Transfer of Technology From Statistical Journals to the Biomedical Literature

Past Trends and Future Predictions

Douglas G. Altman, Steven N. Goodman, MD, PhD

Objective.—To investigate the speed of the transfer of new statistical methods into the medical literature and, on the basis of current data, to predict what methods medical journal editors should expect to see in the next decade.

Design.—Influential statistical articles were identified and the time pattern of citations in the medical literature was ascertained. In addition, longitudinal studies of the statistical content of articles in medical journals were reviewed.

Main Outcome Measures.—Cumulative number of citations in medical journals of each article in the years after publication.

Results.—Annual citations show some evidence of decreasing lag times between the introduction of new statistical methods and their appearance in medical journals. Newer technical innovations still typically take 4 to 6 years before they achieve 25 citations in the medical literature. Few methodological advances of the 1980s seem yet to have been widely cited in medical journals. Longitudinal studies indicate a large increase in the use of more complex statistical methods.

Conclusions.—Time trends suggest that technology diffusion has speeded up during the last 30 years, although there is still a lag of several years before medical citations begin to accrue. Journals should expect to see more articles using increasingly sophisticated methods. Medical journals may need to modify reviewing procedures to deal with articles using these complex new methods.

(*JAMA*. 1994;272:129-132)

THE INFLUX of statistical methods into the medical literature has increased over more than 60 years. Over the same period, statistics itself has undergone major changes, so that not only is the use of statistics in medical research much more common, but the methods used have become progressively more complex. Although some of the methods being introduced in medical research were developed in other contexts, many statistical advances have arisen as solutions to problems arising in medical research. Changes in the type of statistical methods being used in medical articles have implications for editors, referees, and readers.

We report herein a study of citations to investigate the transfer of new statistical methods into the medical literature. We predict some new methods that medical journal editors should expect to see in the next decade.

From the Medical Statistics Laboratory, Imperial Cancer Research Fund, London, England (Mr Altman), and Oncology Center, Division of Biostatistics, The Johns Hopkins University, Baltimore, Md (Dr Goodman).

Presented in part at the Second International Congress on Peer Review in Biomedical Publication, Chicago, Ill, September 10, 1993.

Reprint requests to Medical Statistics Laboratory, Imperial Cancer Research Fund, PO Box 123, Lincoln's Inn Fields, London, England WC2A 3PX (Mr Altman).

METHODS

Influential statistical articles published after 1950 were identified from two books that reprinted important statistical articles,[1,2] from a list of the most cited articles in medical journals, and from personal knowledge (Table 1). Several articles relate to survival analysis[6,9,11,13,14] or meta-analysis,[5,7] two of the strongest growth areas (in both medicine and medical statistics) in recent years. Unfortunately, in some important areas of statistical methods there was no key article that could be widely cited by a large proportion of users, such as logistic regression and sample size calculations for clinical trials. We have included some articles that were published in medical journals (notably, cancer journals) when these seemed to be the primary source of the new method, and also one book.

For each article, the time pattern of citations in the medical literature was ascertained. Citations prior to 1971 were obtained by hand searching of printed volumes of the *Science Citation Index*,[23] as were citations for a few of the later articles with relatively few citations. Citations from 1971 to 1992 were obtained using computer searches of the SciSearch database (Institute of Scientific Information, Philadelphia, Pa). These searches were carried out in July and August 1993, by which time citations for 1992 should have been virtually complete. We did not search for articles that had incorrect citations of the articles of interest. It is our impression that the rate of incorrect citations of these articles was about 10% (excluding errors in titles). Some minor inconsistency between the two methods of searching may have arisen through problems in identifying what constitutes a medical journal. For comparison, similar citation analyses were performed for two heavily cited expository statistical articles published in medical journals.[21,22]

We also sought evidence from longitudinal studies of the statistical content of articles in medical journals to examine changes in the methods used over time.

RESULTS

Figure 1 shows cumulative numbers of citations for the articles listed in Table 1 divided into four decades—the 1950s, 1960s, 1970s, and 1980s. The article by Cox[14] was excluded because it has been cited much more often than the other articles. It is shown in Fig 2, together with the article by Kaplan and Meier.[6] These two articles are frequently cited together in articles reporting the results of survival analyses. They were published 14 years apart, and Fig 2 shows that the citations for the earlier article have risen in parallel with those for the Cox article, but about 14 years later in relation to the year of publication. These are now two of the most heavily cited articles in medical journals. The rise in citations for the article by Kaplan and Meier[6] is especially marked given that it received only six citations in medical journals in the first 10 years after publication.

