

Principles for the ethical analysis of clinical and translational research

Jonathan A. L. Gelfond,^{a*†} Elizabeth Heitman,^b Brad H. Pollock^a and Craig M. Klugman^c

Statistical analysis is a cornerstone of the scientific method and evidence-based medicine, and statisticians serve an increasingly important role in clinical and translational research by providing objective evidence concerning the risks and benefits of novel therapeutics. Researchers rely on statistics and informatics as never before to generate and test hypotheses and to discover patterns of disease hidden within overwhelming amounts of data. Too often, clinicians and biomedical scientists are not adequately proficient in statistics to analyze data or interpret results, and statistical expertise may not be properly incorporated within the research process. We argue for the ethical imperative of statistical standards, and we present ten nontechnical principles that form a conceptual framework for the ethical application of statistics in clinical and translational research. These principles are drawn from the literature on the ethics of data analysis and the American Statistical Association Ethical Guidelines for Statistical Practice. Copyright © 2011 John Wiley & Sons, Ltd.

Keywords: ethics; translational research; data analysis

1. Introduction

Translational research ‘translates’ scientific discoveries from basic science (T1—‘bench to bedside’) or clinical studies (T2) into implementation in health-care practice [1], thereby magnifying the effects of research on the populace. Statistics plays a critical role in translation by helping to identify new interventions that benefit public health. Statistical analysis also safeguards the public from harm and acts as a sentinel of emerging health threats. For example, analyses of Phase IV studies instigated the controversy over rofecoxib, estimated to have caused tens of thousands of serious adverse events [2–4]. Hence, researchers have an obligation to perform and interpret analyses properly, although numerous studies suggest that analytical errors and deficiencies are prevalent and poorly understood [5–9].

Biased or otherwise faulty analyses of clinical and translational research are particularly harmful for several reasons [10]. First, the results of incorrect analyses can directly harm study participants and, later, the public at large. Although research subjects are protected by institutional oversight prescribed by federal regulations (e.g., Title 45 of the US Code of Federal Regulations Part 46), there are no formal legal guidelines for potentially high-impact analyses (e.g., meta-analyses) that do not affect research subjects directly [11]. Second, human subjects are always placed at some risk in research, and analytical errors waste the voluntary assumption of that risk and reduce the possibility of social benefit. Third, economic resources are squandered when analyses are invalid or uninterpretable. Fourth, erroneous results misdirect future research and impede the progress of biomedical science [12]. Fifth, lapses in scientific integrity can promote public distrust of science and medicine [13] that potentially hinders participation in clinical trials and patient adherence. Sixth, the misapplication of statistical methods harms the reputation of the field of statistics and devalues the profession [14, 15].

^aDepartment of Epidemiology and Biostatistics, UT Health Science Center, San Antonio, TX 78229, USA

^bCenter for Biomedical Ethics and Society, Vanderbilt University Medical Center, Nashville, TN 37203, USA

^cCenter for Medical Humanities & Ethics, UT Health Science Center, San Antonio, TX 78229, USA

*Correspondence to: Jonathan A. L. Gelfond, Department of Epidemiology and Biostatistics, UT Health Science Center, San Antonio, TX 78229, USA.

†E-mail: gelfondjal@uthscsa.edu

Statistical analysis is an important and complex task that needs substantial expertise. Physicians are among the best statistically trained translational researchers, but even they may be ill equipped to meet the demands of ever more complicated methodologies [6, 7, 16–23]. Most clinical research training programs offer only abbreviated instruction in statistical methods. Although these programs impart useful knowledge and skills, students who receive insufficient statistical training are at risk of not satisfying current standards for analysis.

The need for statistical expertise is currently underappreciated. In 2009, the Education and Career Development Key Function Committee of the national Clinical and Translational Science Award (CTSA) consortium, together with the National Center for Research Resources at the National Institutes of Health, defined a set of core competencies for masters-level translational researchers [24]. One such competency is that trainees be able to ‘differentiate between the analytic problems that can be addressed with standard methods and those requiring input from biostatisticians and other scientific experts.’ This statement implies that statistical consultation is required only in ‘nonstandard’ situations. However, the ‘standard methods’ of statistics in clinical studies are now quite complex [25]. Since translation also thrusts basic scientists into the clinical realm, they too may be faced with unfamiliar statistical and epidemiological challenges that lead them to invalid conclusions [26].

