An assessment of the statistical procedures used in original papers published in the *SAMJ* during 1992

P. J. Becker, E. Viljoen, L. Wolmarans, C. B. IJsselmuiden

Objective. To assess the statistical procedures used in original papers published in the *SAMJ*.

Design. Descriptive study based on a random sample of 100 papers from the 153 papers with methodological content that were published in the *SAMJ* during 1992.

Results. This review showed that 34% (95% CI (25%; 43%)) of papers used no statistical procedure at all or used simple descriptive statistics only. In sampling methods, there was a predominance of the use of the period sampling method as opposed to probability sampling methods. Inappropriate statistical methods were used in 15% (6%; 24%) of papers, while in 16% (9%; 23%) statistical procedures and in 13% (6%; 20%) the sampling methods used could not be identified. Inaccurate graphical methods were used in 17% (6%; 28%) of papers. Confidence intervals and power calculations are used far too infrequently, in 33% (19%; 47%) and 11% (3%; 19%) of appropriate papers respectively. If the *Journal's* readers are at least familiar with simple descriptive statistics, contingency table analysis, simple epidemiological statistics, *t*-test procedure and confidence interval calculation and interpretation, they will have a complete understanding of the statistical content of 60% of original articles published in the *Journal*.

Conclusion. Guidelines for the statistical treatment of reported data and the statistical review of articles before publication will assist substantially in improving the quality of statistical analysis. More intensive use of available biostatistical and epidemiological expertise at the study design and analysis stages is needed to shift the emphasis from descriptive research to analytical investigation.

S Afr Med J 1995; **85:** 881-884.

Centre for Epidemiological Research in Southern Africa, Pretoria

P. J. Becker, PH.D.

E. Viljoen, B.SC. HONS

Johnson & Johnson Health Management Institute, Centurion, Gauteng

L. Wolmarans, B.SC.

Department of Community Health, Medical University of Southern Africa, PO Medunsa, 0204

C. B. IJsselmuiden, ARTS EXAMEN NEDERLAND, F.F.C.H. (S.A.), M.P.H., D.T.M.&H., D.P.H.

Biostatistical analysis is an essential tool in the interpretation of biomedical research, yet few adequate training courses are provided for those most likely to engage in biomedical research. In most undergraduate medical curricula in South Africa, epidemiology, biostatistics, research philosophy and research techniques are either not taught or are taught in one block at a time when the student's attention is focused on mastering clinical skills, or even before the student has any grasp of the clinical relevance of research, in cases where the course is offered in the pre-clinical years. Biostatistical expertise and proficiency are not among the skills examined at the end of medical training, even though a large part of medical research in this country is being conducted by medical practitioners and continuing medical education is to a great extent based on the reading of biomedical literature in which biostatistics abound.

Biostatistical experience is therefore often only gained in the process of conducting research for a postgraduate qualification, in which case actual biostatistical experience may be substituted by delegating all study design and statistical analysis issues to a biostatistician or epidemiologist or, even worse, to statistical software. Personal computers have become standard pieces of equipment in the research environment. The 'user-friendly' nature of current statistical software has brought statistical data analysis within easy reach of the biomedical researcher, resulting in the frequent use and, knowingly or unknowingly, abuse of biostatistics. Among statisticians and epidemiologists there is a growing concern about the incorrect use of biostatistics in both published and unpublished reporting.

Researchers handle their own data analyses for a variety of reasons. Some have an adequate knowledge of biostatistics, and are therefore perfectly capable of performing their own analyses. Some are unaware of the assistance biostatisticians and epidemiologists can provide in preparing and analysing their research projects. Others experience financial or geographical constraints and cannot afford or access biostatistical or epidemiological support, while still others are unwilling to consult for fear of extending the list of authors of publications resulting from their research.

The use of biostatistics in medical research is widespread and a regular assessment of the statistical content of journal articles can be very informative in respect of changes in statistical methods that are employed in published papers,[1] in respect of the review process of journal articles in terms of statistical content[2] and the abuse of statistics,[1,3] and in respect of the impact of the statistical knowledge of readers on the accessibility of professional journal articles.[4]

The purpose of this paper is to reflect on some important aspects of the statistical content of original articles published in the *SAMJ* during 1992.

Methods

A descriptive (*sic*) study was undertaken of papers published in the *SAMJ* during 1992. Review papers, case studies and papers without any methodological content were not included in a study population of 153 papers. A random sample of 100 papers was drawn. This sample size ensured that if the proportion of papers with statistical errors was 60%[3] then we would be at least 90% confident that the estimate resulting from our study of errors occurring in published articles would lie between 55% and 65%. To reduce measurement bias, the names and particulars of the authors were removed from all selected papers before the statistical content was assessed by three of us (P.J.B., E.V., L.W.).

