Proceedings of the

CENTRE FOR MATHEMATICS
AND ITS APPLICATIONS

AUSTRALIAN NATIONAL UNIVERSITY

Volume 28, 1991

WORKSHOP ON DESIGN OF
LONGITUDINAL STUDIES AND ANALYSIS
OF REPEATED MEASURES DATA
(1991: CANBERRA, ACT)

Edited by S.R. Wilson

PROCEEDINGS

OF

THE CENTRE FOR MATHEMATICS AND ITS APPLICATIONS

AUSTRALIAN NATIONAL UNIVERSITY

VOLUME 28, 1991

WORKSHOP ON DESIGN OF LONGITUDINAL STUDIES AND
ANALYSIS OF REPEATED MEASURES DATA
(1991 : CANBERRA, A.C.T.)

EDITED BY S.R. WILSON

PREFACE

This Volume presents the general Proceedings of the Data Analysis Workshop II, on the **Design of Longitudinal Studies and Analysis of Repeated Measures Data**, held at the Centre for Mathematics and Its Applications, ANU, July 1991.

The aim of this Workshop series is to define the state of the art in applied statistics. This Workshop brought together, in Australia, scientists and researchers with experience in the design of longitudinal studies, in the development of analytic methods for repeated measurements data and in the application of these techniques. The Workshop's Special Guest and Keynote Speaker, Professor David Hand, provided four data sets to stimulate discussion, and these were made available to participants prior to the Workshop.

The Workshop was (partly) supported by the Department of Industry, Technology and Commerce (DITAC) Bilateral Science & Technology Programs.

The Proceedings is presented in three sections which follow the general format of the Workshop. Reflecting the informal structure of the Workshop, this summary refers to the papers by presenting author/s (in italics). The first section, **General Overviews and Recent Developments**, commences with an overview paper by the keynote speaker, *David Hand*. This paper identifies a useful set of factors for choosing which of the various methods are best suited to the data and research issues at hand, and then describes the methods. Repeated measures data commonly arise from either designed experiments or longitudinal surveys, and there is a general paper highlighting issues of concern for each of these foci: *Roger Mead* overviews both general design principles and consequent statistical principles, motivating this discussion with real data examples; *Ray Chambers* gives a strategy for analysis of longitudinal survey data where complex, highly stratified sample designs are used, along with partial sample rotation and accommodating non-response. How to deal with correlated errors is a fundamental problem in longitudinal studies, and *Chris Glasbey* gives an overview of the general difficulties. The section concludes with

Murray Aitkin describing multilevel variance component models, motivated by the analysis of school effectiveness survey data. Papers appearing later in this Volume contain applications of general theoretical developments which have appeared elsewhere in the literature. *Ari Verbyla* and *Brian Cullis* apply their general approach to modelling repeated measures data to three complex sets, including the two most complex provided for the Workshop. *Richard Huggins* applies his robust estimation and outlier detection procedure for repeated measures to the first Workshop data set.

The core of the Workshop was the data sets mentioned above, and these are given in the section **Workshop Data and Analyses, including Applications of Recent Research**. The data can be briefly summarised as follows: (1) A study of the effect of a vitamin E diet supplement on the growth of guinea pigs (three groups, continuous variable at six time points); (2) Part of a clinical trial of a treatment for sprains (two groups, categorical ordinal scale with missing values, four presentations); (3) Part of a larger study to explore the effects of predictability and control on the anxiety experienced by phobics (two groups, two continuous measures at six presentations); (4) A study of patients who suffer from panic attacks and control subjects (a more complicated continuous variable data set with missing values and three variables measured at 11 times, with a structure of conditions over time).

Five written contributions appear for data set (1): *Richard Huggins* uses robust estimation combined with outlier detection; *Ross Cunninngham* outlines a series of steps, including a graphical technique exploring the covariance structure as well as readily comprehensible testing procedures; *Magdalena Mok* (with Rod McDonald) uses a multilevel model which, however, ignores the more complex nature of the correlations arising from this type of repeated measuring of the same subject over time; *Roger Mead* uses a less model-oriented approach and *Sue Wilson* gives some exploratory graphics. The Conclusions vary: Are there outlying observations and, if so, which ones? Is there an effect due to treatment, and, if so, what is its nature? Simple graphs are also provided with the original data to assist the reader's evaluations of the various approaches.

