

CLINICAL RESEARCH AND STATISTICAL METHODS IN THE UROLOGY LITERATURE

CHARLES D. SCALES, JR., REGINA D. NORRIS, BERCEDIS L. PETERSON,
GLENN M. PREMINGER AND PHILIPP DAHM*

From the Departments of Surgery (Division of Urology), and Biostatistics and Bioinformatics (BLP), Duke University Medical Center, Durham, North Carolina

ABSTRACT

Purpose: We provide a systematic assessment of the quality and accuracy of statistical reporting in the urology literature.

Materials and Methods: All original research publications with adult human subjects in a single issue (August 2004) of 4 leading urology journals were identified for formal review. A standardized evaluation form was developed in consultation with an experienced biostatistician and subsequently tested. Two independent reviewers with at least 1 year of formal training in research design and biostatistics who were blinded to authors and institutions reviewed each article. Discrepancies were settled by consensus and/or adjudication by the biostatistician.

Results: Of the 169 articles screened 97 met eligibility criteria for review. Cohort (43 of 97 or 44%) or cross-sectional (28 of 97 or 29%) designs comprised the majority of these studies. Only 10 randomized clinical trials (12.4%) were identified. Statistical tests were identified in 83 studies (93%). Overall 69 of 83 studies (71%) providing statistical comparisons had at least 1 statistical error, including using the wrong test for the data type in 28%, inappropriate use of a parametric test in 22% and failure to account for multiple comparisons in 65%. In studies applying multivariate analysis (29%) over fitting the model with too many variables was the most common statistical flaw (39%).

Conclusions: This formal review suggests that statistical methods are often used inappropriately in the urology literature, thereby, potentially undermining the validity of study results and conclusions. An effort to raise the awareness of appropriate statistical techniques through postgraduate education appears indicated.

KEY WORDS: statistics, research design, urology

The incorporation of statistical methods into the design, presentation and analysis of clinical research has been one of the major forces in the transition from opinion based to evidence based medicine. Modern medical research relies heavily on sound statistical methodology to transform ambiguous raw data into meaningful results. In the last several years this methodology has evolved tremendously with increasing complexity of trial design and biostatistical techniques.^{1,2}

Statistical hypothesis testing in particular has a major impact on medical research by providing investigators with a formalized framework to test scientific assumptions. This development has been hastened by the ubiquity of powerful computers and the widespread availability of statistical software.³ While the application of statistical methods is no longer the domain of biostatisticians alone, it is uncertain to what extent urologists appropriately apply descriptive and comparative statistical techniques in clinical investigations.

In the research literature of other specialties reviews have shown frequent errors in trial design and statistical methods. In 1 study 25% of randomized trials in surgery journals failed to mention eligibility criteria.⁴ Of 50 reports in general medical journals 58% used an incorrect method of subgroup comparison.⁵ Reviews of this nature have led to improvements in the quality of clinical research. For example, after the publication of such a review in the obstetrics/gynecology litera-

ture complete statistical reporting and the use of appropriate statistical methods improved.⁶

To our knowledge there has been no systematic review to date of the use of statistics in the urology literature. However, a recent report showed that 70% of negative clinical trials were inadequately powered, demonstrating that the urology literature may not be immune to flaws in clinical and statistical methods.⁷ In this context the current study offers the first systematic assessment of the use of statistical methods in urology clinical research.

METHODS

The August 2004 issues of the Journal of Urology, Urology, British Journal of Urology International and European Urology were screened. All original research articles with adult human subjects were selected for review. A physician independent of the evaluation team screened articles for review.

A standardized evaluation form was developed with the assistance of an experienced biostatistician. A total of 80 data elements were recorded per article, including information about the study (eg design) and statistical tests reported. The quality of statistical testing and result reporting was evaluated. Statistical testing was evaluated for common errors, such as an inappropriate test reported for the type of data analyzed. Evaluation criteria for reporting were based on the standards published by the International Committee of Medical Journal Editors,⁸ the Transparent Reporting of Evaluations with Nonrandomized Designs Statement⁹ and Instructions for Authors in the relevant journals. Two physicians with at least 1 year of formal training in research design and

Submitted for publication December 20, 2004.

