EDITORIALS

How to present statistics in medical journals

In the 1990s the movement for 'evidence-based medicine', that medical decisions should be based on objective evidence rather than simply on experience, took off. Since then the number of medical studies involving the analysis of data has rapidly increased. Data arise from a variety of types of studies: clinical trials, observational studies, meta-analyses and others.

There is a range of readily accessible sources of recommendations on how to present statistics in medical journals. The instructions for authors for this journal (http://www.oxfordjournals.org/our_journals/ageing/for_authors/) refer to two sources: the Uniform Requirements of the International Committee of Medical Journal Editors (ICMJE) (http://www.icjme.org, 2010) and the CONSORT statement [1].

The Uniform Requirements of the ICMJE include basic guidelines for presenting statistical information and state, among other things, that statistical methods should be described 'with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results'.

The CONSORT statement is specifically directed at the reporting of parallel group randomised trials. The guidelines comprise 25 items covering the whole range of the report, almost all of them having a bearing on statistics. Many of these items also apply to other types of study and will be discussed below. Those which have particular relevance in a randomised trial include guidance on reporting the method of randomisation and blinding.

Other guidelines available include the PRISMA statement [2], for reporting systematic reviews and meta-analyses, the STARD statement [3], for reporting studies of diagnostic accuracy, and the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement [4].

The existence of the STROBE guidelines is particularly important. Observational studies are the most common type of study used in medical research, but are also more likely than randomized trials to lead to erroneous conclusions, typically because of confounding. A well-known example was the finding, based on observational studies, that hormone replacement therapy (HRT) provided protection against coronary heart disease [5]. This was subsequently contradicted following several large randomised controlled trials [6]. The earlier incorrect finding was probably due to the likelihood that both risk of coronary heart disease and use of HRT are associated with socioeconomic status [7]. The most well-conducted observational study can produce misleading results if important confounders are not known. The STROBE checklist for reporting observational studies includes specific

recommendations relating to potential confounders. It also suggests that in Discussion authors should 'discuss limitations…taking into account sources of potential bias' and 'give a cautious overall interpretation of results'. There are STROBE checklists specifically for cohort, case–control and cross-sectional studies.

For readers to be able to evaluate the strength of the conclusions which have been based on statistical evidence presented in a journal article, they need to know the aims of the study, how the data were collected, and how they were analysed. Statistical reviewers will be checking (as far as possible) the validity of the statistical methods used, and whether they are adequately explained and whether the conclusions are reasonable.

The first step in describing how the data were collected is to name the study design. This should appear in the title or the abstract. For an observational study, this will usually be one of cohort (longitudinal), cross-sectional or case–control. Sometimes adjustments are made for confounding variables in the design of a study by matching the treatment groups with respect to the confounders. You should state whether a study is matched, and, if so, which matching variables were used, and whether the matching is, for example, pair wise. A study which is simply described as 'matched' could mean that the matching is paired, implying that a special form of analysis must be applied. Alternatively, the matching could be at the group level, carried out to ensure that the groups have a similar overall composition in terms of confounding variables and lead to a different analysis. It is important to avoid overmatching, where a matching variable is not a true confounder, or is redundant because it is highly correlated with other matching variables. An example of overmatching is given by Marsh *et al.* [8]. They investigated the relationship between cumulative radiation dose at Sellafield nuclear power station and mortality from leukaemia using a case–control study, matching on a number of variables including date of birth and date of entry. The relationship was not found to be significant. The conclusion was reversed when date of entry was omitted from the analysis, the rationale being that, by matching on both date of birth and date of entry, workers in the same matched sets had similar radiation doses. This obscured the relationship between dose and mortality.

In evaluating the statistical evidence in a research paper, it can be difficult to distinguish between errors in presentation and errors in analysis. Poor presentation, apart from blurring findings, could suggest that are serious problems

in the analysis. Some fairly typical errors we have encountered in the course of reviewing a number of submissions to this journal over the last few years could be classified as errors of interpretation, of analysis or of presentation (or a mixture). A quite common example of incorrect interpretation is to state that a lack of statistical significance implies no association.

