

LONGITUDINAL STUDIES BASED ON  
CLINICAL DATA COLLECTIONS

*ROSS LAZARUS*

**INTRODUCTION**

This is an entirely non-technical paper for which no further apology will be made. It is written from the perspective of a public health researcher interested in using statistical techniques which are currently under development, but lacking the training and skills required to contribute directly to their development. One purpose of this paper is to describe an area where longitudinal statistical methods will be increasingly applied, in the hope of arousing some interest in this area among professional theoretical and applied statisticians.

Approximately 8% of Australia's GNP is expended in activities collectively termed "health". The focus of most of these activities is in fact illness, but this perverse nomenclature is traditional and is likely to remain with us for some time. In general terms, most modern medical methods are based on observation. Clinicians informally accumulate observations about their patients and gradually refine their recollections of similar patterns of illness into diagnostic groups. With the addition of relatively recent technological advances (eg microscopy, radiographic imaging and so on) many of these clinical entities have been refined considerably and a biochemical or other mechanistic explanation is now available for some organic diseases.

However, it is important to understand that remarkably few treatment methodologies in modern western medical practice have been subject to formal, experimental, controlled (or even uncontrolled) clinical trials which would enable confident assertions about the relative effectiveness of alternative treatment modalities to be made. Clinicians must often choose from among alternative treatments on the basis of their accumulated experience, rather than from experimental evidence. While

experimental clinical trials are the ideal way to expand knowledge in the area of clinical practice, they are expensive and likely to remain relatively uncommon. On the other hand, the raw materials for descriptive research in the form of clinical records are relatively abundant but underused.

One major issue in observational research is the organisation of the collection of observations available for study. Most clinicians use the time honoured method of formally recording observations in written case notes which are held in a hospital or practice record storage system. These are important and rich sources of data but are cumbersome and expensive to study. Electronic storage of clinical data is becoming increasingly common as the computing resources required for this task become more widely available. As a consequence, a number of large clinical data collections are now becoming available for formal study. Some of these derive from purposive longitudinal studies such as the Framingham study being conducted in North America. Others have been mandated by funding agencies for program monitoring such as the one described in the appendix below. There is a great deal of activity in North America, applying a variety of innovative methods to some very large clinical data collections<sup>1</sup>.

### **CHARACTERISTICS TYPICAL OF CLINICAL DATA COLLECTIONS**

For practical reasons, clinical data collections differ in a number of important respects from the kinds of data sets that most statisticians are currently used to dealing with. Moreover, they often suffer from a number of defects which most statisticians would prefer to avoid.

1) *Clinical data are often accumulated with no clearly articulated research goals in mind.* Many existing electronic clinical data collections were designed and implemented with relatively vague aims for future research use. Nearly all clinicians dream of some far reaching discovery leaping out of their precious clinical experience. Of those clinicians storing clinical data electronically, relatively few have formal hypotheses in mind. Most hypothesis directed collections exist in a formal (often funded) research environment and are outside the scope of this paper.

The vision of a white-coated clinician breaking down the door with a floppy disk in one hand begging for some sense to be made of a scrappy data set probably represents some kind of archetypical nightmare in the collective unconscious of statisticians. While not wishing to defend the indefensible, the danger of the baby being tossed out with the bath water remains ever present. One of the themes of this paper is that this nightmare actually presents the statistician with a creative opportunity !

2) *Clinical data sets are almost inevitably longitudinal.* A typical clinical service has a defined or definable "catchment" population. Individuals come to the attention of the service through a variety of means. An individual seen once is likely to be seen again at some time in the future. This is particularly true of clinical services dealing with specific subgroups of the catchment population such as diabetics.

Unfortunately for the statistician, observations on a given individual are likely to be made at intervals determined by the natural history of the disease process or processes rather than by some formal experimental protocol. It is an unusual luxury for a typical clinical service to be able to perform the kinds of timed, fixed interval follow up which would be preferred from an analytic perspective. Further, the intervals between observations are not likely to be random - for example, an individual with more advanced disease is likely to be seen more often. This may be an important source of bias in clinical data collections.

3) *The records of a clinical service may be incomplete accounts of the total use of a given type of service by an individual.* There are usually a number of similar services available to the individual consumer and there is always the possibility that the clinical services used by an individual from a given catchment area will be provided by a distant clinical service. The lack of some uniform identifier which might permit linkage between sets of clinical data is a substantial barrier for Australian researchers. This may be a subtle source of bias.

4) *The data held by a clinical service may contain a substantial proportion of*

*missing data*. Sadly, filling in data collection forms is often a relatively low priority in the day to day activities of a busy clinician. Rarely are staff dedicated to the task of completing data collection forms. In addition, data may be missing for systematic reasons and this is another potent source of bias.

### **SOME MODELS OF POTENTIAL IMPORTANCE.**

Some of the most important potential "value adding" activities exploiting the longitudinal aspects of clinical data include event modelling. For example, models of transitions between health states. Markov Chain and Markov Process models are already in use in the context of decision analysis<sup>2</sup> and developments from these may well prove to be the most important tools in this endeavour. Unfortunately, expertise and suitable computer software are not widely available (at least not to clinicians) at present, and there is much to be done by theoretical and other professional statisticians in this regard.

These models are potentially useful for a variety of purposes, including health services planning, improving our understanding of the natural history of various diseases<sup>3</sup>, clarifying the relative effectiveness of alternative treatment methods<sup>4</sup> and clarifying the relative costs and benefits of alternative interventions. An economic analogy of interest to health economists would be models of transitions between dependency/cost states in common and expensive illnesses (such as dementia).

