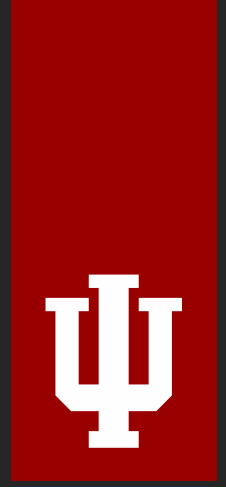


SCHOOL OF PUBLIC HEALTH BLOOMINGTON

Design, statistical analysis, and reporting considerations when using animals in geroscience research

David B. Allison, Ph.D.

INDIANA UNIVERSITY BLOOMINGTON

Disclosures

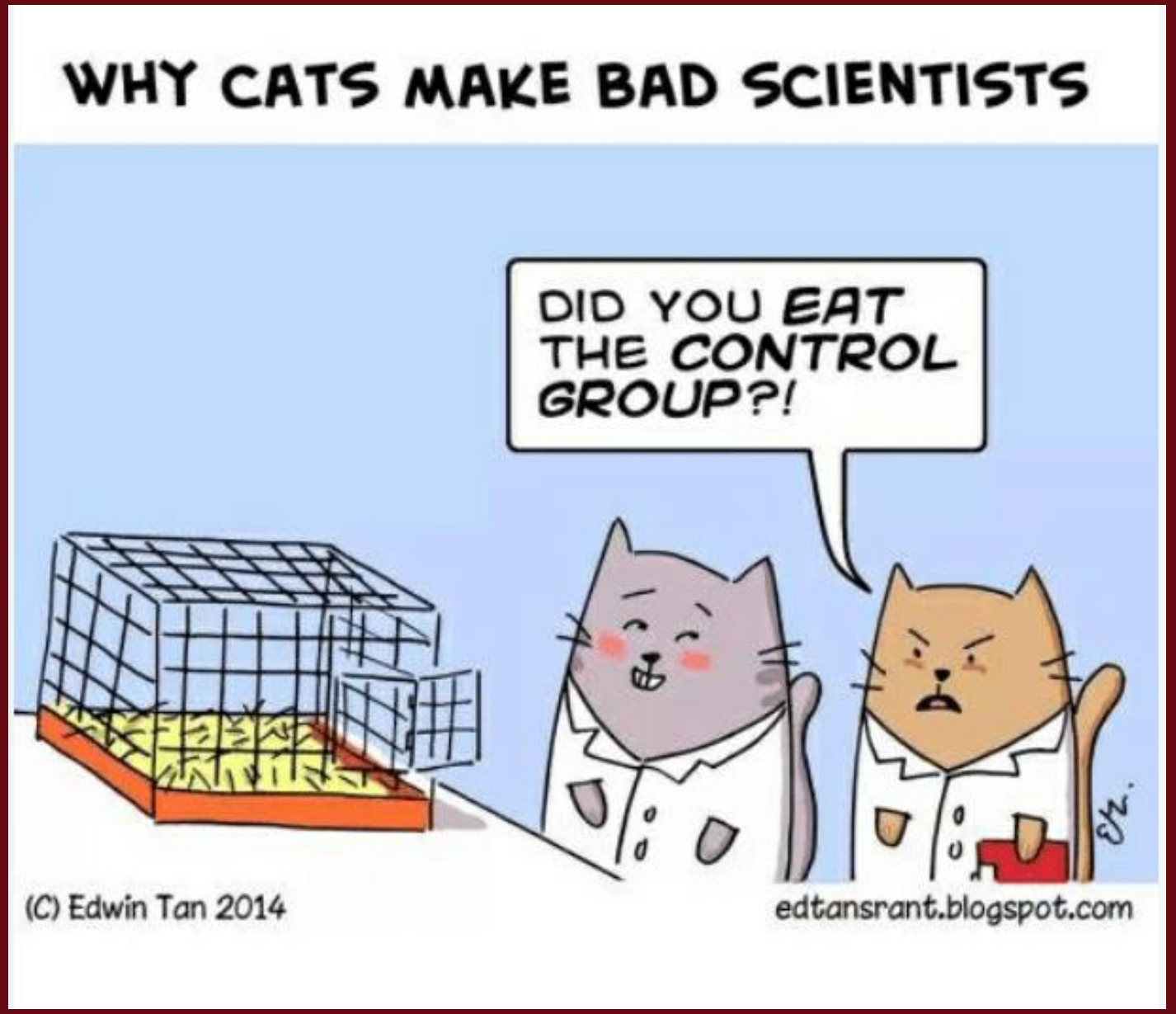
Dr. Allison has received personal payments or promises for same from: American Society for Nutrition; Alkermes, Inc.; American Statistical Association; Biofortis; California Walnut Commission; Clark Hill PLC; Columbia University; Fish & Richardson, P.C.; Frontiers Publishing; Gelesis; Henry Stewart Talks; IKEA; Indiana University; Arnold Ventures (formerly the Laura and John Arnold Foundation); Johns Hopkins University; Law Offices of Ronald Marron; MD Anderson Cancer Center; Medical College of Wisconsin; National Institutes of Health (NIH); Medpace; National Academies of Science; Sage Publishing; The Obesity Society; Sports Research Corp.; The Elements Agency, LLC; Tomasik, Kotin & Kasserman LLC; University of Alabama at Birmingham; University of Miami; Nestle; WW (formerly Weight Watchers International, LLC). Donations to a foundation have been made on his behalf by the Northarvest Bean Growers Association. Dr. Allison was previously a member (unpaid) of the International Life Sciences Institute North America Board of Trustees.

Dr. Allison's institution, Indiana University, has received funds to support his research or educational activities from: NIH; USDA; Soleno Therapeutics; National Cattlemen's Beef Association; Eli Lilly and Co.; Reckitt Benckiser Group PLC; Alliance for Potato Research and Education; American Federation for Aging Research; Dairy Management Inc; Arnold Ventures; the Gordon and Betty Moore Foundation; the Alfred P. Sloan Foundation; and numerous other for-profit and non-profit organizations to support the work of the School of Public Health and the university more broadly.

Acknowledgments

Some slides were inspired by, made in collaboration with, or borrowed from colleagues, and acknowledgments are made throughout. However, the content reflects my thoughts, and not necessarily those of my colleagues or any organization. Special thanks to Colby Vorland, Jasmine Jamshidi-Naeini, Stella Lartey, Luis Mestre, Daniella Chusyd, Andrew Brown, and Keisuke Ejima.