Annual citations for the articles published in the four decades do show some evidence of decreasing lag times between the introduction and widespread use of new statistical methods. Newer technical innovations still typically take 4 to 6

Table 1.—Statistical Articles Included in This Study

Source, y	Topic
Methodological articles	
Cornfield,[3] 1951	Odds ratio
Cochran,[4] 1954	χ^2 Trend test
Woolf,[5] 1955	Combining 2×2 tables
Kaplan and Meier,[6] 1958	Survival curve
Mantel and Haenszel,[7] 1958	Stratified 2×2 table
Cohen,[8] 1960	κ Statistic
Mantel,[9] 1963	Survival analysis
Box and Cox,[10] 1964	Transformations
Mantel,[11] 1966	Survival analysis
Elston and Stewart,[12] 1971	Heredity
Peto and Peto,[13] 1972	Log rank test
Cox,[14] 1972	Proportional hazards regression
Dempster et al,[15] 1977	EM algorithm
Efron,[16] 1979	Bootstrap
Hanley and McNeil,[17] 1982	Receiver operating characteristic curve
Geman and Geman,[18] 1984	Gibbs sampling
Breiman et al,[19] 1984	Classification and regression trees
Zeger and Liang,[20] 1986	Longitudinal data
Expository articles	
Peto et al,[21] 1977	Log rank test
Bland and Altman,[22] 1986	Method comparison

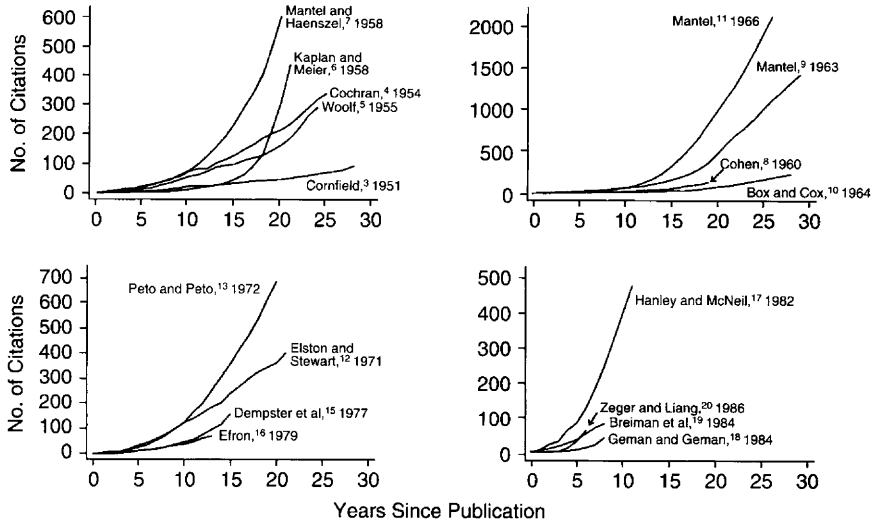

Fig 1.—Cumulative citations in medical journals for selected articles published in 1950 through 1959 (top left), 1960 through 1969 (top right), 1970 through 1979 (bottom left), and 1980 through 1989 (bottom right).

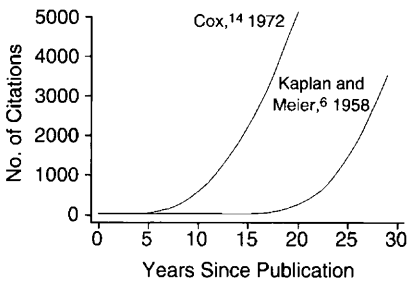

Fig 2.—Cumulative citations in medical journals for two heavily cited articles on survival analysis methods.

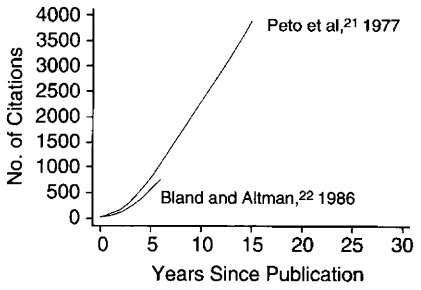

Fig 3.—Cumulative citations in medical journals for two expository articles.

years before they achieve 25 citations in the medical literature. Few methodological advances of the 1980s seem yet to have been widely cited in medical journals. By contrast, expository articles in medical journals can reach 500 citations within 4 to 5 years (Fig 3). Citations for one of the two expository articles[21] have leveled out, with a roughly constant number of citations each year. Most of the methodological articles (notably, the heavily cited articles) have increasing numbers of citations each year.

