had the coauthorship and imprimatur of the distinguished head of a Harvard department (clearly a huge embarrassment). Similar cases have followed.

The committee agreed on a definition of authorship and on a statement of criteria that their journals would henceforth require all authors to sign.[2] All authors must state that they have made substantial contributions to each of three activities: (1) conception and design, or analysis and interpretation; (2) drafting the article or revising it critically for important intellectual content; (3) approval of the final version to be published. This Journal has adhered to the committee's requirements.

It is no secret among authors and scientists, and now even among editors—as is perhaps fitting we seem to be the last to know, or at least to confess to knowing—that these requirements are often flouted,[3–6] solemn signed statements notwithstanding. Indeed, I must admit to failing to persuade first authors who were my coauthors and juniors to observe the spirit of the rules in enlisting authors.

One suggestion in the committee's revised Uniform Requirements is that editors "may ask authors to describe what each contributed."[2] Such a request may not secure any tighter control of authorship claims than do more general signed statements. I must emphasize here that the signatories to these various claims are generally not dishonorable prevaricators. They and we are caught in the toils of a social subsystem in which survival demands accommodation of the truth.

The Uniform Requirements add, however, that the information describing author's contributions "may be published." In this permissive cadenza, we see a little more promise of an effective measure. The Committee B on Professional Ethics of the American Association of University Professors took a similar stance.[7]

Exercised for the integrity of both the Journal and our authors, we prepared to ask that authors, in their signed statements, describe their exact contributions for publication in a footnote. But we stayed our hand when we came upon an editorial by Richard Smith, editor of the *British Medical Journal,* reporting on a meeting about these questions.[8] The view propounded is that the system of authorship is broken and that no tinkering with requirements will fix it.

Instead, Smith offers a modest proposal—less savage than Jonathan Swift's, but as radical—to abolish authorship altogether. Authors would be replaced by "contributors," each of whose contributions would be specified, as in the credits for movies. With loss of authorship would go not only pride but responsibility and accountability. The suggested remedy is to have "guarantors" who would take responsibility for the paper overall—and also, one presumes, for those aspects that are unspecified or are not subject to specification. For the latter, one example cited is the possibility of fraudulent data.

The *British Medical Journal* does not intend to adopt such a plan forthwith. Its temperate course is to solicit the opinions of its readers. In light of this moderate approach, we conclude that we should do no less. So for the moment we shilly-shally and ask our readers, and potential authors in particular, to let us have their thoughts. Those with an interest at stake need to take note. The concern arising from the literature and the related correspondence, as well as from the experience of editors and authors, will surely fracture the status quo as, indeed, the Journal was on the verge of doing. □

***Mervyn Susser***
***Editor***

## References


1. Merton RK. Priorities in scientific discovery. *Am Sociol Rev.* 1957;22:635–639.
2. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. *JAMA.* 1997;227:927–934.
3. Shapiro DW, Wenger NS, Shapiro MF. The contributions of authors to multiauthored biomedical research papers. *JAMA.* 1994; 271:438–442.
4. Goodman N. Survey of fulfillment of criteria for authorship in published medical research. *BMJ.* 1994;309:1482.
5. Shulkin DJ, Goin JE, Rennie D. Patterns of authorship among chairmen of departments of medicine. *Acad Med.* 1993;68:688–692.
6. Bohpal R, Rankin J, McColl E, et al. The vexed question of authorship: views of researchers in a British medical facility. *BMJ.* 1997;314:1009–1012.
7. Knight J. Multiple authorship. *Science.* 1992;275:461. Letter.
8. Smith R. Authorship: time for a paradigm shift. *BMJ.* 1997;314:997.


# Editorial: Statistics in the Journal—Significance, Confidence, and All That

Certain statistical problems in papers submitted to the Journal regularly require editorial advice. Guidelines for presenting results in general in medical journals,[1] and for clinical trials in particular,[2–4] can be found in medical journals. Here, we guide authors about the preferences of this journal in statistical matters, as we continue to inform readers on methods and procedures used by the editors.[5–12] For the work published to be given its due, we must adhere to modern-day statistical standards. These may sometimes go beyond the skills of a prospective author. If so, then the author should of course enlist statistical help.