* Correspondence: Division of Urology, Department of Surgery, Box 2626, MSRB Suite 477, Duke University Medical Center, Durham, North Carolina 27710 (telephone: 919-684-5873; FAX: 919-681-7414; e-mail: dahm0001@mc.duke.edu).

biostatistics evaluated each article independently. Reviewers were blinded to the authors and institutions from which the study originated. Discrepancies were settled by consensus and/or adjudication by the biostatistician. When authors specified multiple statistical tests for multiple data types, tests were assumed to have been applied correctly in the absence of contrary evidence. Retrospective cohort studies were differentiated from case series by the presence of a comparison group. Blinding was accomplished by removing author and institution information from the subtitles and text of each report. Results were entered into a database after consensus review and/or adjudication. Data analysis was performed with SPSS 12.0 (SPSS, Chicago, Illinois). The current study was designed to be purely descriptive in nature, and so no statistical comparisons are reported.

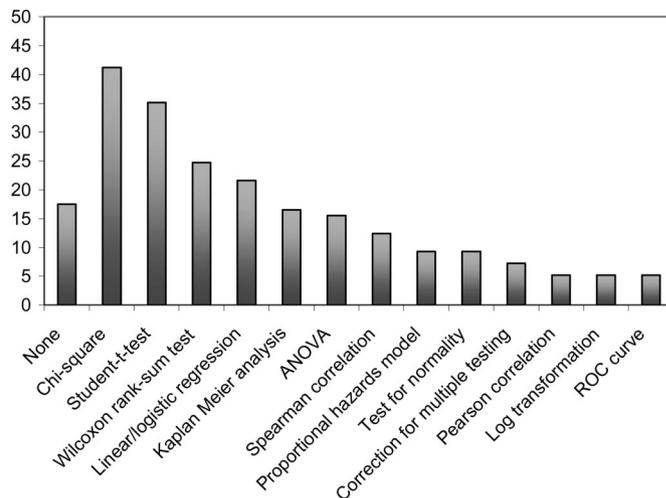
RESULTS

Of the 169 articles screened 97 met eligibility criteria. The most common designs were cross-sectional (28 studies or 29%) and retrospective cohort (28 or 29%). Prospective cohort studies composed 16% of the sample. Clinical trials were the least frequently reported design (12%) (table 1). The greatest number of articles was from the Journal of Urology (46 or 47%). The figure shows the distribution of topics and statistical tests identified.

Of the 97 articles reviewed 80 (83%) included an analysis section in the description of materials and methods (table 2). Excluding studies in which statistical comparison was not done, 6 of 85 articles (7%) failed to identify statistical tests in the methods or text. Only 33 articles (34%) identified the software used for analysis. While it is not a major error, software reporting is recommended by International Committee of Medical Journal Editors standards. Reporting the ethical/institutional review was low with only 38 articles (39%) indicating a review of the research by an ethics or institutional review board (IRB). Similarly only 33 articles (34%) showed informed consent (or approved waiver of consent) by study subjects.

In terms of the analysis reported almost all studies (91 or 94%) included descriptive statistics, eg the mean and variance (table 3). Most articles also included statistical comparisons of 2 groups (75 or 77%). In studies that described statistical tests the chi-square test used to analyze the association of categorical variables was the single most commonly identified statistical test, followed by parametric (Student's t test) and nonparametric (Wilcoxon rank sum test) comparisons of continuous outcome variables. Almost a third of the studies (28 or 29%) showed regression analysis, such as fitting a linear, logistic or proportional hazards model. Other analyses included a comparison of 3 or more groups (29 studies or 30%), association between 2 variables (20 or 21%) and nonparametric tests (31 or 32%).

