Identifying Key Statistical Papers From 1985 to 2002 Using Citation Data for Applied Biostatisticians

Michael J. SCHELL

Dissemination of ideas from theory to practice is a significant challenge in statistics. Quick identification of articles useful to practitioners would greatly assist in this dissemination, thereby improving science. This article uses the citation count history of articles to identify key papers from 1985 to 2002 from 12 statistics journals for applied biostatisticians. One feature requiring attention in order to appropriately rank an article's impact is assessment of the citation accrual patterns over time. Citation counts in statistics differ dramatically from fields such as medicine. In statistics, most articles receive few citations, with 15-year-old articles from five key journals receiving a median of 13 citations compared to 66 in the *Journal of Clinical Oncology*. However, statistics articles in the top 2%–3% continue to gain citations at a high rate past 15 years, exceeding those in JCO, whose counts slow dramatically around 8 years past publication. Articles with the highest expected applied uses 20 years post publication were identified using joinpoint regression. In this evaluation, the fraction of citations that represent applied use was defined and estimated. The false discovery rate, quantification of heterogeneity in meta-analysis, and generalized estimating equations rank as the ideas with the greatest estimated applied impact.

KEY WORDS: Applied fraction; Citation count; Statistical practice.

1. INTRODUCTION

The primary purpose of this article is to describe a surveillance approach to identify key statistical advances. As one usage of this method, the top applied articles from 12 statistical journals that are highly valued by applied biostatisticians are identified and organized into 10 analytical arenas.

Applied statisticians properly bear a number of responsibilities in their analytical work. Several of these, presented in the *Encyclopedia of Statistical Sciences* by Brian Joiner, were summarized as follows by Kenett and Thyregod (2006):

(1) have a broad knowledge and true understanding of statistical and scientific methods,
(2) be able to locate or develop good statistical procedures in a timely fashion, and
(3) be able to keep abreast of developments in statistics.

Unfortunately, there is a significant gap between these desiderata of statistical consulting practice and what actually occurs in practice. As a statistics profession, we need to place greater emphasis on reducing this gap. Doug Altman (1980, 1994), a leading voice in this endeavor, has regularly written provocative articles decrying the problem, arguing that the "misuse of statistics is unethical." In an early article on the responsibilities of a practicing statistician, W. E. Deming (1965) noted the obligation of a statistician "to apply statistical methods appropriate to the problem."

Applied statisticians often have little time to keep abreast of statistical advances, and many continue to use methods learned when they were trained, even when better methods have subsequently been developed. Moreover, the training itself often fails to teach the best practical analytical approaches, as many educators spend little of their time engaging in applied research, or if they do, suffer from time crunches themselves. Thus, a surveillance system which tracks major advances in practice can alert both the busy applied statistician and the educator of applied coursework to noteworthy statistical developments.

One example of a major advance that has lacked full appreciation is the multiple comparison procedure by Sture Holm (1979). Holm's article, published in the *Scandinavian Journal of Statistics*, has over 4000 citations, including over 1100 since 2008. Under general conditions, it improves upon the classic Bonferroni bound for multiple comparisons. In spite of the fact that it is a highly cited work, is 30 years old, and improves upon a previous bound, my discussions with recent graduates suggest that it is still not routinely taught in statistics programs. The claims for desirability of use of most statistical advances are much weaker than that of the Holm procedure, both in terms of the clarity of the advance, and ease of use. In spite of these difficulties, translation of statistical practice improvements should be important to both the developer of new statistical methods, and the applied statistician who strives to apply the strongest methods on real problems that he or she encounters.

2. CITATION COUNTS: COMPARING RESULTS FROM STATISTICS AND MEDICAL ONCOLOGY

Theoharakis and Skordia (2003) provided a ranking of statistical journals on perceived quality and relevance, based on a survey of statisticians. Among the 207 of 1149 surveyed Ph.D. academics who self-identified as biostatisticians, the 40 identified journals can be divided into the top six, the next six, and the remainder, with the breaks being the only locations in which the ranking index dropped by at least 1/3 between successive journals in the list. The Web of Science, a component of the Web of Knowledge (*www.isiknowledge.com*), provides the number of

Michael J. Schell is Senior Member and Chair, Biostatistics Department, Moffitt Cancer Center & Research Institute, University of South Florida, Tampa, FL 33612 (E-mail: *michael.schell@moffitt.org*). The author thanks Dr. Ji-Hyun Lee, the reviewers, and the editor for constructive advice, Bhupendra Rawal, Jimmy Fulp, and Rachna Kapoor with their assistance with the abstraction and analysis of data, and Jane Beley for administrative assistance.

citations that a given research article receives from other journal articles whose citations are tracked. On March 15, 2007, an initial study done to assess the distribution of citations in five of the six leading journals among biostatisticians (*Journal of the American Statistical Association* (*JASA*), *Biometrika* (*Bmka*), *Biometrics* (*Bmcs*), *Journal of the Royal Statistical Society, Series B* (*JRSS-B*), and *Statistics in Medicine* (*Stat Med*)) was obtained. (*The Annals of Statistics* was not included, as fewer applied articles were expected to be there.) Between 1985 and 2005, 11,681 research articles were published from these journals, with the number published annually ranging from 392 to 671. For comparative evaluation, 8452 articles from a leading clinical journal in which biostatisticians contribute significantly, the *Journal of Clinical Oncology* (*JCO*), were studied for the same years. Due to the time lag in accruing and databasing citations, for this section articles published in 2005 were considered to be one year since publication, up to articles from 1985 being 21 years since publication.