Few authors have studied changes over time in the use of statistical methods in one journal. Hayden[24] gave a brief summary of the rise in the use of simple statistical methods in *Pediatrics* from 1952 to 1982, while Felson et al[25] described similar changes in *Arthritis and Rheumatism* from 1967 to 1968 vs 1982. The most detailed information we are aware of relates to the *New England Journal of Medicine*. Articles published in 1978 and 1979,[26] 1989,[27] and 1990[28] have been reviewed using the same set of categories.[26] A large increase was noted during this period in the use of most statistical methods, especially the more complex methods (Table 2). It is notable that survival analysis and logistic regression were found in almost a third of original articles published in 1989 and 1990.

COMMENT

Citation studies are rightly criticized as a means of grading researchers,[29] but we think they provide a valuable measure of the impact of a new methodological development on medical research. Figure 1 suggests that technology diffusion may have speeded up during the last 30 to 40 years, although there is still usually a lag of several years before medical citations begin to accrue.

We used cumulative citations rather than annual citations, as we feel the total impact is more relevant in this context and that fluctuations in the annual counts obscure the trends. For the purposes of documenting technology transfer, it is not the actual number of citations but the shape of the citation curve that is most informative. This shape seems not to have changed greatly during four decades. Almost all of the curves for these classic articles have a dormant early phase followed by a somewhat dramatic takeoff. The general shape does not seem to vary in relation to how heavily cited an article is. There are, however, a few exceptions to this pattern, notably the article by Hanley and McNeil[17] (Fig 1). Developments that have probably contributed to the more rapid diffusion of statistical methods into the medical literature are the increasing number of statisticians working in medicine, the accessibility of powerful desktop computers to medical researchers, and the more rapid development and dissemination of software to implement new statistical methods.

Our analyses took no account of the large increase in the number of articles being published each year in medical journals (1730 journals published in 1950, increasing in 10-year intervals to 2800, 4420, 6780, and 9480) (Ulrich's International Serials database, Bowker Electronic Publishing). However, this increase has been almost linear since 1970, so adjustment for the increasing size of the literature would not greatly alter the shapes of the curves. Furthermore, such adjustment is not appropriate if, as seems likely, researchers today need to access many more articles in a greater number of journals than their predecessors. Huth[30] found a large increase between 1950 and 1985 in the number of

Table 2.—Statistical Methods Used in Original Articles in the *New England Journal of Medicine* in 1978 and 1979[26] and 1989 and 1990[27,28]

Topic	1978-1979, % (n=332)	1989-1990, % (n=215)
Any statistical analysis	73	88
t Tests	44	39
Contingency tables	27	33
Pearson correlation	12	18
Survival methods/ logistic regression	11	31
Nonparametric tests	11	23
Epidemiologic statistics	9	18
Analysis of variance	8	17
Simple linear regression	8	13
Transformation	7	7
Multiple regression	5	10
Multiway tables	4	8
Nonparametric correlation	4	5
Multiple comparisons	3	7
"Other methods" (not on original list)	3	14

Table 3.—Newer Statistical Methods That May Be Seen More Often in the Coming Years

Method	Description	Purpose
Bootstrap (also called resampling; related to the jackknife)[17]	Multiple new data sets are generated by random sampling "with replacement" from the original data	To calculate SEs or assess the stability of a statistical model, often when standard assumptions are unreliable or the sampling distribution is unknown
Gibbs sampling[18,34]	Random sampling from conditional distributions within a complex structure	Bayesian estimation of complex models
Generalized additive models[35]	Nonparametric smoothing of explanatory variables in regression	To replace regression when assumptions are not tenable
Classification and regression trees[19,36] (also known as recursive partitioning)	Division of a set of subjects by combinations of characteristics, to minimize the differences within groups and to maximize the differences between groups	To find combinations of variables of predictive importance
Models for longitudinal data ("general estimating equations")[20]	Modeling repeated measurements of an outcome variable while allowing for covariates	Regression for multiple assessments of outcome
Models for hierarchical data (also called multilevel models)[37]	Fitting mixed linear models to hierarchical data using iterative generalized least squares	Modeling data with more than one level of variation (eg, within and between patients)
Neural networks[38]	Nonparametric modeling of complex data	To provide nonlinear approximations to multivariable functions or for classification

different journals being cited in articles published in the *New England Journal of Medicine*.

