Longevity Studies be Both Short and Powerful?

Henry T. Robertson,<sup>1</sup> Daniel L. Smith,<sup>2,3</sup> Nicholas M. Pajewski,<sup>1</sup> Richard H. Weindruch,<sup>4</sup> Theodore Garland Jr.,<sup>5</sup> George Argyropoulos,<sup>6</sup> Alex Bokov,<sup>7</sup> and David B. Allison<sup>1,2,3</sup>

<sup>1</sup>Section on Statistical Genetics, Department of Biostatistics,

<sup>2</sup>Nutrition Obesity Research Center and <sup>3</sup>Department of Nutrition Sciences, University of Alabama at Birmingham.

<sup>4</sup>Department of Medicine, University of Wisconsin at Madison.

<sup>5</sup>Department of Biology, University of California, Riverside.

<sup>6</sup>Weis Center for Research, Geisinger Medical Center, Danville, Pennsylvania.

<sup>7</sup>Epidemiology and Biostatistics Department, University of Texas Health Science Center at San Antonio.

Address correspondence to David B. Allison, PhD, Professor of Biostatistics, University of Alabama at Birmingham, 1665 University Boulevard, RPHB 420C, Birmingham, AL 35294-0022. Email: dallison@uab.edu

Many rodent experiments have assessed effects of diets, drugs, genes, and other factors on life span. A challenge with such experiments is their long duration, typically over 3.5 years given rodent life spans, thus requiring significant time costs until answers are obtained. We collected longevity data from 15 rodent studies and artificially truncated them at 2 years to assess the extent to which one will obtain the same answer regarding mortality effects. When truncated, the point estimates were not significantly different in any study, implying that in most cases, truncated studies yield similar estimates. The median ratio of variances of coefficients for truncated to full-length studies was 3.4, implying that truncated studies with roughly 3.4 times as many rodents will often have equivalent or greater power. Cost calculations suggest that shorter studies will be more expensive but perhaps not so much to not be worth the reduced time.

**Key Words:** Longevity—Rodent studies—Proportional hazards—Survival analysis—Sample size.

Received May 10, 2010; Accepted September 23, 2010

Decision Editor: Rafael de Cabo, PhD

RODENT longevity studies remain a staple of experimental aging research, having been used to evaluate the effects of diets, drugs, genetic factors, toxins, or other factors on life span. One of the greatest obstacles with such experiments is their long duration, typically requiring 3.5 years or more to observe the complete life span of all study rodents (1). In addition to the significant financial burden for such studies, the investment in terms of time until answers can be obtained and can represent a significant fraction of an investigator's research career, limiting the number of experiments an investigator can perform in their lifetime.

Interventional effects on longevity are commonly estimated through Cox proportional hazards (PHs) regression models (2) or such parametric formulations as the Gompertz model (3). Under the Cox PH model, the relative hazard rate (the instantaneous probability of death occurring at a given moment, conditional on death not happening prior to that moment) between two factors is assumed to have a fixed ratio over time. It is important to note that the Cox PH model does not restrict the hazard for a particular group to be constant as a function of time, instead entailing the inherent flexibility to be congruent with the biological expectation of increasing hazard during the course of normal aging.

Similarly, the Gompertz model assumes that the relative acceleration in mortality is constant as function of age. Because these models assume constant hazard ratios (HRs)/relative increases in mortality rate over time, their use implicitly assumes that any relative effects on mean or median life span are also present for so-called "maximum life span," a quantity of great interest to gerontologists (4).

Recognizing the significant time investment required for full longevity studies along with the application of models that assume constant effects across time invites the question as to whether the rodent mortality experience in these settings is consistent with the assumption of PH. Put another way, is it possible to reduce the length of longevity experiments by running studies for only  $\tau$  years, where  $\tau$  is less than full follow-up and still achieve comparable effect estimates to what one would obtain if one continued the study until all rodents died? Truncating longevity experiments prior to observing death for all rodents reduces statistical power due to the inherent loss of information. Thus, a related question is whether this loss of power can be overcome by increasing sample size, permitting shorter studies in terms of calendar time that still achieve the same expected inference concerning effects on mortality rate.

The goal of the present study is to empirically evaluate the assumption of PH using a collection of 15 rodent longevity experiments. Motivated by the design of long-term rodent toxicology (typically carcinogenicity) experiments, we also consider the situation of truncating longevity studies at 2 years of follow-up. We evaluate whether these shorter term studies could be expected to yield, on average, the same result with respect to effects on mortality rate that would be obtained from a full lifetime study. For situations in which the observed effects on mortality rate would be roughly the same in a shorter versus life-long follow-up study (ie, PH is a valid assumption), we estimate the required increase in sample size needed to achieve the same statistical power and precision as would a full lifetime study. Finally, we also compare the total study costs for a truncated follow-up experiment versus a full-length longevity study.