Piecewise linear regression (also known as joinpoint regression) models were fit for different cumulative citation percentiles separately for the statistics and *JCO* articles, where the independent variable, number of years, was defined as 2006 minus publication year. The presence of up to four joinpoints was tested for using the joinpoint method (Kim et al. 2000) and software of Kim et al. (*www.srab.cancer.gov/joinpoint/*). However, in all instances, at most a single joinpoint was needed. Figure 1 shows the year-by-year data and joinpoint fits for the statistics articles, with Table 1 providing estimated cumulative citation counts at milestone years since publication. A median article accrued 1.2 citations per year for years 1 to 8 before settling down to a total of 15 over 20 years. The 75% and 90% articles exhibited linear gains over the entire 20-year period, gaining 1.9 and 4.3 citations per year, respectively. For 97.5% or higher articles, the cumulative citation gains were linear over most of the 20-year period, before increasing; a phenomenon likely due to articles gaining a foothold in statistical practice. It is these top accruing articles which are the central focus of this article.

Table 1. Estimated cumulative citation counts of statistics journals at milestone years by percentile ranking of article.

Percentile ranking	Milestone years 1	5	10	15	20
25	0.1	2.4	4.5	5.1	5.7
50	0.9	6	10	13	15
75	3	10	20	30	39
90	4	21	43	64	85
95	4	33	69	105	141
97.5*	5	49	103	158	297
99*	9	78	164	250	546

*Year 11 excluded as an outlier.

The cumulative citation count profile is notably different for *JCO* articles. For all percentiles, the citation rate dropped between years 5 and 8 (Figure 2). While the estimated number of citations for a median *JCO* article is five times higher than for the statistics journals (66 versus 13) at 15 years, the 99th percentile counts (374 versus 310) are fairly similar. Thus, the statistical literature yields relatively few high-impact articles compared to the *Journal of Clinical Oncology*.

3. IDENTIFICATION OF THE TOP APPLIED BIOSTATISTICAL ARTICLES

3.1 Definition and Estimation of the Applied Fraction

It is more critical that the applied statistician keep abreast of statistical ideas that are ready for practical use than research ideas that essentially remain in the "statistical workshop." As articles may appear in the upper percentiles of the citation distributions while failing to achieve appreciable practical use, it

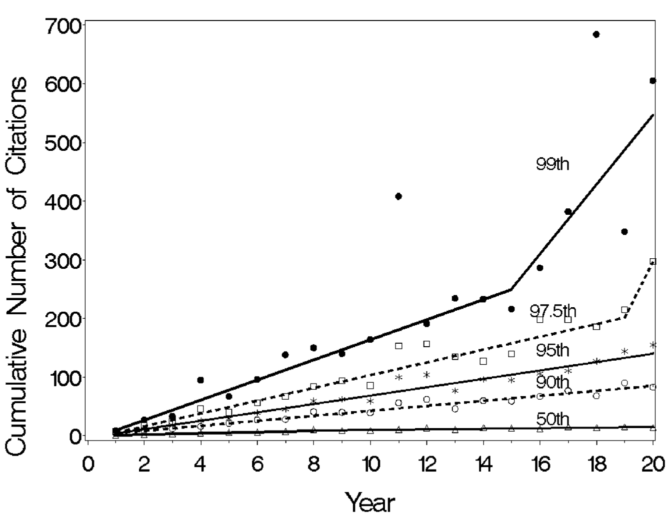

Figure 1. Plot of the cumulative number of citations by year since publication for selected percentiles (50th through 99th) for five influential statistical journals. Trends across years were fitted using joinpoint regression. For 97.5th and 99th percentile fits, year 11 was excluded as an outlier. The online version of this figure is in color.

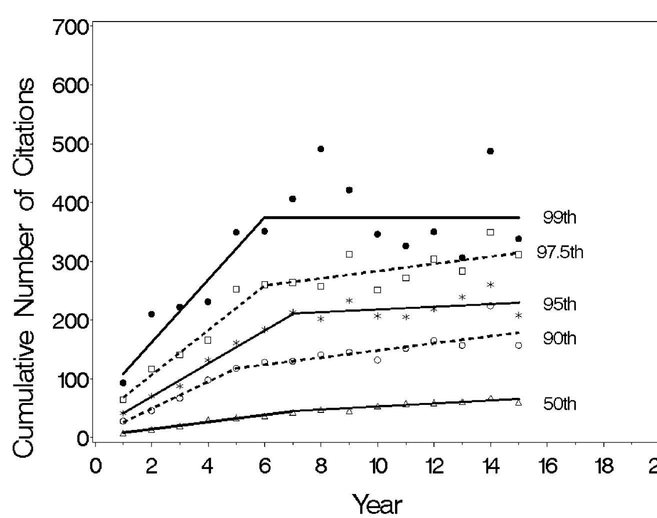

Figure 2. Plot of the cumulative number of citations by year since publication for selected percentiles (50th through 99th) for the *Journal of Clinical Oncology*. Trends across years were fitted using joinpoint regression. The 99th percentile regression fit was constrained to be monotonic. The fits go through 15 years, as the increase in citation counts in *JCO* over time renders extension beyond 15 years problematic. The online version of this figure is in color.

is useful to have a reasonable estimate of the number of applied citations over some period of time. This can be obtained by estimating the fraction of citing articles that represent applications, which will be called the *applied fraction*, and multiplying it by the number of citations in the time period.

For a given article, a list of citations from the leading journals is provided by the Web of Science, beginning with the most recent entries. We selected the 136 (1.4% of the 9926 total) statistical articles from 1985 to 2002 with ≥15 citations/year. To estimate an article's applied fraction over time, we sampled three groups of 25 citations (the Web of Science produces lists of 10, 25, or 50 citations/page) that were 1/6, 3/6, and 5/6 of the total number of citations. For instance, if an article has 305 citations, citations 51–75, 151–175, and 251–275 were examined, and the samples called "recent," "intermediate," and "early," respectively. The following two criteria were adopted for classifying an article:

(1) the title of the article addresses the applied science findings, not the statistical methods, and
(2) the article is in a nonstatistical journal.

The article was classified as "applied," "semi-applied," or "methodological" if both, one, or neither of the criteria was true. Clearly, some judgment is needed in classifying articles, especially for the first criterion. In close calls, if the keywords or abstract largely reflect statistical methodology, the article was not considered "applied." (Recently, I have identified a more objective and simpler approach that yields very similar results. One can simply exclude the articles in the subject areas "Statistics & Probability" and others that include the word "Mathematics" or "Statistics.")