Consequently, there is an ethical obligation to seek such expertise in even seemingly mundane circumstances. Mistakes made by untrained researchers constitute a type of neglect with a potential for serious negative effects on public health. As a parallel, a researcher who performs a ‘standard’ surgery without proper surgical training is acting unethically, regardless of good intentions. Beyond failing to meet the technical standards of practice, this non-surgeon researcher risks violating the ethical precepts of surgical practice, such as informed consent and patient privacy. Likewise, inexperienced statistical analyses are unethical because of the potential damage to human health, disruption of the scientific process, and violations of established ethical codes.

2. Background

The literature on common statistical and analytical errors shows that most problems are clustered around omissions, simple mistakes, incompetence, and ethical lapses [5, 27–31]. Most of these errors can be understood as violations of a few basic principles. Haynes *et al.* [32] have found that surgical teams at all levels of competence, facilities, skill, and experience substantially improved surgical outcomes by assuring the completion of 19 important tasks in every surgery. In the spirit of their surgical safety checklist [32], we present a list of key principles as a checklist for ethical integrity in statistics for translational researchers. Although our principles are not discrete actions, we propose that considering and adhering to them will improve the ethical integrity of statistical analysis, its interpretation, and thereby, the quality of research. Harris [33] argues that such checklists may be useful in preventing mistakes but cautions that they cannot serve as axioms from which all else is derived.

We want to establish our list as guideposts for avoiding ethical lapses and analytical errors. The distinction between what constitutes a simple mistake and negligence or an ethical breach is important. An ethical violation occurs when one violates a code of conduct, rule, or formal obligation. Non-statistician investigators need to be aware of the ethical issues in statistics because they often have executive control over the research process. Although unfamiliar to non-statisticians, formal ethical guidelines for statistical practice have been articulated by the American Statistical Association (ASA) [34, 35]. Whereas the ASA guidelines on ethics concern general research topics, our discussion focuses on ethical issues specific to the conduct of statistical analysis applied to clinical and translational research, and we address frequently cited and important failings.

Other notions of ethical conduct in statistical analysis derive from standards of practice. Standards of practice are the procedures and due diligence that would be commonly acceptable by statisticians who practice in the relevant domain. We define substandard practice as a failure to use the accepted level of care in practice either through intention or neglect. A good faith effort by researchers may not sufficiently fulfill the standards of practice, nor does the process of peer review guarantee its fulfillment.

3. Fundamental Principles

3.1. *Statistics is a profession with its own standard and code of ethics*

It is important that all statistical practitioners recognize their potential impact on the broader society and the attendant ethical obligations to perform their work responsibly. ASA Guidelines [35]

All practitioners, including novices, should strive to follow the ethical codes of their field. The ethical guidelines of the ASA are organized around several key topics:

- A. Professionalism
- B. Responsibilities to funders, clients, and employers
- C. Responsibilities for publication and testimony
- D. Responsibilities to research subjects
- E. Responsibilities to research team colleagues
- F. Responsibilities to other statistical practitioners
- G. Responsibilities regarding allegations of misconduct
- H. Responsibilities of employers

Below, we label the specific considerations within each of the major topics (A–H) as [Topic Letter. Consideration #] to highlight the applicable link to the ASA ethical guidelines.

3.2. *Clinical and translational research teams require multidisciplinary expertise (statistical and scientific) [A.1, A.4]*

The experiment should be conducted only by scientifically qualified persons. The highest degree of skill and care should be required through all stages of the experiment of those who conduct or engage in the experiment. The Nuremberg Code [36]

Non-statisticians who conduct statistical analyses must have more than simply an instrumental understanding of how to execute the statistical methodology; rather, analysts must understand the statistical foundations and assumptions of methods as well as the science being considered [F.6]. Formal education is not always required for expertise, but statistics and bioinformatics are vast fields with various subspecialties so that training in general statistics may not provide sufficient expertise for specialized applications. The need to formalize the concept of expertise is seen in the ASA's decision to develop accreditation standards for individual statisticians [37], including: (1) statistical training and knowledge; (2) proven experience and competence; (3) continual professional development; (4) agreeing to ethical standards; and (5) effective communication.