## METHODS

### Data Sets

We utilized mortality data from a convenience sample of 15 rodent life-span studies available to us from our own work or provided by close collaborators (Table 1). These life-span studies evaluated a broad range of factors, including sex, genotype, and drug and diet, exercise, and surgical interventions, though we do not claim that they represent an exhaustive sampling of rodent longevity studies. Available studies, where the intervention changed during the course of the study, were excluded, such as the methionine study of Miller and colleagues (4). Of the 15 studies included, their average study length was 3.6 years ( $SD = 0.8$ ) with an average sample size of 363 rodents ( $SD = 50$ ).

### Statistical Methods

For each available study, we estimated the mortality effects of the various treatments, interventions, and other factors considered in the original experiment via Cox PH regression. In order to access evidence for deviations from the assumption of PH, we performed statistical tests of the PH assumption by including time-based effects in the regression model for each study (5). The inclusion of time-varying effects for each predictor variable is a standard statistical approach to assess the evidence for changing HRs changes over time, such as could occur if a given treatment leads to increased early mortality yet offers later protective effects (6). For each study, we performed an overall  $F$  test for all time-based effects in the full study data to examine the PH assumption. We utilized a jackknife procedure to conduct more focused tests for differences between the HRs obtained from analyzing the full longevity data versus truncating each study at 2 years (Appendix 1). We report uncorrected  $p$  values, so that readers may choose to apply the multiple testing or false discovery rate correction of

their choice (7) or simply apply a nominal  $\alpha = .05$  significance level (8).

To estimate the required increase in sample size for truncated life-span studies, we assumed that, asymptotically, the variance of estimated effects is proportional to  $1/N$ . That is, if we double the sample size, then the standard errors of estimated effects will shrink by a factor of  $1/\sqrt{2}$ . If the data from the full study yielded an estimate with variance  $VF$  and the truncated data yielded an estimate  $VT$ , then a truncated study will need a larger sample size by a factor of  $\sim VF/VT$  to achieve equivalent power.

Based on the estimated sample size increases for truncated longevity studies, we also estimated the costs associated with performing a hypothetical larger truncated experiment for each of the 15 studies. Based on the National Center for Research Resources rate-setting manual (9) and consulting with experts on rodent life-span experiments, we assumed the following in cost calculations. Both mice and rats incur an initial cost of \$20.00 per animal to purchase, with loaded cage maintenance costs of \$1.27 and \$2.14 per day for mice (up to five per cage) and rats (up to three per cage), respectively. We also assume that a lab technician (salary of \$40,000 per year) can manage up to 1,000 rodents. Due to the fact that, in practice, surviving rodents are not reassigned to new cages, once cage mates have deceased, these costs of full survival studies are likely underestimated at some level. All analyzes were performed using SAS v9.1.3 (SAS, Cary, NC).

## RESULTS

In Table 2, we report the uncorrected results of testing for departures from PH. After considering either the Bonferroni or Holm's (10) multiple-testing corrections, none of the studies gave statistically significant evidence of deviation from a PH model (Table 2). Two studies, 5 and 8, did exhibit nominal departure from PH. For Study 5, which tested selectively bred high-runner mice at different exercise levels against control mice, the control mice initially died off at a slower rate than the other two treatment groups but then died off rapidly toward the end (11). For Study 8, the mice that were administered thyroxine started the study with similar survival rates but then experienced accelerating death rates later in the study (12).

Table 3 displays the results of truncating each of the studies at 2 years, censoring each rodent at 730 days for those with deaths that occurred after that point. The jackknife tests for differences in the estimated HRs between the truncated and full-length studies suggest that in 43 of 49 comparisons (88%), the value of the estimated coefficients did not differ significantly between the truncated and full-length studies at the  $\alpha = .05$  level; at the Bonferroni corrected  $\alpha$  level of  $.05/49$ , none of the differences were significant. Of the six treatment groups whose coefficients differed at the nominal  $.05$  level, 4 of 6 were for dietary interventions.

