

Truths, lies, and statistics

Matthew S. Thiese, Skyler Walker, Jenna Lindsey

Rocky Mountain Center for Occupational & Environment Health, Department of Family and Preventive Medicine, School of Medicine, University of Utah, Salt Lake City, UT 84108, USA

Correspondence to: Matthew S. Thiese, PhD, MSPH. Rocky Mountain Center for Occupational & Environment Health, Department of Family and Preventive Medicine, School of Medicine, University of Utah, 391 Chipeta Way, Suite C, Salt Lake City, UT 84108, USA. Email: matt.thiese@hsc.utah.edu.

Abstract: Distribution of valuable research discoveries are needed for the continual advancement of patient care. Publication and subsequent reliance of false study results would be detrimental for patient care. Unfortunately, research misconduct may originate from many sources. While there is evidence of ongoing research misconduct in all its forms, it is challenging to identify the actual occurrence of research misconduct, which is especially true for misconduct in clinical trials. Research misconduct is challenging to measure and there are few studies reporting the prevalence or underlying causes of research misconduct among biomedical researchers. Reported prevalence estimates of misconduct are probably underestimates, and range from 0.3% to 4.9%. There have been efforts to measure the prevalence of research misconduct; however, the relatively few published studies are not freely comparable because of varying characterizations of research misconduct and the methods used for data collection. There are some signs which may point to an increased possibility of research misconduct, however there is a need for continued self-policing by biomedical researchers. There are existing resources to assist in ensuring appropriate statistical methods and preventing other types of research fraud. These included the “Statistical Analyses and Methods in the Published Literature”, also known as the SAMPL guidelines, which help scientists determine the appropriate method of reporting various statistical methods; the “Strengthening Analytical Thinking for Observational Studies”, or the STRATOS, which emphasizes on execution and interpretation of results; and the Committee on Publication Ethics (COPE), which was created in 1997 to deliver guidance about publication ethics. COPE has a sequence of views and strategies grounded in the values of honesty and accuracy.

Keywords: Research misconduct; falsification; fabrication; analytical methods

Submitted Aug 19, 2017. Accepted for publication Aug 28, 2017.

doi: 10.21037/jtd.2017.09.24

View this article at: <http://dx.doi.org/10.21037/jtd.2017.09.24>

Introduction

Dissemination of quality research findings are necessary for the widespread improvement of patient care. Without quality research, practice improvement would depend upon individual clinicians own observations and desire to improve. This is particularly true for surgical medicine where dissemination of quality research exponentially improves the treatment of patient populations. The goal of medical research should be to determine scientific truth regarding a treatment, exposure, or outcome. Clinicians rely

upon peer-reviewed, published literature to improve patient care and continue to make informed treatment decisions while considering the increasing complexity in medical care including new treatments, procedures, guidelines and related concerns (1).

Research, surgical or otherwise, must have sound design, execution, and analysis to be considered quality. Study design and execution can often have shortcomings, which impact the quality of research. Examples of these include the use of a poor comparison treatment, lack of blinding, poor randomization, and small sample size. In general,

Table 1 Indicating activities or methods of research misconduct

| |
|---|
| Study design and data collection |
| Inappropriate design |
| Distortion of design |
| Carelessness or incompetence |
| Fabrication of data |
| Not following protocol for safety or informed consent |
| Not obtaining institutional review board approval prior to starting the study |
| Analytical methods |
| Falsification of data |
| Improper analysis |
| Multiple comparisons not reported or adjusted |
| Misrepresentation of statistical methods/analysis |
| Post Hoc analyses not identified |
| Exclusion of outliers |
| Publication |
| Selective reporting |
| Failure to publish/agreement not to publish |
| Over interpretation of results |
| Study weaknesses not described |
| Duplicate or near identical publications |
| Undisclosed or meaningful conflicts of interest |
| Will not provide raw data |

most of these elements can be detected and considered with thorough reading of a study's methods and results. Many criteria can assist in the identification of quality evidence, which are employed when creating treatment guidelines (2,3). Conversely, analysis of study data can be manipulated to achieve desired results. In general, most studies report relatively little if any statistical analysis decisions or assessments of validity (4-6). Authors often only state the analytical methods employed and the statistical analysis package used to conduct the analyses.