The analyst's *authenticity* is critical [29]; that is, the analyst must understand and have confidence in the correctness and the validity of results. Analysts should have a broader understanding of the scientific context in which the data arise as well as an understanding of the approach of the research team so that unreasonable values or otherwise suspicious data can be detected. As a whole, research teams should comprise collaborating individuals with both statistical and scientific expertise. These teams should exercise due skepticism and scrutiny of both data and analytical methodology. For example, after a proteomic marker was reported to have 100% sensitivity and 100% specificity for detecting ovarian cancer [38], experienced statisticians recognized that such extraordinary claims required close scrutiny, and they later revealed substantial bias [39]. This example is not isolated, and deficits in experimental design and misapplication of epidemiological principles and statistical techniques have plagued applications of new technologies in translational research [26].

3.3. *Objectivity [H.1]*

Objectivity cannot be equated with mental blankness; rather, objectivity resides in recognizing your preferences and then subjecting them to especially harsh scrutiny. Stephen Jay Gould [40]

Analyses should not be conducted to unduly favor one set of hypotheses. *Equipose* or uncertainty about the best clinical action is often considered to be required to ethically justify a randomized trial [41], and analogous analytical objectivity is required to make unbiased inferences. That is, the data analysis plan should respect the uncertainty within the scientific community, and the analyst should be an advocate for the data and the scientific process, not for a particular result. Merely using a statistically unbiased estimation procedure does not establish the objectivity of the results [42]. Subjective decisions such as Type I and Type II error rates or Bayesian priors should be made before acquiring the data and reported alongside conflicts of interest.

Experimental design or analysis may reflect the sponsor's motivation [43, 44], and most of the publications referenced here exemplify problematic behavior that reflect bias in the researchers' beliefs. This

is not to say that these authors had biased intentions, rather, that it is arguably human nature to be less skeptical of anticipated findings. Nonetheless, confirmation bias (i.e., searching for the preconceived result in the face of findings to the contrary) should be avoided.

There are innumerable ways that analytical objectivity can be corrupted, including multiple tests, inaccurate computation, incomplete disclosure of data, and misinterpretation to name a few. Statisticians are not immune to the scientific biases of their close collaborators, and ideally, investigators may improve objectivity by working with an independent or blinded statistician.

3.4. *Openness and transparency while nevertheless maintaining research participant privacy*

They [scientists] have held free and open communication to be the most important requirement for their work. Sissela Bok [45]

All relevant data [C.5] and analyses [E.4] must be presented so that readers and reviewers may fairly evaluate the quality of the analysis. Openness facilitates the detection of data dredging and other potential biases [28,29], thereby mitigating their negative impacts. The disclosed details of the analysis should be sufficient to allow reproduction. In a recent effort to improve openness and reproducibility, journals such as *Annals of Internal Medicine* now encourage the sharing of both data and statistical code [46]. The concept of reproducibility is further discussed with respect to the accuracy principle in Section 3.6.

The VIGOR (Vioxx Gastrointestinal Outcomes Research) study of rofecoxib [47] is an example of a failure of openness and disclosure [48] in reporting cardiovascular events [49] that suggested harmful side effects. The authors of the original study failed to report the total number of cardiovascular events, provoking an ‘expression of concern’ from the journal [50].

3.5. *Verification of assumptions [C.2, C.8]*

The man of science has learned to believe in justification, not by faith, but by verification. Thomas H. Huxley [51]

Statistical assumptions should be verified. Almost every statistical method has verifiable assumptions such as the additivity of effects and Gaussian distributions, and violation of these assumptions could invalidate the results. It is common for readers of research articles to take the verification of assumptions for granted, but verification is often not reported [25]. Nevertheless, it requires substantially more expertise to test statistical assumptions than it does to implement a statistical method [52], and unsubstantiated or unrealistic statistical assumptions are often a source of erroneous results [53–56].

3.6. *Accuracy of the primary data and computation is critical [C.6, C.14]*

Science is simply common sense at its best that is, rigidly accurate in observation, and merciless to fallacy in logic. Thomas H. Huxley [57]

The fabrication or improper imputation of data [C.7] is clearly unethical practice, but so too are errors of neglect. The use of spreadsheets to collect, manage, and analyze data with no audit trail is quite common among inexperienced researchers. Statistical procedures performed with spreadsheets developed for accounting purposes are notoriously inaccurate compared with dedicated statistical software [58]. A recent clinical genomics study had a data manipulation error in MICROSOFT EXCEL (Microsoft, Seattle, WA) [9] that led to the mislabeling of many genes, which in turn made the published interpretation of the results questionable [59].