Table 1. List of Rodent Longevity Experiments Included in Current Study\*

Study	Authors	Title	Primary Intervention	Number of Rodents	Species	Strain	Sex
1.	Minor RK, Smith DL Jr, Sosong AM, Kaushik S, Poosala S, et al.	Chronic ingestion of 2-deoxy-d-glucose induces cardiac vacuolization and increases mortality in rats. <i>Toxicol Appl Pharmacol.</i> 2010;243(3):332–9.	Drug and Diet: 30% CR, Metformin, 2-deoxyglucose	360	Rat	Fischer-344	Male
2.	Vasselli JR, Weindnuch R, Heymsfield SB, Allison DB, et al.	Intentional weight loss reduces mortality rate in a rodent model of dietary obesity. <i>Obesity Research.</i> 2005;13(4):693–702.	Diet: after establishment of diet-induced obesity	226	Rat	Outbred Sprague–Dawley IGS	Male/female
3.	Redmann SM, Argyropoulos G.	AgRP-deficiency could lead to increased lifespan. <i>Biochemical and Biophysical Research Communications.</i> 2006;351(4):860–864	Genotype/sex	37	Mouse	C57BL/6J wild type	Male/female
4.	Lee CK, Pugh TD, Klopp RG, Edwards J, Allison DB et al.	The impact of $\alpha$ -lipoic acid, coenzyme Q <sub>10</sub> and caloric restriction on life span and gene expression patterns in mice. <i>Free Radical Biology and Medicine.</i> 2004;36(8):1043–1057.	Diet: controls fed 15% below AL, CR 41% reduced	598	Mouse	C57BL/6 $\times$ C3H (B6C3F1)	Male
5.	Vaanholt LM, Daan S, Garland T, Visser GH	Exercising for life? Energy metabolism, body composition, and longevity in mice exercising at different intensities. <i>Physiological and Biochemical Zoology.</i> 2010;83(2):239–251.	Exercise	180	Mouse	Hsd:ICR	Male
6.	Randy S, Miller RA, Astle CM, Floyd RA, Flurkey K, et al.	Nordihydroguaiaretic acid and aspirin increase lifespan of genetically heterogeneous male mice. <i>Aging Cell.</i> 2008;7(5):641–650.	Drug: control and aspirin, nitroflurbiprofen, 4-OH-PBM, NDGA	1,843	Mouse	UM-HET3: CBxB6F1/J, JAX stock #100009 mothers, C3D2F1/J, JAX stock #100004 fathers	Male/female
7.	Miller RA, Harper JM, Galecki A, Burke DT.	Big mice die young: early life body weight predicts longevity in genetically heterogeneous mice. <i>Aging Cell.</i> 2002;1(1):22–29.	Weight	547	Mouse	UM-HET3 (same as above study) Cross between CB6F1 females and C3D2F1 males	Male/female
8.	Vergara M, Smith-Wheelock M, Harper JM, Sigler R, Miller RA.	Hormone-Treated Snell Dwarf Mice Regain Fertility but remain long lived and disease resistant. <i>J Gerontol Biol Sci.</i> 2004;59A(12):1244–1250.	Genotype/hormone: dwarf mice with or without growth hormone or GH + T4 treatment	219	Mouse	(DW C3H)F1 including dw/dwJ, dw/+ , +/dwJ, and +/- from "Snell Dwarf"; F1 progeny of DW/J-Pit1 <sup>dw/+</sup> females and C3H/HeJ-Pit1 <sup>dw/+</sup> male heterozygote breeders	Male/female
9.	Baur JA, Pearson KJ, Price NL, Jamieson HA, Lerin C, et al.	Resveratrol improves health and survival of mice on a high-calorie diet. <i>Nature.</i> 2006;444(7117):337–342.	Diet/drug: standard, high calorie (coconut oil) and HC + Resveratrol	499	Mouse	C57BL/6NIA	Male
10.	Perez VI, Van Remmen H, Bokov A, Epstein CJ, Vijg J, et al.	The overexpression of major antioxidant enzymes does not extend the lifespan of mice. <i>Aging Cell.</i> 2009;8(1):73–75.	Genotype: antioxidant enzyme overexpression	153	Mouse	C57BL/6J	Male
11.	Bokov AF, Lindsey ML, Khodr C, Sabia MR, Richardson A.	Long-lived ames dwarf mice are resistant to chemical stressors. <i>J Gerontol A Biol Sci Med Sci.</i> 2009;64(8):819-827.	Genotype	102	Mouse	Ames Prop-1; C57Bl6	Male

Table 1. (Continued)

Study	Authors	Title	Primary Intervention	Number of Rodents	Species	Strain	Sex
12.	Weindruch R., Walford R., Fligiel S., Guthrie D.	The retardation of aging in mice by dietary restriction: longevity, cancer, immunity and lifetime energy intake. <i>J Nutr.</i> 1986;116:641-654	Diet: calorie and/or protein restriction	237	Mouse	C3B10RF1	Female
13.	Pugh TD, Oberley TD, Weindruch R.	Caloric restriction but not dehydroepiandrosterone sulfate increases lifespan and lifetime cancer incidence in mice. <i>Cancer Res.</i> 1999;59(7):1642-1648.	Diet—Are the supplement groups not included?	151	Mouse	C57BL/6	Male
14.	Weindruch, Richard, Sohal, Rajindar S.	Caloric Intake and Aging. <i>N Engl J Med.</i> 1997;337(14):986-994.		137	Mouse	C3B10RF1	Male
15.	Muzumdar R, Allison DB, Huffiman DM, Ma X, Atzmon G, et al.	Visceral adipose tissue modulates mammalian longevity. <i>Aging Cell.</i> 2008;7(3):438-440.	Diet/surgery	161	Rat	Sprague-Dawley	Male

Note: \*Reported sample sizes may differ from published figures due to missing entries in the data provided.