The development of a statistical analysis plan, adherence to that plan, the specific statistical analysis used, decisions made during analysis, assumptions made regarding the data, and subsequent results are influenced by factors including quality of data, appropriate choice and implementation of statistical analysis methods, rigorous evaluation of the data, and truthful interpretations of the results (7,8).

Biomedical research has become increasingly complex, particularly in surgical clinical research (1). Review of literature finds statistical analyses consisting primarily of t-tests and descriptive statistics (i.e., means, standard deviations, range) (5,9). Recent advances in statistical methods and the increasing computational power have given rise to increasingly robust analytical tools that can be used in clinical research. Although, statistical analysis methods have become more robust, basic statistical tests employed commonly in the seventies continue to be employed as the main or often only analytical tool in surgical research (10). It is still relatively common to find incorrect statistical evaluations performed for the given study design and/or type of data. Basic parametric tests continue to be used often, even though most data are not normally distributed (11). A review of 91 published comparative surgical papers found most of which (78%) contained potentially meaningful errors in the application of analytical statistics. Common errors included not performing a test for significance when indicated, providing p-values without reference to a specific comparison, and inappropriate application of basic statistical methods (12). Another study assessing 100 orthopedic surgery papers reported 17% of the results did not support the overstated conclusions and 39% performed the incorrect analysis altogether (13). Reviews of other peer-reviewed literature found approximately half of the clinical research have one or more statistical errors, a few of which influenced results and interpretation of study findings (14-16).

The goal of this paper is to describe the common statistical errors in published literature and how to avoid and detect these errors.

Research misconduct

Research misconduct, often referred to as fraud, can encompass a range of activities including: fabrication of data or results, plagiarism of ideas or text without giving appropriate credit, falsification of research methods or results (e.g., omitting data or significantly deviating from the research protocol), and manipulation of the peer review process. Research misconduct generally does not include unintentional errors, but rather intentional misrepresentations of research data, processes, and/or findings. *Table 1* includes research practices that are indicators of potential research misconduct, if not research misconduct itself (17-19).

It is difficult to identify the true prevalence of research misconduct, particularly in clinical trials. There have

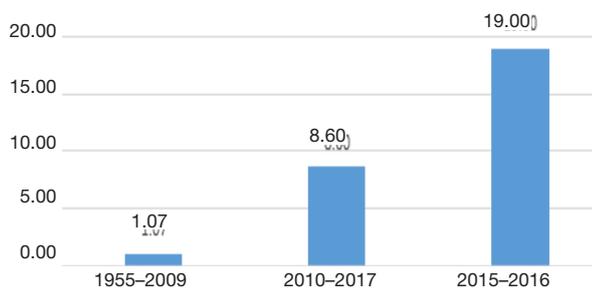


Figure 1 Mean number of retractions per year in published neurosurgical literature adapted from Wang *et al.* 2017 (26).

been some attempts to quantify the prevalence of research misconduct; however, the different studies are not readily comparable due to varying definitions of research misconduct. Additionally, the population studied is difficult to identify and analyze. Theoretically, research misconduct could be committed by anyone conducting research, from research assistants on small pilot studies, to the principal investigator (PI) in a large multicenter trial. Identification of the PI is relatively easy; however, enumeration of all researcher workers involved in a study are generally not feasible. Additionally, it is logical to believe that individuals at certain career levels may have higher likelihood of committing research misconduct (e.g., to publish in order to obtain tenure). The largest problem in identifying the prevalence of research misconduct is the response bias toward under-reporting, even in anonymous surveys. Research in this area has tried to address this by asking non-self-reporting questions such as “do you know anyone who has committed research misconduct?” instead of self-reporting questions such as “have you ever committed research misconduct?”. It is likely that reported prevalence of individuals answering the self-reporting question is meaningfully different from the unbiased or non-self-reporting prevalence. In spite of these shortcomings, there is a small but growing body of literature assessing the prevalence of research misconduct in both self-reported and non-self-reported studies. Among self-reported prevalence of research misconduct, the estimated mean prevalence is approximately 2%, with a possible slight decreasing downward trend over time (20). Conversely, overall prevalence of non-self-reported research misconduct was slightly higher than 14% (20).

