British Journal of Anaesthesia **115** (3): 337–9 (2015)
Advance Access publication 29 May 2015 · doi:10.1093/bja/aev149

Misuse of 'trend' to describe 'almost significant' differences in anaesthesia research

N. M. Gibbs[1,*] and S. V. Gibbs[2]

[1] Department of Anaesthesia, Sir Charles Gairdner Hospital, Nedlands, WA 6009, Australia, and
[2] The University of Western Australia, Crawley, WA 6009, Australia

*Corresponding author. E-mail: neville.gibbs@uwa.edu.au

There are many definitions of 'trend', but none apply to differences that have already been found to be non-significant in a statistical test. Yet, there appear to be many examples in the anaesthesia literature of the use of trend to describe differences that have been found by the authors to be 'almost' but not quite statistically significant (e.g. *P*=0.06). The implication appears to be that there is a subset of non-significant *P* values that suggest, support or represent a trend, by being 'almost significant'. In this editorial we explain that describing non-significant differences as a trend is an error, and argue that it is neither a trivial nor merely semantic error. We also report an audit that suggests that this form of error is not uncommon in the anaesthesia literature and may be increasing in frequency.

The noun trend is defined as a 'general direction in which something is developing or changing' or a 'fashion' by the Oxford Dictionary,[1] and as 'a general direction of change', 'a way of behaving or proceeding', 'something that is developing and becoming more common', 'a tendency' or 'something that is currently popular or fashionable', by the Miriam-Webster's dictionary.[2] To our knowledge, however, no dictionary or any other authoritative source defines trend as 'a difference that is almost, but not quite statistically significant'. Most commonly, 'trend' is used as a general term in both scientific and non-scientific literature to describe apparent changes as per dictionary definitions. On the other hand, formal statistical tests are available when required to estimate the probability that observed changes in an apparent trend (e.g. in a time series) represent true differences rather than chance findings.[3–7] These include the χ^2 test for linear trend, the Cochran-Armitage test, and the Mann-Kendall trend test, to name only a few.[3–7]

Applying the term trend to almost significant differences demonstrates a misunderstanding of the meaning of *P* values. A *P* value describes the probability of obtaining the observed result or one more extreme given that the null hypothesis is true.[8 9] If the probability is less than a pre-specified value (α, acceptable type I error) set by the authors, the null hypothesis is rejected. Typically this value is 0.05, but could be 0.01 or 0.1 or another value determined by the authors. The outcome of an inferential test is either rejection of the null hypothesis, or failure to reject the null hypothesis. There is no other outcome. In particular, there is no 'almost rejected' category when *P* values approach but are slightly greater than the pre-set α. To imply that there is an 'almost rejected' category is an obvious statistical error.

The overwhelming majority of *P* values >0.05 (or other pre-set α) in the anaesthesia literature (and indeed the scientific literature) are reported as non-significant, without any mention of trend, no matter how close they are to being significant. How then do some 'almost significant' *P* values suggest or support a trend, but not others? What range of *P* values are considered 'almost significant'? To be consistent, either *all P* values within this hypothetical range suggest or support a trend or *none*. Describing some 'almost significant' *P* values as a trend but not others introduces a large element of subjectivity.

Describing a *P* value close to but not quite statistically significant (e.g. 0.06) as supporting a trend toward statistical significance has the same logic as describing a *P* value that is only just statistically significant (e.g. 0.04) as supporting a trend toward *non*-significance.[10] Yet *P* values that are only just statistically significant are rarely if ever described as supporting a trend toward non-significance. Also, using trend to suggest 'a general direction of change' or 'tendency' once a finding has already been found to be non-significant (i.e. consistent with a chance finding) is a form of begging the question. The *purpose* of inferential testing (i.e. obtaining a *P* value) is to assess the likelihood that observed differences suggest a 'general direction of change' or 'tendency' as opposed to chance findings.

To estimate how frequently this error occurs in the anaesthesia literature, we conducted an audit of three anaesthesia journals. Using word recognition software we identified all uses of the term trend in all articles (including editorials, excluding case reports) published in the *British Journal of Anaesthesia*, *Anesthesia and Analgesia*, and *Anaesthesia and Intensive Care* in 1990, 2000, and 2010. There was a total of 2143 articles and in 258 articles trend was used at least once. We scrutinized each use of trend to ascertain whether it described a difference that had been found to be non-significant on the basis of the authors' own a priori specified alpha error (e.g. describing a non-significant difference as a 'trend toward statistical significance', a 'non-significant trend', or as 'showing a trend' despite the non-significance). We did not discriminate between primary and secondary outcomes. We found at least one example of this incorrect use in 28 of the 833 articles published by the three journals in 2010 (3.4%). This represented an average of about one example of misuse for each issue of each journal in 2010. For 2000, we found 14 examples among 817 articles (1.7%), and for 1990, 8 examples among 493 articles (1.6%). The 10 examples from the *British Journal of Anaesthesia* are shown in Table 1. [Details of the incidence across the three journals and a list of all 50 examples are available in Supplementary Appendices.]