Three written contributions appear for data set (2): *Bob Murison* develops, and applies, a model for analysing repeated measures of

ordinal data; *Roger Mead* and also *Sue Wilson* use less formal approaches. Again, conclusions are not unequivocal: Is there a difference between the two groups? Do the two groups vary at the first sampling time? Is there a difference between the rate of improvement for the two groups?

There are two written contributions for data set (3): *Roger Mead* uses a straightforward, readily comprehensible, approach and concludes the two groups are very similar. *Ari Verbyla* and *Brian Cullis* apply their general modelling approach which provides much insight into the data, including the possibility of outliers and interesting second-order structure, and also conclude any differences between the groups are inconclusive. Simple graphs also accompany the original data to assist the reader.

There is only one written contribution for the most complex data, set (4): *Ari Verbyla* and *Brian Cullis* show the flexibility of their approach in handling the greater complexity, including missing values, and demonstrate the value of modelling the covariance structure, concluding that there are some differences between the two groups.

The final section, **Specific Applications : Medical, Agricultural & Social Sciences / Statistical Software**, contains a subset of the contributions presented at the Workshop. Some of these illustrate points made in the first, overview, part: *Judy Simpson* discusses the issues of comprehensibility and software availability versus mathematical sophistication and statistical optimality; *Ken Russell* asks how many variables compared with numbers of subjects should one have in longitudinal studies; *Jenny Porteous* highlights some of the difficulties associated with examining complex human behaviours in longitudinal studies. Three further medical applications are concerned with multistate models and "adding value" to large clinical data collections, namely by the development and application of appropriate statistical techniques: *Ross Lazarus* gives the clinician's perspective and outlines the importance, to Australia, of dealing more adequately with these data resources; *Anthony Wohlers* discusses the problems and some methodology, using a longitudinal study in cystic fibrosis as his paradigm; *Chip Heathcote* develops probabilistic theory which is appropriate for dealing with some of this type of data. In the area of agricultural science, *Ari Verbyla* and *Brian Cullis* discuss the analysis of

data from a complex longitudinal experiment evaluating varieties of apples grown under different pruning schemes, and *Ann Cowling* describes factors to be considered in an economic evaluation of an agricultural program. There are two papers describing important Australian government projects: *Peter Boal* (with Geoff Parkinson) details the Australian Longitudinal Survey program which is now the responsibility of the Department of Employment, Education and Training, and highlights the need for comprehensibility of statistical results for a non-technical audience; *Oleh Lukomskyj* outlines plans for a longitudinal survey of immigrants to Australia to gather information on the performance of Government policies, programs and services that affect immigrants.

Finally in this section, *Yvonne Pittelkow* overviews some of the generally available micro computer software for the analysis of repeated measures data.

The scope of this collection of papers is wide, yet, even so, it is important to note that there are other aspects of longitudinal studies that are not addressed here. This collection of papers is focused on <u>data analysis </u>for repeated measures and longitudinal studies, and covers the major facets, difficulties and concerns that are commonly encountered by applied statisticians. So this Volume will be of most interest to consultant statisticians and researchers who are involved with the collection of repeated measures data and the subsequent analyses.

A major benefit of any Workshop is the Discussion sessions. So authors were requested to summarise those parts that were relevant to their contribution, and this has increased the value of many of the papers appearing in this Volume.

Although a uniform style sheet was distributed to all contributors, this was not so uniformly followed. However, no greater information content would follow from insisting on exact style conformity which, moreover, could only be achieved at considerable expense, including time.

Finally we thank Marilyn Gray and Jill Smith for their assistance with the Workshop and the preparation of this Volume.

Sue Wilson, Editor
November, 1991