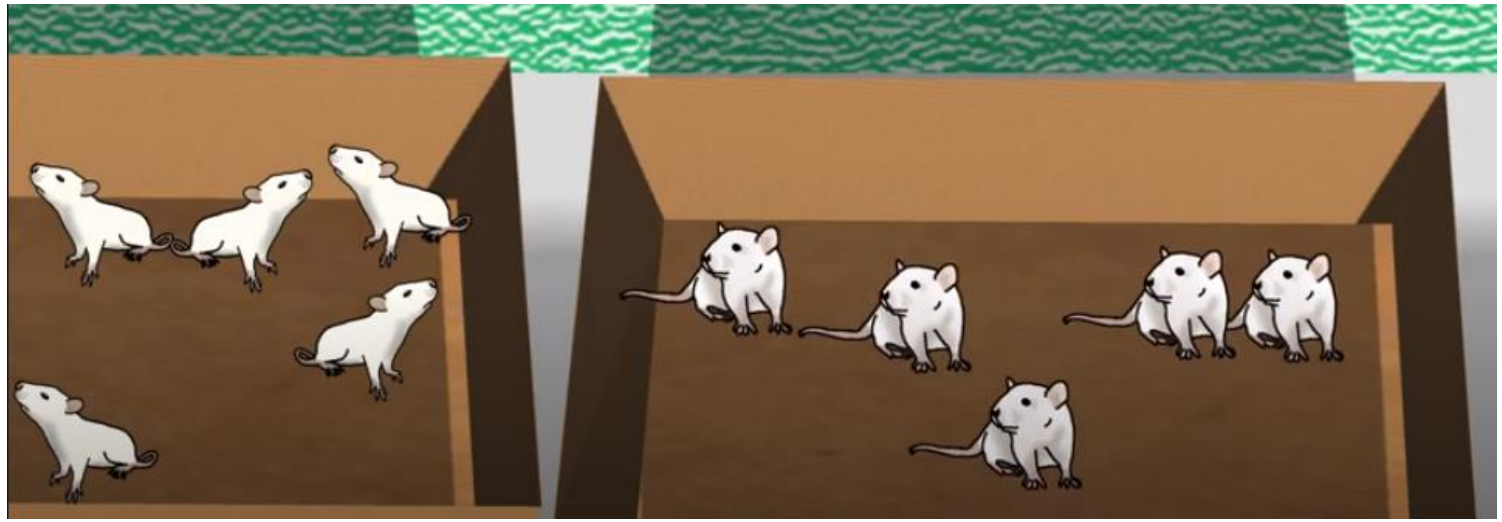
Slides available upon request: Allison@IU.edu

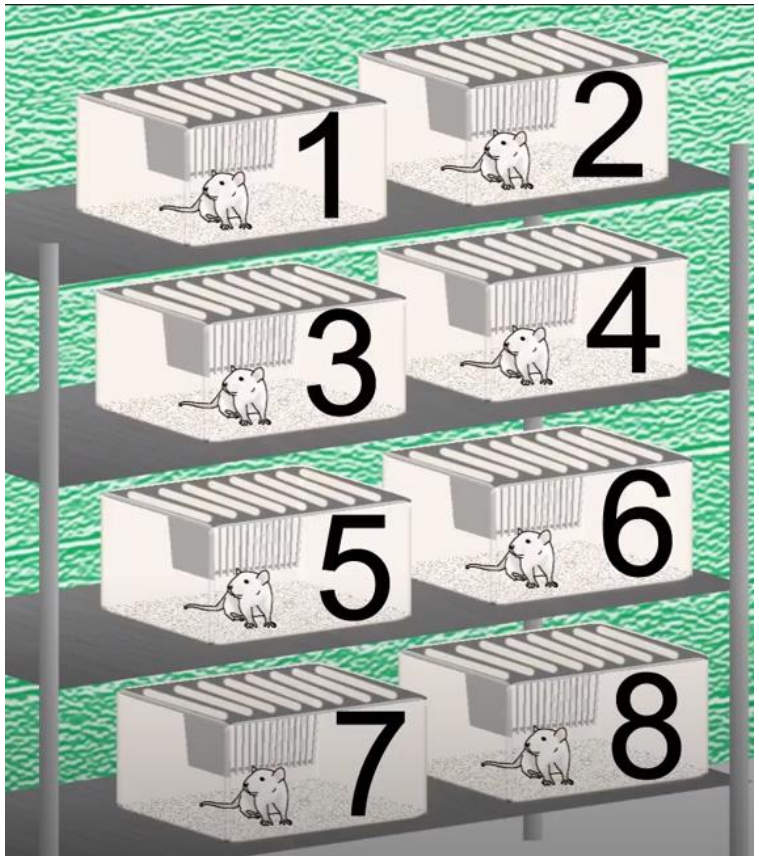


From: https://me.me/i/why-cats-make-bad-scientists-did-you-eat-the-control-1696661

Randomize whenever possible

- Randomization: The only method that controls for both known and ***unknown*** confounders
- Avoid potential bias from unconscious or conscious selection of animal characteristics
- Avoid potential bias from animal positioning in cages/room
- Avoid potential batch effects for outcome measurements