Figure 1 displays variance ratios for the estimated HRs from truncated and full-length studies, defined as  $(SE_{2yr}/SE_{full})^2$ . Based on truncating studies at 2 years, the variances of the estimates became inflated by factors ranging from 1.2 to 34.0. As expected, the proportion of rodents that died before 2 years was strongly correlated with the level of variance inflation (Figure 1). On average, the variance of the estimates increased by a factor of 5.1 ( $SD = 5.6$ , median = 3.4). The largest increases in the variance ratios were in studies where less than 20% of rodents died before 2 years due to long life or small variation in life span. Under the assumption that truncated studies with five times as many rodents allocated to each arm would have equivalent power to a full-length study with one fifth of the total sample size, Appendix 2 displays the estimated total study cost for each type of design for each of the 15 data sets considered. The costs of running shorter experiments ranged from 1.8 to 4.0 times the cost of the full-length study (Table 4). Full details of the cost computations are shown in Appendix 2.

## DISCUSSION

Our analysis indicates that, in most cases, interventions or other factors that influence rodent longevity induce effects consistent with PH models. Truncating the study length to 2 years did not significantly affect the estimated effect; only the variance of this estimate and hence its statistical significance were influenced, a result of reduced sample size and power. This implies that the Cox PH regression model (13) is sufficient for detecting differences in longevity, even when studies are cut short. We propose that increasing the number of rodents in the study by a factor derived from Figure 1, by a factor near 5 on average (mean) but in half the cases by a factor of no more than 3.4 (median), can offer power equivalent to a full-length study.

Table 2. Statistical Tests of the Proportional Hazards Assumption by Experiment

Study	N	% Censored*	Variable of Interest	p Value†
1	400	0.00	Drug	.149
2	226	0.00	Diet	.148
3	37	2.70	Genotype/sex	.969
4	598	0.30	Diet	.454
5	180	0.00	Exercise	.029
6	1,843	0.00	Drug	.379
7	547	0.00	Weight	.418
8	219	5.90	Genotype/hormone	.006
9	499	21.20	Diet	.178
10	133	0.00	Genotype	.378
11	99	0.00	Genotype	.956
12	237	0.00	Diet	.294
13	151	0.00	Diet	.660
14	137	5.10	Diet	.497
15	161	0.00	Diet/surgery	.934

Notes: N = total sample size.

\* Denotes the percentage of right-censored survival times in full-length experiment.

† p Values obtained by performing an overall F test of time-based effects.

Table 3. Comparative Results of the Full Versus Truncated Analyses From Each Data set and Their Implications for the Comparative Direct Costs of 2-Y Versus Full Life-Span Study\*

Study	Treatment Group	N	$\beta_{2y}$	$\beta_{full}$	Difference	p Value	Variance Ratio	Cost Ratio
1	A	45	-1.249	-0.614	0.6352	.055	3.1	3.9
1	B	45	-0.644	0.008	0.6523	.006	2.2	
1	C	45	-1.386	-0.308	1.0776	.002	3.4	
1	D	45	1.589	2.083	0.4937	.001	1.2	
1	E	45	-0.872	-0.444	0.4286	.123	2.5	
1	F	45	-0.742	-0.327	0.4151	.139	2.4	
1	G	45	-0.927	-0.323	0.6039	.064	2.6	
2	A	41	-0.968	-0.759	0.2086	.168	1.6	4.0
2	B	39	-0.437	-0.720	-0.2830	.027	1.2	
2	C	49	-0.405	-0.540	-0.1346	.217	1.3	
2	D	48	-0.303	-0.441	-0.1377	.192	1.3	
3	A	16	0.797	0.756	-0.0406	.693	1.2	1.8
4	A	98	-0.205	-0.039	0.1660	.506	3.4	3.7
4	B	99	-0.111	0.020	0.1302	.557	3.3	
4	C	99	0.289	0.171	-0.1179	.459	2.7	
4	D	99	-0.311	0.084	0.3954	.090	3.6	
4	E	97	0.249	0.073	-0.1764	.332	2.7	
5	A	60	-0.866	-0.160	0.7054	.008	2.9	2.9
5	B	60	-0.030	-0.047	-0.0174	.873	2.0	
6	A	298	0.003	-0.030	-0.0334	.671	3.0	4.0
6	B	308	-0.310	-0.070	0.2402	.043	3.6	
6	C	308	-0.385	-0.218	0.1678	.135	3.8	
6	D	308	-0.189	-0.002	0.1874	.096	3.4	
7	A	206	-0.063	0.286	0.3485	.114	5.9	3.6
7	B	57	0.308	0.241	-0.0667	.747	4.7	
7	C	51	0.518	0.106	-0.4127	.114	4.0	
8	A	47	-1.307	-2.148	-0.8414	.646	2.3	2.6
8	B	28	-0.778	-1.141	-0.3624	.905	2.2	
8	C	35	-1.711	-2.233	-0.5214	.398	7.0	
8	D	38	0.256	0.088	-0.1675	.129	6.5	
8	E	32	0.524	0.481	-0.0427	.450	5.9	
9	A	60	-0.187	-0.240	-0.0535	.947	4.2	3.9
9	B	55	-0.224	-0.160	0.0641	.848	4.2	
9	C	54	-0.605	-0.300	0.3046	.586	5.3	
9	D	55	-0.599	-0.510	0.0888	.878	5.3	
9	E	55	-1.369	-0.790	0.5792	.547	8.4	
9	F	55	-0.851	-0.510	0.3412	.576	5.9	
9	G	55	0.997	0.590	-0.4068	.317	3.0	
9	H	55	-0.016	-0.100	-0.0839	.878	4.2	
10	A	44	-0.512	0.027	0.5385	.507	11.6	1.8
10	B	45	0.185	-0.251	-0.4365	.585	7.9	
11	A	51	-0.041	-0.144	-0.1035	.927	24.0	2.1
12	A	57	-0.996	-1.150	-0.1535	.696	6.4	2.4
12	B	60	-2.696	-3.719	-1.0236	.317	12.6	
12	C	71	-2.168	-3.212	-1.0444	.165	7.7	
13	A	74	-0.221	-0.897	-0.6759	.137	7.1	2.3
14	A	67	-1.818	-0.773	1.0450	.447	34.0	2.1
15	A	53	-2.161	-2.055	0.1059	.916	4.4	2.8
15	B	51	-0.521	-0.713	-0.1922	.848	2.2	