In medical research, continuous variables are frequently converted into categorical variables by grouping values into a few categories. This can have the apparent advantage of simplifying interpretation but can have serious drawbacks. It can be difficult to choose both the number of categories and suitable cutpoints to determine them. Gronbaek *et al.* [9] investigated the relation between alcohol consumption and mortality in elderly people. The authors had a large dataset and grouped weekly alcohol intake into seven categories. They were thus able to demonstrate a U-shaped relation between alcohol and mortality. With a smaller number of categories, this effect may not have been seen. In particular, with only two categories (low and high alcohol), it is possible that no relation at all between the 2 variables would be apparent [10].

In medical studies generally, analyses of subgroups of participants are quite often carried out. Sometimes this seems to have been done from the point of view of simplifying the presentation, for example, by carrying out separate analyses for males and females. An interaction test should be carried out first to investigate whether the level of response to the exposure variable depends on gender. The probability of a false-positive finding rises with the number of subgroup analyses performed [11].

Commendably, multivariable methods frequently appear now in submitted papers, including multiple linear, logistic and Cox regression. Errors that arise with these methods include misuse of automated methods of selection of covariates, either not including or misinterpreting interaction terms, and not checking assumptions (e.g. the assumption of linearity between the dependent and the explanatory variables in linear regression and the proportional hazards assumption of Cox regression). Another detail which is sometimes omitted is an assessment of goodness of fit. This is important because the validity of the conclusions depends on the model providing an adequate fit [12].

Care should be taken with graphical presentation. In one case, response at 0, 2, 4 and 8 weeks was plotted showing equal intervals between the time points and thus exaggerating the appearance of the rate of change between 4 and 8 weeks. This could have happened through using a 'line plot' rather than a 'scatter plot' in Excel! Other errors in presentation encountered include inadequate description of the control group, number of participants in categories not given, incorrect use of statistical terminology (or using statistical jargon such as 'likelihood' in a statistical context when the common English meaning was intended), and inconsistencies in tables of results, for example, confidence intervals not consistent with P values. However, if any apparent inconsistency is valid, an explanation should be given.

Statistical analysis forms an essential part of a large proportion of medical investigations. The need is to use the most appropriate statistical method, which if possible is also not over-complex, and to interpret the results clearly. A number of journals have published guidelines for carrying out commonly used statistical methods, including the British Medical Journal which has published an excellent series of >60 short articles on a wide range of statistical topics [13]. The various published guidelines for reporting studies are clear and easy to follow. With increasing awareness of these, the presentation of statistics will continue to improve.

Key points

- Statistics.
- Journals.
- Ageing.

E. Cheek[1], C. Rajkumar[2],*
[1]Department Computing, Engineering and Mathematics, University of Brighton, Brighton, UK
[2]Department of Medicine, Brighton and Sussex Medical School, University of Sussex, Brighton, UK
*Address correspondence to: C. Rajkumar, Email: raj.rajkumar@bsuh.nhs.uk; da.elderlycare@bsuh.nhs.uk Tel: (+44) 1273 523360; Fax:(+44) 1273 523366.

References

1. Moher D, Schulz K, Altman D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. BMJ 2010; 340: c332.
2. Moher D, Liberati A, Tetzlaff J, Altman D. Preferred reporting items for systematic reviews and meta-analyses (PRISMA). J Clin Epidemiol 2009; 62: 1006–12.
3. Bossuyt P, Reitsma J, Bruns D *et al*. Standards for reporting of diagnostic accuracy. Toward complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. standards for reporting of diagnostic accuracy. BMJ 2003; 326: 41–4.
4. von Elm E, Altman D, Egger M, Pocock S, Gøtzsche P, Vandenbroucke J. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007; 370: 1453–7.
5. Psaty BM, Heckbert SR, Atkins D *et al*. A review of the association of oestrogens and progestins with cardiovasculardisease in postmenopausal women. Arch Intern Med 1993; 153: 1421–7.
6. Beral V, Banks E, Reeves G. Evidence from randomised trialson the long-term effects of hormone replacement therapy. Lancet 2002; 360: 942–4.

7. Potshuma WF, Westendorp RG, Vandenbroucke JP. Cardioprotective effect of hormone replacement therapy in postmenopausal women: is the evidence biased? BMJ 1994; 308: 1268–9.