No pronouncement about science and scientific procedures can pretend to be final, and our thinking about statistical presentations continues to evolve. The diverse readership of the Journal requires us to cross disciplines and find common ground in the reporting of results, but also arms us with variety in methods. We do not wish to forgo this richness by rigid adherence to narrow requirements. In what follows, we deal in sequence with statistics as they appear in the standard sections of a scientific paper.

The introduction to a paper sets out the ways in which the work reported might bring new knowledge or complement existing knowledge. Naturally, for papers on methods per se, statistical considerations will first appear here. The Journal publishes such papers when they are novel and meaningful for public health.

Otherwise, statistical considerations first appear in the methods section. Careful descriptions of the study design and of the data are central. If primary sources were used, what data were collected and in what manner? What population or what sample was studied? If secondary sources were used, how were these particular data sufficient for the analysis at hand?

Editors look to see whether the statistical methods are adequately described and appropriate to the question being addressed. A rationale for the sample size is necessary in studies that test a priori hypotheses.

The Journal advocates no particular statistical techniques. It does call, however, for explicit statements about the statistical models used and about the parameters of interest (e.g., regression coefficients) along with a thoughtful interpretation. Clarity regarding the assumptions of an analysis is essential and helps to identify the strengths and weaknesses of a study.

Potential threats to validity include systematic underrepresentation or overrepresentation of certain groups and any of a number of measurement biases. Whenever these are apparent, statistical and conceptual detective work is needed. Readers need to have a complete accounting of all study participants. Response rates are essential. In longitudinal studies, subjects lost to follow-up, records with missing data, and individuals ineligible for given reasons must be accounted for in the tally. Comparison of the retained population with losses from the sample helps assess the possible extent of selection bias. We especially look for a comparison of known characteristics among nonrespondents, participants with missing data, and participants with complete data.

Sensitivity analyses are helpful in gauging the resulting uncertainty; they establish the effects on the results of assigning values to the missing data that assume a maximum of bias. In substantial samples, single-digit percentages of data missing at random are often negligible, however, and no more than liberal/conservative bounds on estimates and *P* values might be needed.

In longitudinal studies, complete data collected at baseline or early follow-up can be used to weight or otherwise adjust for incomplete data at subsequent follow-ups. Where data loss is substantial, resorting to multiple imputation may be advisable.[13] Imputation techniques repeatedly assign hypothetical values to missing data on various assumptions as an aid to judging sensitivity, variability, and the possible extent of bias.[14,15] While these various devices for handling missing data may require no more than brief summary statements in the final, published version of a paper, at submission it is well to include (within the text or in appendices or supplements) sufficient material for editors and reviewers to judge the issues.

The results section should contain a factual statement of what was found. Wherever feasible, authors should provide estimates of central tendency, for example, point prevalence estimates, along with appropriate indicators of measurement error or uncertainty, such as confidence intervals. The Journal welcomes quotations of likelihood ratios, where feasible, as quantitative assessments of the weight of evidence for competing causes or explanations.[1,16] Similarly, measures of public health impact, such as attributable risks,[10] are encouraged.

We favor the provision of data in forms that enable comparisons with other studies, for instance, statistics in a form suitable for combining in a meta-analysis, that is, a point estimate such as a log odds ratio together with the standard error, study design, and adjustment variables. In the execution of a meta-analysis, the salient features of the method used, for example, fixed effects or random effects models,[17] are for each author to decide.

When the analysis involves regression models, authors should provide at least the regression coefficients and standard errors. From these, readers can construct a confidence level at any percentile desired—usually 99%, 95%, or 90%—as well as a *P* value. Authors may also present either a confidence interval or a *P* value, but normally not both. (A *P* value alone, without a standard error, is seldom acceptable.) For linear regressions, the standard error of regression, the variance explained ($R^2$), and the model F statistic with degrees of freedom are preferred. For logistic regressions, the goodness-of-fit statistic for grouped data and an appropriate statistic for assessing goodness of fit for individual data are advisable.[18] Where authors use esoteric or complex methods, it is permissible to cite published references with only a brief description of the essential assumptions and methods of the analysis. On occasion, an appendix for statistical reviewers may prove useful.

For presenting central results, a point estimate (the best available indication of the true value), together with a confidence interval (which brackets the true value with high probability), does a good job of summarizing the data and describing the degree of uncertainty in the result. For those ancillary results that need to be reported, however, a point estimate plus or minus a standard error will often suffice and avoid a clutter of confidence intervals in tables and text. The debate of *P* values vs confidence intervals[19–22] (or for that matter, likelihood ratios and posterior probabilities) lies at the very foundation of statistical inference and is not one we wish to enter. Competent use of any of these methods is acceptable if it advances understanding.