Independent evidence for genuine changes in the use of statistics comes from studies that have looked at the same journals across time. The few such studies that we are aware of have shown large increases in the use of statistical methods and a tendency to use more complicated methods.[24-28] Thus, there is clearly a strong component of increased use and complexity of statistics independent of the total journal expansion. It is relevant that the number of original articles published per year by the *New England Journal of Medicine* decreased during the period of the studies summarized in Table 2.

Cumulative citations for the methodological articles considered generally curve upward, indicating that the annual number of citations keeps increasing. By contrast, the two expository articles considered show a much more rapid accrual of citations (starting in the year of publication) but near-linear cumulative citation curves, indicating a fairly steady annual citation rate. Expository statistical articles in medical journals can reach 500 citations within 4 to 5 years (Fig 3). Both articles we considered[21,22] described methods previously published in statistical journals[13,31] without achieving many citations in medical journals. These citation figures suggest that expository articles are valuable, especially for topics that are not usually included in medical statistics textbooks. Indeed, the International Committee of Medical Journal Editors guidelines state, "References for study design and statistical methods should be to standard works (with pages stated) when possible rather than to papers in which the designs or methods were originally reported."[32] Expository articles cowritten by a statistician and medical researcher may be especially helpful—a recent example considers receiver operating characteristic curves.[33] Unfortunately, such crossover articles require a considerable amount of work, and such activity (being a form of teaching) may not be helpful to the statistician's or researcher's career in comparison with either more methodological or medical articles.

Several complex statistical methods introduced in the 1980s are beginning to be seen more frequently. Although it is not possible to identify recent articles that will turn out to be major breakthroughs, most of the newer methods are sophisticated. Journals should expect to see growing numbers of articles using them. Methods likely to be seen more often are described briefly in Table 3. Software is available for all of these techniques, and some are beginning to be included in well-known statistics packages. It is worth noting that by the time a topic reaches medical journals there may be a large methodological literature. Ripley[38] notes that there are already more than a dozen journals and at least 15 texts devoted to neural networks.

The evidence of time trends within one major journal (Table 2) supports the idea that there is an ever-increasing variety of statistical methods appearing in medical articles. The speed with which new methods are introduced may pose problems for statistical referees, for the physicians who read the published work, and for the journals themselves. Referees may not be able to judge new methods that they have not yet learned. Physicians may feel that they have no chance of understanding the new methods (even if they are comfortable with more traditional methods) and will have to take the results of such studies on faith. The journals, in whom that faith is being entrusted, may bear an increasing burden to ensure that the methods are indeed valid, since most of their audience will be unable to assess that for themselves.

We think that the following developments are possible and may be desirable in the future:

- Authors using complex methods will be asked to supply additional supporting material for referees but not for publication. This might take the form of a formal appendix in the submitted manuscript, which is peer reviewed (and possibly modified) but not published. It should be supplied by authors to readers on request.
- Because statistical refereeing will be a more difficult process (because of both the novelty and the complexity of methods), medical journals may need to recruit panels of methodological reviewers who specialize in specific methods.
- Editors of medical journals should encourage or actively solicit more crossover (expository) articles on new methods, perhaps with both medical and statistical authors.
- More postgraduate training for medical researchers should be developed, with formal accreditation, both in basic statistical methods and also to help those who wish to keep abreast of newer methods.

It is likely that the statistical education of physicians, already poor,[39,40] will in the future lag even further behind the

methods that are used in medical journals. Already the standard methods taught in an introductory course would leave a reader unable to judge a high percentage of articles published in the *New England Journal of Medicine*, and that proportion is likely to increase with time.

We thank Scott Zeger, PhD, for suggesting this topic of investigation. We are grateful to Will Russell-Edu for carrying out the computer citation searches.