In this convenience sample there were 12 clinical trials, of which 10 (83%) were randomized. In these clinical trials reporting methods and results often lacked critical details (table 4). For example, only 5 trials (42%) defined a clinically significant difference in the primary outcome measure. Half of the trials (6 or 50%) failed to show a power calculation to



Frequency of statistical techniques used in 97 articles. ROC, receiver operating characteristics.

TABLE 2. Topics reported in Methods section

Topic	No. Applicable Studies	No. Reporting Topic (%)
Analysis section	97	80 (83)
Statistical tests	85	79 (93)
Statistical software	97	33 (34)
Identification	83	39 (47)
Test sidedness	83	20 (24)
IRB review	97	38 (39)
Consent or waiver	97	33 (34)

TABLE 3. Analysis types reported

Analysis Type	No. (% articles)
Descriptive	91 (94)
2 Group comparison	75 (77)
3 or More group comparison	29 (30)
Association between 2 variables	20 (21)
Prediction/control via regression	28 (29)
Nonparametric testing	31 (32)
Time to event (Kaplan-Meier)	17 (18)
Diagnostic test evaluation	9 (9)

Percents do not total 100% since more than 1 type of analysis may have been reported.

support the sample size. Only 3 randomized trials (30%) showed the randomization method.

The quality of the analysis reported was less than optimal. Overall 69 of 83 studies (71%) providing statistical tests had at least 1 statistical error. Of the 80 articles in which the data type and statistical test applied could be specifically identified 22 (28%) used an inappropriate test for the data type. When multiple tests were reported without specifying which test was applied to which data, authors were assumed to have used the tests correctly. In reports with inappropriate statistical tests the most common errors were failing to apply

TABLE 1. Study design of articles reviewed by journal

Study Design	No. Articles (%)				
	J Urol	Urology	Eur Urol	BJU Int	All Journals
Clinical trial	9 (20)	2 (8)	0 (0)	1 (7)	12 (12)
Prospective cohort	8 (17)	4 (16)	2 (18)	1 (7)	15 (16)
Retrospective cohort	13 (28)	7 (28)	3 (27)	5 (33)	28 (29)
Cross-sectional	10 (22)	7 (28)	5 (46)	6 (40)	28 (29)
Case series	6 (13)	5 (20)	1 (9)	2 (13)	14 (14)
Totals	46 (47)	25 (26)	11 (11)	15 (16)	97 (100)

TABLE 4. *Design features of 12 clinical trials*

Design Feature	No. (%)
IRB approval	7 (58)
Informed consent process	9 (75)
Clinically significant difference detailed	5 (42)
Power calculation	6 (50)
Randomized trials:	
All randomized	10 (83)
Randomization method described	3 (30)
Treatment concealment (blinding):	
Study subject	5 (42)
Study personnel	4 (33)
Pt + study personnel	4 (33)
Intent to treat analysis	7 (58)
Flow diagram	3 (25)

or incorrectly applying nonparametric tests (17 of 77 or 22%), failing to apply or incorrectly applying paired statistical tests (6 of 71 or 9%) and applying an inappropriate test for categorical or continuous data (eg a t test for categorical data) (table 5).

Other, less serious statistical errors were also common. For example, 47 of 72 articles (65%) with multiple testing of outcomes failed to account for the effects of multiple testing on the probability of a type I error, that is finding apparent differences by chance when in fact none exist. Similarly of 27 articles in which 3 or more groups of subjects were compared the authors of 9 (33%) reported apparent differences among subsets without prior testing for overall differences, again increasing the probability of a type I error.

Regression analysis in particular was a frequent source of error. Of the 28 articles providing regression results 23 (82%) had methodological errors. The most frequent regression error was over fitting, ie allowing fewer than 10 to 15 observations/events per candidate variable in the model.¹⁰ This error occurred in 11 of these regression results (39%). Other regression errors were inappropriate univariate screening for predictor variables, post hoc categorization of continuous predictor variables and a lack of validation of regression models when sample size permitted construction of a validation set (data not shown).