Detection of misrepresentations, including inadvertent error and intentional misconduct should result in a retraction of the published article(s), which is uncommon

with estimates of retracted articles at 0.07% (21). A study in 1999 identified retracted articles indexed in Medline from 1966 to 1996 (22). The study identified 198 statements of retraction of publication, retracting 235 articles. These were subsequently classified according to the primary reason for the retraction, with most (38.7%) being retracted for errors followed closely by (36.6%) retracted for misconduct (22). Many of these retracted articles were then cited by 1,893 subsequent articles after the retraction was made, with nearly all citations being either explicitly (14.5%) or implicitly positive (77.9%) and only a small proportion (7.5%) acknowledging the retraction (22). A recent study evaluating retractions specifically in orthopedic research found that a majority of the retractions were for fraud (26.4%) or plagiarism (22.7%) (23).

There are multiple studies which demonstrate that retraction rates are increasing (24). An article evaluating retractions between 2001 and 2011 reported a 10 fold increase in retractions for that time period (25). *Figure 1* shows data adapted from Wang *et al* depicting the increase in retraction frequency of neurosurgical literature (26). The study of retractions in orthopedic research found similarly increasing trends of the numbers of retractions per year (26).

Some solutions have been suggested and progress has been made in trying to prevent research fraud in clinical research. Potential solutions include statutory regulation (19) and requiring sharing of raw data for independent evaluation and replication of results (18).

Unintentional statistical misrepresentations

In many specific instances it may be difficult or impossible to identify whether or not an issue with a study was intentional and therefore considered research misconduct, or unintentional and attributed to an honest error, difference of opinion, or other benign cause. It is likely that prevalence estimates of unintentional errors in research is less biased than estimates of research misconduct, although both are still underestimates. Unintentional errors in data analysis account for only 15% of retracted publications (21). Reported unintentional causes for retractions include improper reporting of cases, inadvertent exclusion of outliers leading to misinterpretations, and unrecognized biases that came to light at the end of the study or after publication.

Detecting and avoiding statistical untruths

There are multiple resources to assist in avoiding

Table 2 Statistical signs indicating potential research misconduct

| |
|--|
| Employing incorrect statistical test |
| Oversimplification of analyses |
| Exclusion of data |
| Exploratory analyses |
| Multiple tests performed but few statistically significant |
| Pattern of effect size inconsistent |
| P values not adjusted for multiple comparisons |
| Low statistical power/high type II error |
| Incorrect identification of the study design |
| Did not follow an a priori analysis plan |
| Performing only one-sided tests for statistical significance without justification |

inappropriate statistical procedures and presentation. One of which is the “Statistical Analyses and Methods in the Published Literature”, also known as the SAMPL guidelines which assist researchers in the proper manner of reporting various statistical methods (27). Another valuable resource is the “Strengthening Analytical Thinking for Observational Studies”, or the STRATOS, which focuses on the guidance of execution and interpretation of results (28). Lastly, the Committee on Publication Ethics (COPE) was established in 1997 to provide guidance regarding publication ethics. COPE has a series of tenets and guidelines, which are based on the principles of honesty and accuracy (29). Honesty, in the form of transparency, disclosure, and critical self-appraisal in surgical research, prohibits intentional misrepresentations. Accuracy prevents unintentional statistical misrepresentations. Inaccuracy demonstrates methodological, analytical, or interpretational understanding or attention to detail.

Clinical research requires self-policing and holding peers to high rigorous standards in order to maintain credibility (30). Select studies identified actions taken against research misconduct. In one study, 46% of the 78 non-self-reported cases of research misconduct had some action taken (31). In another study of non-self-reported cases of research misconduct, more than half of the individuals reporting non-self-misconduct, also reported confronting the individual committing the misconduct. Furthermore, many also reported the misconduct to the supervisor (36.4%) or the Institutional Review Board (IRB) (12.1%) (32).