Another reporting matter involves one-tailed versus two-tailed procedures. Almost always, two-sided procedures are favored. One-sided tests might be acceptable when accompanied by a strong defense that an effect in the opposite direction to that hypothesized is entirely inconsequential.[23(pp27–29)]

## *Some Commonly Encountered Analytic Issues: Intent to Treat, Complex Sampling, and Levels of Analysis*

First, for group comparisons in randomized clinical trials, the strongly recommended analysis is "intent to treat"; that is, the denominators of the outcomes include all participants randomized in their originally assigned treatment groups, irrespective of losses after randomization and before outcome is measured because of dropout, crossover, noncompliance, and so on.

An alternative, "efficacy analysis," includes in the denominators only participants who actually received the intervention. This approach restricts the measurement of the effect of the intervention to those who received it and puts aside those who did not. Although important exceptions arise, efficacy analysis is subject to bias, may produce invalid results, and is seldom acceptable in the Journal unless complementary to intent-to-treat analysis. Should an efficacy analysis be proffered, it will need to be clearly reported and defended,[3] first of all by comparison with an intent-to-treat analysis. For example, in a randomized trial, if large-scale but well-documented crossover of subjects undermines the statistical power of an intent-to-treat analysis, the Journal may consider efficacy analysis. More likely than not, however, such a circumstance will be deemed a failed trial.

Second, with designs complicated by the sampling of clusters and overrepresentation of certain subgroups, adjustments will be needed to obtain unbiased population estimates. Weighting programs are available for the purpose. When the relationships between factors rather than population estimates are of interest, how-

ever, subgroup analysis without such a weighting program is often permissible.

Third, authors should take care to perform the analysis at the appropriate level, group or individual, at which the data are collected. For example, if communities are randomized to receive a certain prevention, the community, and not the individual, is the proper unit of analysis.

## *Tabulation, Figures, and Text*

The Journal prefers that most of the numerical results be presented in tables rather than text. Delays in publication often occur because in the final stages of acceptance of a paper, tables still need refinement. Attention to the following points helps ensure that Journal requirements for tables are met:

- Each table should stand alone (i.e., readers need not refer to the text for understanding). This requires that the titles fully identify the data displayed (who? when? where? how?).
- Displays should be brief and clear with a minimum of mathematical symbols.
- Point estimates should be accompanied, where appropriate, by measures of uncertainty, for example, standard errors with means and regression coefficients, 95% confidence intervals with rate ratios and odds ratios.
- All but universal abbreviations and acronyms should be spelled out in tables and, if unavoidable, should be explained in footnotes.
- The total numbers (n) relevant to the columns in a table should be presented at the top of each column. Within the individual cells of a column, percentages alone should appear whenever the numbers can be reconstructed from percentages and the total numbers, except where numbers are small. Seldom do public health data call for more than one decimal place, or at most two.
- The text *explains* the data, while tables *display* the data. That is, text pertaining to the table reports the main results and points out patterns and anomalies, but avoids replicating the detail of the display.
- Only results essential to the main thesis should be presented, for example, the regression coefficient for the main effect of interest, but not the regression coefficients of other main effects unless they are essential to interpretation, as occurs with interactions. For lack of space, with rare exceptions we avoid publishing appendices. Instead, we suggest a footnote to indicate their availability either upon request from the author or from the National Auxiliary Publication Service.

Figures are valuable for summarizing complex results, emphasizing patterns across groups, and highlighting magnitudes of differences or slopes. The visual display of data has a great impact on the communication of results.[24] Graphs should have all axes clearly labeled and should begin with logical origin values. Labels directly indicating the category for each line on the graph are best and require no key. If such labeling on a graph is difficult, a key to the symbols must be provided.

We encourage authors to submit additional material for the convenience of referees and editors, for example, diagnostic graphics or flowcharts, to support what is omitted from the paper itself, especially for complex analyses. Before a final version goes to press, however, we ask authors to eliminate redundancy between text, tables, and figures.

In the discussion section, authors should indicate how their study adds to the existing literature and the direction of research: in other words, suggest how the results complement, challenge, or subvert previous work. Authors should repeat only those results key to the argument; they should not introduce results not shown or noted earlier except for good cause, for instance, to counter a possible charge of confounding or bias. Interpretation of the findings should not extend beyond what the data allow.