Computational accuracy refers to the numerical precision of statistics computed from a dataset. Whereas simple statistical procedures such as a *t*-test will likely be performed accurately, more computationally intensive techniques such as high-dimensional integration or Bayesian Monte Carlo analyses may require expert diligence [60,61].

Reproducibility of each step of computation from the raw data to the published results [62] is an important component of computational accuracy. Reproducible analysis can have a striking impact on the quality of science in many fields that rely upon statistical analyses [63]. Given the original data and the programs, another researcher should get the *exact same* result as in the publication. This approach has

two advantages. First, automation removes a degree of human error especially if, for any reason, the analysis must be repeated. Second, the programs provide an open and transparent record, reinforcing the discussion in Section 3.4. The concept of analytical reproducibility may not be apparent to trainees when prominent research training texts contain antiquated statistical tables that are relatively clumsy and error prone, but reproducible results from computer programs are becoming the standard for professional biostatisticians [62]. Reproducibility requires programming skills, reinforcing the need for the multidisciplinary expertise mentioned in Section 2.1.

3.7. *Appropriate study design and sample size [D.2]*

The importance of this [design] stage cannot be overemphasized since no amount of clever analysis later will be able to compensate for major design flaws. Doug Altman [64]

Sample size estimation should be used to minimize the risk to study subjects while maintaining scientific value, and data monitoring plans should be employed to protect participants throughout the course of the study. The need for power calculations is seen in point #2 of the Nuremberg Code: ‘The experiment should be such as to yield fruitful results for the good of society’ [36]. Underpowered studies are not likely to yield results with practical translational value; they put subjects at unnecessary risk and waste resources. Likewise, studies that have too many subjects incur unnecessary risk and are also wasteful.

There is more to good study design than the determination of sample size. Ransohoff [26] points out that translational studies have been prone to biased designs for identifying molecular markers for cancer. Many such biases have resulted from observational study designs in which the cases and controls differ by more than their disease status. For example, the case samples might have been stored for years whereas the control samples were not. Any systematic differences between the groups may not be attributed solely to cancer, and such non-biological, systematic differences may easily corrupt the data to the extent that statistical adjustment cannot remove the inherent bias.

Research participants should not be put at risk if the statistical evidence already produced by the study is conclusive [65]. Data monitoring plans are supported by the Nuremberg Code’s point # 10: ‘The scientist in charge must be prepared to terminate the experiment at any stage, if he has probable cause to believe, in the exercise of the good faith, superior skill, and careful judgment required of him, that a continuation of the experiment is likely to result in injury, disability, or death.’ The TGN1412 trial is an example of a poorly executed study design [66]. The compound had never been tested in humans, but it first was administered to six healthy males who all immediately (<3 h) became seriously ill. Clearly, the implemented design should have halted the trial before six consecutive bad outcomes occurred. However, there was insufficient time allotted between dosing additional subjects to collect highly informative data.

3.8. *Parsimonious model selection weighed against bias, precision, and validity*

Goodness [of a model] is often defined in terms of prediction accuracy, but parsimony is another important criterion: simpler models are preferred for the sake of scientific insight into the x - y relationship. Efron *et al.* [67]

Model selection is a critical and often difficult decision that goes beyond issues of validity. The choice between a complex model versus simple model (with few components) or an elegant model (mathematically attractive) must be justified, and complexity at the cost of interpretation or stability of estimation should be avoided [B.1, C.9]. For example, higher complexity is justified when it reduces bias or variance. However, if the sample size is too small relative to the number of parameters, then a simpler model may be preferable. A simple model is not preferable if it is invalid, biased or lacks statistical efficiency or power [E.5]. There are cases when simple models are invalid, and although the complex model is valid, the dataset has insufficient information for stable estimation. In these cases, the limitations of statistical analyses should be acknowledged.

A good example of how to balance complexity and bias is provided by Johnson *et al.* [68], who estimated the effects of topiramate on alcohol dependence. Some patients did not return for visits, possibly due to relapse, resulting in missing data. The authors used a relatively simple imputation method and also had an independent statistician conduct an additional, more complex analysis. By presenting the agreement between the two models, the authors demonstrated robustness and accurate estimation.