Identification of research misconduct often comes

from many sources including: research peers, reviewers, IRB auditors, and even study participants. One study reported that nearly one quarter (24.4%) of the 115 cases of misconduct were reported by a study participant (30). There are also some common statistical signs which may indicate potential research misconduct (4,8,17,30). These are outlined in *Table 2*.

Choice of statistical test

Selecting the correct type of statistical test to use in the analytical plan is chosen by the type of data and the study question being addressed (33). Study questions generally come in two different forms, those assessing differences between groups and those assessing similarities/equality between groups. For example, if the study question is to evaluate if there is a statistical difference between two surgical approaches then there are a select number of statistical tests that may be employed to assess for a difference between the surgical approaches. However, if the study question aims to demonstrate that two surgical approaches are equivalent then a different set of statistical tests should be employed, dependent upon the type of data being analyzed (11). Proper identification of the correct study question and data type are needed before making an analysis plan in order to minimize significant mistakes and potential misinterpretation of study results. The results of employing the incorrect statistical test could become a minor limitation or, worst case, completing invalidating the study results (34).

Incorrect statistical evaluations often fall into one of two categories. The most common category is testing to evaluate if there is a difference between two groups (35) and the data are assumed to be normally distributed, when in fact they are not. Tests for differences may be either one-sided test, that is assessing if there is a difference between groups in a single direction (only improved, only higher, etc.), or a two-sided test which analyze for statistical differences in either direction (improved or worse, higher or lower, etc.). One-sided tests are less conservative than two-sided tests and should be well justified and outlined in an analysis plan prior to conducting the study. The other type of statistical tests are equivalence and non-inferiority tests, which assess if two groups or interventions are statistically equivalent (36,37). These tests are not stating that two groups are simply not statistically different (37). Moreover, they are used to evaluate if two things are statistically interchangeable and are often used to evaluate new assessment methods, surgical approaches, or treatments (38). An example of a non-inferiority test designed to assess if a new

surgical approach has the positive outcomes as the current surgical approach. The application of testing for differences as opposed to testing for similarities may be inappropriately applied, resulting in false conclusions.

Identification of the correct type of data (numeric or categorical) and distribution (normal or not normal) dictate the specific tests to use (35), and is often oversimplified in surgical trials. Randomized trials habitually used paired sample t-tests if the same patient was assessed both before and after the intervention, or independent sample t-tests in traditional experimental *vs.* control trial. However, t-tests assume normal distribution of the data, and biomedical data are not usually normally distributed. When data are not normally distributed, use of non-parametric tests (e.g., Chi-Squared test, Mann-Whitney U Test, Wilcoxon Sign Rank Test, Wilcoxon Rank Sum Test) should be employed (39). Inappropriate use of these parametric tests instead of their non-parametric equivalent has likely resulted in false results. Assumption that all data are normally distributed are an indication of either carelessness, ignorance, or incompetence (11) and may be a symptom of a larger, more pervasive research misconduct. Many prior studies outline the appropriate analysis type given the data type and study design (5,11,13,27,39-41).

Oversimplification of analyses

Statistical complexity in clinical surgery research is growing (1); however, simple and basic statistical tests and models are still commonly employed in analysis of complex data. An evaluation of 240 surgical peer-reviewed publications found that meaningful proportions of studies had one or more signs of potential research misconduct. Of these publications, 60% used rudimentary parametric statistics with no test for normality reported, 21% did not report a measure of central tendency (mean, median or mode) for the primary measures, and 10% did not identify the type of statistical test was used to calculate a P value (10).

