Analysis of longitudinal data

Beyond MANOVA

B. S. EVERITT

Background Longitudinal data arise frequently in psychiatric investigations, and are most often analysed by multivariate analysis of variance (MANOVA) procedures. However, as routinely applied, the method is not satisfactory, particularly when the data are affected by subjects dropping-out of the study. More suitable methods are now available.

Method Problems with the MANOVA approach are discussed and the advantages of alternative procedures stressed.

Results Using MANOVA on complete cases to analyse unbalanced longitudinal data can be seriously misleading. More recently developed methods are far more suitable, but only if the missing values are non-informative.

Conclusions Routine use of MANOVA for the analysis of longitudinal data, particularly when there is a substantial proportion of drop-outs, is ill advised. Statisticians have considerably enriched the available methodologies during the past decade, and psychiatric researchers dealing with such data should be aware of the advantages of the newer methods.

Longitudinal data arise when each member of one or more cohort or panel of subjects provides a measurement on a number of occasions. The cohort may, for example, be defined in terms of the date of birth of its members, the time of onset of a disease or, in the case of a clinical trial, the beginning of treatment or time of randomisation. The repeated measurements recorded may be quantitative or qualitative. The main objectives of studies which give rise to such data (longitudinal studies) are to characterise change in the response variable(s) over time and to examine covariates which contribute to that change. The clinical trial is perhaps the most common type of longitudinal study. Analysis of longitudinal data is not straightforward because of problems such as the serial correlations between the repeated measurements, and patients who drop-out permanently or miss a scheduled visit. The aim of this paper is to give a non-technical account of the most appropriate methods for the analysis of such data and to point out how the techniques currently most commonly used by psychiatric researchers are at fault.

LONGITUDINAL STUDIES

Longitudinal studies involve the repeated measurement of the same dependent variable on a subject or patient on a number of different occasions. In many such studies the subjects will have been randomly allocated to one of several treatments, and often the repeated measures will include several taken prior to randomisation in addition to those taken after treatment begins. Frequently a number of other covariates, some of which may vary over time, will also be of interest.

The design of most longitudinal studies specifies that all subjects are to have the same number of repeated measurements made at the same time points. Such data are generally referred to in this context as 'balanced'. Although balanced longitudinal data are aimed for, unbalanced data, in which subjects may have different numbers of repeated measurements made at different time points, frequently arise for a variety of reasons. Occasionally, the data are unbalanced or incomplete by design; an investigator may, for example, choose in advance to take measurements every hour on one-half of the subjects and every two hours on the other half. In general, however, the main reason for the appearance of unbalanced data in a longitudinal study is the occurrence of missing values, in the sense that intended measurements are not taken, are lost or are otherwise unavailable. Here the data are incomplete because, for example, readings were taken incorrectly, subjects missed appointments, or subjects dropped-out of the study prematurely.

Longitudinal data that are unbalanced, for whatever reason, raise technical difficulties. Some methods of analysis can only cope with data for which measurements are made at the same set of time points for each subject. Unplanned missing values also raise deeper conceptual issues, since it has to be asked why the values are missing and, more specifically, whether their being missing has any bearing on the practical questions to be posed about the data.

MISSING VALUES IN LONGITUDINAL STUDIES

Here attention will be restricted to missing values that arise because of a subject dropping-out of the study, so that once an observation at a particular time point is missing, so are all subsequent planned observations. (The distinction between missing values that occur because of drop-outs and those that arise when data are missing intermittently is really one of convenience for description rather than one of fundamental importance.)

In the statistical literature three types of drop-out have been distinguished:

(a) Missing completely at random (MCAR) – the probability of drop-out does not depend on the subject's observed measurements or those that would have been available had they not been missing. Hence, the observed values effectively constitute a simple random sample from the distribution of the repeated measurements. The missing values would have the same distribution as the values observed.

https://doi.org/10.1192/bjp.172.1.7 Published online by Cambridge University Press

(b) Missing at random (MAR) – the probability of drop-out depends on the outcomes observed in the past (and possibly other covariates) but not on the unrecorded values of the response variable following drop-out.

(c) Non-ignorable (informative) – the probability of drop-out depends on the unobserved values of the response variable.

The distinction between the three types is described in detail in more technical terms in Little & Rubin (1987) and Diggle *et al* (1994).