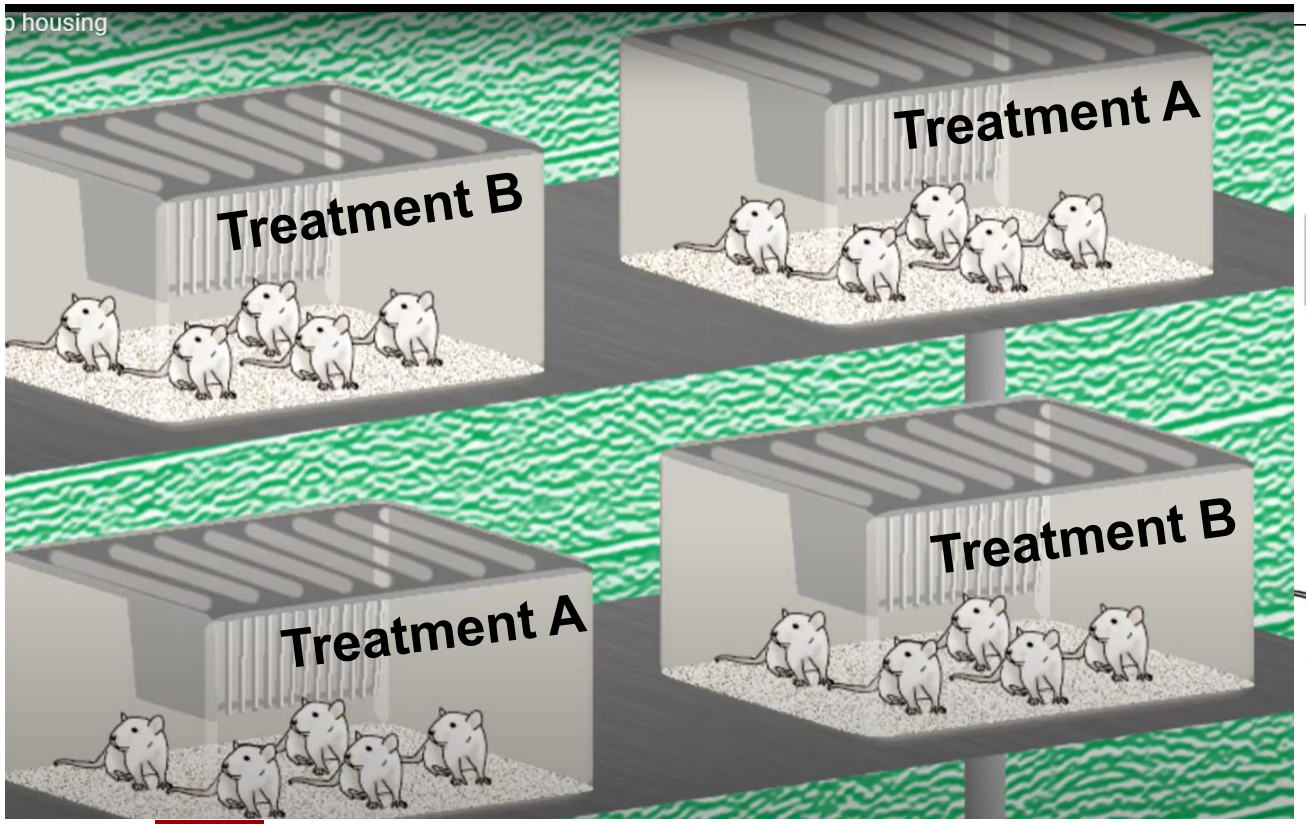


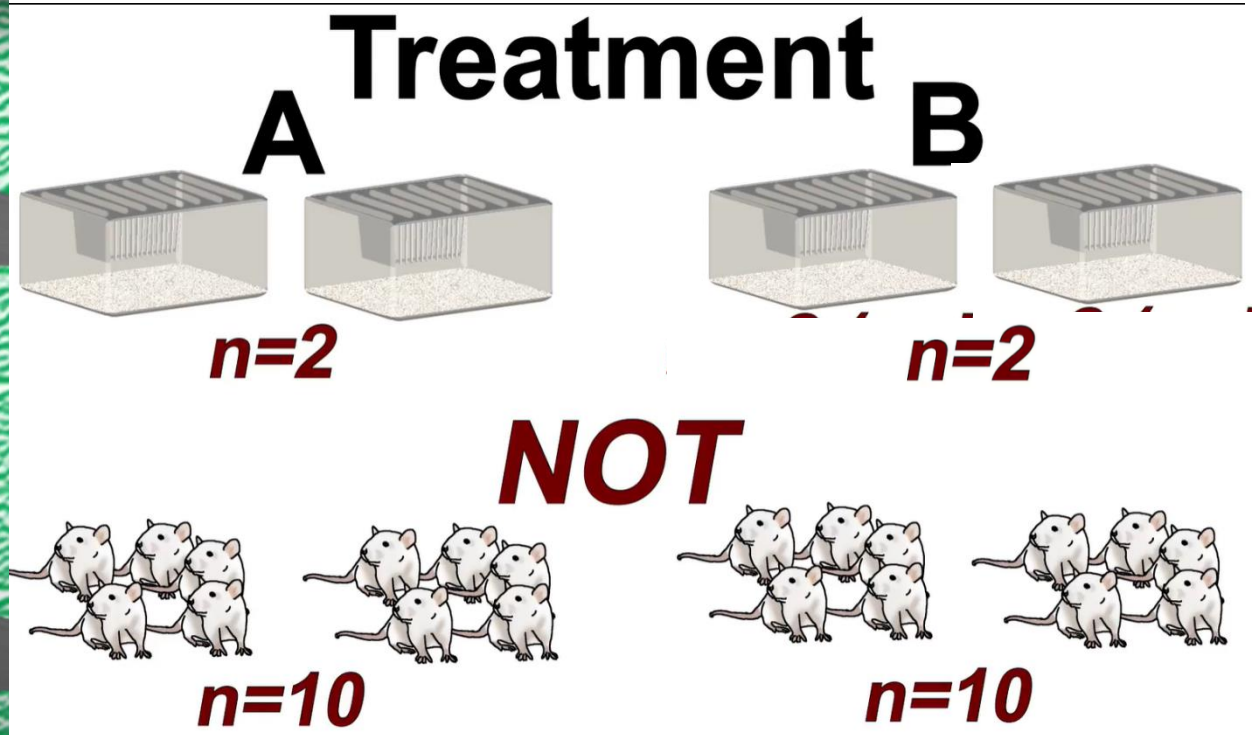
Use a random number generator

- Get help from a professional biostatistician

Account for clustering and nesting

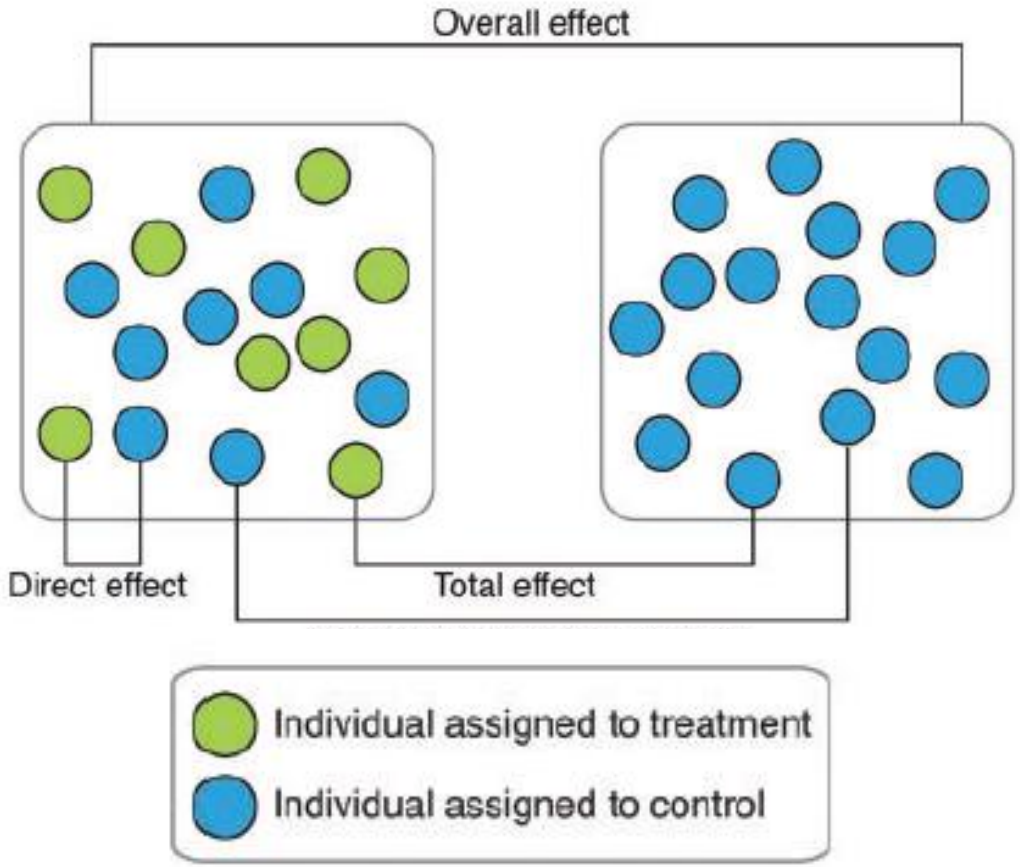
- For treatments applied to cages with multiple animals per cage, the unit of analysis is the cage NOT the individual animals
- Power and sample size are functions of cages AND animals
- Account for clustering and nesting in the statistical model





"Stats in the Lab" Vignettes
https://www.youtube.com/watch?v=dm0LBbN1i0c

Account for the interference effects



Adapted from Benjamin-Chung et al. 2018

Ample evidence that interference occurs in senescence and aging research:

"the developmental dietary history of an animal regulates its own longevity and that of its conspecific neighbors" *Obata et al. 2020*

"results suggest that physical interference between neighboring individuals due to crowding negatively affects growth and reproduction in daphniids" *Ban et al. 2012*

"It has been shown that senescence and SASP production can trigger senescence in neighboring cells via paracrine signaling, a phenomenon that has been referred to as bystander senescence" *Childs et al. 2014*

Consider environmental and design heterogeneity

Highly standardized conditions are often used in experimental studies.

Could the outcome of the study, e.g., mean life span, be different:

1. Had the animals been left in their natural habitat?
2. If the conditions or the design are slightly altered?

How does this affect reproducibility of results in animal studies?

Naked mole-rat

In the wild

In the lab

(Image credit: Shutterstock)

PLOS BIOLOGY

OPEN ACCESS PEER-REVIEWED

META-RESEARCH ARTICLE

Reproducibility of preclinical animal research improves with heterogeneity of study samples

Bernhard Voelkl, Lucile Vogt, Emily S. Sena, Hanno Würbel

Published: February 22, 2018 • https://doi.org/10.1371/journal.pbio.2003693

The call:

1. More comparative empirical work needed to ascertain and bridge any differences, if existing.