Exclusion of data and treatment of outliers

Excluding data should rarely occur. Furthermore, it is only justifiable when there are documented errors in the data collection process. The best situation is proper planning of possible data collection errors are considered before collecting data and written into an analysis plan (12,20) Protocol failure, testing error, lab error, or equipment failure are unfortunately common in research and are all reasonable reasons to exclude data, given that they are documented

sources of erroneous data (11). Possible circumstances which may result in incorrect and inaccurate individual datum should be considered prior to data collection, and should either lessened through a study protocol (e.g., standardizing physicians or calibrating equipment before every assessment) or memorialized in a plan to identify and handle these issues. One circumstance in which data exclusion is often done without explanation in the publication is outliers. It is wise to have a set definition for what constitutes an outlier (i.e., 3 standard deviations, or 1.5 IQR) before analysis and clearly include this information in the methods section. Without considering word count restrictions or giving too much information it is best to be as transparent as possible by including analyses of data with outliers and without. Analyses plans are becoming more commonly available through secondary sources such as clinicaltrials.gov or requested by journal editors prior to publishing an article. Appropriate explanation for excluding each datum should result from a documented protocol deviation or lab error, not merely data points which are “beyond what is expected” (e.g., 2 standard deviations above the mean). A study looking at research misconduct concluded more than a third (33.7%) of surveyed researchers admitted to poor research methodologies indicative of research misconduct, including exclusion of datum or multiple data due of a “gut feeling that they were inaccurate” and misleading or selective reporting of study design, data or results (20).

Detection of fraud

There are three distinct types of fraud detection. These include the oversight by trial committees, on site monitoring and central statistical monitoring (42). Each type is important for detecting different aspects of fraud. The oversight by trial committee members is best for preventing flaws in the study design as well as the interpretation of results (42). On-site monitoring is useful for ensuring that no procedural errors occur during data collection at participating centers (42). Lastly, the statistical monitoring is essential for eliminating data errors as well as incidence of faulty equipment or sloppiness. Unfortunately, most studies only implement one fraud detection type, the on-site monitoring (43). While this type of fraud detection is useful for ensuring that data collection is being done efficiently and correctly (i.e., ensuring that all participants are giving consent, etc.) Unfortunately, this type of fraud detection does not ensure source data verification without being costly and timely (44-46). However, source data

verification by central statistical monitoring is necessary for making sure the data collected is trustworthy in answering your central question, detecting fraud, and any abnormal patterns, especially in multicenter trails (17,47,48). In some circumstances, mainly to reduce cost of constant source data verification, targeted data audits may be recommended in detecting fraud in addition to promoting other self-policing activities (44-46,48).

Principles of statistical monitoring

Statistical Monitoring is made possible by the structured nature of clinical data. In regard to multicenter trails often data is collected utilizing similar collection forms, making it easy to run analyses evaluating any abnormalities in all variables or specific variables of one center compared to others (49-52). Randomized trails may also take advantage of the structured nature of variables for statistical monitoring. In theory, baseline variables should not differ significantly between all randomized groups. Dates of visits by participants can also be monitored for abnormalities, including a suspiciously high number of weekend visits (42,48,53). Another fraudulent act that researchers do is the copy and pasting of data from one participant to another. A test for similarity can be utilized to ensure honesty in data reporting. Conversely, some researchers take a lengthier approach to fabricating data by making up values for missing variables. However, humans are not good at making completely random numbers, using Benford's Law on the distribution of first digits investigators can see if researchers have invented numbers (54).

Conclusions

Research misconduct may originate from many sources. It is often difficult to detect and little is known regarding the prevalence or underlying causes of research misconduct among biomedical researchers. Prevalence estimates of misconduct are likely underestimates, ranging from 0.3% to 4.9%. There are signs which may indicate research misconduct and continued self-policing by biomedical researchers is needed. Additionally, there are some best practices, including development and dissemination of an analysis plan prior to collecting data, which may minimize the opportunity for research misconduct.