The proportion of papers in the complete study population that had the support of a statistician or an epidemiologist or both was assessed by C.B.I. who was 'semi-blinded' to the extent that he had no knowledge which of the 153 papers had been selected for inclusion in the study. Studies that emanated from or had co-authors from research institutions, or from departments of statistics and engineering, were assumed to have had statistical or epidemiological support or both.

To help classify the statistical content of the papers included in the sample, partial use was made of the *British Medical Journal* guidelines for statistical review.[5] Findings are reported mainly in terms of point estimates. However, approximate 100 (1-α)% confidence intervals for a proportion of interest can be determined from

$$(p - z_{1-\alpha/2}\,\text{SE};\ p + z_{1-\alpha/2}\,\text{SE}),$$

where p is the sample proportion, SE the approximate standard error of p given by $\sqrt{p(1-p)/n}$ for a sample of size n and $z_{1-\alpha/2}$ is the 1-α/2 probability point of the standardised normal distribution.

Results

In the total study population of 153 papers, 16 (11%) papers were supported by a statistician, 26 (17%) by an epidemiologist, and 12 (8%) by both a statistician and an epidemiologist, giving a total of 54 papers (35%) that had received methodological support. The study designs and sampling methods used in the 100 papers sampled are presented in Table I.

Table I. Study design and sampling methods used in 100 original articles published in the *SAMJ* during 1992

Study characteristics	Number (%)
Study objectives adequately described	92
Study design used	
Descriptive study	61
Cohort study	3
Case-control study	10
Clinical trial	23
Not applicable	3
Sampling technique*	
Simple random sample	15
Stratified sample	2
Systematic sample	4
Cluster sample	5
Other methods	
Volunteer	4
Postal survey	2
Census/period	47
Convenient	5
Consecutive	6
Insufficient information	13
Not applicable	4

* Seven studies included census/period sampling together with simple random sampling (4), systematic sampling (2) and a volunteer sample (1).

The statistical methods used in the 100 original *SAMJ* papers are shown in Table II together with the results of two similar studies of papers published in the *New England Journal of Medicine* during 1978 - 1979[4] and 1990[1] respectively.

Table II. Statistical procedures used in 100 articles published in the *SAMJ* in 1992, compared with results from two similar studies in the *New England Journal of Medicine* (%)[1,4]

Procedure	*N Engl J Med* 1978/9 (*N* = 332)	*N Engl J Med* 1990 (*N* = 100)	*SAMJ* 1992 (*N* = 100)
No statistical method or descriptive statistics only	27	11	34
t-test	44	39	17
Contingency tables	27	30	26
Pearson correlation	12	17	3
Nonparametric test	11	25	4
Any survival analysis/ logistic regression	11	32	5
Basic epidemiological statistics*	9	13	23
Simple linear regression	8	18	0
Analysis of variance	8	14	6
Transformation	7	8	3
Multiple regression	5	6	1
Nonparametric correlation	4	9	5
Multiway tables	4	7	0
Multiple comparisons	3	5	4
Adjustment and standardisation	3	1	7
Other methods:	3	19	29
Confidence intervals†			14
Power calculations‡			6
Test for normality			1
Factor analysis			1
Discriminant analysis			1
Measure of agreement			3
Methodology development			1
Bootstrapping			1
Demographic statistic			1

* Basic epidemiological statistics: relative risk and odds ratio.
† Confidence interval analysis was judged necessary in 42 papers.
‡ Power calculations were judged necessary in 56 papers.

Incorrect use of some or all of the statistical procedures used was observed in 53% of the papers. Details are given in Table III. No assessment was made of the adequacy of the study design used in relation to the study objectives or research questions, and hence no comments can be made on this important aspect.

Table IV represents a somewhat arbitrary guide to determining how accessible, from a biostatistical point of view, the 1992 *SAMJ* articles were. The first column lists the number of papers that used the corresponding biostatistical method(s). The second column expresses the statistical accessibility of *SAMJ* articles, in which an article is regarded as accessible if all the statistical methods used in an article can be understood and interpreted by a reader. If, for example, a reader's understanding was limited to descriptive statistics only, then 34% of papers were accessible. If, however, a reader is familiar with both descriptive statistics and with the use and interpretation of contingency tables, then 42% (34% using only descriptive statistics plus 8% using contingency tables only or in combination with descriptive statistics) were accessible. Similarly, a reader with understanding of simple descriptive statistics, contingency tables and common epidemiological statistics would have been able to access 53% of papers, and so on.

