What Your Doctor Should Know about Statistics
(but Perhaps Doesn't...)

Suzanne S. Switzer and Nicholas J. Horton

As students and teachers of statistics, we are curious about how statistics is applied in other fields. The frequency of use of statistical methods in *The New England Journal of Medicine* (*NEJM*) has been studied by J. D. Emerson and G. A. Colditz, who wrote an article titled "Use of Statistical Analysis in *The New England Journal of Medicine*" that was published in *NEJM*. We recently updated their work to address the question of what statistical techniques are commonly used in research articles 15 years later. What follows is a summary of our findings.

The main findings are in Table 1, which displays the frequency of methods used in original articles, as opposed to editorials or reviews, published during 1978–1979, 1989, and 2004–2005. The 1978–1979 and 1989 columns are from Emerson and Colditz, who originally defined the categories; the 2004–2005 column is our work.

In the most recent period, the percentage of articles containing no statistical methods or only simple descriptive statistics did not substantially change from the 1989 survey. A number of categories remained relatively constant, or usage decreased slightly between 1989 and 2004–2005.

There were substantial increases in the percentages of articles using contingency tables (53%), epidemiological statistics (35%), survival methods (61%), multiple regression (51%), multiple comparisons (23%), and power (39%). Fifteen articles utilized count models (Poisson/negative binomial regression), while 70 articles utilized logistic regression models.

The "accumulated by" column describes how many articles were "fully accessible" with knowledge of that method and all the preceding methods (using a hierarchy established by Emerson and Colditz). Of particular note is the entry for multiple regression. When knowledge of this method is added, the percentage of "accessible" articles jumps from 24%–39% (see Figure 1).

Table 2 describes the increase in the number of methods used over time. For papers that used a method beyond descriptive statistics, the number of methods used per paper has more than doubled over the past 25 years.

Consistent with prior research, we found there was a continued trend toward increased use of newer and more sophisticated statistical methods by journal authors. Readers with knowledge of only the topics typically included in introductory statistics courses may not fully comprehend a large fraction of the statistical content of original articles in the *NEJM*. We concur with the conclusion of Emerson and Colditz that "an acquaintance with a few basic statistical techniques cannot give full statistical access to research appearing in the journal."

As an example of the challenges in interpreting new research results and the need to have a solid understanding of statistics, we consider assessing the risk profile of Vioxx, a prescription medication used to treat pain associated with arthritis. A randomized control study regarding the use of rofecoxib, the scientific name for Vioxx, was published by *NEJM* in November of 2000, shortly after the drug was approved

Table 1—Statistical Content of *The New England Journal of Medicine* Original Articles over Time

Procedure*	**Notes**	**Original Articles Containing Methods 1978–1979 n=332**	**Original Articles Containing Methods 1989 n=115**	**Original Articles Containing Methods 2004–2005 n=311**	**Accumulated by Article** 2004–2005 %**
No statistical methods, or descriptive statistics only	Summary statistics and confidence intervals	27%	12%	13%	13%
t-tests		44%	39%	26%	14%
Contingency tables		27%	36%	53%	15%
Nonparametric tests		11%	21%	27%	17%
Epidemiological statistics	Relative risk, measures of association, sensitivity, etc.	9%	22%	35%	18%
Pearson correlation		12%	19%	3%	18%
Simple linear regression	Regression with one predictor and one response variable	8%	9%	6%	18%
Analysis of variance		8%	20%	16%	20%
Transformation		7%	7%	10%	20%
Nonparametric correlation		4%	1%	5%	21%
Survival methods	Includes logistic regression	11%	32%	61%	24%
Multiple regression	Includes stepwise regression and smoothing	5%	14%	51%	39%
Multiple comparisons	Includes interim analyses	3%	9%	23%	41%
Adjustment and standardization		3%	9%	1%	41%
Multiway tables	Mantel-Haenszel procedure, log-linear models	4%	10%	13%	44%
Power	Includes sample-size calculations	3%	3%	39%	68%
Cost-benefit analysis		1%	0%	0%	68%
Sensitivity analysis***		0%	0%	6%	72%
Repeated measures analysis*	Includes longitudinal and clustered regression	-	-	12%	80%
Missing data methods		-	-	8%	87%
Noninferiority trials	Equivalence trials			4%	91%
Receiver operating characteristic	Area under curve/ sensitivity and specificity	-	-	2%	93%
Resampling inference		-	-	2%	94%
Principal component/ factor/cluster analysis		-	-	2%	96%
Other methods****		3%	9%	4%	100%

* Any item in bold print indicates changes from the previous descriptions utilized by Emerson and Colditz (1992).
** The accumulation by article indicates the number and percentage of articles including only that procedure and those above it (nothing used from further down in the hierarchy).
*** Sensitivity analysis was not reported for the 1978–1979 and 1989 survey of Original Articles, so a value of 0 was assumed.
**** Methods listed as “other” include family-based association tests (n=2), Hardy-Weinberg equilibrium (2), meta-analysis (2), bagging algorithm, Christmas tree correction (1), event charts (1), Markov chain random walk algorithm (1), Monte Carlo simulation (1), propensity score (1), and scan statistics (1).