DISCUSSION

This study shows that most clinical investigations in the urology literature have provided little detail about the type of statistical analysis performed and 1 of 5 articles provided no information at all. The frequency of statistical errors in the urological literature was high (71%) but comparable to the rate in other statistical reviews of the biomedical literature.¹¹ Few studies described how the choice of parametric vs nonparametric test was informed by informal (eg graphic plots) or formal (eg the Kolmogorov-Smirnoff test) methods. Less than a third of studies (28 of 97 or 29%) used multivariable analysis in the form of linear or logistic regression,

and/or the proportional hazards model to analyze the data, although it appeared that this would have been feasible in more of them.

Detailed review of these studies demonstrated that statistical techniques were often applied incorrectly. Examples are use of the chi-square test to compare continuous outcomes (eg age) or alternatively use of the Student t test for categorical outcome variables (eg disease recurrence: yes vs no). The results of such testing must be considered uninterpretable.¹² Another common mistake was the use of parametric measures of distribution (eg mean \pm SD) and tests (eg Student's t test) to describe and subsequently analyze small samples (eg fewer than 20 observations) or variables with skewed non-Gaussian distributions (eg health related quality of life scores with ceiling effects). While such testing does not necessarily lead to incorrect results or conclusions, it threatens the validity of these studies.¹²

Finally, the flawed application of statistical methods in these investigations raises concerns over an increased likelihood of type I error, defined as the probability of rejecting the null hypothesis (there is no difference) in favor of the alternative hypothesis (there is a difference) when in fact there is no difference.¹³ Examples of these flawed applications are performed, the uncritical use of multiple testing of several outcome variables (65%)³ as well as unmotivated subset analysis, eg testing for difference between subgroups with Student's t test when ANOVA demonstrates no difference among groups. Assuming a commonly accepted α of 0.05, defined as the probability of the false inference described, the probability of a type I error is 5% per test. Thus, if 5 statistical tests are performed, the probability of at least 1 type I error could be as large as $1 - 0.95^5$ or 22.6%.¹⁰ Therefore, it would appear important for investigators to be cognizant of this potential threat to the validity of their conclusions. Multiple testing should be avoided or specifically addressed as a potential threat. Similarly post hoc subset analysis should be considered as hypothesis generating, rather than as confirmatory.⁹

An additional aspect of this study was the evaluation of study designs in the urology literature. Overall prospective and retrospective cohort studies were the most frequent design in this sample, accounting for 44% of all reported studies. Only 10 studies were randomized clinical trials (RCTs), which are considered the gold standard of study designs and offer the potential for yielding the highest level of evidence. A relative paucity of RCTs has been previously noted in the surgical literature and it is likely of multifactorial etiology. Barriers to RCTs in surgical subspecialties are the lack of surgeon and patient equipoise about a certain form of therapy, the difficulty of standardizing the quality of a given surgical procedure and accounting for various skill levels, limited funding mechanisms for such studies and potentially the fact that few urologists have had the opportunity to receive formal training in the concepts of clinical trial design.¹⁴⁻¹⁶ This may also be reflected in the low quality of

TABLE 5. *Specific statistical errors identified on systematic review*

Biostatistical Error Type	No. With Error	No. Applicable Studies (%)
Any statistical error*	69	83 (71)
Inappropriate data test (any)*	22	80 (28)
Categorical (continuous error)	6	80 (8)
Paired (unpaired error)	6	71 (9)
Parametric (nonparametric error)	17	77 (22)
Pairwise comparison without prior testing for difference among all groups	9	27 (33)
Failure to account for multiple testing	47	72 (65)
Regression errors:		
All	23	28 (82)
Model over fitting	11	28 (39)

* Each article was only counted once even if multiple errors were present and, therefore, the number of articles with any error may be less than the sum of individual error types.

reporting in the 12 clinical trials that we identified when applying criteria defined by the Consolidated Standards of Reporting Trials group.¹⁷ With regard to statistical methods, it was notable that only half of the studies provided any information about power calculations at the time of study design. The issue of unsupported claims of equivalence by under powered studies of inappropriate design was recently raised^{7,18} and it was noted again in the current study. In addition, many studies failed to provide an intent to treat analysis, which represents another well recognized threat to the validity of these study conclusions.¹⁷

