British Journal of Social Work (2003) **33,** 567–573

ADVICE NOTE

Conducting and Presenting Social Work Research: Some Basic Statistical Considerations

Nigel Smeeton and David Goda

Nigel Smeeton is lecturer in Medical Statistics, Department of Public Health Sciences, King's College London. David Goda is senior lecturer in Statistics, University of Wolverhampton. Each has acted for many years as a statistical adviser to the British Journal of Social Work.

Correspondence to David Goda, SCIT, 35/49 Lichfield Street, Wolverhampton WV1 1EL, UK.

Editors' Note: An earlier version of this paper appeared in the *Journal* in 1993 and has formed the basis of our advice to intending contributors ever since. Ten years later, we are very grateful to the authors for providing an updated version, which continues to provide clear, accessible authoritative advice in the statistical field.

Introduction

A substantial proportion of the papers submitted to the *British Journal of Social Work* involve the presentation and statistical analysis of data. The guidelines presented some years ago (Goda and Smeeton, 1993) are in need of revision to give greater assistance to those involved in quantitative research.

Statistical advice should be obtained at the stage of initial planning of an investigation so that, for example, the method of sampling and design of questionnaire are appropriate. Study deficiencies at this stage can very rarely be rectified later. In particular, even the most sophisticated statistical methods cannot produce meaningful results from poorly designed studies.

The following paragraphs attempt to address the problems that, in our experience as Statistical Advisers to the *Journal,* are commonly found in submitted papers. More comprehensive advice in a medical context is given by Altman *et al*. (2000).

Conduct of the study

1 A detailed search of the literature should be performed for the current evidence on the issue under investigation, preferably using electronically available

sources. Unnecessary replication and the production of findings that are not original should then be avoided.

2 In general, it is only possible to study a proportion of the individuals of interest due to limitations of time and money. If this is the case, the method used to obtain the sample should be described. Quota sampling (where the investigator haphazardly questions a predetermined number of cases) and self-selected sampling (where, for example, clients might be encouraged to answer a questionnaire from a stock of forms that has been placed in a waiting room) require the least effort. The samples obtained are likely to be unrepresentative of the population under study, however, as those who have strong views about the subject are more likely to agree to take part. Ideally, subjects for the study should be randomly selected from the complete list of eligible cases. This involves the use of random numbers, obtained either from a computer package or calculator, or given in some text books.

   For studies that assess the characteristics of clients using a particular service, systematic sampling—interviewing (say) every fifth caller at a reception desk—may be appropriate. Researchers should be aware that this method of sampling can introduce bias, particularly with a rigid appointment system. For instance, the first client of the day may be more likely to be in paid employment and calling en route to work.

   Where individuals are sampled in clusters, for example as school classes, it cannot be assumed that subjects are independent of each other. Highly sophisticated statistical methods are then required to take account of similarities between subjects in the same class. This method of sampling is therefore best avoided; if it really is essential, the problem should be pointed out.

3 If ethical approval is required before a study can proceed, the ethical committee concerned is likely to require justification of the sample size proposed in terms of the chance that the study will demonstrate the existence of a real difference, if one exists. This is a technical area and advice from a statistician is strongly recommended. In general, the number of cases needed for a worthwhile study will depend on the type of measurement involved, the variability between cases, the likely size of any difference between groups and so on. If, however, fewer than, say, 100 individuals (or institutions) are involved, differences between groups will need to be striking in order for them to be statistically significant. Referees will give special consideration when subjects are difficult to obtain, for example in studying extreme old age, but the basic principle still applies. This is not to rule out case studies in appropriate circumstances, but statistical considerations are not relevant to these.

4 In order to make the best use of resources, variables should be measured and analysed at the highest appropriate level. For example, age should if possible be recorded in years rather than in intervals of ten years, say.

5 The use of previously validated questionnaires is preferred. If a new questionnaire is required, its design is obviously crucial. Expert assistance at this stage

can be particularly helpful. Pilot testing is essential; several rounds of testing and amendment are usually required.

6 The lower the response rate in a study, the greater is the doubt about the extent to which those responding are representative of the population of interest. This is of particular concern if less than 75 per cent of the sample responds. Basic information, perhaps on age and gender, might have already been recorded for all those sampled, in which case responders and non-responders should be compared. If the two groups are similar in terms of that basic information, the results from the responders are more likely, though not certain, to carry over to the study group as a whole. Where the responders and non-responders differ substantially in terms of routinely recorded information, the results are questionable.