Editors check the discussion section (as well as the abstract) for adequate qualification of results and conclusions. Listing strengths and weaknesses, including those of the statistical approach, is a common and useful practice. In this Journal, it is also appropriate to outline public health and policy implications of the findings.

In the face of the information overload characteristic of our era, research findings need to be reported clearly and concisely. Advances in computing allow one to perform sophisticated analyses with efficiency. They also permit the facile use of highly complicated procedures. These capabilities do not obviate the need for a clear statement of the statistical question at hand and the selection of appropriate methods for addressing it, but make these all the more important. Properly used, these are essential elements for building a body of evidence and advancing understanding of crucial public health issues. □

***Mary E. Northridge, Deputy Editor***
***Bruce Levin, Consulting Statistical Editor***
***Manning Feinleib, Associate Editor***
***Mervyn W. Susser, Editor***

## References

1. Bailar JC, Mosteller F. Guidelines for statistical reporting in articles for medical journals. *Ann Intern Med.* 1988;108:266–273.
2. Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature. Call for comments on a proposal to improve reporting of clinical trials in the biomedical literature: a position paper. *Ann Intern Med.* 1994;121:894–895.
3. The Standards of Reporting Trials Group. A proposal for structured reporting of randomized controlled trials. *JAMA.* 1994; 272:1926–1931. Correction appears in *JAMA.* 1995;273:776.
4. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA.* 1996;276:637–639.
5. Susser M. Serving all. *Am J Public Health.* 1992;82:521–522.
6. Susser M, Yankauer A. Prior, duplicate, repetitive, fragmented, and redundant publication and editorial decisions. *Am J Public Health.* 1993;83:792–793.
7. Northridge ME, Susser M. The paper route for submissions to the Journal. *Am J Public Health.* 1994;84:717–718.
8. Northridge ME, Susser M. Seven fatal flaws in submitted manuscripts. *Am J Public Health.* 1994;84:718–719.
9. Levin B. Annotation: accounting for the effects of both group- and individual-level variables in community-level studies. *Am J Public Health.* 1995;85:163–164.
10. Northridge ME. Annotation: public health methods—attributable risk as a link between causality and public health action. *Am J Public Health.* 1995;85:1202–1204.
11. Susser M, Northridge ME. Editor's note: reviewing for the Journal. *Am J Public Health.* 1996;86:161.
12. Susser M. International submissions to the Journal. *Am J Public Health.* 1996;86:298–299.
13. Rubin DB. *Multiple Imputation for Nonresponse in Surveys.* New York, NY: John Wiley & Sons Inc; 1987. See especially Chap 3, Section 1, and Chap 5, Section 3.
14. Heitjan DF, Basu S. Distinguishing 'missing at random' and 'missing completely at random.' *Am Statistician.* 1996;50:207–213.
15. Heitjan DF. What can be done about missing data? approaches to imputation. *Am J Public Health.* 1997;87:548–549.
16. Goodman SN, Royall R. Evidence and scientific research. *Am J Public Health.* 1988;78:1568–1574. Correction appears in *Am J Public Health.* 1989;79:273.
17. Fleiss JL. The statistical basis of meta-analysis. *Stat Methods Med Res.* 1993;2: 121–145.

18. Hosmer DW, Lemeshow S. *Applied Logistic Regression.* New York, NY: John Wiley & Sons Inc; 1989.
19. Walker AM. Reporting the results of epidemiologic studies. *Am J Public Health.* 1986;76:556–558.
20. Fleiss JL. Significance tests have a role in epidemiologic research: reactions to A.M. Walker. *Am J Public Health.* 1996;76:559–560.
21. Fleiss JL. Confidence intervals vs. significance tests: quantitative interpretation. *Am J Public Health.* 1986;76:587–588. Letter.
22. Walker AM. Significance tests represent consensus and standard practice. *Am J Public Health.* 1986;76:1033–1034. Letter.
23. Fleiss JL. *Statistical Methods for Rates and Proportions.* New York, NY: John Wiley & Sons Inc; 1981.
24. Tufte ER. *Visual Display of Quantitative Information.* Cheshire, Conn: Graphics Press; 1983.