Table 2—Average Number of Statistical Methods Used per Article

	1978–1979	1989	2004–2005
Articles that used methods beyond descriptive statistics	241	101	272
Article method uses	546	297	1271
Average method uses	2.3	2.9	4.7

for use by the FDA. The study was designed to compare the effect of Vioxx on gastrointestinal (GI) toxicity—a potentially life-threatening outcome—to that of a competing drug (naproxen). The researchers found that Vioxx greatly reduced the risk of GI problems (relative risk of 0.5; 95% confidence interval = 0.3–0.6). The study also reported the incidence of myocardial infarction (heart attack) in patients, noting "the incidence of myocardial infarction was lower among patients in the naproxen group than among those in the rofecoxib group (0.1% vs. 0.4%; relative risk, 0.2; 95% confidence interval, 0.1–0.7)." Among these subjects who had rheumatoid arthritis but few other complications (i.e., no unstable medical condition, history of cancer, substance abuse, cerebrovascular events, myocardial infarction, or coronary bypass), the risk of having a heart attack for those subjects taking Vioxx was nearly five times that for those taking naproxen.

While doctors judge risks based on their severity and how often they occur, a five-fold risk for heart attack typically attracts notice. Was this result understood to be an acceptable risk, justifiable given the benefit of decreased gastrointestinal events? Would readers have made a different judgment if the relative risk had been reported as 5.0 (comparing Vioxx to naproxen), as opposed to 0.2 (comparing naproxen to Vioxx)? These values are mathematically equivalent, but they may be perceived differently. In any case, Vioxx was pulled off the market by Merck in the fall of 2004, due in large part to this increased risk of heart attack, and numerous lawsuits have since ensued.

Many research studies require sophisticated statistical methods to fully describe a complicated state of nature. Are these subtleties fully comprehended, given the competing demands on a clinician's time? Do their eyes glaze over when they reach the methods and results sections? As some doctors have confessed to us, they may read only the abstract (or title) as a way to keep track of current research. It is important to realize that many times the information revealed in those short summaries will be incomplete. For the Vioxx example, the title was "Comparison of Upper Gastrointestinal Toxicity of Rofecoxib and Naproxen in Patients with Rheumatoid Arthritis" and the entire conclusion of the abstract was, "In patients with rheumatoid arthritis, treatment with rofecoxib, a selective inhibitor of cyclooxygenase-2, is associated with significantly fewer clinically important upper GI events than treatment with naproxen, a nonselective inhibitor." Neither the title nor the abstract conclusion hinted at the increased myocardial infarction risk.

THE LANCET]　　　　MATHEMATICS

THE LANCET

LONDON: SATURDAY, JANUARY 2, 1937

MATHEMATICS AND MEDICINE

STATISTICS are curious things. They afford one of the few examples in which the use, or abuse, of mathematical methods tends to induce a strong emotional reaction in non-mathematical minds. This is because statisticians apply, to problems in which we are interested, a technique which we do not understand. It is exasperating, when we have studied a problem by methods that we have spent laborious years in mastering, to find our conclusions questioned, and perhaps refuted, by someone who could not have made the observations himself. It requires more equanimity than most of us possess to acknowledge that the fault is in ourselves. A colleague of a famous contemporary statesman has been quoted as complaining that he used figures as though they were adjectives. The medical profession as a whole would not find it easy to counter a similar charge. For most of us figures impinge on an educational blind spot. We have never been taught to recognise them for what they are. This is a misfortune, because simple statistical methods concern us far more closely than many of the things that we are forced to learn in the six long years of the medical curriculum. Many of our problems *are* statistical; and there is no other way of dealing with them. In preventive medicine this is so obvious that it has acquired general recognition. In laboratory work, though recognition has come more slowly, it is now widely realised that it is very unsafe to base conclusions on statistically inadequate data. In clinical medicine recognition is coming more slowly still; so slowly that many avoidable errors, and a sad waste of material, still hinder progress.