2. Collaborations between bio-gerontologists and field biologists.
3. Encourage data sharing and design regulations that allow for data sharing across laboratories and different fields of studies.

Utilize the appropriate between-group statistical comparison in randomized experiments

Conclusion of "Nordic walking for geriatric rehabilitation: a randomized pilot trial" is based on faulty statistical analysis and is inaccurate

David B. Allison, Michelle S. Williams, Gregory A. Hand, John M. Jakicic & Kevin R. Fontaine

Pages 1692-1693 | Published online: 19 Jan 2015

Download citation http://dx.doi.org/10.3109/09638288.2014.1002580 Check for updates

The American Journal of CLINICAL NUTRITION

Best (but oft forgotten) practices: testing for treatment effects in randomized trials by separate analyses of changes from baseline in each group is a misleading approach[1]

doi: 10.3945/ajcn.115.119768

J Martin Bland[2,*] and Douglas G Altman[3]

Incorrect statistical method in parallel-groups RCT led to unsubstantiated conclusions

David B. Allison, Lisa H. Antoine and Brandon J. George

Lipids in Health and Disease 2016 15:77 | DOI: 10.1186/s12944-016-0242-3 | © Allison et al. 2016

Received: 3 December 2014 | Accepted: 21 March 2016 | Published: 15 April 2016

Abstract

The article by Aiso et al. titled "Compared with the intake of commercial vegetable juice, the intake of fresh fruit and komatsuna (*Brassica rapa* L. var *perviridis*) juice mixture reduces serum cholesterol in middle-aged men: a randomized controlled pilot study" does not meet the expected standards of *Lipids in Health and Disease*. Although the article concludes that there are some significant benefits to their komatsuna juice mixture, these claims are not supported by the statistical analyses used. An incorrect procedure was used to compare the differences in two treatment groups over time, and a large number of outcomes were tested

Plan your power and sample size

Power and sample size for survival analysis under the Weibull distribution when the whole lifespan is of interest