Acknowledgements

Funding: This study has been funded, in part, by grants

from the National Institute for Occupational Safety and Health (NIOSH/CDC) Education and Research Center training grant T42/CCT810426-10.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Kurichi JE, Sonnad SS. Statistical methods in the surgical literature. *J Am Coll Surg* 2006;202:476-84.
2. Harris JS, Sinnott PL, Holland JP, et al. Methodology to update the practice recommendations in the American College of Occupational and Environmental Medicine's Occupational Medicine Practice Guidelines, second edition. *J Occup Environ Med* 2008;50:282-95.
3. Melhorn JM HK. Methodology. In: Melhorn JM, Ackerman WE. editor. *Guides to the Evaluation of Disease and Injury Causation*. AMA Press; 2008.
4. Strasak AM, Zaman Q, Pfeiffer KP, et al. Statistical errors in medical research--a review of common pitfalls. *Swiss Med Wkly* 2007;137:44-9.
5. Feinstein AR. Clinical biostatistics. XXV. A survey of the statistical procedures in general medical journals. *Clin Pharmacol Ther* 1974;15:97-107.
6. Williams JL, Hathaway CA, Kloster KL, et al. Low power, type II errors, and other statistical problems in recent cardiovascular research. *Am J Physiol* 1997;273:H487-93.
7. Thiese MS, Arnold ZC, Walker SD. The misuse and abuse of statistics in biomedical research. *Biochem Med (Zagreb)* 2015;25:5-11.
8. Cassidy LD. Basic concepts of statistical analysis for surgical research. *J Surg Res* 2005;128:199-206.
9. Emerson JD, Colditz GA. Use of statistical analysis in the New England Journal of Medicine. *N Engl J Med* 1983;309:709-13.
10. Oliver D, Hall JC. Original Articles: Usage of statistics in the surgical literature and the 'Orphan P' Phenomenon. *Aust N Z J Surg* 1989;59:449-51.
11. Greenland S, Senn S, Rothman K, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol* 2016;31:337-50.
12. Hall JC, Hill D, Watts JM. Misuse of statistical methods in the Australasian surgical literature. *Aust N Z J Surg* 1982;52:541-3.
13. Parsons NR, Price CL, Hiskens R, et al. An evaluation of

- the quality of statistical design and analysis of published medical research: results from a systematic survey of general orthopaedic journals. *BMC Med Res Methodol* 2012;12:60.
14. Gore SM, Jones IG, Rytter EC. Misuse of statistical methods: critical assessment of articles in *BMJ* from January to March 1976. *Br Med J* 1977;1:85-7.
 15. Kim JS, Kim DK, Hong SJ. Assessment of errors and misused statistics in dental research. *Int Dent J* 2011;61:163-7.
 16. White SJ. Statistical errors in papers in the *British Journal of Psychiatry*. *Br J Psychiatry* 1979;135:336-42.
 17. George SL, Buyse M. Data fraud in clinical trials. *Clin Investig (Lond)* 2015;5:161-73.
 18. Myers PO. Open data: can it prevent research fraud, promote reproducibility, and enable big data analytics in clinical research? *Ann Thorac Surg* 2015;100:1539-40.
 19. Smith R. Statutory regulation needed to expose and stop medical fraud. *BMJ* 2016;352:i293.
 20. Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One* 2009;4:e5738.
 21. Moylan EC, Kowalczyk MK. Why articles are retracted: a retrospective cross-sectional study of retraction notices at *BioMed Central*. *BMJ Open* 2016;6:e012047.
 22. Budd JM, Sievert M, Schultz TR, et al. Effects of article retraction on citation and practice in medicine. *Bull Med Libr Assoc* 1999;87:437-43.
 23. Yan J, MacDonald A, Baisi LP, et al. Retractions in orthopaedic research. *Bone Joint Res* 2016;5:263-8.
 24. Bean JR. Truth or consequences: the growing trend of publication retraction. *World Neurosurg* 2017;103:917-8.
 25. Van Noorden R. The trouble with retractions. *Nature* 2011;478:26.
 26. Wang J, Ku JC, Alotaibi NM, et al. Retraction of neurosurgical publications: a systematic review. *World Neurosurg* 2017;103:809-14.e1.
 27. Lang TA, Altman DG. Basic statistical reporting for articles published in *Biomedical Journals: The "Statistical Analyses and Methods in the Published Literature" or The SAMPL Guidelines*. In: Smart P, Maisonneuve H, Polderman A, eds. *Science Editors' Handbook*, European Association of Science Editors, 2013.
 28. Sauerbrei W, Abrahamowicz M, Altman DG, et al. STRENGTHENING analytical thinking for observational studies: the STRATOS initiative. *Stat Med* 2014;33:5413-32.
 29. Wager E, Barbour V, Yentis S, et al. Retractions: Guidance from the committee on publication ethics (COPE). 2009;50:532-5.
 30. Titus SL, Wells JA, Rhoades LJ. Repairing research integrity. *Nature* 2008;453:980-2.
 31. Tangney JP. Fraud will out--or will it? *New Scientist* 1987;115:62-3.
 32. Kattenbraker MS. Health education research and publication: Ethical considerations and the response of health educators. Southern Illinois University at Carbondale, 2007.
 33. Hulley SB, Cummings SR, Browner WS, et al. *Designing clinical research*. Lippincott Williams & Wilkins, 2013.
 34. Jamart J. Statistical tests in medical research. *Acta Oncol* 1992;31:723-7.
 35. Woolson RF, Clarke WR. *Statistical methods for the analysis of biomedical data*. John Wiley & Sons, 2011.
 36. Altman DG, Bland JM. *Measurement in medicine: the analysis of method comparison studies*. *The statistician* 1983;32:307-17.
 37. Wellek S. *Testing statistical hypotheses of equivalence and noninferiority*. CRC Press; 2010.
 38. D'Agostino RB, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues--the encounters of academic consultants in statistics. *Stat Med* 2003;22:169-86.
 39. Hollander M, Wolfe DA, Chicken E. *Nonparametric statistical methods*. John Wiley & Sons, 2013.
 40. Gore A, Kadam Y, Chavan P, et al. Application of biostatistics in research by teaching faculty and final-year postgraduate students in colleges of modern medicine: a cross-sectional study. *Int J Appl Basic Med Res* 2012;2:11-6.
 41. Little RJ, Rubin DB. *Statistical analysis with missing data*. John Wiley & Sons; 2014.
 42. Baigent C, Harrell FE, Buyse M, et al. Ensuring trial validity by data quality assurance and diversification of monitoring methods. *Clinical trials* 2008;5:49-55.
 43. Morrison BW, Cochran CJ, White JG, et al. Monitoring the quality of conduct of clinical trials: a survey of current practices. *Clinical Trials* 2011;8:342-9.
 44. Eisenstein EL, Collins R, Cracknell BS, et al. Sensible approaches for reducing clinical trial costs. *Clinical Trials* 2008;5:75-84.
 45. Christina Reith MB C, Martin Landray MB C, Granger CB, et al. Randomized clinical trials--removing unnecessary obstacles. *N Engl J Med* 2013;369:1061-5.
 46. Grimes DA, Hubacher D, Nanda K, et al. The good clinical practice guideline: a bronze standard for clinical research. *Lancet* 2005;366:172-4.
 47. Food and Drug Administration. *Guidance for industry: oversight of clinical investigations--a risk-based approach*