The kind of problem that demands statistical treatment is familiar to all of us. What is the therapeutic value of antipneumococcal or of anti-streptococcal serum? What is the expectation of life of patients who are operated on for carcinoma at various sites and in various stages of development? What are the relative merits of artificial pneumothorax or thoracoplasty in the treatment of phthisis? In all such problems, and they could be multiplied ad nauseam, the unknown, or uncontrollable, variables are so many that no sound conclusion can be based on the observation of a few cases, or on the uncritical assessment of many. Clinical science, like any other, is essentially experimental. We do things to our patients and we see what happens. It has become a truism that the history of a science is the history of its technique; and the technique of seeing what happens when we attempt to control disease is as yet very poorly developed. Because it is poorly developed we still do many things that we ought not to do,

Text from *The Lancet*, 1937

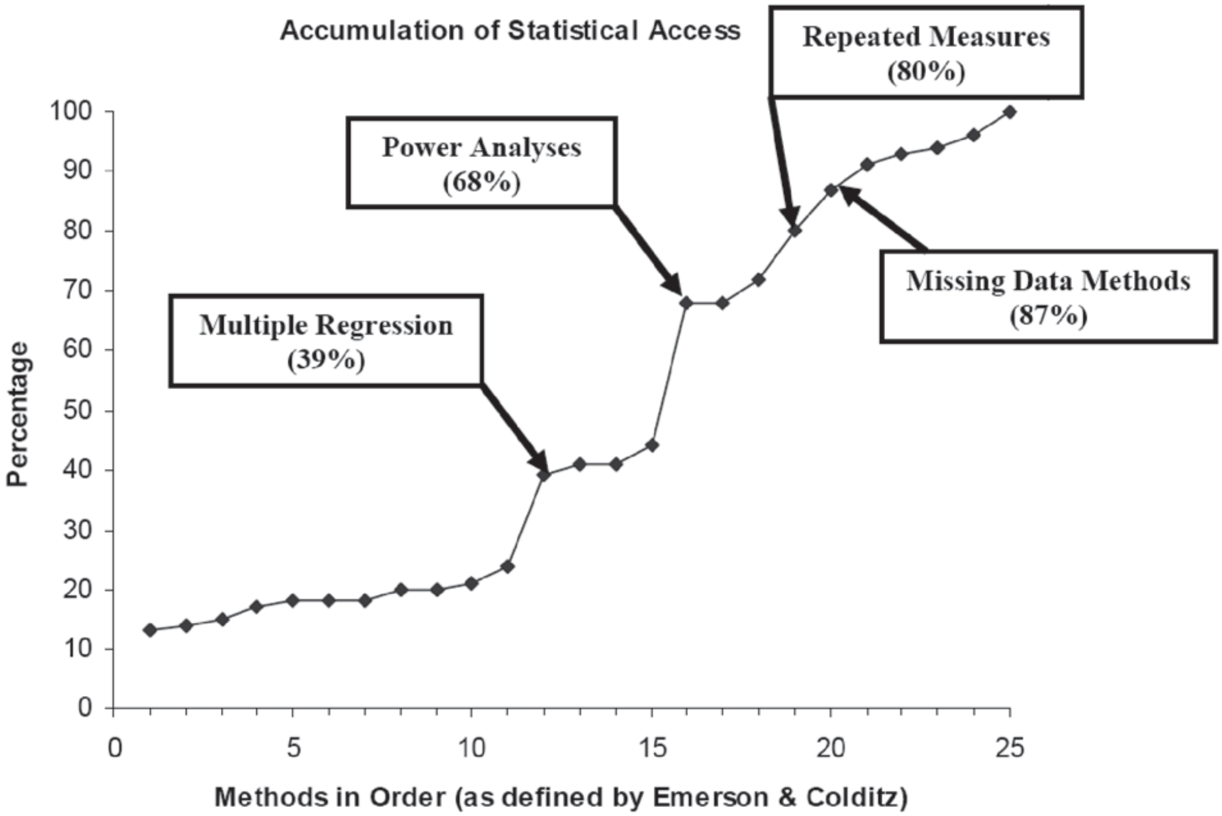

Figure 1. Accumulated by percentage as a function of methods

We do note that some of the changes we report over time may be driven by improvements in clinical trial reporting. The medical profession has widely adopted standard guidelines for reporting clinical trials. Those guidelines arise from the CONSORT statement (*www.consort-statement.org*), a checklist to improve the reporting of study design, conduct, analysis, and interpretation of randomized clinical trials. The checklist specifies that power calculations and multiple comparison procedures be specified. The adoption of CONSORT guidelines is likely to have increased the mention of these items in articles more recently published. Furthermore, advances in statistical software may have contributed to some of the variation in statistical methods used over time, as the software now supports and contributes to the ease of carrying out complex analyses. The statistical methods used in handling repeated measures and missing data, for instance, were unavailable or relatively undeveloped at the time of the previous two reviews, and articles were therefore less likely to use them.