In the setting of increasing public scrutiny about medical research we found it notable that only 39% and 58% of all studies and clinical trials, respectively, made explicit mention of a study approval by an IRB or ethics committee. Informed consent or the appropriate waiver thereof was likewise explicitly mentioned in only 25% of all studies. A higher proportion of clinical trials reported consent (75%) but this proportion is less than optimal. In accordance with suggested guidelines for good publishing practice more consistent reporting should be encouraged in the urology literature.^{8,9}

We believe that several features of our study design strengthen the validity of our findings. 1) A detailed evaluation comprising 80 items was developed (see Appendix) and tested a priori in close consultation with an experienced biostatistician. 2) Each reviewer had a least 1 year of formal didactic training in biostatistics and clinical research methods. Reviewers evaluated the articles independently and were blinded to authors and institutions. 3) All 97 articles were evaluated in their entirety, thereby, allowing us to account for descriptions of statistical methods not mentioned in the methods section but placed elsewhere in the text, such as in legends or in author comments on potential limitations in the discussion section of the article.

Several limitations of this study also deserve critical mention to the reader. 1) We recognize that our study was based on a convenience sample of articles in 4 leading urology journals published in a single month. Although we have no reason to believe that our sample is not representative, a lower quality of reporting than is typical for the urology literature would overestimate the frequency of statistical and other methodological

errors. 2) Almost half of the articles (47%) originated from the Journal of Urology, which was the only journal to include statistical guidelines in the Information for Authors. If standards of statistical review differ among journals, this weighting would bias these results. However, post hoc analyses of our results were not suggestive of different error rates among journals in the study sample. 3) Our sample was relatively small, as mandated by the work intense nature of evaluating 97 entire articles. 4) We recognize that statistics are a rich field in which to find mistakes, and yet not all statistical errors distort the conclusions of a study.¹⁹ To avoid an overly critical approach we chose to make the conservative assumption that authors applied statistical methods appropriately when in doubt. While recognizing these limitations, we nevertheless believe that this study makes an important contribution in the growing effort toward evidence based medicine, of which statistical literacy is an integral part.

The results of this study demonstrate that the urology literature, like other biomedical literature, is susceptible to errors in biostatistical methods and reporting. Studies in other specialties have attributed these types of errors to various causes, including a lack of statistical and clinical research knowledge among authors and reviewers, a lack of involvement of biostatisticians in clinical investigations,²⁰ and the time and cost constraints involved in the editorial review process, which appear prohibitive of a formal review of every manuscript by a biostatistician prior to publication.²¹ However, research efforts documenting these flaws and subsequent educational endeavors have led to improvements in other biomedical specialties without the need for including biostatisticians on every study, which could prove infeasible.⁶ We anticipate similar improvement in the urology literature as a result of the current study and ongoing educational efforts.

CONCLUSIONS

This formal review suggests that statistical methods are often used inappropriately in the urology literature, which potentially undermines the validity of study results and conclusions. An effort appears indicated to raise the awareness of appropriate statistical methods among authors and reviewers through postgraduate education and among resident urologists through graduate education.