- to monitoring. Silver Spring, MD: FDA 2013.
48. Buyse M, George SL, Evans S, et al. The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials. *Stat Med* 1999;18:3435-51.
 49. Edwards P, Shakur H, Barnetson L, et al. Central and statistical data monitoring in the Clinical Randomisation of an Antifibrinolytic in Significant Haemorrhage (CRASH-2) trial. *Clinical Trials* 2014;11:336-43.
 50. Kirkwood AA, Cox T, Hackshaw A. Application of methods for central statistical monitoring in clinical trials. *Clinical Trials* 2013;10:783-806.
 51. Lindblad AS, Manukyan Z, Purohit-Sheth T, et al. Central site monitoring: results from a test of accuracy in identifying trials and sites failing Food and Drug Administration inspection. *Clinical Trials* 2014;11:205-17.
 52. Venet D, Doffagne E, Burzykowski T, et al. A statistical approach to central monitoring of data quality in clinical trials. *Clinical Trials* 2012;9:705-13.
 53. Buyse M, Evans SJ. *Fraud in clinical trials*. Wiley statsref: statistics reference online, 2005.
 54. Al-Marzouki S, Evans S, Marshall T, et al. Are these data real? Statistical methods for the detection of data fabrication in clinical trials. *BMJ* 2005;331:267-70.

Cite this article as: Thiese MS, Walker S, Lindsey J. Truths, lies, and statistics. *J Thorac Dis* 2017;9(10):4117-4124. doi:10.21037/jtd.2017.09.24