The comprehension of current research in the medical field is of particular importance to physicians. In order to comprehend, physicians must understand the underlying statistical concepts used to quantify the effectiveness of new treatments or medications. Our findings indicate that complex statistical methods are increasingly common in medical journals, and we contend that physicians need more fundamental understanding of those methods. Therefore, medical and statistical educators need to consider how to prepare future health professionals to be able to comprehend statistical methodology that exceeds what is typically presented in semester-long introductory courses.

As an example, many articles in *NEJM* incorporated complex multiple regression models, often featuring correlated and/or incomplete data, relatively novel survival methods, or resampling-based inference. Full interpretation and comprehension of the assumptions, results, and limitations of these articles and their medical implications requires knowledge and understanding of a wide range of statistical concepts. In "Doctors' Ignorance of Statistics," published in the *British Medical Journal*, D. R. Matthews and K. McPherson caution that, "Innumerate doctors...are doomed to have to accept without reservation the statements made in summaries, discussions, or conclusions, and their clinical practice may thus be altered on the basis of flimsy or inconclusive evidence."

The statistical education reform movement makes focus on concepts a priority, but the increasing sophistication of statistical methods poses serious challenges for clinicians, journal editors, and medical educators, among others. In addition to medicine, changes are needed more generally in approaches to introductory statistics education. In "Improving Doctors' Understanding of Statistics," published in the *Journal of the Royal Statistical Society-Series A, Applied Statistics*, D. G. Altman and J. M. Bland address ways to specifically improve doctors' understanding of statistics. Though 15 years old, the article and related discussion remain relevant. Altman and Bland suggest offering new forms of continuing education, workshops, and tutorial papers to augment knowledge of statistical concepts that medical professionals already have received. Another option is to alter the material covered in statistics courses required of medical professionals to include the statistical concepts they are most likely to encounter when reading or conducting research in their field. It also may be appropriate to require training in statistics as a prerequisite to medical school.

An editorial published in *The Lancet* nearly 70 years ago states, "...simple statistical methods concern us far more closely than many of the things that we are forced to learn in the six long years of the medical curriculum." We heartily concur. C

Further Reading

Altman, D.G. and Bland, J.M. (1991). "Improving Doctors' Understanding of Statistics." *Journal of the Royal Statistical Society-Series A, Applied Statistics*, 154:223–267.

Bombardier, C.; Laine, L.; Reicin, A., et al. (2000). "Comparison of Upper Gastrointestinal Toxicity of Rofecoxib and Naproxen in Patients with Rheumatoid Arthritis." *The New England Journal of Medicine*, 343:1520–1528.

Cobb, G. (1991). "Teaching Statistics: More Data, Less Lecturing." *UME Trends*, 3:3–7.

Curfman, G.D.; Morrissey, S.; and Drazen, J.M. (2006). "Expression of Concern Reaffirmed." *The New England Journal of Medicine*, 354(11):1196–1199.

Emerson, J.D. and Colditz, G.A. (1983). "Use of Statistical Analysis in *The New England Journal of Medicine*." *The New England Journal of Medicine*, 309:709–713.

Emerson, J.D. and Colditz, G.A. (1992). "Use of Statistical Analysis in *The New England Journal of Medicine*." In *Medical Uses of Statistics, 2nd Ed*. (Bailar, J.C. and Mosteller, F., Eds.) Boston: *NEJM* Books.

Horton, N.J. and Switzer, S.S. (2005). "Statistical Methods in the Journal." *The New England Journal of Medicine*, 353:1977–1979.

"Mathematics and Medicine." (1937). *The Lancet*, i, 31.

Matthews, D.R. and McPherson, K. (1987). "Doctors' Ignorance of Statistics." *British Medical Journal*, 294:856.

Moher, D.; Schulz, K.F.; and Altman, D.G. (2001). "The CONSORT Statement: Revised Recommendations for Improving the Quality of Reports of Parallel Group Randomized Trials." *BMC Medical Research Methodology*, 1:2.

Moore, T.L. (2000). *Teaching Statistics: Resources for Undergraduate Instructors*. Washington, DC: Mathematical Association of America and the American Statistical Association.

Pagano, M. and Gauvreau, K. (2000). *Principles of Biostatistics, 2nd Ed.* Pacific Grove: Duxbury Press.