Note: \*Some studies had more than two groups, which yields more than one estimated treatment effect. We have denoted each comparison within a study as A, B, etc.

In situations where time is critical and the impact of discovery is large, the additional cost burden of a shorter experiment may be justified. It may also be worth commenting on whether diet, strain, environment, cohort, etc. might have the most impact on achieving similarity of results in truncated versus full longevity studies, particularly given that each of these factors can affect both the overall life-span trends and specific mortality trajectories at earlier time-points. Of course, the truncated approach will not provide

full information on diseases of aging compared with a full longevity/disease experiment.

The current results suggest empirical adherence of mouse and rat longevity studies to PH, implying similar effects on mortality rate across the full life span. Thus, although the risk of death (hazard rate) may certainly accelerate with advanced age, it appears that differences in acceleration between groups commonly occur by a proportional factor. This further suggests that cases where early,

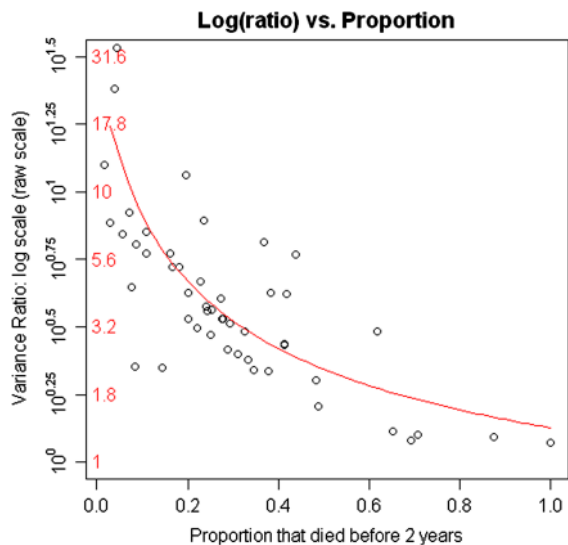


Figure 1. Ratios of variance in effect estimates in truncated versus full life-span studies as a function of proportion of animals dying in truncated study. \*The curved line represents a least squares second order polynomial fitted to the data. \*\* The numbers on the left side of the y-axis represent the log scale, whereas the numbers on the right side of the y-axis represent the absolute scale.

mean, and median life span are extended, extensions in so-called maximal life span would also be the expected norm. However, exceptions to this general rule may be possible, in principle and observation. It is commonly observed and cited that in addition to mean and median life span, maximal life span is extended in response to calorie restriction, setting it apart from interventions that may increase mean or median life span independent of maximum life span (14,15). It is theoretically possible that an intervention that extends both mean and median life span has no benefit on maximal life span or that maximal life span may be increased independent of mean or median benefits. What is less commonly reported is the case where higher early- to mid-life mortality is followed by a subsequent extension of maximal life span. Although such an instance has been reported (4) and others may exist, it should be noted that the constant effects on longevity would only be expected in response to a single intervention that was maintained for the duration of the longevity study (16). Therefore, the alteration of a single intervention, as occurred with a methionine restriction protocol that showed increased early-life mortality (using the initial dietary formulation) followed by subsequent extension (as the diet was reformulated at two interim study points), does not contradict the expected proportionality of life span (4). Whether other studies that show potential differences in either early-life or late-life mortality effects that might contradict the predicted proportionality has not been formally tested. For example, calorie-restricted wild-derived mice were not extended at early- and mid-life but were extended at the 90th percentile (17); resveratrol was reported to increase the early- and

mid-life span of high fat-fed mice, with no benefit on maximal life span (18).