Overall, the applied fractions (Table 2) from the five journals increased from 51% to 57% to 62% for early to intermediate to recent citations ($p < 0.0001$ for recent compared to both early and intermediate applied fractions; paired t-tests). Overall, for these journals, this trend demonstrates that the transition of an idea to practice takes longer to occur than does methodological follow-up. The recent applied fractions (RAF) were highest for *Statistics in Medicine* (90%), 72% for both *Biometrics* and *Biometrika*, and lowest for *JASA* (47%) and *JRSS-B* (40%). The semi-applied fractions for articles are consistently around 10% overall, in the 12–15% range for *JASA* and *JRSS-B*, usually between 7% and 9% for *Biometrics* and *Biometrika*, and about 4% for *Statistics in Medicine*.

Table 2. Average applied and semi-applied fraction percentages for early, intermediate, and recent citations by journal for highly cited articles.*

		Applied fraction %			Semi-applied fraction %		
Journal	*N*	Early	Inter.	Recent	Early	Inter.	Recent
Stat Med	27	81	85	90	4	4	3
Bmcs	27	62	71	72	13	9	8
Bmka	13	58	63	72	8	6	7
JASA	46	33	40	47	12	12	12
JRSS-B	23	32	37	40	13	14	15
Overall	136	51	57	62	11	9	9

*136 articles from 1985 to 2002 with ≥15 citations/year as assessed on 3/15/07. Inter. = Intermediate.

3.2 The Top Applied Biostatistics Articles From 1985 to 2002

In order to identify the top applied articles among the 171 articles published between 1985 and 2002 with ≥15 total citations per year as accessioned on March 2, 2009, the search was widened further on March 10, 2010 by accessioning 48 additional articles from the other seven journals with highest familiarity and perceived quality for biostatisticians (*The Annals of Statistics*, *Technometrics*, *Journal of the Royal Statistical Society, Series A* (*JRSS-A*), *The American Statistician* (*TAS*), *Journal of the Royal Statistical Society, Series C*, *Statistical Science*, and *Biostatistics*). The numbers of citations for each year since publication were obtained (in an Excel spreadsheet) from the Web of Science, and the RAF was estimated from the 25 most recent citations. The yearly citations for each article were fit using the joinpoint regression method of Kim et al. (2000), with ≤1 joinpoint for articles ≤8 years old or ≤2 joinpoints for older articles. Figure 3 shows four different patterns; the slope for Benjamini and Hochberg (1995) increased twice, with the second slope at a staggering increase of 281 citations/year. Guo and Thompson (1992) citations have shown steady growth. Phillips and Perron (1988) citations increased for seven years, stabilized, before a recent uptick. Finally, Gelfand and Smith (1990) shifted to a slow decline phase in year 9. Among the top accruing articles, types Figure 3(a) and (b) are most common. As the annual citation rate tends to increase with years since publication, a fairer comparison is to estimate the rate for a common year. The estimated applied count for year 20 (AC20), which serves as the basis for the ranking, is obtained by multiplying the RAF by the average of the estimated numbers of citations from 2008, the most recent year's data, and for year 20 obtained from the joinpoint fit. The rationale for the averaging is that extrapolating to year 20 may strain the fit, especially as the article may plateau before then.

Table 3 lists the top applied articles between 1985 and 2002. The top article is on the false discovery rate by Benjamini and Hochberg (1995). The joinpoint fit gives a year-1 estimate of 2.3 (not shown) and an initial increase by 6.2 citations/year. In year 7, the increase shifted up to 122.4 and since year 10 has increased to 280.9 citations/year. Overall, 26 articles had at least one joinpoint, with 17 having only single slope changes, perhaps indicating key dissemination timepoints to applied use.

4. PRINCIPAL RESEARCH AREAS OF THE TOP ARTICLES

Meta-Analysis. This is currently the strongest area for biostatistical practice innovations. Eight articles on meta-analytical issues are among the top articles, with four of them appearing in 2002. The Medical Research Centre in Cambridge, England is making a cottage industry of these articles with Higgins and Thompson being authors on two articles, including "Quantifying heterogeneity in a meta-analysis" which ranks second, and a third article by Parmar and colleagues. Specific topics are: quantifying and explaining heterogeneity (Rank in Table 3, #2), publication-bias (#11, 36, 53), extracting summary statistics for survival endpoints (#22), meta-regression analysis (#34, 39), and meta-analysis of cluster randomized trials (#57).