APPENDIX

Reviewer:	1 — Scales	2 — Dahm			
Journal:	1 — J Urol	2 — Urology	3 — Eur Urol	4 — BJU Int	
Study design:	1 — Clinical trial	2 — Prospective cohort	3 — Retrospective cohort	4 — Cross-sectional	5 — Case series
Study topic:	1 — Oncology	2 — Stones/Endourology	3 — Laparoscopy	4 — Trauma/Reconstruction	
	5 — Voiding dysfunction	6 — Infection/Inflammation	7 — Infertility/Erectile dysfunction		
Analysis section part of Materials and Methods	0 — no	1 — yes			
Statistical tests used described	0 — no	1 — yes			
Statistical tests reported to have been used					
None or N/A	0 — no	1 — yes			
Chi-square/Fisher's exact	0 — no	1 — yes			
Student t test	0 — no	1 — yes			
ANOVA	0 — no	1 — yes			
Time to event/Kaplan-Meier	0 — no	1 — yes			
Cox regression analysis	0 — no	1 — yes			
Test for distribution	0 — no	1 — yes			
Pearson correlation	0 — no	1 — yes			
Spearman correlation	0 — no	1 — yes			
Mann-Whitney/Wilcoxon rank sum	0 — no	1 — yes			
Kruskal-Wallis	0 — no	1 — yes			
McNemar's test	0 — no	1 — yes			
Bonferroni/other correction	0 — no	1 — yes			
Linear/logistic regression	0 — no	1 — yes			
Statistical software identified	0 — no	1 — yes			
If identified:	1 — SAS (SAS Institute, Cary, North Carolina)	2 — SPSS	3 — other: _____		
Primary end point identified/analyzed	1 — single	2 — multiple	N/A		

Sample/cohort size	__ patients			
If applicable, number of events	—			
Type of outcome	1 — Time to event	2 — Categorical	3 — Ordinal	4 — Continuous
Identification of α :	0 — no	1 — yes		
Sidedness of testing:	0 — not specified	1 — 1-sided	2 — 2-sided	
Nonsignificant p values detailed:	N/A	0 — no	1 — yes	
IRB review specified	0 — no	1 — yes		
Consent specified	0 — no	1 — yes		
Time to event analysis				
Patients at risk over time presented:	0 — no	1 — yes		
Length of followup detailed:				
Overall	0 — no	1 — yes		
By subgroup	0 — no	1 — yes		
Length of followup characterized by	1 — mean	2 — median	3 — both	
Number of patients censored presented:	0 — no	1 — yes		
Percent of patients overall:	—			
Multivariate Cox regression analysis	0 — no	1 — yes		
Number of independent variables included in model	—			
Justification of stratification cutoffs for independent variables	0 — no	1 — yes		
Reporting of HR with CI	0 — no	1 — yes		
Logistic regression				
Justification of dichotomization (if applicable)	0 — no	1 — yes		
Clinical trial				
Randomized	0 — no	1 — yes		
Method of randomization described	0 — no	1 — yes		
Study personnel blinded	0 — no	1 — yes		
Participants blinded	0 — no	1 — yes		
Clinically significant difference defined	0 — no	1 — yes		
Power analysis performed	0 — no	1 — yes		
Intent to treat analysis	0 — no	1 — yes		
Study flow sheet	0 — no	1 — yes		
Type of analysis reported				
Descriptive (central tendency \pm distribution)	0 — no	1 — yes		
Comparison of 2 groups	0 — no	1 — yes		
Comparison of 3 or more groups	0 — no	1 — yes		
Association between 2 variables	0 — no	1 — yes		
Prediction/Control via regression (log/linear/Cox)	0 — no	1 — yes		
Transformation	0 — no	1 — yes		
Nonparametric testing	0 — no	1 — yes		
Time to event analysis	0 — no	1 — yes		
Evaluation of diagnostic test	0 — no	1 — yes		
Biostatistical Errors				
Wrong test for data type (e.g. chi-square for continuous data)	N/A	0 — no	1 — yes	
Use of mean \pm SD for small/non-Gaussian distribution	N/A	0 — no	1 — yes	
Use of SEM to describe variable distribution	N/A	0 — no	1 — yes	
Use of parametric test for small/nonGaussian distribution	N/A	0 — no	1 — yes	
Pearson's correlation coefficient for nonparametric data	N/A	0 — no	1 — yes	
Use of paired testing for unpaired data	N/A	0 — no	1 — yes	
Use of unpaired testing for paired data	N/A	0 — no	1 — yes	
Unmotivated subgroup analysis	N/A	0 — no	1 — yes	
Univariate screening for predictor variable	N/A	0 — no	1 — yes	
Failure to account for multiple testing (mathematically or in discussion)	N/A	0 — no	1 — yes	
Unsupported claim of equivalence (H_0 assumed true without specific testing)	N/A	0 — no	1 — yes	
Unsupported claim of difference (H_0 assumed to be false without specific testing)	N/A	0 — no	1 — yes	
Lack of power calculation	N/A	0 — no	1 — yes	
Post hoc categorization of variables	N/A	0 — no	1 — yes	
Failure to define clinically relevant effect size	N/A	0 — no	1 — yes	
Over fitting of variable models	N/A	0 — no	1 — yes	
Inclusion of discontinuous effect variables in Cox PH model	N/A	0 — no	1 — yes	
Lack of multivariate analysis when appropriate	N/A	0 — no	1 — yes	