3.9. Interpretable quantification of evidence

This only is certain, that there is nothing certain. Pliny [the Elder] [69]

If a parameter is estimated, then its uncertainty must be considered and quantified through interpretable measures like p -values, confidence intervals, or posterior probabilities. For example, the predictive accuracy of new genomic prognostic models is often reported with invalid confidence intervals [70]. The study by Conrads *et al.* [38] mentioned in an earlier paragraph did not provide confidence intervals alongside its claims of 100% sensitivity and 100% specificity, resulting in a positively biased estimate that defies proper interpretation.

The multiplicity of hypotheses underlies many issues in reporting and interpreting analyses [71], and this challenge is increasingly difficult and frequent with genomic studies [72]. Multiple tests (e.g., many variables or many groups compared) must be correctly described and adjusted for in presenting evidence [73]; otherwise, the analysis is biased towards false discovery. In the ISIS-2 randomized trial for the treatment of myocardial infarction [74], the authors demonstrated how fallacies may arise from multiple testing in arbitrary subgroups: They found that patients born under the astrological signs of Gemini and Libra did not appear to benefit from aspirin, unlike participants born under other astrological signs.

3.10. Avoidance of misinterpretation

When distant and unfamiliar and complex things are communicated to great masses of people, the truth suffers a considerable and often a radical distortion. The complex is made over into the simple, the hypothetical into the dogmatic. Walter Lippman [75]

Claims not supported by the strength of evidence or unreasonable extrapolations are potentially harmful. In order to be scientifically valid, the results of a statistical method need proper interpretation [C.12, C.15]. For example, an article in *Nature* [76] reported that women will overtake men in the 100-m sprint based upon a dubious linear extrapolation [77]. Whenever possible, association should be distinguished from causation [A.5] [73]. The important feature of interpretation is that it occurs *after* the quantitative analyses, and interpretation is performed by not only a statistician but also by collaborating scientists, reviewers, practitioners, and journalists. Each individual who constructs an interpretation of the statistical results has a responsibility to fellow researchers, patients, and the public to be precise, measured, and correct.

A recent study of published randomized controlled trials [8] found that researchers used various ‘spin strategies’ in interpreting the data when the primary efficacy outcome yielded results that were not statistically significant. The biased reporting took various forms: (1) focusing on significant results for subgroups or secondary endpoints; (2) claiming equivalence of therapies without establishing statistical equivalence; and (3) focusing on effectiveness of therapies instead of the failed attempt at demonstrating comparative superiority.

4. Conclusion

The ethics of biomedical research extends beyond the treatment of research subjects as discoveries are ‘translated’ into practice that affects millions of patients. Statistical analyses play a pivotal gatekeeper role, tipping the scales of evidence away from equipoise, but analyses are vulnerable to negligent or intentionally substandard practice. Through our list of principles, we aim to raise awareness of analytical challenges faced by clinical and translational researchers, and to demonstrate the ethical imperative to incorporate statistical expertise in data analysis. The inclusion of essential statistical ethics in the knowledge base of researchers should enhance their statistical skills and the ultimate value of their work.