These results confirm that there is a subset of articles in the anaesthesia literature in which trend is being misused to

Table 1 The 10 examples of the use of trend to describe an almost significant difference in the *British Journal of Anaesthesia* in 1990, 2000, or 2010. The use of bold type and information in large brackets are added by the editorial authors

Although not statistically significant, there **was a trend** which suggested that bupivacaine with adrenaline performed best in almost every variable tested; 1990: 65: 648–653
The **trend toward reduced motor block** together with a significantly increased need for supplementary analgesics . . . confirm findings from a recent study . . . [There was no significant difference in relation to motor block]. 2000; 84: 826–7
Results **showed a trend** for a lower rate of bacterial infection with the use of tranexamic acid when compared with placebo (*P*=0.12). 2010; 104: 23–30.
A trend was also seen towards a reduction in myocardial damage after operation. However, this trend did not reach statistical significance. 2010; 104: 305–12
Patients with the combination of both active drugs **showed a trend** to request opioids later than those with a single analgesic (*P*>0.05). 2010; 104: 761–7
However, none of the LA treatments significantly influenced epidermal or inflammatory cell wound MIF levels, although there **was a trend** towards increased overall wound MIF levels in animals treated with bupivacaine 0.5%. 2010; 104: 768–73
There **was a trend** to less pain at rest in the F group. [*P*=0.09]. 2010; 105: 185–95
There **was a non-significant trend** for non-achievers to have higher sedation scores at the time of study entry. 2010; 105: 326–33
The study was not powered for subgroup analysis, but there **was a trend** towards reduced hospital mortality in the cell saver group[48]. [Reference 48 reports that there was no statistically significant reduction in overall hospital mortality in the cell saver group, *P*=0.07]. 2010; 105: 401–16
The worst postoperative chronic pain score (VAS/NRS) was reported in one trial[24] **showing a trend** for a better outcome 12 months after surgery (*P*=0.14). 2010; 105: 842–52

describe 'almost significant' differences. Moreover, we found an increase over the three index years consistent with a true trend (Cochran-Armitage test for trend, *P*=0.021, Fig. 1). This observation serves to highlight the ambiguity that is introduced if trend is used for statistical findings that have *not* been subject to a specific test for trend, and moreover, which have been found to be *non-significant* in another test. Furthermore, this incorrect use of trend represented only a small proportion of the total uses of trend in our audit. The majority of uses were for correct purposes (e.g. in relation to dictionary definitions or specific statistical tests for trend). This majority correct use is undermined by the ambiguity introduced by the small subset of misuse.

Although we audited only three anaesthesia journals, we have no reason to suspect that our findings are not typical of the broader range of anaesthesia journals. The three journals are published in different regions indicating that this is not a local issue. Moreover, two of the journals have a high impact factor and wide readership, and would be considered to be amongst the mostly highly regarded in the anaesthesia literature.

It is likely that the use of trend to describe almost significant differences is mostly an innocent error, and that the intention is to imply only that the observed differences, although non-significant, may be worthy of further investigation in subsequent more highly powered studies. This may be an entirely appropriate interpretation, as a negative finding is never proof of 'no difference'. On the other hand, misuse of trend to describe almost significant differences could be misinterpreted by less informed readers as suggesting a real trend, which would be misleading.

In summary, the use of trend to describe 'almost significant' differences is an error both in word usage and statistical inference. It introduces both inconsistency and ambiguity. It promotes a misunderstanding of *P* values and undermines the many correct uses of the term. More importantly, it may be misleading if readers assume that a real trend has been suggested, supported or demonstrated. Our audit findings indicate that this error is not uncommon in anaesthesia research and may be increasing. We recommend that trend should not be used to describe any subset of non-significant differences and should be reserved only for the currently accepted dictionary or scientific definitions of the term, or in relation to specific statistical tests for trend.

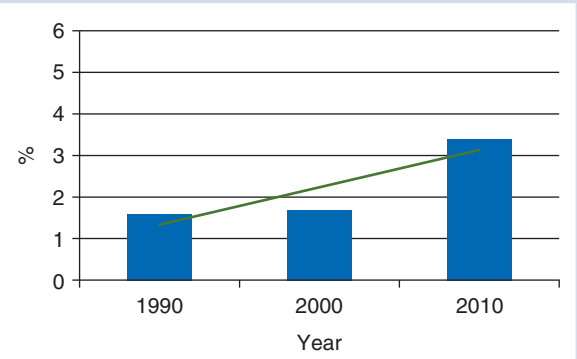

Fig 1 Use of trend to describe almost significant differences in the *British Journal of Anaesthesia*, *Anesthesia and Analgesia*, and *Anaesthesia and Intensive Care* in 1990, 2000, and 2010. Bars indicate the number of articles using trend to describe an almost significant difference as a percentage of all scientific articles (including editorials, excluding case reports) published by the three journals for the yr. *P* value for trend=0.021 (Cochran-Armitage test).

Supplementary material

Supplementary material is available at *British Journal of Anaesthesia* online.

Acknowledgement

The authors wish to thank Dr William Weightman for performing the Cochran-Armitage test for trend.