Drop-outs that are MCAR or MAR are often referred to as 'ignorable', but the use of the term does not imply that the subjects with missing data can simply be disregarded; rather, the term ignorable is used to indicate that it is not necessary to take account of the drop-out mechanism in some of the methods of analysis to be mentioned later. It is the non-response process that can be ignored, not the subjects with missing data. This is not a trivial point since, as will be seen in the next section, one of the most frequent methods for dealing with missing data in longitudinal studies remains the analysis of only those subjects having a complete set of observations. Fortunately, the more recently developed methods of analysis outlined later can accommodate non-informative missing values satisfactorily and thus make use of subjects whose data are incomplete.

A simple (non-longitudinal) example taken from Diggle *et al* (1994) may help to clarify why this taxonomy of missing values has implications for data analysis:

> Suppose that we want to compare the mean concentration of a hormone in blood samples taken from each of ten subjects in each of two groups; one a control, the other an experimental treatment intended to suppress production of the hormone. We are presented with ten assayed values of the hormone concentration from the control subjects, eight assayed values from the treated subjects, and are told that the values of the other two treated subjects are 'missing'. If we knew that this was because somebody dropped the test tubes on the way to the assay lab, we would probably be happy to proceed with a two-sample t-test using the 18 non-missing values. If we knew that the missing values were from subjects whose hormone concentrations fell below the sensitivity threshold of the assay, ignoring the missing values would mask the very effect we are looking to detect.

In the first case the observations are likely to be MCAR but in the second they are informative.

ANALYSIS OF LONGITUDINAL DATA

A model for longitudinal data has two components: a regression model for the dependence of the response variable (which, for the moment, will be assumed to have a normal distribution) on time and the covariates of interest (one of which, particularly in clinical trials, will be treatment group); and a model for the serial correlation (or autocorrelation) among the repeated observations for an individual. In most cases, it is the parameters of the regression model that are of prime interest but estimates of these and their standard errors, and consequently inferences about them, will be affected by the model chosen to describe the serial correlations.

The most frequently used methods for analysing longitudinal data from psychiatric studies appear, from an examination of the psychiatric literature, to be univariate analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA) with 'time' being regarded as a within-subjects factor. If it is assumed that the only covariate involved is treatment group, then the regression model component and serial correlation model component of these two approaches are as follows:

> *ANOVA* – the regression model consists of group, time and group × time effects, and the serial correlation structure is constrained to be compound symmetry, implying that in each treatment group the variances of the observations at each time point are the same and the covariances of the observations at each pair of time points are also equal.

> *MANOVA* – the regression model again consists of group, time and group × time effects, but no constraints are now placed on the form of the serial correlation structure of the repeated measures.

Both the ANOVA and MANOVA approaches to the analysis of longitudinal data are described in detail in Dunn & Everitt (1992).

Although ANOVA is still occasionally used for the analysis of longitudinal data, most researchers are now well aware of its limitations. The assumed compound symmetry structure for the observed correlations is very unrealistic since observations taken closer together in time are likely to have a larger correlation than those taken further apart. Also, the variances of the observations taken at each time point will often differ. Consequently, MANOVA is now used almost routinely. Unfortunately, it too has drawbacks of which most users seem to be unaware. The prime difficulty is that of 'overparameterisation' (i.e. allowing an excess number of parameters in the model for the serial correlations). Diggle (1988) suggests that this will lead to inefficient estimation and potentially poor assessment of the standard errors of the model parameters of most interest, namely the treatment group and group × time interaction effects. Consequently, the MANOVA approach to the analysis of longitudinal data can lead to erroneous inferences about these effects.

Fortunately, it is now relatively simple to build models for longitudinal data which allow consideration of a much richer variety of correlational structures for the observed serial correlation than the unrealistic compound symmetry structure implied by the use of ANOVA and the unconstrained but overparameterised structure used by the MANOVA procedure. Such models are described in detail in Jennrich & Schluchter (1986) and Diggle *et al* (1994), and are implemented in software such as BMDP5V (Statistical Solutions, Cork Technology Park, Model Farm Road, Cork, Ireland) and Proc Mixed in SAS (SAS Institute Inc., Box 8006, Cary, NC 27511-8000, USA). Using such models enables a correlation structure to be determined that describes the observed serial correlations with as few parameters as possible, thus helping to prevent the misleading inferences about the treatment group effects mentioned above.