Acknowledgements

The authors would like to thank the editor, two anonymous referees, Keith Baggerly, Joel Michalek, and Christopher Loudon whose insights have greatly improved the manuscript. J.A.L.G. was supported by CTSA Award KL2 RR025766 from the National Center for Research Resources. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Woolf SH. The meaning of translational research and why it matters. *JAMA-Journal of the American Medical Association* 2008; **299**:211–213.
2. Bottone FG, Barry WT. Postmarketing surveillance of serious adverse events associated with the use of rofecoxib from 1999–2002. *Current Medical Research and Opinion* 2009; **25**:1535–1550.
3. Sawicki PT, Bender R, Selke GW, Klauber J, Gutschmidt S. Assessment of the number of cardio- and cerebrovascular events due to rofecoxib (Vioxx (R)) in Germany between 2001 and 2004. *Medizinische Klinik* 2006; **101**:191–197.
4. Topol EJ. Failing the public health—Rofecoxib, Merck, and the FDA. *New England Journal of Medicine* 2004; **351**:1707–1709.
5. Strasak AM, Zaman Q, Pfeiffer KP, Gobel G, Ulmer H. Statistical errors in medical research—a review of common pitfalls. *Swiss Medical Weekly* 2007; **137**:44–49.
6. Windish DM, Huot SJ, Green ML. Medicine residents' understanding of the biostatistics and results in the medical literature. *JAMA-Journal of the American Medical Association* 2007; **298**:1010–1022.
7. Darmon M, Tabah A. Commission Jeune SCWS. Basic concepts of biostatistics are poorly understood by intensivists. *Intensive Care Medicine* 2009; **35**:0077.
8. Boutron I, Dutton S, Ravaut P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA-Journal of the American Medical Association* 2010; **303**:2058–2064.
9. Hutson S. Data handling errors spur debate over clinical trial. *Nature Medicine* 2010; **16**:618–618.
10. Altman DG. Statistics and ethics in medical-research .1. Misuse of statistics is unethical. *British Medical Journal* 1980; **281**:1182–1184.
11. Roseman M, Milette K, Bero LA, Coyne JC, Lexchin J, Turner EH, Thombs BD. Reporting of conflicts of interest in meta-analyses of trials of pharmacological treatments. *JAMA-Journal of the American Medical Association* 2011; **305**:1008–1017.
12. Ioannidis JPA. Why most published research findings are false. *PLoS Medicine* 2005; **2**:696–701.
13. Jacobs AK. Rebuilding an enduring trust in medicine—a global mandate—Presidential Address American Heart Association Scientific Sessions 2004. *Circulation* 2005; **111**:3494–3498.
14. Pantula SG, Teugels J, Stefanski L. Feedback: odds are, it's wrong. *AMSTAT News* 2010:25.
15. Siegfried T. Odds are, it's wrong. *Science News* 2010; **177**:26.
16. Estellat C, Faisy C, Colombet I, Chatellier G, Burnand B, Durieux P. French academic physicians had a poor knowledge of terms used in clinical epidemiology. *Journal of Clinical Epidemiology* 2006; **59**:1009–1014.
17. Ferrill MJ, Norton LL, Blalock SJ. Determining the statistical knowledge of pharmacy practitioners: a survey and review of the literature. *American Journal of Pharmaceutical Education* 1999; **63**:371–376.
18. Hellem MA, Gurka MJ, Hayden GF. Statistical literacy for readers of *Pediatrics*: a moving target. *Pediatrics* 2007; **119**:1083–1088.