Declaration of interests

N.M.G. is currently the Chief Editor of *Anaesthesia and Intensive Care*. S.V.G. has no interests to declare.

References

1. *Oxford Dictionary of English*, 3rd Edn. Oxford: Oxford University Press, 2010
2. *The Merriam-Webster Dictionary*. Revised Ed, Springfield: Merriam Webster US, 2004
3. Bland M. *An Introduction to Medical Statistics*, 3rd Edn. Oxford: Oxford University Press, 2000; 243–5
4. Armitage P, Berry G, Matthews JNS. *Statistical Methods in Medical Research*, 4th Edn. xxx: Blackwell Science, 2002; 503–11
5. Chi-squared tests for trend. Available from http://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/8-chi-squared-tests (accessed November 2014)
6. Armitage P. Tests for linear trends in proportions and frequencies. *Biometrics* 1955; **11**: 375–86
7. Cochran WG. Some methods for strengthening the common χ^2 tests. *Biometrics* 1954; **10**: 417–51
8. Myles PS, Gin T. *Statistical Methods for Anaesthesia and Intensive Care*. Edinburgh: Elsevier, 2000; 19–23
9. Kusuoka H, Hoffman JIE. Advice on statistical analysis for circulation research. *Circ Res* 2002; **91**: 662–71
10. Mansfield L. The reading, writing, and arithmetic of the medical literature, part 2: critical evaluation of statistical reporting. *Ann Allergy Immunol* 2005; **95**: 315–22

British Journal of Anaesthesia **115** (3): 339–42 (2015)
Advance Access publication 1 June 2015 · doi:10.1093/bja/aev154

Information technology innovation: the power and perils of big data

G. P. Giambrone[1], H. C. Hemmings[1], M. Sturm[2] and P. M. Fleischut[1,*]

[1] Department of Anesthesiology, Weill Cornell Medical College, 525 East 68th Street, Box 124, New York, NY 10065, USA, and [2] Department of Information Services, NewYork-Presbyterian Hospital, 333 East 38th Street, New York, NY 10065, USA

*Corresponding author. E-mail: pmf9003@med.cornell.edu, pfleischut@gmail.com

The global health-care system is consistently under tremendous pressure to lower health-care costs, maintain high efficiency and quality of care, and remain up-to-date technologically in an era of instantaneous information exchange. In the UK, around 8.4% of the gross domestic product is spent on health care (approximately 0.19 trillion GBP).[1] In the USA, this number is 17.9% of gross domestic product, or 2.7 trillion USD.[2] With the introduction of health-care reform and a shift in payment structure to pay-for-performance, further pressure has been placed on the health-care system to reduce costs and increase health-care quality. Additionally, a shift in patient characteristics to an ageing population and improved access to care have increased the number of patients seeking care.[3] Compounding the situation is a shortage of key practitioners, including nursing staff, in the medical workforce.[3 4] As a result of staff shortage and external pressure and regulations from government agencies to reduce costs, the health-care system must find a way to improve the quality of patient care for more patients with fewer resources. With these difficulties in mind, and the additional challenges that lie ahead, the health-care system, including anaesthetists, must continue to use innovative medical technologies and become more efficient in the collection and analysis of this information to drive cost-effective clinical practice.

As discussed in the article by Simpao and colleagues,[5] technological advancements in health care have led to an explosion in data collection, increasing storage and analysis needs. In 2011, there were 1.8 zettabytes of data created globally.[6] In the same year, it was estimated that data from the US health-care system reached 150 exabytes.[6 7] This number will continue to grow to reach zettabyte (10^{21} gigabytes) followed by yottabyte (10^{24} gigabytes) levels over time.[6] Data of this magnitude are known as 'big data,' defined as electronic data sets so large and complex that they are difficult or impossible to manage with traditional software, hardware, or both; nor can they be easily managed with traditional or common data-management tools and methods.[7] There are three primary characteristics of big data: volume (the amount of data generated by organizations, individuals, or machines), variety (data in all forms; structured, unstructured, and semi-structured), and velocity (the speed of data generation, delivery, or processing).[7] The creation of these massive data sets with varying formats is a result of the proliferation of electronic health records (EHRs). The EHRs have vastly improved the maintenance of health information and have promoted the collection and sharing of information among providers across all health-care disciplines, leading to a more collaborative approach to patient care. In the field of anaesthesia, the EHRs, also known as Anaesthesia Information Management Systems (AIMS) or Anaesthesia Information Systems (AIS), have decreased inaccuracies, incompleteness, biases, and inherent errors.[8] However, implementation of these EHRs has created data sets that can be difficult to analyse for quality control or research purposes. In one study of AIMS, event recording dependent on user input can have a low sensitivity (38%), leading to under-reporting of key clinical events,[8] demonstrating that EHR systems are in need of improvements and analytics to identify issues. Further adding to the predicament of using big data in health care is the development and use of low-cost, non-invasive, wearable health-monitoring systems that allow for continuous monitoring of patients' vital signs and mobility from external locations rather than the traditional approach of hard-