8. Marsh JL, Hutton JL, Binks K. Removal of radiation dose response effects: an example of over-matching. BMJ 2002; 325: 327–30.

9. Gronbaek M, Deis A, Becker U, Hein HO, Schnohr P, Jensen P *et al.* Alcohol and mortality: is there a U-shaped relation in elderly people? Age and Ageing 1998; 27: 739–44.

10. Altman DG. Categorizing Continuous Variables. Encyclopedia of Biostatistics. Chichester: John Wiley & Sons, 2005.

11. Wang R, Lagakos S, Ware J, Hunter D, Drazen J. Statistics in medicine – reporting of subgroup analyses in clinical trials. N Engl J Med 2007; 357: 2189–94.

12. Bender R. Logistic regression models used in medical research are poorly presented. BMJ 1996; 313: 628.1.

13. Altman DG, Bland JM. Parametric v non-parametric methods for data analysis. BMJ 2009; 338: a3167.

The human immunodeficiency virus and ageing

Introduction

HIV, first described clinically in 1981 and identified in 1983, rapidly emerged to be a major cause of morbidity and premature mortality. Prior to the introduction of combination antiretroviral therapy (cART) in 1996, life expectancy was poor—typically <2 years after initial diagnosis—with the majority of deaths arising secondary to AIDS defining illnesses [1]. Presently, for HIV-positive individuals receiving cART in accordance with recommended guidelines, life expectancy has improved significantly, with many studies suggesting it to be near normal [2–3]. The majority of deaths now occurring in this population are secondary to non-AIDS defining illnesses [1].

The United Kingdom: a changing epidemiological picture

In 2012, the Health Protection Authority (HPA) published epidemiological estimations, predicting there were 96,000 people in United Kingdom living with HIV (95% credible interval: 90,800–102,500), of whom, 73,660 are diagnosed and in care [1]. As the number of new diagnoses per year remains at best stable, and as outcomes improve, this total figure continues to increase year on year. Largely as a result of this improved survival, the demographics of those accessing care is changing. In 2011, one in five adults (22%, 16,550) receiving HIV care in the United Kingdom was aged over 50, whereas in 2002 the comparable figure was one in nine (12%, 3,640) [1]. In the United States, it has been estimated that by the year 2015, over 50% of HIV-positive patients in care will be over the age of 50 [4].

It is important to note that the short-term mortality in those diagnosed late (with an initial CD4 count <350 cells/mm^3) remains high despite the availability of cART, with an observed 10-fold increase in mortality when compared with those diagnosed early (CD4 count > 350 cells/mm^3) [1]. Multiple data have shown that older persons are significantly more likely to be diagnosed later than younger persons and thus are at significantly higher risk of short-term mortality [5]. It is imperative therefore that geriatricians consider HIV infection within their differential diagnosis and recommend testing particularly in those with clinical indicator diseases, as recommended in national and international HIV testing guidelines (http://www.bhiva.org/documents/Guidelines/Testing/GlinesHIVTest08_Tables1-2.pdf) [6] (Figure 1).

The impact of ageing on HIV

Prior to the advent of cART, it was demonstrated that older individuals recently acquiring HIV infection had a significantly faster progression to AIDS or death [7], although this effect appears to be reduced or even lost since the advent of effective therapy. On initiation of cART, overall it appears that older patients have a similar or improved virological response—likely driven by improved adherence to therapy, although the immunological response to cART may be impaired [8, 9]. Some guidelines have recommended earlier initiation of cART amongst older individuals, defined by their considered cardiovascular risk [10].

The impact of HIV on ageing

It has been proposed that HIV infection is associated with accelerated or premature ageing. First, in the ageing and the HIV population a similar disruption to normal immune function, described as immunosenescence, is seen—this is summarised in Table 1 [11].

Second, it is postulated that, despite viral suppression by cART, there remains a persistent low grade systemic immune activation and inflammation which contributes to an accelerated ageing pattern [11]. Third, it has been proposed that this accelerated or premature ageing may be a consequence of either toxicities resulting from long-term use of antiretroviral