7 If two types of intervention are to be compared, new and control methods for instance, subjects should be allocated to one or other of the interventions in a random manner. If two groups of similar size are required, random numbers should be used to produce a method equivalent to tossing a coin and selecting the group on the basis of a head (group 1) or a tail (group 2). The allocation of alternate individuals to each treatment is *not* random and is not acceptable in a controlled experiment.

Statistical analysis

1 Some commonly used statistical techniques, if not already familiar to the reader, are covered in readily available basic texts and need not be described in detail: the chi-squared test for contingency tables, with Fisher's exact test as appropriate; the paired and unpaired *t* tests; correlation and linear regression; one-way analysis of variance; and simple rank tests such as Mann–Whitney. For less familiar techniques, their purpose should be briefly explained. A reference should be given, ideally to a widely available text. The manual of a computer package should not be cited, as the explanation is likely to be from the viewpoint of implementation of the technique within the package. In addition, access to the manual will be difficult for those without the package.

   There is a lack of satisfactory modern statistical books aimed at social work in particular, or at the social sciences more generally. The choice of book will depend on both the background of the researcher and the study itself and it might be useful to ask a statistician to suggest an appropriate text. Swinscow and Campbell (2002) provides a readable introduction to statistics from a medical perspective.

2 In a non-parametric analysis such as the Mann–Whitney method, it is the test that is non-parametric, not the data.

3 The ongoing development of computer packages has put in reach many statistical methods for which calculation by hand is not a practical option. Examples are Fisher's exact test for general contingency tables, logistic regression, and repeated measures analysis of variance. Many packages can also deal with multi-

variate techniques such as discriminant analysis and factor analysis. Although the appropriate use of these techniques is encouraged, the principles behind them need to be understood to avoid meaningless results being produced by use of an unsuitable method. The computer package used for any advanced technique should be stated and referenced.

4 The assumptions underlying the statistical tests used should be considered—although their satisfaction may not be paramount. For example, the chi-squared test is commonly applied without thought being given to the underlying assumptions. It should not be used for a 2×2 contingency table where any expected value is less than five. For larger tables, most texts state that no more than 20 per cent of the expected values may lie between one and five, and none should be less than one. For the unpaired t test, the data are assumed to come from normal distributions with equal variances; a moderate degree of skewness should not influence the results unduly but a modification of the unpaired t test that does not require the equal variances assumption is available with many computer packages.

5 Simple tables should be presented in such a manner that a careful reader is able to check the calculations. The statistical assessor can then demonstrate how incorrect calculations should have been performed. It is not sufficient to give percentages only; base numbers will make clear the group under discussion. It is not necessary to present the individual values used to calculate means and standard deviations, etc., but it may be appropriate for authors to supply more detail for the referee than is to be included in the published paper, avoiding the delay inevitable if the referee has to ask the editor to request additional information.

6 The hypotheses to be tested should be set up before the study is carried out. If they are made following a perusal of the data, it may be possible by grouping data in various ways to produce results that appear to be statistically significant when there are no genuine underlying patterns. Findings obtained by this process are questionable.

7 In a study of randomly allocated interventions in which some participants have a change of treatment following the commencement of the study, the analysis should be performed on the original groups, or the 'intention to treat' basis. Doing otherwise can introduce bias into the study as those who change treatment are generally not typical of the subjects as a whole.

8 It is important to distinguish between statistical significance, in the established use of the term, and practical importance. Large studies have a tendency to produce differences that, despite being too small to be of practical importance, are nevertheless statistically significant. Small studies can highlight differences that would be large enough to influence management decisions but lack statistical significance due to a modest sample size. A finding that is not statistically significant may not necessarily be irrelevant. The reporting of only those results that are statistically significant encourages publication bias due to an over-representation of 'significant' studies in the literature.

9 If a paper is initially drafted using incomplete data, the manuscript should be checked against the final analyses that include the late returns. Failure to do so is likely to lead to inconsistent figures and arithmetic errors in the submitted article.

