American Journal of Epidemiology

Vol. 161, No. 5
Printed in U.S.A.
DOI: 10.1093/aje/kwi073

SPECIAL ARTICLE

An Analysis of *American Journal of Epidemiology* Citations with Special Reference to Statistics and Social Science

J. Michael Oakes

From the Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN.

Received for publication June 7, 2004; accepted for publication October 14, 2004.

In an effort to inform the ongoing discussion about the purpose, purview, theoretical orientation, and viability of epidemiology, this paper considers the contemporary epistemological foundations of the discipline by analyzing article citations. Two principal questions are the following: 1) What research do *American Journal of Epidemiology* (AJE) authors rely on to support, inform, and frame their investigations? and 2) to what extent do such authors use social scientific and statistical citations? The data used appear to be superior to those used in previous efforts because they contain complete citations for all articles published, along with complete within-article citations, for all AJE articles published from January 1981 to December 2002. The most frequent AJE citations are statistically oriented works. About 9% of citations are to AJE articles, 15% are to a larger set of eight epidemiologic journals, 15% are to a select set of eight medical journals, 3% are to (bio)statistics journals, and just 0.2% are to social science journals. Trend analysis reveals little change during the 22-year study period. The principal implication is that AJE authors are overlooking a vast literature that could inform their understanding of how exposures emerge and are maintained.

history; knowledge

Abbreviations: AJE, *American Journal of Epidemiology*; JAMA, *Journal of the American Medical Association*.

Epidemiology is experiencing a disciplinary debate over its purpose, the proper level of analysis, the utility and scope of epidemiologic theory, and the standards for causal explanation (1–17). The stakes are high if one goal of epidemiology is to improve the public's health. While it is perhaps unsettling, I believe that the current debate is an indication of disciplinary health, especially when answers are empirically informed.

Presumably, one kind of useful empirical information is data on the contemporary epistemological foundations of the discipline. Citation analyses are helpful in this regard. By tabulating and categorizing citations, one may document a discipline's institutional or virtual knowledge and perhaps delineate communication networks within and across disciplines (18). Citation analyses also reveal aspects of intellectual history and so help identify not only a discipline's "normal science" (19, 20) but emergent paradigmatic shifts.

Two previous citation analyses have been published in (and on) this journal (21, 22). Both were focused on the use of clinical medicine by epidemiologists and the use of epidemiologic research by clinicians; extensive and growing interactions were found. Other analyses have examined trends in article content, such as the use of race and class (23–26).

No such study has systematically considered epidemiology's reliance on statistical or social science research. This is surprising since (bio)statistics is such an important part of epidemiologic training and activity, epidemiology and social science share a common origin, and there is growing interest in social epidemiology (27–29).

This paper aims to contribute empirical information on the contemporary epistemological foundations of epidemiology. According to Merton's paraphrase of Newton (30), on whose shoulders do epidemiologists stand? That is, what works do

Correspondence to Dr. J. Michael Oakes, Division of Epidemiology, University of Minnesota, Minneapolis, MN 55454-1015 (e-mail: oakes@epi.umn.edu).

TABLE 1. Ten most frequent citations in the *American Journal of Epidemiology*, 1981–2002*

No. of times cited	Cited article/book
312	Breslow NE, Day NE. Statistical methods in cancer research. Vol I. The analysis of case-control studies. Lyon, France: International Agency for Research on Cancer, 1980. (IARC scientific publication no. 32)
233	Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research: principles and quantitative methods. New York, NY: Van Nostrand Reinhold Company, 1982
227	Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959; 22:719–48
221	Rothman KJ. Modern epidemiology. Boston, MA: Little, Brown, 1986
218	Cox DR. Regression models and life tables (with discussion). J R Stat Soc (B) 1972;34:187–220
215	Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York, NY: John Wiley & Sons, Inc, 1981
155	Schlesselman JJ, Stolley PD. Case-control studies: design, conduct, analysis. New York, NY: Oxford University Press, 1982
132	Mantel N. Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. J Am Stat Assoc 1963; 58:690–700
127	Snedecor G, Cochran W. Statistical methods. Ames, IA: Iowa University Press, 1967
122	Miettinen OS. Stratification by a multivariate confounder score. Am J Epidemiol 1976;104:609–20

* Frequency for texts includes all subsequent editions.