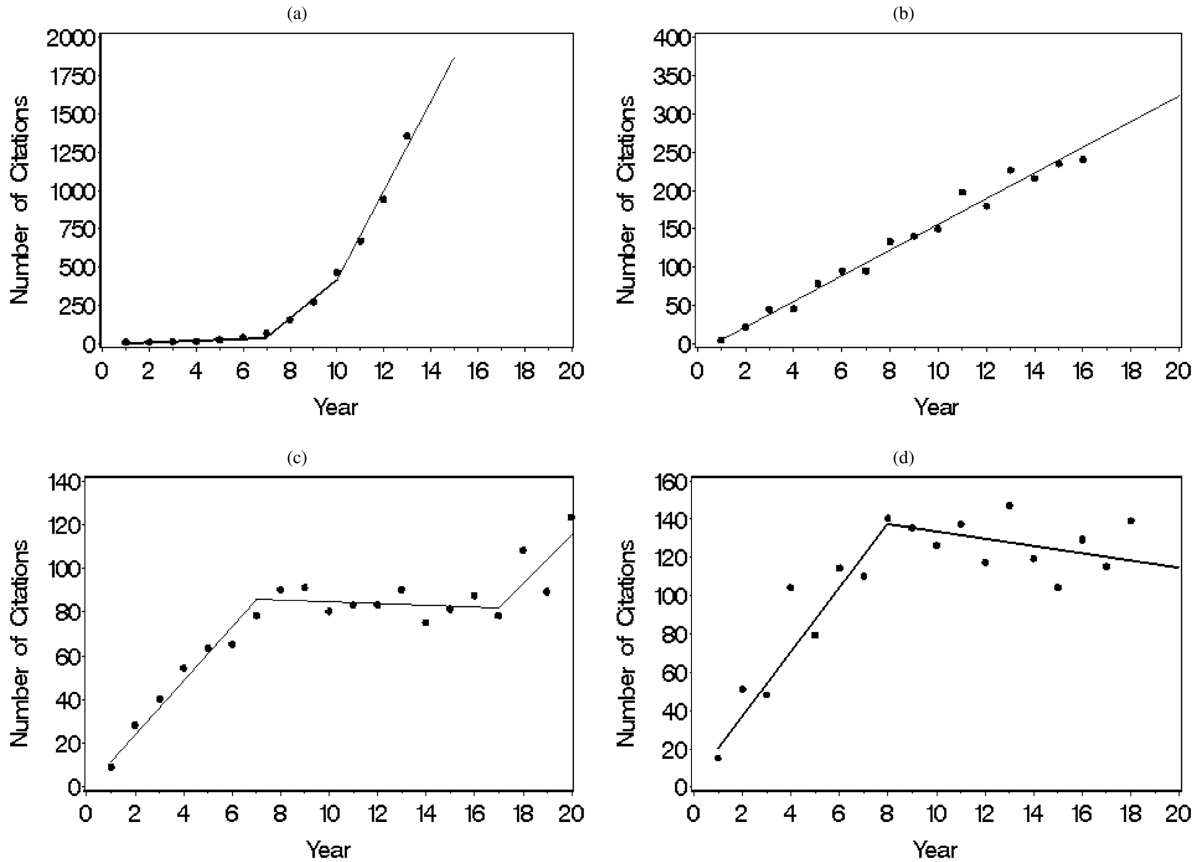

Figure 3. Annual citation counts for four highly cited statistical articles: (a) Benjamini and Hochberg (1995), (b) Guo and Thompson (1992), (c) Phillips and Perron (1988), and (d) Gelfand and Smith (1990).

Regression Fitting. While at Duke, Frank Harrell and colleagues authored the 10th ranked article by providing tools and guidelines for variable selection. Other authors (#24, 48) contributed additional insights. Jerry Friedman authored three popular *Annals* articles on flexible regression models (#28, 51, 52). Other regression-related articles in this area either add modeling flexibility (joinpoint regression (#40), nonlinear regression (#44), loess (#45), or smoothing parameter estimation (#59)) compared to simple linear regression analysis, or examined diagnostic tests for logistic regression (#47).

Correlated Data, Especially Longitudinal Data Analysis. This was arguably the major research area leading to changes in biostatistical practice in the late 1980s, led by John Hopkins University professors Liang and Zeger. Their two seminal 1986 articles on longitudinal data analysis both rank among the top 10 (#3, 9), while a third article with Albert also appeared in 1988 (#37). Other topics are: robust variance estimation for cluster-correlated data (#14), and small sample inference for fixed effects from REML (#29).

Multiple Comparisons, False Discovery Rate. The 1995 article "Controlling the false discovery rate—A practical and powerful approach to multiple testing," by Benjamini and Hochberg, both at Tel Aviv University, ranks as the top applied article. Beginning in year 10, the citations have increased by 281 per year. Clearly that pace must slow down at some point, but their 2008 count of 1249 citations is more than double the next highest total. The authors also share the 54th ranked article, Benjamini also contributed to the #12 paper, while Hochberg authored the #18 paper. J. D. Storey's (Stanford University) rival approach to false discoveries ranks as eighth best, while Sankoh et al. (#50) compared multiple adjustment methods commonly used in clinical trials.

Genetics Data. Genetic data analysis is currently a hot area for practical research. Four articles form this series, with three of them having been published from 1999 to 2002. The oldest article, "Performing the exact test of Hardy–Weinberg proportion for multiple alleles" by Guo and Thompson at the University of Washington, is the sixth highest rated article. Other contributions have been in case-control studies (#25, 46), and microarray analysis (#43).

Bayesian Methods. Bayesian methods are being used increasingly often in practical situations, as the articles in this grouping demonstrate. The article simply entitled "Bayes Factors" by Robert Kass of Carnegie Mellon and Adrian Raftery of the University of Washington has skyrocketed since 2007, resulting in a fifth-place rank. A 2002 article by Spiegelhalter from the Medical Research Centre, Cambridge and colleagues (#7) on model complexity in hierarchical models also places among the top 10. Other articles include cluster analysis issues (#27), MCMC methods (#35), and Bayesian model averaging (#41).

Table 3. Ranking of the top articles for applied biostatisticians, 1985–2002.