# Annotation: Disparity in Cancer Survival and Alternative Health Care Financing Systems

The primary objective of the interesting and clever study by Gorey et al. in this issue of the Journal is to compare medical outcomes in two metropolitan areas: one, Toronto, Ontario, operating under Canada's single-payer system, and the other, Detroit, Mich, existing within the "insurance-driven US system."[1] The authors hypothesize that for persons of low socioeconomic status (SES), survival is better in Toronto than in Detroit. A secondary hypothesis is that the direct relationship between SES and cancer survival seen in many Western nations should be relatively weaker in Toronto than Detroit. The survival figures for most major cancer sites corroborate both hypotheses. The authors conclude that a likely explanation for these findings is greater access to preventive and therapeutic care in Toronto.

It's not easy to test mega-hypotheses concerning optimum strategies for organizing and financing health care. A controlled trial involving, say, the randomization of cities or neighborhoods to one or another health care financing scheme is infeasible at present. We are left, as Gorey et al. are, with devising observational studies to evaluate alternative health care systems. Cancer survival, which is sensitive to prompt diagnosis and therapeutic quality, makes a conceptually reasonable end point. The availability of population-based cancer survival data is a particular advantage.

Gorey et al. reason that people of low socioeconomic status are likely to be more affected by health care financing issues than their more affluent counterparts. That is, in a system with substantial barriers to care, the affluent may be better able to maneuver their way to early diagnosis and high-quality oncologic treatment. It is this focus on SES that makes this study especially informative and differentiates it from an earlier study by the US General Accounting Office, which found little cancer survival difference between Canada and the United States.[2]

Studies of SES and cancer face a number of problems, not the least of which is the absence of individual-level socioeconomic information in cancer registries. The authors resort to an ecological (census tract–based) measure of SES. Although ecological fallacy is a generic worry, a number of recent studies have established the value of such ecological measures of SES,[3,4] especially when researchers—like Gorey et al.—limit their inferences to areal variables, such as residence in a low-, middle-, or high-income area, rather than individual income status.

Does "low socioeconomic status" mean the same thing in Toronto as in Detroit? The authors compare *relative* socioeconomic tertiles, that is, categories derived from low-, middle-, and high-income areas within countries. In *absolute* income terms, though, the two cities differ greatly. For the critical low-income census tracts, the median income (in US dollars) was $30 400 in Toronto and $17 800 in Detroit. It may well be absolute, rather than relative, income that primarily determines that mix of lifestyle, physical environmental, and even health services factors that affects cancer survival. One might ask, with respect to the intercountry comparison, whether the Toronto "low-income" areas are loaded with truly higher-SES tracts with more favorable survival experience. Or, similarly, whether a direct relationship between SES and survival in Toronto is obscured by not making the "low-income" group low enough. Gorey et al. address these questions. They perform some finer quintile analysis and note in the discussion that the nonsignificance of the association between SES and survival in Toronto obtains. Moreover, Toronto's survival advantage is maintained in a comparison of the Canadian city's poorest quintile (median income = $28 000) with Detroit's second poorest (median income = $26 300).

Socioeconomic status reflects a host of biological, behavioral, and social systemic factors, some individual-level, others aggregate in nature. The question arises whether residents of Detroit's low-SES tracts differ from residents of Toronto's low-SES tracts in characteristics other than medical care that influence cancer survival. These characteristics might include smoking, body size, diet, alcohol intake, physical activity, use of medication, chemical exposure, immune status, and so on—our knowledge of the factors influencing survival from various malignancies is far from complete. One could even speculate that there is something about the social environment of Toronto, compared with Detroit, that confers a survival advantage through some as yet unrecognized cascade of psychological, neurological, endocrine, or immune phenomena that somehow influences the behavior of residual malignant cells and precancerous lesions. Although differential access to health care is a reasonable, even likely, explanation for the survival advantage of low-SES residents of Toronto, it is difficult to rule out some of these alternative explanations. To argue, as the authors do, that overall smoking rates are comparable for Canada and the United States does not preclude the possibility that smoking prevalence is higher in Detroit's low-SES areas than in Toronto's. As they point out, few country- and SES-specific data are available on smoking and other characteristics potentially linked to both SES and cancer survival. The authors' ongoing efforts to incorporate individual-level data on prognostic and treatment-related variables may help bolster the argument that the intercountry survival differential primar-

**Editor's Note.** See related article by Gorey et al. (p 1156) in this issue.