19. Herman A, Notzerl N, Libman Z, Braunstein R, Steinberg DM. Statistical education for medical students—concepts are what remain when the details are forgotten. *Statistics in Medicine* 2007; **26**:4344–4351.
20. Leyland AH, Pritchard CW. What do doctors know of statistics? *Lancet* 1991; **337**:679–679.
21. Scheutz F, Andersen B, Wulff HR. What do dentists know about statistics? *Scandinavian Journal of Dental Research* 1988; **96**:281–287.
22. Swift L, Miles S, Price GM, Shepstone L, Leinster SJ. Do doctors need statistics? Doctors' use of and attitudes to probability and statistics. *Statistics in Medicine* 2009; **28**:1969–1981.
23. Wulff HR, Andersen B, Brandenhoff P, Guttler F. What do doctors know about statistics? *Statistics in Medicine* 1987; **6**:3–10.
24. CTSA Education Core Competency Work Group. *Core competencies in clinical and translational research*, 2009. Available at <http://www.ctsaweb.org/index.cfm?fuseaction=home.showCoreComp> (accessed 1 October 2010).
25. Strasak AM, Zaman Q, Marinell G, Pfeiffer KP, Ulmer H. The use of statistics in medical research: a comparison of *The New England Journal of Medicine* and *Nature Medicine*. *American Statistician* 2007; **61**:47–55.
26. Ransohoff DF. Opinion—Bias as a threat to the validity of cancer molecular-marker research. *Nature Reviews Cancer* 2005; **5**:142–149.
27. Altman D. Statistics and ethics in medical research. *British Medical Journal* 1981; **282**:990–990.
28. DeMets DL. Statistics and ethics in medical research. *Science and Engineering Ethics* 1999; **5**:97–117.
29. Kromrey JD. Ethics and data analysis. *Educational Researcher* 1993; **22**:24–27.
30. Marco CA, Larkin GL. Research ethics: ethical issues of data reporting and the quest for authenticity. *Academic Emergency Medicine* 2000; **7**:691–694.
31. Sterba SK. Misconduct in the analysis and reporting of data: bridging methodological and ethical agendas for change. *Ethics & Behavior* 2006; **16**:305–318.
32. Haynes AB, Weiser TG, Berry WR, Lipsitz SR, Breizat AHS, Dellinger EP, Herbosa T, Joseph S, Kibatala PL, Lapitan MCM, Merry AF, Moorthy K, Reznick RK, Taylor B, Gawande AA. Safe Surgery Saves Lives Study G. A surgical safety checklist to reduce morbidity and mortality in a global population. *New England Journal of Medicine* 2009; **360**:491–499.
33. Harris J. In praise of unprincipled ethics. *Journal of Medical Ethics* 2003; **29**:303–306.
34. Ellenberg JH. Ethical guidelines for statistical practice—a historical-perspective. *American Statistician* 1983; **37**:1–4.
35. American Statistical Association Committee on Professional Ethics. *Ethical Guidelines for Statistical Practice*, 1999. Available at <http://www.amstat.org/committees/ethics/index.cfm> (accessed 1 October 2010).
36. Shuster E. The Nuremberg Code: Hippocratic ethics and human rights. *Lancet* 1998; **351**:974–977.
37. Johnstone I. Board approves accreditation guidelines. *AMSTAT News* 2010:10–11.
38. Conrads TP, Fusaro VA, Ross S, Johann D, Rajapakse V, Hitt BA, Steinberg SM, Kohn EC, Fishman DA, Whiteley G, Barrett JC, Liotta LA, Petricoin EF, Veenstra TD. High-resolution serum proteomic features for ovarian cancer detection. *Endocrine-Related Cancer* 2004; **11**:163–178.