Part of the question regarding maximal life-span effects may be related to the lack of proper statistical analyzes of reported maximal longevity results. Clearly, truncating studies at any point prior to the maximal observed life span prevents assessment of any late-life specific effects that might occur. Yet, the limited sample size defining the maximum life span (90th percentile—4 animals of a cohort of 40) results in low-powered comparisons between groups, and it is often unclear whether observed increases in maximum life span are statistically significantly greater rather than simply numerically increased. The application of a standardized method, such as those described in (19) and (20) for maximal life-span assessment between studies, would be useful to determine whether longevity studies that have observed early- and mid-life extension consistently occur independent of late-life (maximal) extension. Another point of interest related to the proportionality of survival/mortality is whether the cause of mortality may differ at various points along the survival curve. This is often likely to be the case, at least for comparisons of some groups (eg, sex differences, and if so, then an intervention that affects a specific cause of death that is isolated to late-life in rodents would be unobserved in a truncated study). Of course, our approach assumes that the observed mortality was due to natural causes in shorter lived mice; in practice, the investigator will need to check this assumption by carefully examining the causes of death.

The current results lend support to several practical applications. The first is the potential for statistical analyzes of existing data sets in order to hypothesize about the expected survival impact that may occur in response to a particular diet, compound, or intervention. In particular, we envision the use of toxicology studies of at least 2-year duration where greater than 20% mortality has occurred, potentially advancing aging research without additional study-related costs. Second, although we have specifically evaluated truncated 2-year life-span experiments in order to parallel existing toxicology resources, our results in no way imply that 2-year studies or designs with a fixed stopping point are optimal. Rather than specifying a deterministic study length, an alternative implication from PH would be the potential for applying group sequential or similar clinical trial designs that permit interim analyzes of a life-span experiment (21). For example, one could envision a life-span study employing planned interim analyzes at 3 to 6-month intervals beginning at 2 years using such approaches as conditional power methodology (22) or continual monitoring using Bayesian decision-theoretic techniques to evaluate the need to continue the experiment (23). This methodology may be particularly beneficial, given that recent publications indicate that exploration of factors affecting life span in rodents remains a staple of experimental

aging research (24–28). Although similar analyzes have been applied to toxicology experiments in clinical drug testing, our analyzes illustrate the potential to apply similar methodologies to a multifactorial endpoint of longevity. However, a meticulous adherence to the planned study design and proper analytic techniques is certainly advised to prevent improper interpretations, such as could occur by simply ending a study when a statistically significant result is obtained.

#### FUNDING

This work was supported in part by National Institutes of Health grants: P30DK056336, R01DK076771, T32HL072757, T32DK062710, and T32HL079888. T.G. was supported by NSF grant IOB-0543429.

#### DISCLOSURE

D.B.A. has received grants, honoraria, donations, book royalties, and consulting fees from numerous food, beverage, dietary supplement, pharmaceutical companies, litigators and other commercial, government and nonprofit entities with interests in obesity, and related topics.

#### ACKNOWLEDGMENTS

We would like to thank the following individuals for their generosity in sharing rodent data: George Roth, Chief Executive Officer, GeroScience Inc.; Richard Miller, Professor of Pathology, University of Michigan at Ann Arbor; Donald K. Ingram, Professor of Nutritional Neuroscience and Aging Laboratory, Pennington Biomedical Research Center; Julie Mattison, Laboratory of Experimental Gerontology, National Institute on Aging, National Institutes of Health (NIH); Arlan Richardson, Senior Research Career Scientist (STVHCS) Director, University of Texas Health Science Center; and Kyle Grimes, Professor of English at the University of Alabama at Birmingham. We would also like to thank Junior Bazile, Nigel Rozario, and Huichien Kuo for assistance with data compilation and standardization. The opinions expressed are those of the authors and not necessarily those of the NIH or any other organization with which the authors are affiliated.