The development of models of this type would undoubtedly be of considerable interest to a number of funding agencies such as the Commonwealth Department of Health, Housing and Community Services, since this is of importance to their expenditure planning. Exploration of these possibilities may be fruitful activities for statisticians seeking funds for their research efforts in event modelling.

## EVENT HISTORY ASPECTS OF CLINICAL DATA

Assuming that longitudinal aspects of clinical data collections are to be modelled, then we face a number of specific issues. Some of these are currently at the "bleeding edge" of event history analysis and have been summarised by Alison<sup>5</sup>.

1) Competing risks are commonly a problem - for example transitions to the death state compete with other transitions.

2) Some events will change the risk of other events - for example, a single non-fatal stroke is usually associated with an increased risk of a subsequent stroke.

3) Some events are largely independent of each other - for example, the risk of stroke and the risk of lung cancer are largely unrelated (although a patient with lung cancer is unlikely to survive long enough to have a stroke).

4) Many hazards change with time - for example, the risk of dementia of Alzheimer's type increases exponentially with increasing age.

Taken together, these issues substantially complicate the task of modelling this type of longitudinal data. While tools are available for this purpose (such as those described by Chip Heathcote at this conference), they are not widely dispersed or understood at this time.

## CONCLUSION

Clinical data collections represent a potentially rewarding area for "value adding" statistical activities. Statisticians willing to take up the challenge might be able to find funding for their work through sources which they might not normally consider. Understanding and accepting the limitations of clinical data does not necessarily mean a complete loss of integrity. While it may mean working with messy and dirty data, in the long term there may be important results applicable to health service planning and to refining methods of clinical practice. The development and dissemination of appropriate tools and expertise needed for these activities represent a significant challenge for theoretical and applied statisticians.

## **APPENDIX**

### **BRIEF DESCRIPTION OF THE NSW CAP GAT DATABASE**

#### **MAJOR AIMS**

The project was originally instigated in order to satisfy the prescribed requirements for reporting of patient activity by NSW Area Geriatric Teams (GAT) receiving enhancement funding under the Commonwealth Assessment Program (CAP). In addition it was intended to serve as a state wide source of descriptive data on use of GAT services in NSW<sup>6</sup>.

#### **POPULATION AND SAMPLE**

The entire elderly population of NSW is the eligible population. Currently data is being gathered from 39 (of a total of 46) Area Geriatric Services collectively responsible for about 90% of the 70+ population of NSW. Universal coverage for NSW is anticipated by the end of 1991. Participating services typically deal with elderly (poorly defined, generally over 65 years old) patients often with multiple pathology and complex social problems.

The sample comprises all individuals referred to a NSW Area Geriatric Service participating in the data collection exercise - sampling occurs whenever an elderly person is referred to one of these services. An individual may be referred on many separate occasions to a service - related to "random" deterioration (in health status usually - less often because of a change in social/support mechanisms).

## **INITIAL OBSERVATIONS**

A single transaction record is collected from each occasion a patient is assessed by a participating service. Each record contains 72 fields, including demographic data (date of birth, sex, marital status, living arrangements, domicile, COB, language preferred and so on), use of community services at referral and process data (source and reason for referral, crude and limited clinical data, diagnoses, recommendations). In addition, records of the movement of the patient through the various components of the service (for example from acute inpatient bed to day hospital to domiciliary care) are held in a separate database.

A trial of follow up data collection is underway at the Westmead Geriatric Medicine Unit - data is being collected on the cohort consisting of all referrals received in March 1991 and will be completed at the end of June 1991. The feasibility of collecting routine follow up data from all patients in NSW is extremely low as there are many technical and administrative problems.

This is a large data collection - currently about 25000 transactions/year, expected to reach about 28000/year when the entire state is covered. NSW has about (1991 estimate) 460000 persons aged 70 and over (70+) and annual referral rates of about 60 per 1000 70+ population are expected.

## **AVAILABILITY OF DATA**

Data stripped of identifying details may be obtained from the author. Explanatory material including coding rubrics and other details are also available. It is hoped that this large data collection will be of interest to other researchers including statisticians interested in developing techniques for modelling longitudinal observations.

## REFERENCES

1. McDonald CJ and Hui SL, "The Analysis of Humongous Databases: Problems and Promises", *Statistics in Medicine*, 10:511-518 (1991)
2. Beck RJ, "The Markov Process in Medical Prognosis", *Med Decis Making*, 3(4):419-458, 1983.
3. Aalen O, Borgan O, Keiding N and Thormann J, "Interaction between Life History Events. Nonparametric analysis for Prospective and Retrospective Data in the Presence of Censoring", *Scand. J. Statist.*,7:161-171,1980.
4. Kay R, "Multistate Survival Analysis: An Application in Breast Cancer", *Meth. Inform. Med.*, 23(3):157-162,1984.
5. Alison P, "Event History Analysis: Regression for Longitudinal Data", Sage University Papers - Quantitative Applications in the Social Sciences, Sage Publications, California, ISBN 0-8039-2055-5, 1984.
6. Price M, Lazarus R and Benson C, "New South Wales Geriatric Assessment Program Evaluation Unit Report for July to December 1990", NSW Evaluation Unit, C4c Westmead Hospital, 1991 (Available from the Authors).

**Ross Lazarus,**

**Lecturer in Geriatrics, Department of Community Medicine, Sydney University.**

Address Written Correspondence to :-

Dr Ross Lazarus,

GMU, C4c

Westmead Hospital

Westmead NSW 2145

Email : rml@extro.ucc.su.OZ.AU

Tel : (+602) 633 7946