Rank #	First author	Journal, Year	Joinpoint fit[a]	Yrs	RAF[b]	Estimated citations[c] 2008	Year 20 Total	AC20
1	Benjamini	*JRSS-B*, 1995	6.2; 7, 122.4; 10, 280.9	13	84	1249	2232	1875
2	Higgins	*Stat Med*, 2002	48.6	6	100	244	584	584
3	Liang	*Bmka*, 1986	7.3; 5, 29.9	22	72	538	508	366
4	D'Agostino	*Stat Med*, 1998	10.2; 6, 28.4	10	100	167	309	309
5	Kass	*JASA*, 1995	14.0; 11, 59.8	13	60	282	492	295
6	Guo	*Bmcs*, 1992	16.7	16	96	256	289	278
7	Spiegelhalter	*JRSS-B*, 2002	38.6	6	56	213	482	270
8	Storey	*JRSS-B*, 2002	31.2	6	72	174	393	267
9	Zeger	*Bmcs*, 1986	3.8; 4, 13.5	22	100	256	242	242
10	Harrell	*Stat Med*, 1996	13.0	12	100	162	214	214
11	Begg	*Bmcs*, 1994	3.6; 8, 20.6	14	96	149	210	202
12	Benjamini	*AnnS*, 2001	14.4; 3, 27.2	7	60	144	320	192
13	Fine	*JASA*, 1999	4.2; 4, 9.8; 7, 23.2	9	88	90	217	191
14	Williams	*Bmcs*, 2000	0.5; 3, 17.7	8	96	90	196	189
15	DeLong	*Bmcs*, 1988	2.3; 10, 13.6; 18, 31.7	20	96	193	193	185
16	Lo	*Bmka*, 2001	4.0; 5, 16.8	7	92	50	159	147
17	Nagelkerke	*Bmka*, 1991	0.1; 5, 6.6; 13, 17.0	17	96	122	148	142
18	Hochberg	*Bmka*, 1988	5.0; 14, 14.6	20	88	154	154	136
19	Donoho	*Bmka*, 1994	17.9	14	44	237	291	128
20	van Buuren	*Stat Med*, 1999	0.0; 3, 3.8; 6, 17.3	9	80	64	159	127
21	Lin, L. I.	*Bmcs*, 1989	0.3; 4, 3.8; 15, 20.9	19	88	129	139	123
22	Parmar	*Stat Med*, 1998	3.9; 5, 11.1	10	96	72	127	122
23	Gooley	*Stat Med*, 1999	8.7	9	88	87	135	119
24	Tibshirani	*JRSS-B*, 1996	2.4; 7, 24.9; 10, 73.9	12	48	238	536	107
25	Devlin	*Bmcs*, 1999	11.7	9	64	101	165	106
26	Grambsch	*Bmka*, 1994	7.1	14	92	93	114	105
27	Fraley	*JASA*, 2002	14.7	6	56	82	185	104
28	Friedman	*AnnS*, 2000	17.1	8	44	123	226	99
29	Kenward	*Bmcs*, 1997	2.3; 4, 10.7	11	76	82	130	99
30	Newcombe	*Stat Med*, 1998	8.4	10	80	78	120	96
31	Gray	*AnnS*, 1988	1.6; 12, 9.8	20	100	95	95	95
32	Bound	*JASA*, 1995	5.3	13	100	77	95	95
33	Phillips	*Bmka*, 1988	12.4; 7, −0.4; 17,10.7	20	80	114	114	91
34	Thompson	*Stat Med*, 2002	10.0	6	72	56	126	91
35	Green	*Bmka*, 1995	12.8	13	40	158	203	81
36	Duval	*Bmcs*, 2000	6.4	8	96	46	84	81
37	Zeger	*Bmcs*, 1988	6.0	20	68	118	118	80
38	Hurvich	*Bmka*, 1989	−1.1; 3, 2.8; 13, 11.8	19	72	101	107	77
39	van Houwel	*Stat Med*, 2002	8.1	6	68	48	105	71
40	Kim	*Stat Med*, 2000	5.9	8	84	49	84	71
41	Hoeting	*SSci*, 1999	8.1	9	52	88	133	69
42	Lin, D. Y.	*JASA*, 1989	1.0; 8, 5.8	17	92	72	75	69
43	Dudoit	*JASA*, 2002	17.0	6	28	127	246	69
44	Cole	*Stat Med*, 1992	2.2; 13, 10.9	16	84	60	82	69
45	Cleveland	*JASA*, 1988	4.0	20	84	81	81	68
46	Gauderman	*Stat Med*, 2002	5.7	6	88	33	73	64
47	Hosmer	*Stat Med*, 1997	1.4; 7, 7.0	11	88	42	73	64
48	Altman	*Stat Med*, 2000	6.3	6	68	53	91	62
49	Agresti	*TAS*, 1998	6.0	10	68	57	87	59
50	Sankoh	*Stat Med*, 1997	0.6; 4, 5.1	11	96	39	61	59
51	Friedman	*AnnS*, 2001	8.4	7	56	50	105	59
52	Friedman	*AnnS*, 1991	5.6	17	52	104	113	59
53	Macaskill	*Stat Med*, 2001	4.9	7	92	31	63	58
54	Hochberg	*Stat Med*, 1990	0.6; 11, 6.3	18	96	54	60	58
55	Metz	*Stat Med*, 1998	4.0	10	76	50	70	53
56	Besag	*JRSS-B*, 1986	5.8	22	44	123	117	51
57	Donner	*Stat Med*, 2002	5.2	6	88	22	58	51
58	Rubin	*JASA*, 1996	5.1	12	64	59	80	51
59	Wood	*JRSS-B*, 2000	5.3	8	72	38	70	50

[a]The joinpoint fit gives the initial slope of increase in annual citations, and years and slopes at significant shifts.

[b]RAF = "recent applied fraction," defined in Section 3.1.

[c]Estimated number of citations for 2008 from joinpoint fit; estimated year-20 total = (2008 total + year-20 joinpoint estimate)/2; AC20 = RAF/100 × the estimated year-20 total.

Survival Analysis. The key contributions in survival methods are to: competing risks (#13, 23, 31), multiple imputation for missing data (#20), diagnostics (#26), and robust inference (#42).

ROC Curves. The two articles in this group involve comparing correlated ROC curves (#15) and fitting ROC curves to continuously-distributed data (#55).

Time Series. A pair of articles from the late 1980s made key contributions to time series unit-root testing (#33) and model selection in small samples (#38).

Other. Ten articles did not easily fall into one of the areas above. Ralph D'Agostino Jr. addressed bias reduction for propensity scores with the fourth ranked article. The other articles addressed: testing for the number of components in a normal mixture (#16), the coefficient of determination (#17), wavelets (#19), a concordance correlation-coefficient (#21), 2-sided confidence intervals for the single proportion (#30), instrumental variables (#32), approximate binomial confidence intervals (#49), image analysis (#56), and multiple imputation (#58).