39. Baggerly KA, Edmonson SR, Morris JS, Coombes KR. High-resolution serum proteomic patterns for ovarian cancer detection. *Endocrine-Related Cancer* 2004; **11**:583–584.
40. Gould SJ. *The Lying Stones of Marrakech: Penultimate Reflections in Natural History*. Harmony Books: New York, 2000.
41. Freedman B. Equipoise and the ethics of clinical research. *New England Journal of Medicine* 1987; **317**:141–145.
42. Berger JO, Berry DA. Statistical analysis and the illusion of objectivity. *American Scientist* 1988; **76**:159–165.
43. Bhandari M, Busse JW, Jackowski D, Montori VM, Schunemann H, Sprague S, Mears D, Schemitsch EH, Heels-Ansell D, Devereaux PJ. Association between industry funding and statistically significant pro-industry findings in medical and surgical randomized. *Canadian Medical Association Journal* 2004; **170**:477–480.
44. Djulbegovic B, Lacey M, Cantor A, Fields KK, Bennett CL, Adams JR, Kuderer NM, Lyman GH. The uncertainty principle and industry-sponsored research. *Lancet* 2000; **356**:635–638.
45. Bok S. Secrecy and openness in science—ethical considerations. *Science Technology & Human Values* 1982; **7**:32–41.
46. Laine C, Goodman SN, Griswold ME, Sox HC. Reproducible research: moving toward research the public can really trust. *Annals of Internal Medicine* 2007; **146**:450–453.
47. Bombardier C, Laine L, Reicin A, Shapiro D, Burgos-Vargas R, Davis B, Day R, Ferraz MB, Hawkey CJ, Hochberg MC, Kvien TK, Schnitzer TJ, Weaver A, VIGOR Study Group. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. *New England Journal of Medicine* 2000; **343**:1520–1528.
48. Krumholz H, Ross JS, Presler AH, Egilman DS. What have we learnt from Vioxx. *British Medical Journal* 2007; **334**:120–123.
49. Bresalier RS, Baron JA, Trial AP. Adverse cardiovascular effects of rofecoxib—reply. *New England Journal of Medicine* 2006; **355**:204–205.
50. Curfman GD, Morrissey S, Drazen JM. Expression of concern: Bombardier *et al.*, "Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis," *N Engl J Med* 2000;343:1520-8. *New England Journal of Medicine* 2005; **353**:2813–2814.
51. Huxley TH. *On the Advisableness of Improving Natural Knowledge*. London: Chapman & Hall, 1866.
52. Dagostino RB, Belanger A. A suggestion for using powerful and informative tests of normality. *American Statistician* 1990; **44**:316–321.
53. Altman DG. Statistics and ethics in medical-research .5. Analyzing data. *British Medical Journal* 1980; **281**:1473–1475.
54. Goodman NW, Hughes AO. Statistical awareness of research workers in british anesthesia. *British Journal of Anaesthesia* 1992; **68**:321–324.
55. Kanter MH, Taylor JR. Accuracy of statistical methods in transfusion. *Transfusion* 1993; **33**:S30–S30.
56. Macarthur RD, Jackson GG. An evaluation of the use of statistical methodology in the *Journal of Infectious Diseases*. *Journal of Infectious Diseases* 1984; **149**:349–354.
57. Huxley TH. *The Crayfish: An Introduction to the Study of Zoology*. (3d edn). C. K. Paul & co.: London, 1881.
58. McCullough BD, Heiser DA. On the accuracy of statistical procedures in Microsoft Excel 2007. *Computational Statistics & Data Analysis* 2008; **52**:4570–4578.
59. Coombes KR, Wang J, Baggerly KA. Microarrays: retracing steps. *Nature Medicine* 2007; **13**:1276–1277.
60. Jones GL, Hobert JP. Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statistical Science* 2001; **16**:312–334.
61. Natarajan R, McCulloch CE. Gibbs sampling with diffuse proper priors: a valid approach to data-driven inference? *Journal of Computational and Graphical Statistics* 1998; **7**:267–277.
62. Peng RD. Reproducible research and biostatistics. *Biostatistics* 2009; **10**:405–408.
63. Donoho DL. An invitation to reproducible computational research. *Biostatistics* 2010; **11**:385–388.
64. Altman DG. Statistics and ethics in medical research: study design. *British Medical Journal* 1980; **281**:1267–1269.
65. Califf RM, Morse MA, Wittes J, Goodman SN, Nelson DK, DeMets DL, Iafate RP, Sugarman J. Toward protecting the safety of participants in clinical trials. *Controlled Clinical Trials* 2003; **24**:256–271.
66. Suntharalingam G, Perry MR, Ward S, Brett SJ, Castello-Cortes A, Brunner MD, Panoskaltis N. Cytokine storm in a phase 1 trial of the anti-CD28 monoclonal antibody TGN 1412. *New England Journal of Medicine* 2006; **355**:1018–1028.
67. Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. *Annals of Statistics* 2004; **32**:407–451.
68. Johnson BA, Rosenthal N, Capece JA, Wiegand F, Mao L, Beyers K, McKay A, Ait-Daoud N, Anton RF, Ciraulo DA, Kranzler HR, Mann K, O'Malley SS, Swift RM, Topiramate for Alcoholism Advisory Board, Topiramate for Alcoholism Study Group. Topiramate for treating alcohol dependence: a randomized controlled trial. *JAMA-Journal of the American Medical Association* 2007; **298**:1641–1651.
69. Pliny. *Naturalis Historia*. Michael Manuzolus: Tarvisii, 1479.
70. Xu Q, Hua JP, Braga-Neto U, Xiong ZX, Suh E, Dougherty ER. Confidence intervals for the true classification error conditioned on the estimated error. *Technology in Cancer Research & Treatment* 2006; **5**:579–589.
71. Leek JT, Storey JD. A general framework for multiple testing dependence. *Proceedings of the National Academy of Sciences of the United States of America* 2008; **105**:10718–10723.
72. Hunter D, Kraft P. Statistics and medicine: Drinking from the fire hose—statistical issues in genomewide association studies. *New England Journal of Medicine* 2007; **357**:436–439.
73. Altman DG. Statistics and ethics in medical research .7. Interpreting results. *British Medical Journal* 1980; **281**:1612–1614.
74. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Randomized trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet* 1988; **2**:349–360.
75. Lippmann W. *Essays in the Public Philosophy*. (1st edn). Little, Brown: Boston, MA., 1955.
76. Tatem AJ, Guerra CA, Atkinson PM, Hay SI. Momentous sprint at the 2156 Olympics? *Nature* 2004; **431**:525–525.
77. Rice K. Sprint research runs into a credibility gap. *Nature* 2004; **432**:147–147.