Presentation

1 Percentages should be rounded to the nearest whole number unless the sample size is over 100, in which case one decimal place may be appropriate. Values should always be rounded rather than truncated—for example, 21.4 should be given as 21 whereas 21.5 and 21.9 should be given as 22. Percentages relating to categories within a group will sometimes add up to for example 99 or 102 due to rounding. The category percentages should not be adjusted to make them add up to 100, although the total should be shown as 100 per cent.

2 Means should normally be rounded to one significant digit more than the original data, and standard deviations to two additional digits. For example, if each individual produces a whole number score, means should be given to one decimal place and standard deviations to two decimal places.

3 For skewed data, the median and interquartile range may be more appropriate as atypical outliers can heavily influence the mean and standard deviation.

4 In order to keep the paper as readable as possible, detailed statistical results (for example F or t statistics, numbers of degrees of freedom) should be avoided where possible, although appended to the main document they may give useful supplementary information for the referee.

5 P values should be shown where appropriate in or immediately below the relevant tables. Exact p values should be given unless $p < 0.001$ or $p > 0.1$. For very small p values do not state, like some statistical packages, that $p = 0.000$. This gives the incorrect impression that the p value is precisely zero. The traditional $p = 0.05$ threshold between 'significant' and 'not significant' results should be avoided as it leads to very similar findings ($p = 0.049$, $p = 0.051$) being interpreted in entirely different ways. P values should be visualized on a sliding scale indicating the strength of evidence against the 'null' hypothesis (no difference, no linear correlation, etc.). P values close to zero show the strongest evidence to the contrary. Use of 'ns' for 'not significant' should be avoided; stating $p > 0.1$ is more informative.

6 Key findings should be presented with confidence intervals so that the reader can assess the range of plausible true values for a difference, for instance. All plausible values should be considered in terms of statistical significance and practical importance.

7 Care should be taken with the layout of tables, so that they contain the relevant information without intimidating the reader. The numbers in a table should be presented so as to highlight the main patterns in the data, but the table cannot speak for itself: those patterns must be described in the text. Detailed advice is

available from Ehrenberg (1982) or Chapman (1986). Among their recommendations are that labels should be clear and brief; numbers are easier to read downwards than across; numbers that have to be compared should be close together; spacing and horizontal ruled lines should be used to guide the eye, with minimum use of vertical grid-lines; and irregular spacing of rows and columns tends to be particularly distracting. Footnotes may be used to expand brief headings, or to prevent misinterpretation of the data.

8 Reference can be made to internet material but the full web address should be given and the date of access stated.

Discussion

An important aspect of a paper aimed at social workers is the discussion of the results in the context of previous social work research. Results that confirm previous findings add to the support for an established theory and this should be pointed out. Unexpected findings, which might challenge commonly held views, should not be automatically accepted. New findings require confirmation by further research. Note that statistically significant results will, from time to time, arise on the grounds of chance alone, particularly if several tests are carried out and the strengths of evidence against the null hypotheses are marginal. Conversely, an insufficiently large sample can lead to new insights being undetected or an expected positive result failing to emerge. A trend in the expected direction may indicate that a larger study should be conducted.

Concluding comments

As statistical assessors we are aware that it is not always easy for social workers to obtain statistical assistance. A long list of detailed comments should not be received with dismay: working through them in the preparation of a revised manuscript is often the easiest way to produce a document for which the statistical aspects are adequate. In common with social work professionals, we find the publication of well designed and appropriately analysed studies relevant to social work satisfying.

References

Altman, D. G., Gore, S. M., Gardner, M. J., and Pocock, S. J. (2000) 'Statistical guidelines for contributors to medical journals', in Altman, D. G., Machin, D., Bryant, T. N., and Gardner, M. J. (eds), *Statistics with Confidence—Confidence Intervals and Statistical Guidelines*, 2nd edn, London, BMJ Books, pp. 171–90.

Chapman, M. (1986) *Plain Figures*, London, HMSO.

Ehrenberg, A. S. C. (1982) *A Primer in Data Reduction*, Chichester, Wiley.

Goda, D. and Smeeton, N. (1993) 'Statistical considerations in social work research', *British Journal of Social Work*, **23** (3), pp. 277–81.

Swinscow, T. D. V. and Campbell, M. J. (2002) *Statistics at Square One*, 10th edn, London, BMJ Books.