While the top 59 articles are at least 6 years old, seven articles from 2003 to 2004 would place among them. The topics, first authors, and AC20 counts are: a background adjustment for microarrays (Wu et al. 2004; 249); spatially balanced sampling (Stevens and Olsen 2004; 101); continuity corrections in meta-analysis (Sweeting, Sutton, and Lambert 2004; 101); relative survival regression models (Dickman et al. 2004; 73); a measure of discrimination in survival analysis (Pencina and D'Agostino 2004; 56); a breast cancer prediction model (Tyrer, Duffy, and Cuzick 2004; 55); and multiple smoothing parameter estimation for generalized additive models (Wood 2004; 55). These articles are worthy of consideration.

5. DISCUSSION

The top statistical articles from 1985 to 2002 from the 12 statistics journals with the highest familiarity and perceived quality for applied biostatisticians (Theoharakis and Skordia 2003) have been identified using a citation-based surveillance method. Ranking the applied impact of articles is admittedly imperfect, with a number of decisions affecting the assessment of an article's long-term impact. The citation count is clearly not a direct measure of the intrinsic significance of a research idea; it does, however, provide a measure of statistical "technology transfer" (Altman and Goodman 1994) of its impact. In spite of its limitations, the surveillance method proposed here seems quite useful in identifying articles expected to achieve high applied use and worthy of promulgation to practicing statisticians.

The list of top applied articles provided here is most targeted for applied biostatisticians, using the 12 statistical journals rated most highly by them. Overall, 11 of these journals also rank among the top 12 in the Index of familiarity and rank for all statisticians in the survey (Theoharakis and Skordia 2003). However, these 12 do not include five of the top 10 journals ranked by econometricians, six of the top 10 for stochastic process statisticians, and three of 10 for survey methodologists. The most time-consuming aspect of this research involves obtaining the joinpoint fits for the articles and recent applied fraction assessment. A quick approximation to this list can be obtained from the average number of citations per year − 1 (AvgCit). The list includes all five *JRSS-B* articles with AvgCit ≥ 55, five of nine *JASA* articles with AvgCit ≥ 44, five of six *Annals* articles with AvgCit ≥ 34, and all 17 *Bmcs/Bmka* articles and 11 of 12 *Stat Med* articles with AvgCit ≥ 30. Eight of 24 additional *Stat Med* papers have 15 ≤ AvgCit ≤ 30. The other journals contained at most one article in the highly cited applied articles list. Using such a quick approximation approach, articles can be identified from a given journal within minutes, and can be used for specialized topics.

Different fields of study and journals used citation data to identify their most influential articles (van Dalen and Henkens 2001; Aylward et al. 2008), typically using the total number of citations without accounting for time since publication. The prototypical article in science gains citations fairly quickly and then damps down, although *The Annals of Statistics* has been noted as an exception (Glanzel and Moed 2002). The surveillance method presented here accounts for time since publication, which is particularly important for fields whose citation rates increase for many years. As the relative impact of articles such as those presented here surely changes with time, the list will be updated periodically on *www.beststatisticalpractices.org*. The use of a surveillance method can alert practicing statisticians to statistical advances; the method presented here gleans them from the upper percentiles of the citation distributions. To achieve the highest percentiles in the first place, though, the article needed to attract a following. How can this be done? Here are my suggestions. Keep a strong focus on the needs of the potential adopter of the method. As much as possible, simplify the approach. If existing methods are currently in use, clearly show how the new method is better. Identify available software, and actively work with major software developers to incorporate the method. Finally, advertise the advance through lectures or short courses and/or be an early user of it.

If the ultimate value of a statistical idea is its application to real data, what can be said about the vast majority of articles that fail to achieve appreciable applied usage? Perhaps they achieved small niches in the marketplace of ideas. They might be rivulets in a stream of ideas that ultimately leads to a major breakthrough, and those researchers must be content to have fed that stream. Others will be forgotten altogether. Better partnering between methodologists and practitioners, though, may improve the success rates of ideas and benefit society through their usage.

[Received November 2008. Revised October 2010.]

REFERENCES

Agresti, A., and Coull, B. A. (1998), "Approximate Is Better Than 'Exact' for Interval Estimation of Binomial Proportions," *The American Statistician*, 52, 119–126. [314]

Altman, D. G. (1980), "Statistics and Ethics in Medical Research, Misuse of Statistics Is Unethical," *British Medical Journal*, 281, 1267–1269. [310]

——— (1994), "The Scandal of Poor Medical Research," *British Medical Journal*, 308, 283–284. [310]

Altman, D. G., and Goodman S. N. (1994), "Transfer of Technology From Statistical Journals to the Biomedical Literature—Past Trends and Future Predictions," *Journal of the American Medical Association*, 272 (2), 129–132. [315]

Altman, D. G., and Royston, P. (2000), "What Do We Mean by Validating a Prognostic Model?" *Statistics in Medicine*, 19 (4), 453–473. [314]

Aylward, B. S., Roberts, M. C., Colombo, J., and Steel, R. G. (2008), "Identifying the Classics, An Examination of Articles Published in the *Journal of Pediatric Psychology* From 1976–2006," *Journal of Pediatric Psychology*, 33 (6), 576–589. [315]

Begg, C. B., and Mazumdar, M. (1994), "Operating Characteristics of a Bank Correlation Test for Publication Bias," *Biometrics*, 50, 1008–1101. [314]

Benjamini, Y., and Hochberg, Y. (1995), "Controlling the False Discovery Rate—A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society, Ser. B*, 57 (1), 289–300. [312-314]

Benjamini, Y., and Yekutieli, D. (2001), "The Control of the False Discovery Rate in Multiple Testing Under Dependency," *The Annals of Statistics*, 29 (4), 1165–1188. [314]

Besag, J. (1986), "On the Statistical-Analysis of Dirty Pictures," *Journal of the Royal Statistical Society, Ser. B*, 48 (3), 259–302. [314]

Bound, J., Jaeger, D., and Baker, M. (1995), "Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak," *Journal of the American Statistical Association*, 90, 443–450. [314]

Cleveland, W., and Devlin, S. (1988), "Locally Weighted Regression—An Approach to Regression Analysis by Local Fitting," *Journal of the American Statistical Association*, 83, 596–610. [314]