References

1. Greenland S, ed. *Evolution of Epidemiologic Ideas: Annotated Readings on Concepts and Methods.* Chestnut Hill, Mass: Epidemiology Resources Inc; 1987.
2. Kotz S, Johnson N, eds. *Breakthroughs in Statistics: Volume II: Methodology and Distribution.* New York, NY: Springer Publishing Co; 1992.
3. Cornfield J. A method of estimating comparative rates from clinical data: applications to cancer of the lung, breast, and cervix. *J Natl Cancer Inst.* 1951;11:1269-1275.
4. Cochran WG. Some methods for strengthening the common χ^2 tests. *Biometrics.* 1954;10:417-451.
5. Woolf B. On estimating the relation between blood group and disease. *Ann Hum Genet.* 1955; 11:251-253.
6. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc.* 1958;53:457-481.
7. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst.* 1958;22:719-748.
8. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas.* 1960;20:37-46.
9. Mantel N. Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. *J Am Stat Assoc.* 1963;58:690-700.
10. Box GEP, Cox DR. An analysis of transformations (with discussion). *J R Stat Soc B.* 1964;26: 211-252.
11. Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep.* 1966;50:163-170.
12. Elston RC, Stewart J. A general model for the genetic analysis of pedigree data. *Hum Heredity.* 1971;21:523-542.
13. Peto R, Peto J. Asymptotically efficient rank invariant test procedures (with discussion). *J R Stat Soc A.* 1972;135:185-207.
14. Cox DR. Regression models and life tables (with discussion). *J R Stat Soc B.* 1972;34:187-220.
15. Dempster AP, Laird N, Rubin D. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc B.* 1977;39:1-38.
16. Efron B. Bootstrap methods: another look at the jackknife. *Ann Stat.* 1979;7:1-26.
17. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982;143:29-36.
18. Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images. *IEEE Trans Pattern Anal Machine Intell.* 1984; 6:721-741.
19. Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and Regression Trees.* Belmont, Calif: Wadsworth; 1984.
20. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics.* 1986;42:121-130.
21. Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient, II: analysis and examples. *Br J Cancer.* 1977;35:1-39.
22. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1:307-310.
23. Institute of Scientific Information. *Science Citation Index.* Philadelphia, Pa: Institute of Scientific Information; 1951-1971.
24. Hayden GF. Biostatistical trends in *Pediatrics*: implications for the future. *Pediatrics.* 1983;72:84-87.
25. Felson DT, Cupples LA, Meenan RF. Misuse of statistical methods in *Arthritis and Rheumatism*: 1982 versus 1967-68. *Arthritis Rheum.* 1984;27:1018-1022.
26. Emerson JD, Colditz G. Use of statistical analysis in the *New England Journal of Medicine. N Engl J Med.* 1983;309:707-713.
27. Emerson JD, Colditz G. Use of statistical analysis in the *New England Journal of Medicine.* In: Bailar JC III, Mosteller F, eds. *Medical Uses of Statistics.* 2nd ed. Boston, Mass: NEJM Books; 1992: 45-57.
28. Altman DG. Statistics in medical journals: developments in the 1980s. *Stat Med.* 1991;10:1897-1913.
29. Seglen PO. Citation frequency and journal impact: valid indicators of scientific quality? *J Intern Med.* 1991;229:109-111.
30. Huth E. The information explosion. *Bull N Y Acad Med.* 1989;65:647-661.
31. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *Statistician.* 1983;32:307-317.
32. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. *JAMA.* 1993;269:2282-2286.
33. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental tool in clinical medicine. *Clin Chem.* 1993;39:561-577.
34. Gilks WR, Clayton DG, Spiegelhalter DJ, et al. Modelling complexity: applications of Gibbs sampling in medicine. *J R Stat Soc B.* 1993;55:39-52.
35. Hastie T, Tibshirani R. Generalized additive models. *Stat Sci.* 1986;1:297-318.
36. Ciampi A, Lawless JF, McKinney SM, Singhal K. Regression and recursive partition strategies in the analysis of medical and survival data. *J Clin Epidemiol.* 1988;41:737-748.
37. Goldstein H. *Multilevel Models in Educational and Social Research.* London, England: Griffin; 1986.
38. Ripley BD. Statistical aspects of neural networks. In: Barndorff-Nielsen OE, Jensen JL, Kendall WS, eds. *Networks and Chaos: Statistical and Probabilistic Aspects.* London, England: Chapman and Hall; 1993:40-123.
39. Wulff HR, Andersen B, Brandenhoff P, Guttler F. What do doctors know about statistics? *Stat Med.* 1987;6:3-10.
40. Altman DG, Bland JM. Improving doctors' understanding of statistics (with discussion). *J R Stat Soc A.* 1991;154:223-267.