#### REFERENCES

- Swindell WR, Harper JM, Miller RA. How long will my mouse live? Machine learning approaches for prediction of mouse life span. *J Gerontol A Biol Sci Med Sci*. 2008;63:895–906.
- Conover CA, Bale LK, Mader JR, Mason MA, Keenan KP, Marler RJ. Longevity and age-related pathology of mice deficient in pregnancy-associated plasma protein-A. *J Gerontol A Biol Sci Med Sci*. 2010;65(6):590–599.
- Mason JB, Cargill SL, Anderson GB, Carey JR. Transplantation of young ovaries to old mice increased life span in transplant recipients. *J Gerontol A Biol Sci Med Sci*. 2009;64(12):1207–1211.
- Miller RA, Buehner G, Chang Y, Harper JM, Sigler R, Smith-Wheelock M. Methionine-deficient diet extends mouse lifespan, slows immune and lens aging, alters glucose, T4, IGF-I and insulin levels, and increases hepatocyte MIF levels and stress resistance. *Aging Cell*. 2005;4:119–125.
- Klein J, Moeschberger M. *Survival Analysis: Techniques for Censored and Truncated Data*, 2nd ed.. New York: Springer-Verlag; 2003.
- Bellera CA, MacGrogan G, Debled M, de Lara CT, Brouste V, Mathoulin-Pelissier S. Variables with time-varying effects and the Cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer. *BMC Med Res Methodol*. 2010;10:20.
- Gadbury GL, Xiang Q, Yang L, Barnes S, Page GP, Allison DB. Evaluating statistical methods using plasmode data sets in the age of massive public databases: an illustration using false discovery rates. *PLoS Genet*. 2008;4:e1000098.
- Bailar JC III, Mosteller F. Guidelines for statistical reporting in articles for medical journals. Amplifications and explanations. *Ann Intern Med*. 1988;108:266–273.
- McPherson P. Cost Analysis and Rate Setting Manual for Animal Research Facilities, Bethesda, MD: NCRP Office of Science Policy and Public Liaison [serial online] 2010.
- Holm S. A simple sequentially rejective multiple test procedure. *Scand J Stat*. 1979;6(2):65–70.
- Vaanholt LM, Daan S, Garland T, Visser GH. Exercising for life? Energy metabolism, body composition, and longevity in mice exercising at different intensities. *Physiol Biochem Zool*. 2010;83:239–251.
- Vergara M, Smith-Wheelock M, Harper JM, Sigler R, Miller RA. Hormone-treated snell dwarf mice regain fertility but remain long lived and disease resistant. *J Gerontol A Biol Sci Med Sci*. 2004;59:1244–1250.
- Cox D. Regression models and life-tables. *J R Stat Soc Ser B*. 1972;34:187–220.
- Holloszy JO, Schechtman KB. Interaction between exercise and food restriction—effects on longevity of male-rats. *J Appl Physiol*. 1991;70:1529–1535.
- Weindruch R, Walford RL. *The Retardation of Aging and Disease by Dietary Restriction*; Springfield, IL: C.C. Thomas Publisher; 1988.
- Rausser CL, Laurence DM, Travisano M, Rose MR. *Evolution of Aging and Late Life. Experimental Evolution Concepts, Methods, and Applications of Selection Experiments*; Berkeley, CA: University of California press; 2009:551–584.
- Harper JM, Leathers CW, Austad SN. Does caloric restriction extend life in wild mice? *Aging Cell*. 2006;5:441–449.
- Baur JA, Pearson KJ, Price NL, et al. Resveratrol improves health and survival of mice on a high-calorie diet. *Nature*. 2006;444:337–342.
- Wang CX, Li Q, Redden DT, Weindruch R, Allison DB. Statistical methods for testing effects on “maximum lifespan”. *Mech Ageing Dev*. 2004;125:629–632.
- Gao GM, Wan W, Zhang SJ, Redden DT, Allison DB. Testing for differences in distribution tails to test for differences in ‘maximum’ lifespan. *BMC Med Res Methodol*. 2008;8:49.
- Piantadosi S. *Clinical Trials: A Methodologic Perspective*; New York: Wiley-Interscience; 2005.
- Andersen PK. Conditional power calculations as an aid in the decision whether to continue a clinical trial. *Control Clin Trials*. 1987;8:67–74.
- Lewis RJ, Lipsky AM, Berry DA. Bayesian decision-theoretic group sequential clinical trial design based on a quadratic loss function: a frequentist evaluation. *Clin Trials*. 2007;4:5–14.
- Ricklefs RE, Scheuerlein A. Biological implications of the Weibull and Gompertz models of aging. *J Gerontol A Biol Sci*. 2002;57: B69–B76.
- Singh SP, Niemczyk M, Saini D, Sadovov V, Zimniak L, Zimniak P. Disruption of the mGsta4 gene increases life span of C57BL mice. *J Gerontol A Biol Sci Med Sci*. 2010;65(1):14–23.
- Smith DL Jr., Elam CF Jr., Mattison JA, et al. Metformin supplementation and life span in Fischer-344 rats. *J Gerontol A Biol Sci Med Sci*. 2010;65(5):468–474.
- Sun L, Sadighi Akha AA, Miller RA, Harper JM. Life-span extension in mice by preweaning food restriction and by methionine restriction in middle age. *J Gerontol A Biol Sci Med Sci*. 2009;64(7):711–722.
- Zhang Y, Ikeno Y, Qi W, et al. Mice deficient in both Mn superoxide dismutase and glutathione peroxidase-1 have increased oxidative damage and a greater incidence of pathology but no reduction in longevity. *J Gerontol A Biol Sci Med Sci*. 2009;64(12):1212–1220.
- Efron B. *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia, PA: Society for industrial and applied mathematics; CBMS-NSF Regional conference series in applied mathematics, 1982.

**APPENDIX 1. PSEUDOCODE FOR JACKKNIFE TEST OF DIFFERENCES IN FULL-LENGTH AND 2-YEAR TRUNCATED HRs.**

Let  $\theta = \log(\text{HR}_{\text{full}}) - \log(\text{HR}_{2\text{yr}})$ , with corresponding estimator  $\hat{\theta}$  based on the full data set ( $i = 1, 2, \dots, N$  observations). The jackknife procedure then consists of the following steps (29).