Moonseong Heo, Myles S. Faith, David B. Allison *

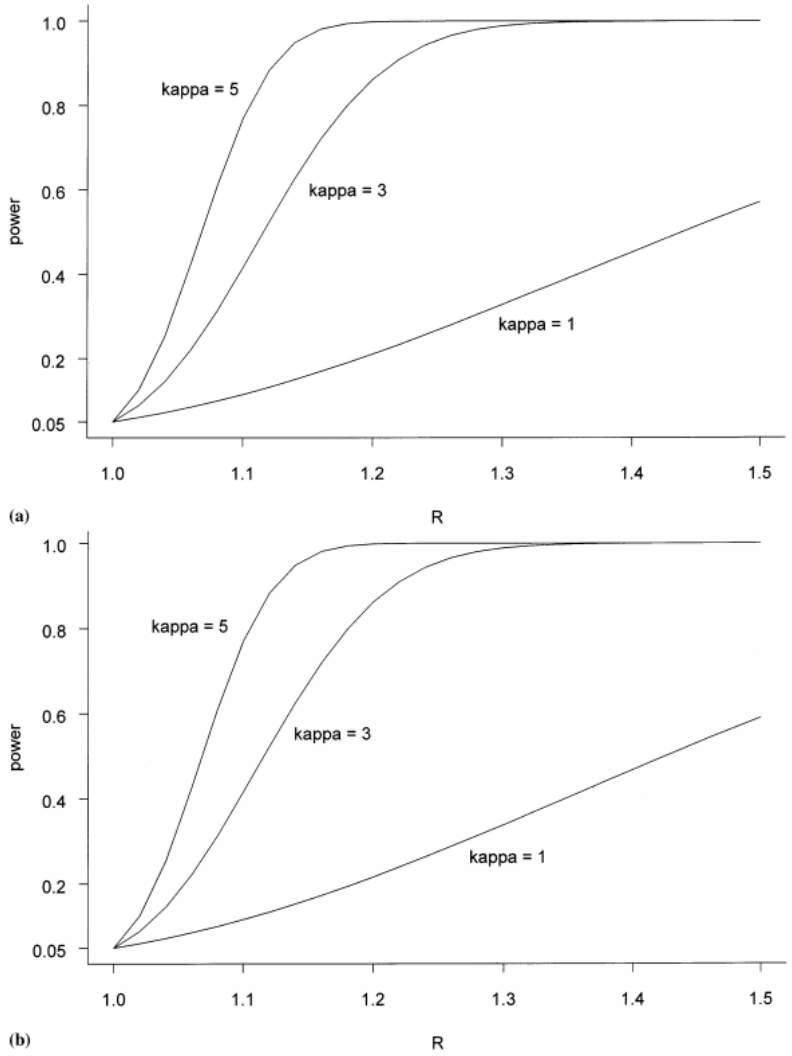


Fig. 2. Power curves for the varying k: (a) = 0, (b) = 1.

Murine genetic models of obesity: type I error rates and the power of commonly used analyses as assessed by plasmode-based simulation

Keisuke Ejima[1,2] · Andrew W. Brown[3] · Daniel L. Smith Jr[4,5,6] · Ufuk Beyaztas[7] · David B. Allison[1]

Accurate and flexible power calculations on the spot: Applications to genomic research

Hemant K. Tiwari*,
Department of Biostatistics, Section on Statistical Genetics, University of Alabama at Birmingham, Birmingham, Alabama, USA, htiwari@uab.edu

Thomas Birkner,
Department of Biostatistics, Section on Statistical Genetics, University of Alabama at Birmingham, Birmingham, Alabama, USA

Ankur Moondan,
Indian Institute of Technology, New Delhi, India

Shiju Zhang,
St. Cloud University, St. Cloud, MN, USA

Grier P. Page,
Research Triangle Institute International, Atlanta, GA, USA

Amit Patki, and
Department of Biostatistics, Section on Statistical Genetics, University of Alabama at Birmingham, Birmingham, Alabama, USA

David B. Allison
Department of Biostatistics, Section on Statistical Genetics, University of Alabama at Birmingham, Birmingham, Alabama, USA

Clinical Nutrition Research Center, University of Alabama at Birmingham, Birmingham, Alabama, USA

Testing for effects beyond the mean

Generalized lambda distribution for flexibly testing differences beyond the mean in the distribution of a dependent variable such as body mass index

K Ejima[1,2,3], G Pavela[3,4], P Li[5] and DB Allison[1,3,5,6]

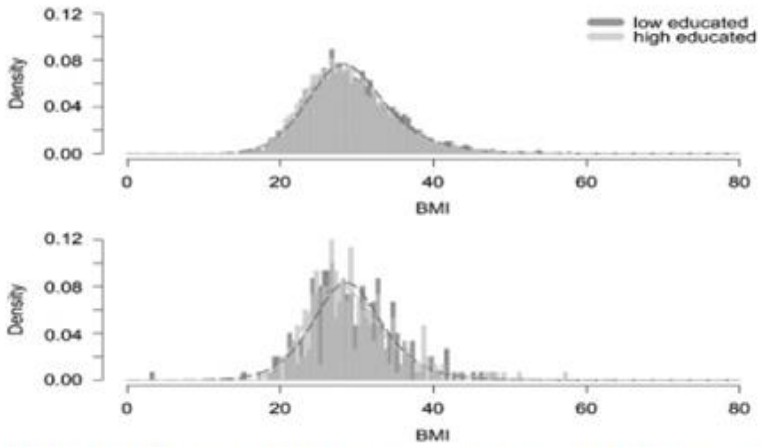


Figure 1. Distribution of BMI by education. The dashed lines show the estimated distributions. Upper panel: complete data set; lower panel: a resampled subsample.

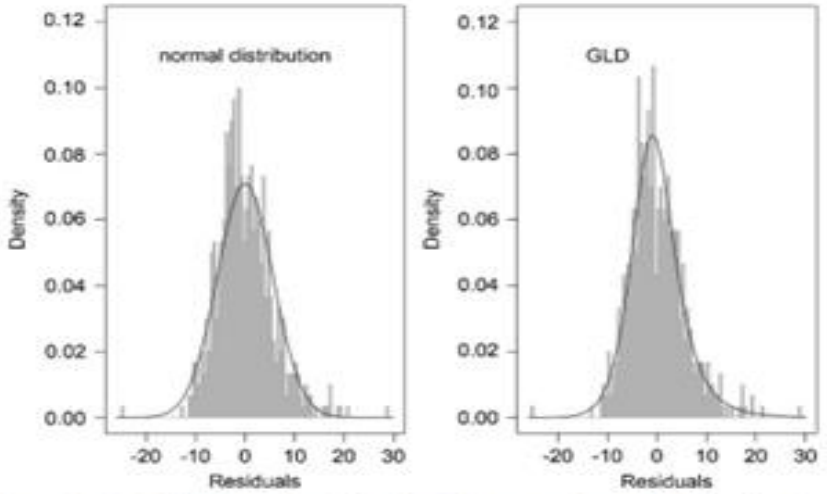


Figure 2. Distribution of residuals. Right panel: normal distribution; left panel: GLD. The lines show the fitted distributions.