Cole, T. J., and Green, P. J. (1992), "Smoothing Reference Centile Curves—The LMS Method and Penalized Likelihood," *Statistics in Medicine*, 11 (10), 1305–1319. [314]

D'Agostino, R. B. (1998), "Propensity Score Methods for Bias Reduction in the Comparison of a Treatment to a Non-Randomized Control Group," *Statistics in Medicine*, 17, 2265–2281. [314]

DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988), "Comparing the Areas Under 2 or More Correlated Receiver Operating Characteristic Curves—A Nonparametric Approach," *Biometrics*, 44, 837–845. [314]

Deming, W. E. (1965), "Principals of Professional Statistical Practice," *Annals of Mathematical Statistics*, 36, 1883–1900. [310]

Devlin, B., and Roeder, K. (1999), "Genomic Control for Association Studies," *Biometrics*, 55 (4), 997–1004. [314]

Dickman, P. W., Sloggett, A., Hills, M., and Hakulinen, T. (2004), "Regression Models for Relative Survival," *Statistics in Medicine*, 23 (1), 51–64. [315]

Donner, A., and Klar, N. (2002), "Issues in the Meta-Analysis of Cluster Randomized Trials," *Statistics in Medicine*, 21 (19), 2971–2980. [314]

Donoho, D. L., and Johnstone, I. M. (1994), "Ideal Spatial Adaptation by Wavelet Shrinkage," *Biometrika*, 81 (3), 425–455. [314]

Dudoit, S., Fridlyand, J., and Speed, T. (2002), "Comparison of Discrimination Methods for the Classification of Tumors Using Gene Expression Data," *Journal of the American Statistical Association*, 97 (457), 77–87. [314]

Duval, S., and Tweedie, R. (2000), "Trim and Fill: A Simple Funnel-Plot-Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis," *Biometrics*, 56 (2), 455–463. [314]

Fine, J. P., and Gray, R. J. (1999), "A Proportional Hazards Model for the Subdistribution of a Competing Risk," *Journal of the American Statistical Association*, 94 (446), 496–509. [314]

Fraley, C., and Raftery, A. E. (2002), "Model-Based Clustering, Discriminant Analysis, and Density Estimation," *Journal of the American Statistical Association*, 97 (458), 611–631. [314]

Friedman, J., Hastie, T., and Tibshirani, R. (2000), "Additive Logistic Regression: A Statistical View of Boosting," *The Annals of Statistics*, 28 (2), 337–374. [314]

Friedman, J. H. (1991), "Multivariate Adaptive Regression Splines," *The Annals of Statistics*, 19 (1), 1–67. [314]

——— (2001), "Greedy Function Approximation: A Gradient Boosting Machine," *The Annals of Statistics*, 29 (5), 1189–1232. [314]

Gauderman, W. J. (2002), "Sample Size Requirements for Matched Case-Control Studies of Gene-Environment Interaction," *Statistics in Medicine*, 21 (1), 35–50. [314]

Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85 (410), 398–409. [312,313]

Glanzel, W., and Moed, H. K. (2002), "Journal Impact Measures in Bibliometric Research," *Scientometrics*, 53 (3), 171–193. [315]

Gooley, T. A., Leisenring, W., and Storer, B. E. (1999), "Estimation of Failure Probabilities in the Presence of Competing Risks, New Representations of Old Estimators," *Statistics in Medicine*, 18 (6), 695–706. [314]

Grambsch, P. M., and Therneau, T. M. (1994), "Proportional Hazards Tests and Diagnostics Based on Weighted Residuals," *Biometrika*, 81 (3), 515–526. [314]

Gray, R. J. (1988), "A Class of k-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk," *The Annals of Statistics*, 16 (3), 1141–1154. [314]

Green, P. J. (1995), "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, 82 (4), 711–732. [314]

Guo, S. W., and Thompson, E. A. (1992), "Performing the Exact Test of Hardy–Weinberg Proportion for Multiple Alleles," *Biometrics*, 48 (2), 361–372. [312-314]

Harrell, F. E., Lee, K. L., and Mark, D. B. (1996), "Multivariable Prognostic Models, Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors," *Statistics in Medicine*, 15 (4), 361–387. [314]

Higgins, J. P. T., and Thompson, S. G. (2002), "Quantifying Heterogeneity in a Meta-Analysis," *Statistics in Medicine*, 21 (11), 1539–1558. [314]

Hochberg, Y. (1988), "A Sharper Bonferroni Procedure for Multiple Tests of Significance," *Biometrika*, 75 (4), 800–802. [314]

Hochberg, Y., and Benjamini, Y. (1990), "More Powerful Procedures for Multiple Significance Testing," *Statistics in Medicine*, 9 (7), 811–818. [314]

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999), "Bayesian Model Averaging: A Tutorial," *Statistical Science*, 14 (4), 382–401. [314]

Holm, S. (1979), "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics*, 6 (2), 65–70. [310]

Hosmer, D. W., Hosmer, T., leCessie, S., and Lemeshow, S. (1997), "A Comparison of Goodness-of-Fit Tests for the Logistic Regression Model," *Statistics in Medicine*, 16 (9), 965–980. [314]

Hurvich, C. M., and Tsai, C. L. (1989), "Regression and Time-Series Model Selection in Small Samples," *Biometrika*, 76 (2), 297–307. [314]

Kass, R. E., and Raftery, A. E. (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90 (430), 773–795. [314]

Kenett, R., and Thyregod P. (2006), "Aspects of Statistical Consulting Not Taught by Academia," *Statistics Neerlandica*, 60, 396–411. [310]

Kenward, M. G., and Roger, J. H. (1997), "Small Sample Inference for Fixed Effects From Restricted Maximum Likelihood," *Biometrics*, 53 (3), 983–997. [314]

Kim, H. J., Fay, M. P., Feuer, E. J., and Midthune, D. N. (2000), "Permutation Tests for Joinpoint Regression With Applications to Cancer Rates," *Statistics in Medicine*, 19 (3), 335–351. [311,312,314]