American Journal of Epidemiology (AJE) authors rely on to support, inform, and frame their research, and to what extent do AJE authors use social scientific and statistical citations? This paper does not address the extent to which AJE or other epidemiologic research is used elsewhere. Special attention is paid to statistics, because it is a tool for epidemiologic inquiry and so is a useful guidepost against which to consider social science. Social science is considered because of its prominent role in the ongoing discussion and debates, and because there is little empirical evidence of its influence.

MATERIALS AND METHODS

Data for this paper were purchased from the leading citation-data firm, Thomson ISI (The Thomson Corporation, Stamford, Connecticut). This firm sells *Journal Citation Reports* used by previous researchers (21, 22). Unlike previous research that relied on journal-level citation data (i.e., *Journal Citation Reports* for the 100 most cited journals), this paper's data are at the citation level and include complete citations for all articles published in AJE, along with complete citations for each within-article citation, for all AJE articles published from January 1981 to December 2002. These data permit the extensive analyses offered here, such as the most frequent citations presented in table 1. Analyses are limited to "research articles" and their "research article citations." Abstracts, discussions, letters, software reviews, notes, editorials, and book reviews are excluded from the working data set.

The data were of high quality, but several typographic errors required correction. For example, several journals had more than one entry: the *American Journal of Epidemiology* was listed as both "Am. J. Epidemiology" and "AJE." Data-cleaning algorithms are available upon request.

To compare citation rates across disciplines, I mapped journals perceived to be "leading disciplinary journals" into four groups: 1) epidemiologic journals, 2) medical journals, 3) statistical and biostatistical (referred to here as "statistics") journals, and 4) social scientific journals (table 3). All other citations, including hybrid or interdisciplinary journals such as the *American Journal of Public Health* and *Social Science & Medicine*, were coded into an "all others" group. Because there is no established function that maps journals and texts to disciplines, the taxonomy used in this paper is subjective. As in previous research (24), I endeavored to enhance reliability by polling colleagues in many disciplines and letting my groups reflect consensus opinion. This approach appears superior to reliance on existing journal-type codes, which appear too broad for the purposes here (21). To assess sensitivity, I defined extremely inclusive groupings for both the social scientific and statistical citations. Among others, the *American Journal of Public Health* was added to the epidemiology journal group; *Social Science & Medicine*, law journals, and social theory text citations were included in the social science group; and the oft cited statistical software reference manual and textbook citations were added to the statistical group. Substantive findings were little changed.

RESULTS

During the 22-year period from 1981 to 2002, a total of 4,835 articles were published in AJE. Within these are 134,774 citations published in 82,874 journals or books. The mean number of articles AJE published per year was 219.77, with a standard deviation of 38.35. Regressing the number of articles published per year on year reveals a well-fit increasing linear slope of 5.13. The mean number of "research article" citations per AJE article was 27.88 (median = 26; minimum = one; maximum = 135; standard deviation = 13.54). Regression analysis shows a slightly increasing linear trend ($\beta = 0.51$) in the mean number of citations per year over the study period.

Table 1 presents the 10 most frequent citations in AJE during the study period. Breslow and Day's classic text is ranked first. Of note is that all but one of these 10 citations are statistical in nature; in fact, all but two of the top 25 citations are statistical.