The advantages of these newer modelling procedures become even more pronounced when dealing with unbalanced longitudinal data since the current practice for dealing with missing values when using MANOVA is particularly poor. Two approaches seem to be in common use. The first, 'complete case analysis', uses only the subset of subjects with no missing values. This procedure is valid only if the values are missing completely at random, but even then it may be highly inefficient, particularly when a large proportion of subjects have to be discarded. If the values are not MCAR, then complete case analysis may be seriously biased and, consequently, seriously misleading. In both cases an analysis

https://doi.org/10.1192/bjp.172.1.7 Published online by Cambridge University Press

that uses the information available in incomplete cases is essential (see Newer approaches, below).

An alternative to complete case analysis when using MANOVA for the analysis of longitudinal data is to replace the missing values with imputed values. The two most common methods for estimating missing values are the use of cross-sectional means (replacing missing values with the cross-sectional mean over all subjects observed at the appropriate time point) or 'last observation carried forward' (substituting each missing value with the subject's last available assessment of the same type). Both approaches have little to recommend them. Cross-sectional means can be misleading if the data are not MCAR, and they are likely to lead to underestimation of standard errors, since uncertainty in the imputed values is not reflected in the analysis. Likewise, last observation carried forward appears to have little in its favour except possibly an appeal to clinicians in that this value records what has been achievable with a particular patient. The procedure involves, however, highly unlikely assumptions, for example that the (unobserved) post-drop-out responses remain frozen at their last observed value.

NEWER APPROACHES

The more recently developed techniques for analysing longitudinal data described in Jennrich & Schluchter (1986) and Diggle *et al* (1994), and implemented in BMDP5V and Proc Mixed in SAS, are specifically formulated so that they can accommodate non-informative missing values (i.e. MCAR or MAR) in the models used. Consequently, such missing values raise no particular problems in the estimation of parameters or their standard errors. (It should be remembered, however, that the percentage of missing data in a study can be considered as one indicator of the quality of the data and, therefore, the quality of the study. It is important not to be seduced into thinking that investigations that are carried out poorly can be rescued by sophisticated statistical analysis.)

Although valid inferences can be obtained when the missing values are non-informative, validity does depend upon fitting the correct model to the data, far more so than when the data are complete. The increased sensitivity to model misspecification in the case of unbalanced data arises because different models will lead to different imputed values for the missing observations. Consequently, data model validation becomes of even greater importance than usual.

Examples of the use of these newer techniques applied to the analysis of unbalanced longitudinal data resulting from psychiatric studies are given in Everitt (1995*a*), Gregoire *et al* (1996) and Kemp *et al* (1997).

In the discussion above it has been assumed that the response variable can be assumed to have a normal distribution. As with many statistical procedures based on the normal distribution, some departure from the assumption is unlikely to be crucial; nevertheless, when non-normality is severe (for example when the response variable observed at each time point is binary) the methods described will not be appropriate. Recently, however, statisticians have developed a number of techniques that are suitable for the analysis of non-normal longitudinal data. These methods are described in Liang & Zeger (1986), Conoway (1990), Davis (1991), Carey *et al* (1993), Fitzmaurice *et al* (1993) and Heagerty & Zeger (1996). Unfortunately, the development of software to make the application of these powerful methods routine has lagged behind, although some software is available, for example STATA (Stata Corp., 702 University Drive East, College Station, TX 77840, USA) and Oswald (Object oriented software for the analysis of longitudinal data in S-plus; Statistics Group, Centre for Applied Statistics, University of Lancaster, Lancaster LA1 4YF).

A further exciting development in methods for the analysis of longitudinal data is the Bayesian approach described in, for example, Spiegelhalter *et al* (1993). Such methods depend heavily on the increased computing power of today's generation of personal computers, relying as they do on computationally intensive techniques such as Gibbs sampling; see Everitt (1995*b*). Again, experimental (but well documented) software is available via the BUGS package (Bayesian inference under Gibbs sampling; MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR).

Finally, what if the missing values are informative, and how can one tell? This is a difficult and, as yet, largely unresolved problem, although work is in progress that should help in the near future (see, for example, Diggle & Kenward (1994), Little & Wang (1996) and Molenberghs *et al* (1997)). Arguably, however, in the presence of informative drop-out a wholly satisfactory analysis of the data will not be feasible.