Issue: Distributional assumptions of conventional statistics methods may be violated in some situations.

Research Letter

QUANTILE REGRESSION—OPPORTUNITIES AND CHALLENGES FROM A USER'S PERSPECTIVE

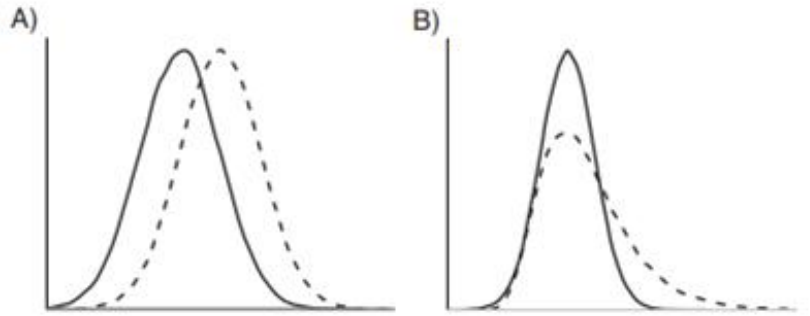
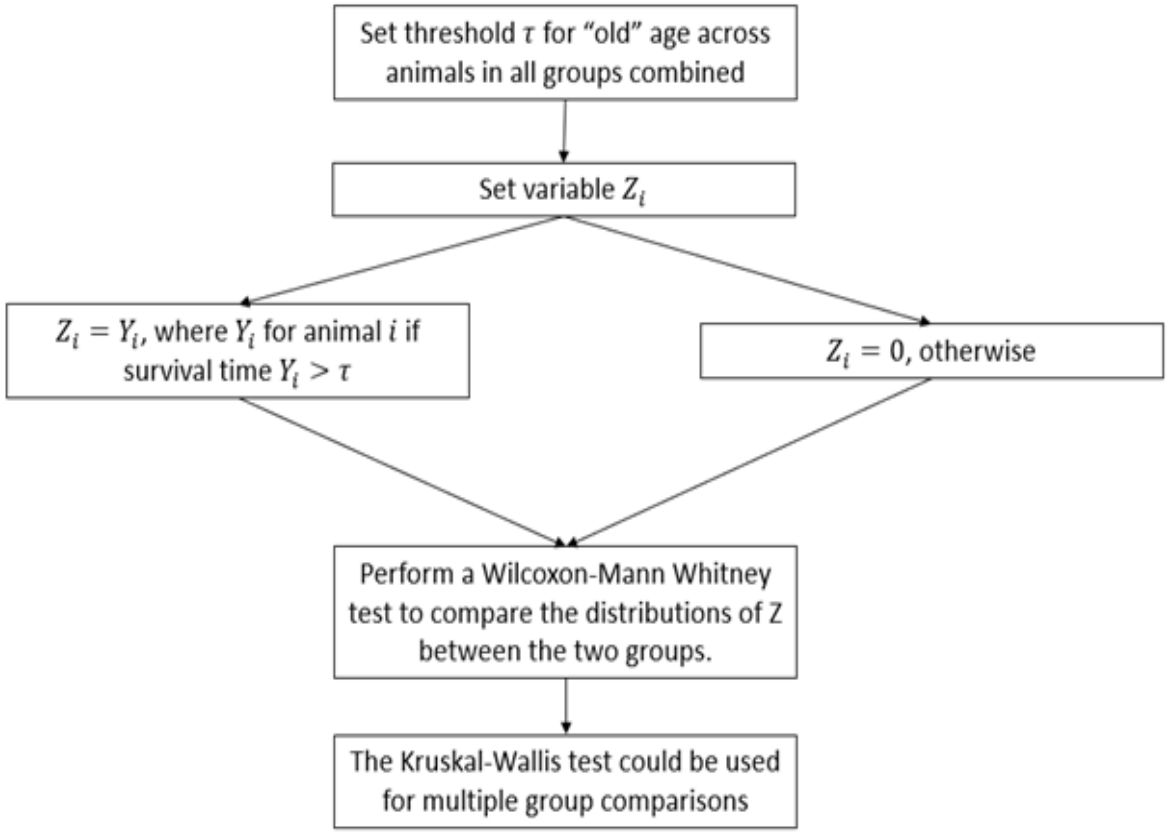


Figure 1. Two distributions may differ with respect to A) their mean only or B) specific quantiles.

Issue: Measures of central tendency only change a little when the distribution changes at the tails.

Maximum Lifespan testing

Objective: To compare longevity among the animals that lived the longest



ELSEVIER

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Mechanisms of Ageing and Development 125 (2004) 629–632

mechanisms of ageing and development

www.elsevier.com/locate/mechagedev

Statistical methods for testing effects on "maximum lifespan"

Chenxi Wang[a,b,1], Qing Li[a], David T. Redden[a,b], Richard Weindruch[c], David B. Allison[a,b,*]

BMC Medical Research Methodology

BioMed Central

Research article

Open Access

Testing for differences in distribution tails to test for differences in 'maximum' lifespan

Guimin Gao[1], Wen Wan[4], Sijian Zhang[5], David T Redden[1,2,3] and David B Allison*[1,2,3]

NOTE: τ is commonly the 90th percentile of survival across the two groups

Tests and power for negligible senescence

- Mortality rate doubling, initial mortality rate, and survival methods have previously been used to test negligible senescence.

- Some of these measures are averred to have failed to accurately capture negligible senescence.

Journal of Gerontology: BIOLOGICAL SCIENCES
1998, Vol. 53A, No. 4, B235–B239

Variations in Senescence and Longevity Include the Possibility of Negligible Senescence

Caleb E. Finch

Andrus Gerontology Center and Department of Biological Sciences, University of Southern California.