Table III. Comments on the use of statistical procedures in a sample of 100 papers published in the *SAMJ* during 1992

Comments	%
On analytical technique used	
Unidentified statistical method(s)	16
Inappropriate methods	10
More tests specified than used	3
Testing irrelevant to objectives	3
Multiple comparisons — each test at α level of significance	8
On presentation and interpretation	
Inadequate measure of location	2
Mean ± 1SD	22
Significant results not associated with *P*-value	4
Inconsistent/incorrect use of terminology	9
Numbers in tables and text do not correspond	8
Correlation coefficient as measure of agreement	1
Infer equality from a 'not significant' finding	2
No adjustment for baseline co-factors	2
On graphical displays used	
Bar chart for means	3
Bar chart instead of histogram or vice versa	4
Frequency polygon not anchored	1

Table IV. Accessibility of statistical content of *SAMJ* articles (%)

Procedure	Papers using method	Papers using this and prior methods
No statistical method or descriptive statistics only	34	34
Contingency tables	26	42
Epidemiological statistics	23	53
t-test	17	57
Confidence intervals	14	60
Nonparametric tests	4	62
Pearson correlation	3	63
Nonparametric correlation	5	66
Analysis of variance	6	67
Multiple comparisons	4	70
Power calculations	6	76
Adjustment and standardisation	7	83
Any survival analysis or logistic regression	5	87
Transformation	3	89
Measure of agreement	3	92
Other methods*	7	99†
Paper — method uses	167	

* Including multiple regression.
† One paper did not identify any method.

Discussion

The most commonly used study designs were descriptive designs (61%) and clinical trials (23%) while the most frequent sampling methods in these studies made use of a 'census' or 'period' sample (47%) (Table I). The predominance of the descriptive study design may reflect several issues, including a general lack of training in analytical study techniques, a greater need for descriptive data of many kinds given the paucity of health information on many groups and health problems in this country, or may be the result of a lack of adequately planned studies.

Similarly, the frequent use of period sampling is likely to result from a lack of study planning, and occurs because retrospective, usually routinely recorded, data are commonly used. The predominance of non-probability sampling methods casts doubt on the validity of many observations and should be reduced. In addition, in 13% of papers the methods section provided too little information to decide what sampling methods had been used, making interpretation of study results in effect impossible.

Except for the use of contingency tables there are few similarities between the *SAMJ* papers and papers reviewed in *N Engl J Med* (Table II) in terms of the frequency with which certain statistical methods were used. The two *N Engl J Med* reviews[1,4] of uses of statistics reflect a change in the use of statistical methods from 1978 - 1979 to 1990 with a shift towards the more frequent use of complex statistical techniques. The distribution of statistical methods used in the *SAMJ* is different from either of the *N Engl J Med* studies and hence these are difficult to compare. Nevertheless, the methods used in the *SAMJ* tend to cluster around the simple descriptive statistics, *t*-test and contingency tables, with another cluster in the use of basic epidemiological statistics (relative risk and odds ratio). A measure of comparison, although perhaps not very informative, is that the rate at which statistical methods other than descriptive statistics were used in the *N Engl J Med* during 1990 was 2,7 ('methods per paper') while for the *SAMJ* this rate was 2,0 ('methods per paper') in 1992.

The frequency with which certain statistical methods are used, particularly techniques for survival analysis and epidemiological statistics, suggests that the type of research being reported by the two journals is very different.

A large proportion of the *SAMJ* papers (34%) do not make use of any statistical methods, or use descriptive statistics only. Confidence intervals facilitate the interpretation of results[6-8] and their use should be encouraged.[9] Confidence intervals were reported in 33% (14/43) of *SAMJ* papers where they could have contributed to a better understanding of study results. Although sample size and *a priori* power calculations could have been done in more of the studies, they were reported in only 11% (6/56) of the *SAMJ* papers. Especially in studies with negative results actual power calculations should be reported. None of these studies in the *SAMJ* showed such calculations.

Fifteen per cent (10/65) of *SAMJ* papers used inappropriate methods of data analysis (Tables II and III). This might be an underestimate since in 16% of the papers not all the tests used were specified, and their appropriateness could not be assessed. In 17% (8/47) of papers the graphic display of data included errors.

Statistical software packages are often to blame for these errors as methods have been incorrectly programmed and researchers accept their outcome in good faith. A specific problem for submission of graphic displays is that different colours reprint in the journal only as black and white. In some of the papers where colour graphs were used in the original submission, graphs could not be interpreted after being published in black and white. Consequently, patterns rather than colour should be used to distinguish between the elements in a graphic display.

Multiple comparisons, each at the overall level of significance, were made in 8 of the *SAMJ* papers and this was noted in Table III. It should be noted, however, that because of the consistency of the unrestricted least significant difference procedure[10] as the method of doing multiple comparisons, this comment is not necessarily valid.

Reporting of the appropriate measures for location and spread is important for any study. The notation $\bar{x} \pm SD$ for reporting the mean and standard deviation, which was adopted by 22 papers, is widely used. This is, however, not correct because it only has any meaning for normally distributed variables, in which case it denotes the largely irrelevant 68% confidence interval ($\bar{x}$ – SD; $\bar{x}$ + SD). Instead, the notation 'mean (SD)' should be adopted, or when descriptive statistics are reported in a table, means and standard deviations can be reported in separate columns, each with its own heading.