TABLE 2. Twenty-five most frequently cited journals in the *American Journal of Epidemiology*, 1981–2002

Rank	Journal name	No. of times cited	% of total citations
1	*American Journal of Epidemiology*	12,576	9.33
2	*Journal of the American Medical Association*	4,242	3.15
3	*New England Journal of Medicine*	4,163	3.09
4	*Lancet*	3,329	2.47
5	*Journal of the National Cancer Institute*	2,766	2.05
6	*Journal of Clinical Epidemiology*	2,570	1.91
7	*British Medical Journal*	2,429	1.80
8	*American Journal of Public Health*	2,212	1.64
9	*American Journal of Clinical Nutrition*	1,905	1.41
10	*International Journal of Epidemiology*	1,835	1.36
11	*Circulation*	1,627	1.21
12	*International Journal of Cancer*	1,241	0.92
13	*Cancer*	1,227	0.91
14	*Annals of Internal Medicine*	1,116	0.83
15	*Journal of Infectious Disease*	977	0.72
16	*American Journal of Obstetrics and Gynecology*	953	0.71
17	*British Journal of Cancer*	884	0.66
18	*Cancer Research*	874	0.65
19	*Biometrics*	835	0.62
20	*Epidemiology*	823	0.61
21–22	*Pediatrics*	754	0.56
21–22	*Journal of the American Statistical Association*	754	0.56
23	*Archives of Internal Medicine*	746	0.55
24	*Statistics in Medicine*	713	0.53
25	*Statistical Methods in Medical Research*	699	0.52
	Total	52,250	38.77

Table 2 lists the 25 most frequently cited journals during the study period, the leading two being AJE and the *Journal of the American Medical Association* (JAMA). Together, the 25 journals account for 38.77 percent of citations. Although classification is difficult, around one third to one half of these 25 journals are medically focused, two are statistical, and none come from the social sciences.

Table 3 tabulates the four journal groups and their constituent journals. Overall, epidemiologic journals, headed by AJE and the *Journal of Clinical Epidemiology* (and the *Journal of Chronic Diseases*, which it replaced in 1985), comprised 14.5 percent of citations. The proportion of citations from medical journals was similar, with JAMA and the *New England Journal of Medicine* in the lead. AJE authors cited statistical journals considerably less. Of note is the relative frequency of biostatistical journals such as *Biometrics*, *Statistics in Medicine*, and *Statistical Methods in Medical Research*, as opposed to formal statistical journals such as the *Journal of the American Statistical Association*. Most striking is the infrequent use of social science citations—only 0.2 percent of the total. The most frequently cited social science article is Leridon's epidemiologic paper on intrauterine mortality (31), followed closely by Robinson's classic, but essentially statistical, paper on the ecologic fallacy (32).

The remaining citations comprise numerous texts and handsful of citations to (nonleading) medical, public health, psychology, and natural science journals. The median number of times any of these "other" citations were cited is unity.

What about citation trends over the study period? Taken together, the top 10 most frequently cited works show a generally decreasing citation frequency after the late 1980s, which is understandable given the lagged nature of citations. Individually, the top 10 citations show the same general decline with the exceptions of Cox's paper and Rothman's text, which continue their strong influence. In terms of journal groups, figure 1 plots group trends. There are a slight but noticeable decline in the proportion of medical journal citations and a slight increase in the proportion of epidemiologic journal citations over time. There was no obvious reason for the 1983 dip in medical citations; it does not

TABLE 3. Proportion of citations in the *American Journal of Epidemiology* by journal-group, 1981–2002

Journal-group	No. of times cited	% of group	% of total
Epidemiology			
American Journal of Epidemiology	12,576	64.17	
*Journal of Clinical Epidemiology**	2,570	13.11	
International Journal of Epidemiology	1,835	9.36	
Epidemiology	823	4.20	
Journal of Epidemiology and Community Health	624	3.18	
Annals of Epidemiology	461	2.35	
Cancer Epidemiology Biomarkers & Prevention	409	2.09	
Epidemiologic Reviews	301	1.54	
Total	19,599	100.00	14.54
Medicine			
Journal of the American Medical Association	4,242	21.19	
New England Journal of Medicine	4,163	20.80	
Lancet	3,329	16.63	
Journal of the National Cancer Institute	2,766	13.82	
British Medical Journal	2,429	12.13	
Cancer	1,227	6.13	
Annals of Internal Medicine	1,116	5.57	
Archives of Internal Medicine	746	3.73	
Total	20,018	100.00	14.85
Statistics			
Biometrics	835	20.40	
Journal of the American Statistical Association	754	18.42	
Statistics in Medicine	713	17.42	
Statistical Methods in Medical Research	699	17.08	
Journal of the Royal Statistical Society (A-D)	572	13.98	
Biometrika	425	10.38	
American Statistician	70	1.71	
International Statistical Review	25	0.61	
Total	4,093	100.00	3.04
Social science			
Demography	59	23.60	
American Sociological Review	53	21.20	
Population Studies	41	16.40	
American Journal of Sociology	36	14.40	
Sociological Methods & Research	23	9.20	
Social Forces	20	8.00	
Social Problems	6	2.40	
American Economic Review	5	2.00	
American Anthropologist	4	1.60	
Journal of Political Economy	3	1.20	
Total	250	100.00	0.19
All other citations	90,814	100.00	67.38
Total citations	134,774		100.00