Lack of multivariate analysis when appropriate	N/A	0 — no	1 — yes
Lack of validation of prognostic model	N/A	0 — no	1 — yes

REFERENCES

- Pocock, S. J., Hughes, M. D. and Lee, R. J.: Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med*, **317**: 426, 1987
- Altman, D. G. and Goodman, S. N.: Transfer of technology from statistical journals to the biomedical literature. Past trends and future predictions. *JAMA*, **272**: 129, 1994
- Moye, L. A.: Random research. *Circulation*, **103**: 3150, 2001
- Hall, J. C., Mills, B., Nguyen, H. and Hall, J. L.: Methodologic standards in surgical trials. *Surgery*, **119**: 466, 1996
- Assmann, S. F., Pocock, S. J., Enos, L. E. and Kasten, L. E.: Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet*, **355**: 1064, 2000
- Welch, G. E., 2nd and Gabbe, S. G.: Statistics usage in the American Journal of Obstetrics and Gynecology: has anything changed? *Am J Obstet Gynecol*, **186**: 584, 2002
- Breau, R. H., Carnat, T. A. and Gaboury, I.: Inadequate statistical power of negative studies in urology literature. *J Urol*, suppl., **171**: 98, abstract 374, 2004
- Uniform requirements for manuscripts submitted to biomedical journals. International Committee of Medical Journal Editors. *N Engl J Med*, **336**: 309, 1997
- Des Jarlais, D. C., Lyles, C., Crepaz, N. and TREND Group: Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health*, **94**: 361, 2004
- Motulsky, H.: Multiple comparisons. In: *Intuitive Biostatistics*. New York: Oxford University Press, pp. 118–126, 1995
- Kanter, M. H. and Taylor, J. R.: Accuracy of statistical methods in TRANSFUSION: a review of articles from July/August 1992 through June 1993. *Transfusion*, **34**: 697, 1994
- Greenhalgh, T.: How to read a paper. Statistics for the non-statistician. I: Different types of data need different statistical tests. *BMJ*, **315**: 364, 1997
- Guyatt, G., Jaeschke, R., Heddle, N., Cook, D., Shannon, H. and Walter, S.: Basic statistics for clinicians: 1. Hypothesis testing. *CMAJ*, **152**: 27, 1995
- Solomon, M. J. and McLeod, R. S.: Surgery and the randomised controlled trial: past, present and future. *Med J Aust*, **169**: 380, 1998
- Evans, C. P.: Evidence-based medicine for the urologist. *BJU Int*, **94**: 1, 2004
- Mansson, W.: Evidence-based urology—a utopia? *Eur Urol*, **46**: 143, 2004
- Moher, D., Schulz, K. F., Altman, D. and CONSORT Group (Consolidated Standards of Reporting Trials): The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA*, **285**: 1987, 2001
- Greene, W. L., Concato, J. and Feinstein, A. R.: Claims of equivalence in medical research: are they supported by the evidence? *Ann Intern Med*, **132**: 715, 2000
- Bacchetti, P.: Peer review of statistics in medical research: the other problem. *BMJ*, **324**: 1271, 2002
- Altman, D. G., Goodman, S. N. and Schroter, S.: How statistical expertise is used in medical research. *JAMA*, **287**: 2817, 2002
- Goodman, S. N., Altman, D. G. and George, S. L.: Statistical reviewing policies of medical journals: caveat lector? *J Gen Intern Med*, **13**: 753, 1998