The variations in senescence observed in different species span an enormous range of rates that may be described by mortality rate doubling times. This review considers examples of very slowly senescing conifers and fish from natural populations in which advanced age may not compromise reproductive functions. There is thus a basis for considering the possibility that some organisms may experience negligible degrees of senescence in certain environments. A tissue bank is urgently needed to provide specimens of long-lived organisms for study of possible anti-aging mechanisms that permit achievement of great ages.

- Thus, there is the need for solid statistical approaches to accurately determine negligible senescence.

The Annals of Statistics
2005, Vol. 33, No. 3, 1109–1137
DOI 10.1214/009053605000000039

TESTING FOR MONOTONE INCREASING HAZARD RATE

BY PETER HALL AND INGRID VAN KEILEGOM[1]

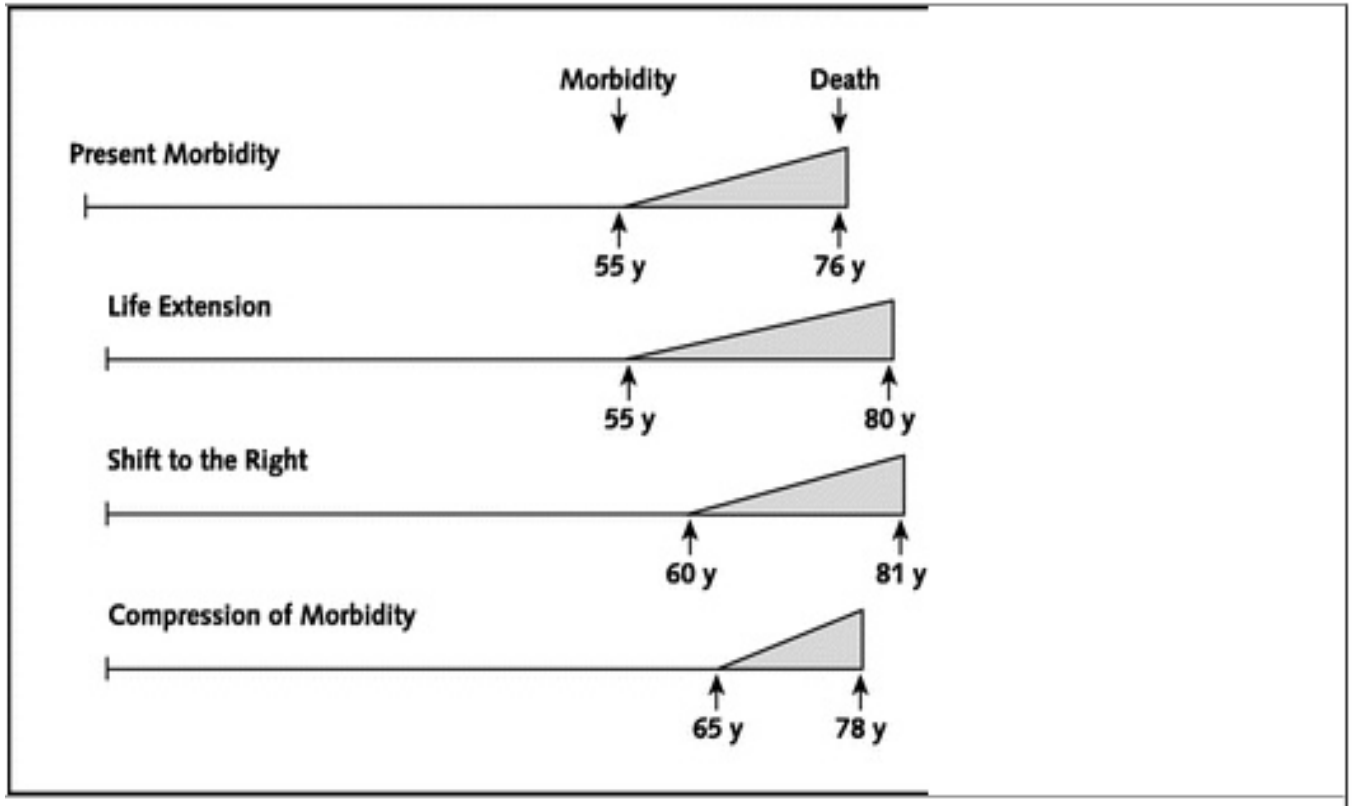
Australian National University, and Australian National University and Université Catholique de Louvain

A test of the null hypothesis that a hazard rate is monotone nondecreasing, versus the alternative that it is not, is proposed. Both the test statistic and the means of calibrating it are new. Unlike previous approaches, neither is based on the assumption that the null distribution is exponential. Instead, empirical information is used to effectively identify and eliminate from further consideration parts of the line where the hazard rate is clearly increasing; and to confine subsequent attention only to those parts that remain. This produces a test with greater apparent power, without the excessive conservatism of exponential-based tests. Our approach to calibration borrows from ideas used in certain tests for unimodality of a density, in that a bandwidth is increased until a distribution with the desired properties is obtained. However, the test statistic does not involve any smoothing, and is, in fact, based directly on an assessment of convexity of the distribution function, using the conventional empirical distribution. The test is shown to have optimal power properties in difficult cases, where it is called upon to detect a small departure, in the form of a bump, from monotonicity. More general theoretical properties of the test and its numerical performance are explored.