* Includes *Journal of Chronic Diseases*.

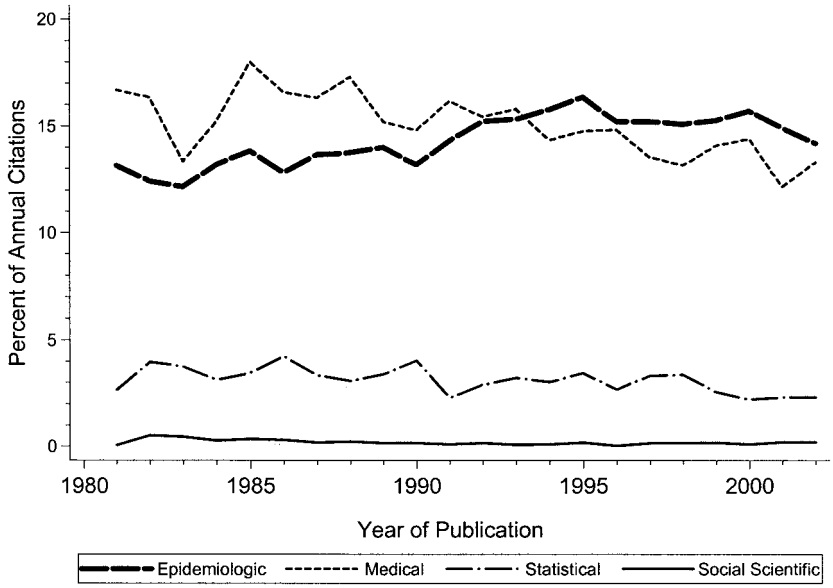

FIGURE 1. Trends in within-*American Journal of Epidemiology* citations by journal-group, 1981–2002.

appear to be a data-processing error. Statistical and social science journal citations were fairly constant.

DISCUSSION

Citation analyses are helpful in identifying the epistemological foundations of a discipline, especially as it evolves and/or confronts its own "scientific revolution" (20). Results show that, at the citation level, the most frequently cited works over the period 1981–2002 are all statistical. That said, it is important to note that data do not permit identification of actual influence, which is a complex task (33). After all, one will not appreciate Newton's influence by looking at physics citations in the last 22 years.

AJE authors rely on other AJE articles about 9 percent of the time (table 2), which is almost three times greater than the next most cited journal. This is not surprising for a leading journal in good health. Somewhat surprising is that the frequency has not changed appreciably since Dannenberg's analysis of 1974–1982 data (21); his data showed a 9.5 percent citation rate, while mine show a 9.33 percent rate. With respect to the most cited journals, my table 2 rankings differ only slightly from Hasbrouck et al.'s table 2 (22) because I excluded nonarticle citations, as discussed above.

Although my taxonomy may be debated, it seems clear that articles published in AJE rely equally on medical and epidemiologic research. While reliance on statistical journals is only about 3 percent, the most frequent citations are statistical in nature, and there are high percentages of statistical texts cited (texts are not included in journal groups). While the discipline of statistics clearly matters to epidemiology, social science is another story. I expected low citation rates but was surprised by just how low they were. Again, even an extremely inclusive definition of social science research revealed only slightly larger percentages.

Trend analysis showed remarkably consistent rates. That AJE authors are increasingly relying on epidemiologic instead of medical citations (at least as defined here) bodes well, I think, for the health of AJE as a journal and epidemiology as a discipline.