Liang, K. Y., and Zeger, S. L. (1986), "Longitudinal Data-Analysis Using Generalized Linear-Models," *Biometrika*, 73 (1), 13–22. [314]

Lin, D. Y., and Wei, L. J. (1989), "The Robust Inference for the Cox Proportional Hazards Model," *Journal of the American Statistical Association*, 84 (408), 1074–1078. [314]

Lin, L. I. (1989), "A Concordance Correlation-Coefficient to Evaluate Reproducibility," *Biometrics*, 45 (1), 255–268. [314]

Lo, Y. T., Mendell, N. R., and Rubin, D. B. (2001), "Testing the Number of Components in a Normal Mixture," *Biometrika*, 88 (3), 767–778. [314]

Macaskill, P., Walter, S. D., and Irwig, L. (2001), "A Comparison of Methods to Detect Publication Bias in Meta-Analysis," *Statistics in Medicine*, 20 (4), 641–654. [314]

Metz, C. E., Herman, B. A., and Shen, J. H. (1998), "Maximum Likelihood Estimation of Receiver Operating Characteristic (ROC) Curves From Continuously-Distributed Data," *Statistics in Medicine*, 17 (9), 1033–1053. [314]

Nagelkerke, N. J. D. (1991), "A Note on a General Definition of the Coefficient of Determination," *Biometrika*, 78 (3), 691–692. [314]

Newcombe, R. G. (1998), "Two-Sided Confidence Intervals for the Single Proportion, Comparison of Seven Methods," *Statistics in Medicine*, 17 (8), 857–872. [314]

Parmar, M. K. B., Torri, V., and Stewart, L. (1998), "Extracting Summary Statistics to Perform Meta-Analyses of the Published Literature for Survival Endpoints," *Statistics in Medicine*, 17 (24), 2815–2834. [314]

Pencina, M. J., and D'Agostino, R. B. (2004), "Overall C as a Measure of Discrimination in Survival Analysis: Model Specific Population Value and Confidence Interval Estimation," *Statistics in Medicine*, 23 (13), 2109–2123. [315]

Phillips, P. C. B., and Perron, P. (1988), "Testing for a Unit-Root in Time-Series Regression," *Biometrika*, 75 (2), 335–346. [312-314]

Rubin, D. B. (1996), "Multiple Imputation After 18+ Years," *Journal of the American Statistical Association*, 91 (434), 473–489. [314]

Sankoh, A. J., Huque, M. F., and Dubey, S. D. (1997), "Some Comments on Frequently Used Multiple Endpoint Adjustment Methods in Clinical Trials," *Statistics in Medicine*, 16 (22), 2529–2542. [314]

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002), "Bayesian Measures of Model Complexity and Fit," *Journal of the Royal Statistical Society, Ser. B*, 64, 583–616. [314]

Stevens, D. L., and Olsen, A. R. (2004), "Spatially Balanced Sampling of Natural Resources," *Journal of the American Statistical Association*, 99 (465), 262–278. [315]

Storey, J. D. (2002), "A Direct Approach to False Discovery Rates," *Journal of the Royal Statistical Society, Ser. B*, 64, 479–498. [314]

Sweeting, M. J., Sutton, A. J., and Lambert, P. C. (2004), "What to Add to Nothing? Use and Avoidance of Continuity Corrections in Meta-Analysis of Sparse Data," *Statistics in Medicine*, 23 (9), 1352–1375. [315]

Theoharakis, V., and Skordia, M. (2003), "How Do Statisticians Perceive Statistics Journals?" *The American Statistician*, 57 (2), 115–123. [310,315]

Thompson, S. G., and Higgins, J. P. T. (2002), "How Should Meta-Regression Analyses Be Undertaken and Interpreted?" *Statistics in Medicine*, 21 (11), 1559–1573. [314]

Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society, Ser. B*, 58 (1), 267–288. [314]

Tyrer, J., Duffy, S. W., and Cuzick, J. (2004), "A Breast Cancer Prediction Model Incorporating Familial and Personal Risk Factors," *Statistics in Medicine*, 23 (7), 1111–1130. [315]

van Buuren, S., Boshuizen, H. C., and Knock, D. L. (1999), "Multiple Imputation of Missing Blood Pressure Covariates in Survival Analysis," *Statistics in Medicine*, 18 (6), 681–694. [314]

van Dalen, H. P., and Henkens, K. (2001), "What Makes a Scientific Article Influential? The Case of Demographers," *Scientometrics*, 50 (3), 455–482. [315]

van Houwelingen, H. C., Arends, L. R., and Stijnen, T. (2002), "Advanced Methods in Meta-Analysis, Multivariate Approach and Meta-Regression," *Statistics in Medicine*, 21 (4), 589–624. [314]

Williams, R. L. (2000), "A Note on Robust Variance Estimation for Cluster-Correlated Data," *Biometrics*, 56 (2), 645–646. [314]

Wood, S. N. (2000), "Modelling and Smoothing Parameter Estimation With Multiple Quadratic Penalties," *Journal of the Royal Statistical Society, Ser. B*, 62, 413–428. [314]

Wood, S. N. (2004), "Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models," *Journal of the American Statistical Association*, 99 (467), 673–686. [315]

Wu, Z. J., Irizarry, R. A., Gentleman, R., Martinez-Murillo, F., and Spencer, F. (2004), "A Model-Based Background Adjustment for Oligonucleotide Expression Arrays," *Journal of the American Statistical Association*, 99 (468), 909–917. [315]

Zeger, S. L., and Liang, K. Y. (1986), "Longitudinal Data-Analysis for Discrete and Continuous Outcomes," *Biometrics*, 42 (1), 121–130. [314]

Zeger, S. L., Liang, K. Y., and Albert, P. S. (1988), "Models for Longitudinal Data—A Generalized Estimating Equation Approach," *Biometrics*, 44 (4), 1049–1060. [314]