

A Helping Hand during t -Testing Times

Siouxie Wiles^{1,2*}

1 Bioluminescent Superbugs Lab, Department of Molecular Medicine and Pathology, University of Auckland, New Zealand, **2** Maurice Wilkins Centre for Molecular Biodiscovery, Auckland, New Zealand

I vividly remember the lecturer who attempted to teach me statistics as an undergraduate. The poor man tasked with imparting such vital knowledge alas had no ability to engage a large class of budding biologists. Each lecture he mumbled his way through his notes, writing out long, incomprehensible equations on the board, as paper aeroplanes bounced off his back. Talking to other biologists, it is clear that my experience wasn't unique.

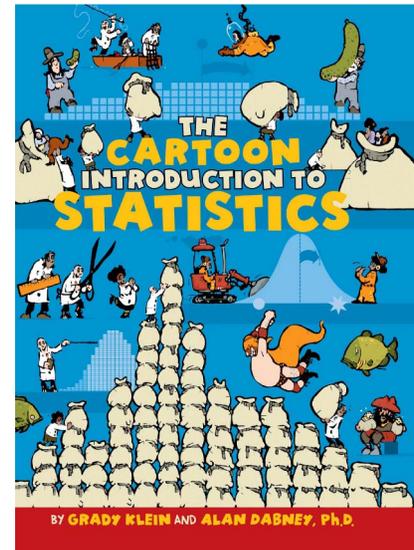
With training like this, perhaps it's no surprise to hear that around half of published biomedical papers are thought to contain mistakes in data presentation and statistical analysis [1–3]. Mistakes include failing to account for multiple comparisons, which can lead to false-positive results, and using tests that assume a normal distribution on data that are skewed. Assumptions of normality are likely to be particularly prevalent when it comes to the analysis of small datasets, such as those generated during research projects involving animals. Such experiments are kept small (but not too small) for ethical reasons, and as such often cannot even be mathematically tested for normality [4].

This fundamental failure to properly address how to best teach statistics and experimental design to biologists has left our community with several generations of principal investigators, journal reviewers, and editors unaware that there is even a problem. Further, the problem is self-perpetuating, as badly presented and incorrectly analysed data often get published in high impact journals, and the methodology emulated by other authors. Perhaps this is how we have arrived at the current situation, where many biologists believe that one-size-fits-all when it comes to performing statistics, the “one size” almost always being the Student's t -test.

With the biological sciences moving rapidly into the “omics/big data” era, which has an inherent need for relatively sophisticated data analysis techniques, we may find that a more maths-savvy generation of biologists arises. In the meantime

though, how can we address this issue? A great start would be to make every biologist read *The Cartoon Introduction to Statistics* by cartoonist Grady Klein and statistician Associate Professor Alan Dabney (Texas A&M University). In their excellent and engaging book, a cast of pirates, aliens, rhinos, flying pigs, and dragon-riding Vikings explain a range of important concepts including bias, means, medians, variation, and hypothesis testing. Like me, Klein and Dabney see no reason to show biologists pages and pages of mathematical formulae to explain these concepts. Instead, Klein has conjured up a “stats machine” that would make Wallace and Gromit proud (“You put your random numbers in here ... slide this thingy over ... and out pops a p -value!”), and Dabney leaves all the technical details for “The Math Cave” at the end of the book.

If nothing else, I would hope that biologists reading Klein and Dabney's book take home just one message: to really look at their data; how did they gather it, how much do they have of it, and how is it distributed? In all honesty, this book won't solve the problem of just how biologists choose the correct statistical test for their data, but hopefully it will pique their interest sufficiently to find out for themselves. As for the wider issue of improving statistical literacy in the biological sciences, it is clear that a range of approaches are currently needed. Lecturers could do to take a leaf out of Klein and Dabney's book and ditch the maths when teaching biologists. Alongside this



Klein G, Dabney A (2013) The Cartoon Introduction to Statistics. New York: Hill and Wang. 240 pp. ISBN 978-0809033591 (paperback). doi:10.1371/journal.pbio.1001785.g001

should be the requirement for engaging stats courses for all post-graduate students carrying out research.

As for principal investigators and journal editors? Perhaps the simplest approach would be to provide clear guidance about what tests are relevant for which kind of data. Indeed, there are a number of easy to follow resources and decision charts available which could be used to either determine the appropriate test to carry out, or in the case of reviewers and editors,

Citation: Wiles S (2014) A Helping Hand during t -Testing Times. *PLoS Biol* 12(2): e1001785. doi:10.1371/journal.pbio.1001785

Published: February 18, 2014

Copyright: © 2014 Siouxie Wiles. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: S.W. is supported by a Sir Charles Hercus Fellowship (09/099) from the Health Research Council of New Zealand (www.hrc.govt.nz). The funders had no role in the preparation of the manuscript.

Competing Interests: S.W. declares that her company, Lucy Ferrin Ltd, supports development of smartphone application versions of the BioStat Decision Tool [8,9] that are available for purchase and will therefore benefit from any profits generated from the sale of the smartphone applications.

* E-mail: s.wiles@auckland.ac.nz

About the Author

Dr. Siouxsie Wiles is an award-winning microbiologist and bioluminescence enthusiast who heads up the Bioluminescent Superbugs Lab at the University of Auckland in New Zealand. Here she combines her twin passions to understand and combat infectious diseases. In a nutshell, Siouxsie and her team make nasty bacteria glow in the dark. Further information about Dr. Wiles' research can be found on her website at <http://superbugslab.org>.

from which they could work backwards to see whether a particular test was used appropriately [5–9].

Nowadays, performing a statistical test merely requires clicking a few buttons in a computer programme. By ensuring biologists are provided with guidance about which buttons to click, and why, I strongly believe that we can raise the standard of data presentation and statistical analysis so that published papers accurately represent the data they describe.

References

1. Cruess DF (1989) Review of use of statistics in The American Journal of Tropical Medicine and Hygiene for January-December 1988. *Am J Trop Med Hyg* 41: 619–626.
2. Welch GE, 2nd, Gabbe SG (1996) Review of statistics usage in the American Journal of Obstetrics and Gynecology. *Am J Obstet Gynecol* 175: 1138–1141.
3. Olsen CH (2003) Review of the use of statistics in infection and immunity. *Infect Immun* 71: 6689–6692.
4. D'Agostino RB, Belanger A, D'Agostino Jr RB (1990) A suggestion for using powerful and informative tests of normality. *Am Stat* 44: 316–321.
5. Millar N (2001) Biology statistics made simple using Excel. *School Sci Rev* 83: 23–34.
6. Handbook of Biological Statistics. “Step-by-step analysis of biological data.” Available: <http://udel.edu/~mcdonald/statstep.html>. Accessed 12 December 2013.
7. Gerwien R. A Painless Guide to Statistics. “Which test do I use?” Available: <http://abacus.bates.edu/~ganderso/biology/resources/statistics.html#whichtest>. Accessed 12 December 2013.
8. Biostat Decision Tree. Available: <http://flexiblelearning.auckland.ac.nz/biostat-tree/>. Accessed 12 December 2013.
9. Wiles S, Bishop AL (2013) BioStat Decision Tree. figshare. <http://dx.doi.org/10.6084/m9.figshare.95778>. Accessed 12 December 2013.