It is worth emphasizing that, except for table 1, my analyses rely on pooled data. This approach is consistent with past efforts and answers the question at hand. However, additional insight may be gained by collapsing the data to the citing AJE article (i.e., the index article) level and examining the percentage of index articles that cited research in at least one social science, statistical, medical, and epidemiologic journal within my taxonomy. Results show that 100 percent of index articles cite at least one epidemiologic journal article; 84.34 percent cite at least one medical journal article; 41.82 percent cite at least one statistics journal article; and 3.04 percent cite at least one social science journal article. No trends are apparent. Thus, the upshot of this alternative analysis is consistent with the main results.

It is worth repeating that the chief limitation of this paper is my journal taxonomy. Journal groups were subjectively defined, and other groupings might yield different results—though I doubt very different. Since many important epidemiologic studies are published in JAMA and other prestigious "medical journals" and since AJE often publishes important statistical papers, such as those by Greenland and Miettinen, classifying journals into groups is challenging.

Nevertheless, the results beg the question of why epidemiologists do not cite more social science articles. Could it be that social science has nothing to offer epidemiology? Or, are epidemiologists not aware of useful social science? Unfortunately, the reasons why scientists cite this or that work are complicated (34–36). If presented with these results, philosophers of science might wonder if epidemiology and social science are *incommensurable*. This is a term

used by Kuhn (19) to convey a sense of differing disciplinary paradigms, the upshot of which is that disciplines talk past, instead of with, each other (37). Yet, as demonstrated by the rise of molecular biology, commensurability has little to do with goals and objectives (i.e., dependent variables) and everything to do with 1) language and vocabulary and 2) standards of evidence for explanations. That social science aims to explain the social system and epidemiology aims to explain population health is thus not the problem. Further, my sense is that standards of evidence are not all that different, especially among the leading journals, and that the vocabulary is similar enough.

The social science citation rate appears to defy the facts that epidemiology is a population science and that exposures are not random but socially mediated, an issue recognized not only in the ongoing debate but also by the first epidemiologists, such as Virchow, whose slogan was "Medicine is a social science and politics nothing but medicine on a grand scale" (Virchow cited in Baldwin (38, p. 13)) (4, 10, 39). Extrapolation is always dangerous, but it seems that many epidemiologists envisage their subjects to be Robinson Crusoe, someone affected only by biologic organisms and Mother Nature, not by other individuals or strategic interaction. Even the recent efforts of social epidemiologists seem to treat "society" and "social structure" as *island weather*, influential but not something to be explained. I believe it is the absence of social science research, especially political economy, that explains the lack of a policy mandate Samet (9) and others seek.

Putting speculation properly aside, this paper shows that AJE authors consistently stand on the shoulders of other epidemiologists, clinicians, and the tools offered by statisticians; epidemiologists do not rely on social science. Just how far epidemiologists can, or will, see and do from this perch remains an issue for discussion.