To conclude, according to Hand & Sham (1995), "statisticians are often frustrated by the slow uptake of statistical ideas by the medical profession" (see also Altman & Bland, 1991). But the fault is not always the clinician's. Statisticians need to make the effort to inform their clinical colleagues of important new developments relevant to their research. This has been the aim of the present paper and it is hoped that it will persuade psychiatric researchers working with longitudinal data (particularly when the data are unbalanced) that the extra effort that may be needed in tracking down the relevant software or the help of a statistician in order to undertake a more appropriate analysis is likely to be well worthwhile. Of course, no account of new techniques such as this will help the clinical researcher who insists on collecting dozens of outcome measures on several different occasions in a longitudinal study and analysing them as if each provided new information. Such researchers often also completely ignore the missing data problem and usually present their results in the form of large and unpallatable tables of *P* values. This is simply poor science and the editors of psychiatric journals and their statistical advisors should do their utmost to persuade authors against such practices.

REFERENCES

Altman, D. G. & Bland, J. M. (1991) Improving doctors' understanding of statistics. *Journal of the Royal Statistical Society (Series A)*, **154**, 223–267.

Carey, V., Zeger, S. L. & Diggle, P. J. (1993) Modelling multivariate binary data with alternating logistic regressions. *Biometrika*, **80**, 517–526.

Conoway, M. R. (1990) A random effect model for binary data. *Biometrics*, **46**, 371–328.

Davis, C. S. (1991) Semi-parametric and non-parametric methods for the analysis of repeated measurements with applications to clinical trials. *Statistics in Medicine*, **310**, 1959–1980.

Diggle, P. J. (1988) An approach to the analysis of repeated measures. *Biometrics*, **44**, 959–971.

—— & Kenward, M. G. (1994) Informative drop-out in longitudinal data analysis. *Applied Statistics*, **43**, 49–93.

——, Liang, K. & Zeger, S. L. (1994) *Analysis of Longitudinal Data*. Oxford: Oxford University Press.

Dunn, G. & Everitt, B. S. (1992) *Applied Multivariate Data Analysis*. London: Arnold.

Everitt, B. S. (1995*a*) The analysis of repeated measures: a practical review with examples. *Statistician*, **44**, 113–135.

https://doi.org/10.1192/bjp.172.1.7 Published online by Cambridge University Press

— **(1995*b*)** *The Cambridge Dictionary of Statistics in the Medical Sciences*. Cambridge: Cambridge University Press.

Fitzmaurice, G. M., Laird, N. M. & Rotnitzky, A. G. (1993) Regression models for discrete longitudinal responses. *Statistical Science*, **8**, 248–309.

Gregoire, A. J. P., Kumar, R., Everitt, B. S., *et al* (1996) Transdermal oestrogen for treatment of severe postnatal depression. *Lancet*, **347**, 930–933.

Hand, D. & Sham, P. (1995) Improving the quality of statistics in psychiatric research. *British Journal of Psychiatry*, **167**, 689–691.

Heagerty, P. J. & Zeger, S. L. (1996) Marginal regression models for clustered ordinal measurements. *Journal of the American Statistical Association*, **91**, 1024–1036.

Jenrich, R. I. & Schluchter, M. D. (1986) Unbalanced repeated measures models with structured covariance matrices. *Biometrics*, **42**, 805–820.

B. S. EVERITT, MSc, Department of Biostatistics and Computing, Institute of Psychiatry, De Crespigny Park, Denmark Hill, London SE5 8AF

(First received 8 April 1997, revised 1 August 1997, accepted 28 August 1997)

Kemp, R., Everitt, B. S., Hayward, P., *et al* (1997) A randomised controlled trial of compliance therapy: an 18-month follow-up. *British Journal of Psychiatry*, in press.

Liang, K. Y. & Zeger, S. L. (1986) Longitudinal data analysis using generalised linear models. *Biometrika*, **73**, 13–22.

Little, R. J. A. & Rubin, D. B. (1987) *Statistical Analysis with Missing Data*. New York: John Wiley.

— **& Wang, Y. (1996)** Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics*, **52**, 98–111.

Molenberghs, G., Kenward, M. G. & Lesaffre, E. (1997) The analysis of longitudinal data with nonrandom drop-out. *Biometrika*, **84**, 33–44.

Spiegelhalter, D. J., Freedman, L. S. & Parmar, M. K. B. (1993) Applying Bayesian ideas in drug development and clinical trials. *Statistics in Medicine*, **12**, 1501–1512.

https://doi.org/10.1192/bjp.172.1.7 Published online by Cambridge University Press