Compression of morbidity versus survival curve-squaring

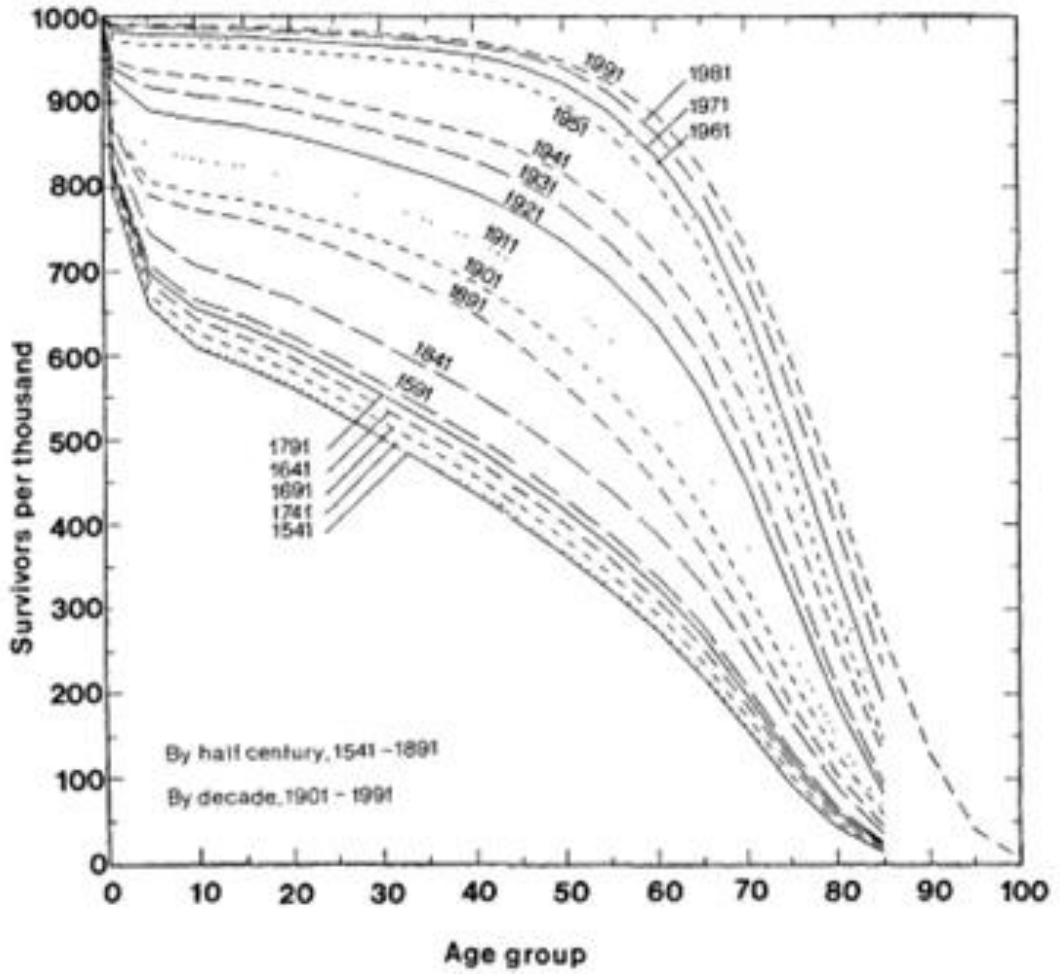
- These are two different concepts.
- However, they are often confused and conflated.

Compression of Morbidity Theory by Dr. James Fries (1980)



Possible scenarios for future morbidity and longevity. Present lifetime morbidity, portrayed as the shaded area, is contrasted with three possible future scenarios.

Squaring the survival curve in human populations



Trajectories of Human Lifespan

Sources: Fries JF. Measuring and monitoring success in compressing morbidity. Annals of Internal Medicine. 139(5 Pt 2):455-9, 2003 Sep 2. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2424092/pdf/0981163.pdf; Rafael de Cabo, Translational Gerontology Branch, National Institute on Aging, NIH.

Report well; follow the ARRIVE Guidelines

Animals	Housing and Husbandry	Diet
• Strain/genotype • Sex • Age • Source	• Type of cage • Type of bedding • Room temperature • Photoperiod • Number of companions	• Background chow • Control diet • Eating frequency • Water quality

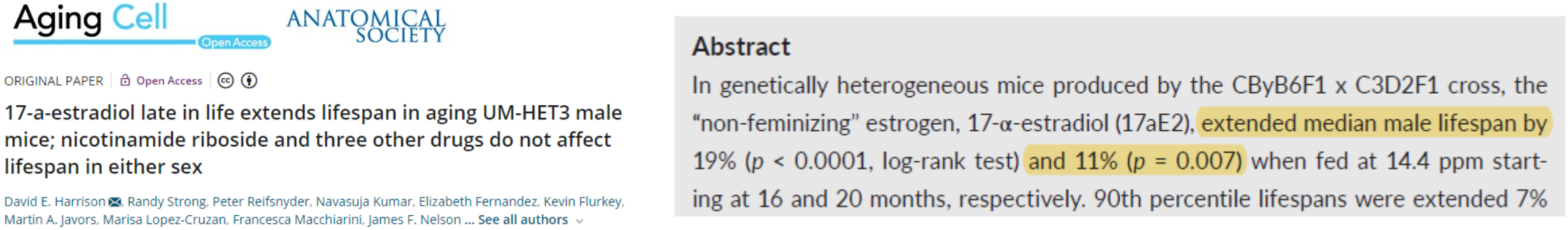
Examples of items in the ARRIVE reporting guidelines

"ARRIVE Essential 10" (the basic minimum)

1. Study design
2. Sample size
3. Inclusion and exclusion criteria
4. Randomization
5. Blinding
6. Outcomes
7. Statistical methods
8. Experimental animals
9. Experimental procedures
10. Results

Report well; use standardized effect size indicators

Not all metrics of effect size are easily interpreted.



Aging Cell Open Access ANATOMICAL SOCIETY

ORIGINAL PAPER | Open Access

17-a-estradiol late in life extends lifespan in aging UM-HET3 male mice; nicotinamide riboside and three other drugs do not affect lifespan in either sex

David E. Harrison, Randy Strong, Peter Reifsnyder, Navasuja Kumar, Elizabeth Fernandez, Kevin Flurkey, Martin A. Javors, Marisa Lopez-Cruzan, Francesca Macchiarini, James F. Nelson ... See all authors

Abstract

In genetically heterogeneous mice produced by the CByB6F1 x C3D2F1 cross, the "non-feminizing" estrogen, 17-α-estradiol (17aE2), extended median male lifespan by 19% ($p < 0.0001$, log-rank test) and 11% ($p = 0.007$) when fed at 14.4 ppm starting at 16 and 20 months, respectively. 90th percentile lifespans were extended 7%

Effect size was reported as 11% median longevity increase from birth. If calculated from time of treatment (20 months), the effect size was 45%.

Standardized mean difference (Cohen's d) , Cohen's f, Eta Squared, Hazard ratio reduction, etc. each has advantages and disadvantages.

For the ideal situation (freedom in reporting and comparison across studies): publicly deposit your raw data

“To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of.”

R.A. Fisher

Contact the NSC for study design and analysis assistance:

- https://nathanshockcenters.org/university-of-alabama
- shockcenter@uab.edu