REFERENCES

1. Krieger N, Zierler S. What explains the public's health?—A call for epidemiologic theory. Epidemiology 1996;7:107–9.
2. Savitz DA. The alternative to epidemiologic theory: whatever works. Epidemiology 1997;8:210–12.
3. Susser M. Does risk factor epidemiology put epidemiology at risk? Peering into the future. J Epidemiol Community Health 1998;52:608–11.
4. McMichael AJ. Prisoners of the proximate: loosening the constraints on epidemiology in an age of change. Am J Epidemiol 1999;149:887–97.
5. Rothman KJ, Adami HO, Trichopoulos D. Should the mission of epidemiology include the eradication of poverty? Lancet 1998;352:810–13.
6. Pearce N. Traditional epidemiology, modern epidemiology, and public health. Am J Public Health 1996;86:678–83.
7. Poole C, Rothman KJ. Our conscientious objection to the epidemiology wars. J Epidemiol Community Health 1998;52:613–14.
8. Weed DL. Beyond black box epidemiology. Am J Public Health 1998;88:12–14.
9. Samet JM. Epidemiology and policy: the pump handle meets the new millennium. Epidemiol Rev 2000;22:145–54.
10. Shy CM. The failure of academic epidemiology: witness for the prosecution. Am J Epidemiol 1997;145:479–84; discussion 485–7.
11. Syme SL. Rethinking disease: where do we go from here? Ann Epidemiol 1996;6:463–8.
12. Skrabanek P. The emptiness of the black box. Epidemiology 1994;5:553–5.
13. Taubes G. Epidemiology faces its limits. Science 1995;269:164–9.
14. Willett W, Greenland S, MacMahon B, et al. The discipline of epidemiology. Science 1995;269:1325–6.
15. Savitz DA, Poole C, Miller WC. Reassessing the role of epidemiology in public health. Am J Public Health 1999;89:1158–61.
16. Kuller LH. Circular epidemiology. Am J Epidemiol 1999;150:897–903.
17. Walker AM. "Kangaroo court": invited commentary on Shy's "the failure of academic epidemiology: witness for the prosecution." Am J Epidemiol 1997;145:485–6.
18. Garfield E. Citation analysis as a tool in journal evaluation. Science 1972;178:471–9.
19. Kuhn TS. The road since structure. Chicago, IL: Chicago University Press, 2000.
20. Kuhn TS. The structure of scientific revolutions. Chicago, IL: University of Chicago Press, 1970.
21. Dannenberg AL. Use of epidemiology in medical specialties: an examination by citation analysis. Am J Epidemiol 1985;121:140–51.
22. Hasbrouck LM, Taliano J, Hirshon JM, et al. Use of epidemiology in clinical medical publications, 1983–1999: a citation analysis. Am J Epidemiol 2003;157:399–408.
23. Muntaner C, Nieto FJ, O'Campo P. The Bell Curve: on race, social class, and epidemiologic research. Am J Epidemiol 1996;144:531–6.
24. LaPorte RE, Cresanta JL. Research trends in the *American Journal of Epidemiology*. Am J Epidemiol 1980;111:137–41.
25. Jones CP, LaVeist TA, Lillie-Blanton M. "Race" in the epidemiologic literature: an examination of the *American Journal of Epidemiology*, 1921–1990. Am J Epidemiol 1991;134:1079–84.
26. Comstock RD, Castillo EM, Lindsay SP. Four-year review of the use of race and ethnicity in epidemiologic and public health research. Am J Epidemiol 2004;159:611–19.
27. Krieger N. Epidemiology and social sciences: towards a critical reengagement in the 21st century. Epidemiol Rev 2000;22:155–63.
28. Berkman LF, Kawachi I. Social epidemiology. New York, NY: Oxford University Press, 2000.
29. Hamlin C. Public health and social justice in the age of Chadwick: Britain, 1800–1854. New York, NY: Cambridge University Press, 1998.
30. Merton RK. On the shoulders of giants: a Shandean postscript. New York, NY: Free Press, 1965.
31. Leridon H. Facts and artifacts in the study of intra-uterine mortality: a reconsideration from pregnancy histories. Popul Stud 1976;30:319–35.
32. Robinson WS. Ecological correlations and the behavior of individuals. Am Sociol Rev 1950;15:351–7.
33. Zuckerman H. Citation analysis and the complex problem of intellectual influence. Scientia 1987;12:329–38.
34. Case DO, Higgins GM. How can we investigate citation behavior? A study of reasons for citing literature in communication. J Am Soc Inform Sci 2000;51:635–45.
35. Merton RK. The Matthew effect in science. Science 1968;159:56–63.
36. Gilbert GN. Referencing as persuasion. Soc Stud Sci 1977;7:

113–22.
37. Godfrey-Smith P. Theory and reality: an introduction to the philosophy of science. Chicago, IL: University of Chicago Press, 2003.
38. Baldwin P. Contagion and the state in Europe, 1830–1899. Cambridge, United Kingdom: Cambridge University Press, 1999.
39. Susser M, Susser E. Choosing a future for epidemiology. II. From black box to Chinese boxes and eco-epidemiology. Am J Public Health 1996;86:674–7.