In two studies equivalence was inferred from 'not significantly different', in other words, the null hypothesis (i.e. equivalence) was not rejected because the sample-related *P*-value exceeded the chosen level of significance, usually a somewhat arbitrary probability of 0,05. When a test is 'not significant' the null hypothesis is 'not rejected' on the basis of a (sample-related) *P*-value only, and it is always good practice to report the actual *P*-values and not adopt notations such as $P < 0{,}05$ for significant and $P > 0{,}05$ or NS for not significant. However, note that when the difference between two groups is not 'statistically significant', clinically or biologically meaningful differences may still exist. Confidence intervals[6-8] are helpful in such cases and give the likely range of possible differences. Confidence intervals should therefore accompany statistical test results, especially when considered to be 'not significant', and only when differences are neither clinically nor biologically important can the two groups be considered similar.

Although many comments on the published papers were reported, only a few dealt with major problems. From review of the papers only, it was not possible to calculate whether or not statistical and graphical errors and inaccuracies would substantially alter the outcome of studies since the descriptions of raw data, sampling frames, study designs and hypotheses were insufficient to allow recalculation. *Guidelines for authors on statistical reporting and a process of statistical review*[1,2] should minimise such errors and inaccuracies and will promote the standardisation and the quality of statistical reporting in published papers. Statistical review, according to this study, can eliminate the incorrect use of statistics in 15% (10/66) of papers and of graphical displays in 17% (8/47) of papers, and can ensure that sufficient information is provided to determine what statistical procedures and sampling methods were used in 16% and 13% of papers respectively. Statistical review is therefore clearly of great importance and any journal will upgrade the quality of its published content by introducing such a process.

The most important way in which the adequacy of statistical treatment of study data can be ensured, is to educate those using and reading them. This review of 1992 *SAMJ* papers provides a clear guideline of what should be taught to enhance understanding of the medical literature, a feature essential in continuing medical education. By ensuring that all those who write and read medical (and allied) literature are up to date with descriptive statistics, *t*-test, contingency tables, basic epidemiological statistics and the interpretation of confidence intervals, the statistical content in 60% of papers in the *SAMJ* can be written and fully understood (Table IV). Additionally, the use and interpretation of power calculations can further enhance understanding and rational use of published study results.

Study protocols should include details of study design, sampling and power calculations, all aspects which can be dealt with in consultation with a biostatistician or an epidemiologist. From this study the question may also be raised as to whether these principles are clearly understood by ethics and protocol review committees.

In conclusion, the biostatistician and the epidemiologist can play a very important role in both the planning and protocol development phase, and also in the data analysis phase, particularly when more complex procedures are to be applied.

Thanks are due to the extremely valuable comments and suggestions of Dr Carl Lombard (MRC, Cape Town) and Professor Mary-Lou Thompson (UCT).

REFERENCES

1. Altman DG. Statistics in medical journals: developments in the 1980's. *Stat Med* 1991; **10:** 1897-1913.
2. Gore SM, Jones G, Thompson SG. The *Lancet's* statistical review process: Areas for improvement by authors. *Lancet* 1992; **340:** 100-102.
3. Felson DT, Cupples LA, Meenan RF. Misuse of statistical methods in *Arthritis and Rheumatisim* 1982 versus 1967- 68. *Arthritis Rheum* 1984; **27:** 1018-1022.
4. Emerson JD, Colditz GA. Use of statistical analysis in the *New England Journal of Medicine*. In: Bailer III JC, Mosteller F, eds. *Medical Uses of Statistics*. 2nd ed. Boston: NEJM Books, 1992.
5. Gardner MJ, Machin D, Campbell MJ. Use of checklists in assessing the statistical content of medical studies. In: Gardner MJ, Altman DG, eds. *Statistics with Confidence*. London: British Medical Journal, 1989; 101-108.
6. Gardner MJ, Altman DG. Confidence intervals rather than *P*-values: estimation rather than hypothesis testing. *BMJ* 1986; **292:** 746-750.
7. Luus HG, Muller FO, Meyer BH. The use and interpretation of confidence intervals. *S Afr Med J* 1989; **76:** 626-629.
8. Luus HG, Muller FO, Meyer BH. Statistical significance versus clinical relevance. Part III. Methods for calculating confidence intervals. *S Afr Med J* 1989; **76:** 681-685.
9. Rush RW. Statistical validation of clinical studies (Letter). *S Afr Med J* 1993; **83:** 802.
10. Saville DJ. Multiple comparison procedures: the practical solution. *American Statistician* 1990; **44** (2): 174-180.

Accepted 4 Nov 1994.