For  $i = 1$  to  $N$

Remove the  $i$ th subject from the data set.

Estimate  $\theta$  in the reduced sample, denote this estimate as  $\hat{\theta}_i$ .

End

The jackknife estimate of  $\theta$  is then

$$\tilde{\theta}' = N\hat{\theta} - (N-1)\bar{\theta}^{N-1},$$

where  $\bar{\theta}^{N-1} = \sum_{i=1}^N \hat{\theta}_i / N$ . The jackknife estimate of the standard error is

$$SD' = \sqrt{\frac{N-1}{N} \sum_{i=1}^N [\hat{\theta}_i - \bar{\theta}^{N-1}]^2}.$$

Thus, a test statistic under the null hypothesis  $\hat{\theta} = 0$  can be constructed as

$$T = \frac{\tilde{\theta}'}{SD'},$$

where  $T$  follows an approximate  $t$  distribution with  $N - 1$  degrees of freedom.

**APPENDIX 2. COST ESTIMATES BY STUDY.**

Study	Length	$N$	Mean Longevity	Maximum Longevity	Total Rodent Days	# Technicians Required	Technician Cost	Rodent Overhead Cost	Rodent Maintenance Cost	Total Cost	Cost Ratio
1	Full	360	765.5	1,065	275,592	1	\$116,632	\$7,200	\$589,767	\$713,599	
1	2 y	1,800	669.7	730	1,205,430	2	\$160,000	\$36,000	\$2,579,620	\$2,775,620	3.9
2	Full	226	656.6	936	148,387	1	\$102,505	\$4,520	\$317,548	\$424,573	
2	2 y	1,130	623.8	730	704,870	2	\$160,000	\$22,600	\$1,508,422	\$1,691,022	4.0
3	Full	37	646.6	879	23,922	1	\$96,208	\$740	\$30,382	\$127,330	
3	2 y	185	628.5	730	116,275	1	\$80,000	\$3,700	\$147,669	\$231,369	1.8
4	Full	598	793.0	1,476	474,219	1	\$161,643	\$11,960	\$602,258	\$775,861	
4	2 y	2,990	673.7	730	2,014,350	3	\$240,000	\$59,800	\$2,558,225	\$2,858,025	3.7
5	Full	180	734.1	1,099	132,139	1	\$120,356	\$3,600	\$167,817	\$291,772	
5	2 y	900	643.8	730	579,435	1	\$80,000	\$18,000	\$735,882	\$833,882	2.9
6	Full	1,843	817.5	1,332	1,506,625	2	\$291,745	\$36,860	\$1,913,414	\$2,242,019	
6	2 y	9,215	679.8	730	6,264,250	10	\$800,000	\$184,300	\$7,955,598	\$8,939,898	4.0
7	Full	547	859.8	1,306	470,309	1	\$143,025	\$10,940	\$597,292	\$751,258	
7	2 y	2,735	699.4	730	1,912,990	3	\$240,000	\$54,700	\$2,429,497	\$2,724,197	3.6
8	Full	219	932.6	1,524	204,245	1	\$166,899	\$4,380	\$259,391	\$430,671	
8	2 y	1,095	681.1	730	745,815	2	\$160,000	\$21,900	\$947,185	\$1,129,085	2.6
9	Full	499	778.1	1,113	388,283	1	\$121,889	\$9,980	\$493,119	\$624,989	
9	2 y	2,495	681.5	730	1,700,290	3	\$240,000	\$49,900	\$2,159,368	\$2,449,268	3.9
10	Full	133	926.7	2,120	123,254	1	\$232,170	\$2,660	\$156,533	\$391,362	
10	2 y	665	710.1	730	472,225	1	\$80,000	\$13,300	\$599,726	\$693,026	1.8
11	Full	99	968.7	1,245	95,902	1	\$136,345	\$1,980	\$121,796	\$260,121	
11	2 y	495	728.7	730	360,700	1	\$80,000	\$9,900	\$458,089	\$547,989	2.1
12	Full	237	1,128.4	1,638	267,440	1	\$179,384	\$4,740	\$339,649	\$523,773	
12	2 y	1,185	718.2	730	851,125	2	\$160,000	\$23,700	\$1,080,929	\$1,264,629	2.4
13	Full	151	948.5	1,362	143,219	1	\$149,158	\$3,020	\$181,888	\$334,066	
13	2 y	755	714.7	730	539,597	1	\$80,000	\$15,100	\$685,288	\$780,388	2.3
14	Full	137	1,048.1	1,410	143,594	1	\$154,415	\$2,740	\$182,364	\$339,519	
14	2 y	685	727.2	730	498,112	1	\$80,000	\$13,700	\$632,602	\$726,302	2.1
15	Full	161	816.7	1,078	131,481	1	\$118,056	\$3,220	\$166,981	\$288,257	
15	2 y	805	692.3	730	557,330	1	\$80,000	\$16,100	\$707,809	\$803,909	2.8