



**Maximising Return from Cohort Studies:
Prevention of attrition and efficient analysis**

Report of a workshop held on Thursday 29 June 2006

at

The Royal Statistical Society, 12 Errol Street,
London EC1Y 8LX

funded by

MRC Population Health Sciences Research Network

Introduction

The MRC, along with other research bodies, invests heavily in large cohort studies as an efficient means of studying disease aetiology. The ageing of the cohorts combined with declining participation rates and population mobility inevitably lead to attrition, with subsequent loss of power and potential for bias. There have been recent statistical and computational advances in the methodology for analysing surveys subject to attrition (e.g. see the ESRC's recent Research Methods meeting on statistical methods for attrition and non-response in social surveys, and the ULSC conference on Methodology of Longitudinal Surveys). However, the structure and volume of data typically collected in cohort studies means that there is still no simple solution regarding best practice.

Against this background the PHSRN funded a workshop "Maximising Return from Cohort Studies: Prevention of attrition and efficient analysis". The aims of the workshop were:

- to bring together researchers working on cohort studies;
- to identify current practices of cohort maintenance; and
- to identify current practices of analysis of cohort data suffering from attrition.

The list of invited participants at the workshop, representing many of the major cohort studies in the UK, is given at the end of this report. The program of the day is also provided.

This report, together with slides for the sessions, is available from http://www.sphsu.mrc.ac.uk/sitepage.php?page=cohort_studies.

Cohort maintenance

There were four presentations in the morning on issues related to prevention of attrition in cohort studies. Three of these presentations covered the current practices used to prevent attrition and lessons learnt in the different studies: The NHLBI Growth & Health Study (NGHS), a cohort of African American adolescents; The ONS Longitudinal Study, a cohort based on record linkage; and the MRC National Survey of Health & Development 1946 birth cohort. The other presentation dealt with the changes in the future management of cohort studies in England and Wales which will be based on the use of the Personnel Demographic Service and the Secondary Uses Service.

Professor Pat Crawford, from the University of California, in her keynote presentation emphasised the need for both traditional and enhanced/focused practices, including the development of tracking system and refusal conversion systems. These systems were seen as the keys to success in achieving the remarkably high retention rates in the NGHS over 10 years of follow-up.

Two key themes emerged from this session on cohort maintenance: issues relating to the methods used, including cost implications; and issues relating to the reporting of attrition. The following summarises the discussion within these two themes.

1. Achieving high response rates

- There was concern about cohorts becoming increasingly unrepresentative from loss to follow-up (particularly if refusers in earlier waves were not followed up at subsequent waves) and from the loss of 'hard to reach' groups. It was felt, however, that the approach to addressing attrition, including the allocation of resources and use of focused strategies, was probably compromised by budgetary constraints.
- Issues regarding design and planning were raised. It was felt that it was important that researchers should consider the implications of attrition early on in the design of the study, including the designing of sub-studies.
- Various strategies for increasing participation were discussed. There was some concern about the ethical implications of the process of consent, including refusal conversion processes. Some felt that too much interaction with subjects may compromise the quality of data collected, potentially reducing objectiveness. It was noted there was a need to recognise that cultural differences in the perception of 'contracts' may require different processes.

- It was unclear whether data relating to the cost of reducing attrition in cohorts was collected but there was general consensus that the cost effectiveness of retention methods should be assessed.
- A lack of papers on issues related to attrition in studies was noted, although the IJ series on cohort profiles were seen as helpful.
- The group agreed that it would be very useful to have guidelines in 'best practice' preventing attrition.

2. Reporting of attrition

- A range of practices were used to calculate attrition rates across studies. Some key parameters were agreed to standardise reporting. The method of derivation should enable comparison across studies, the target population should be defined and assumptions about eligibility in numerators and denominators clearly documented. In long-term follow-up, response rates should allow for respondents moving in and out of sample (e.g. present response rate at each sweep), and for the problems associated with cluster samples.
- The characteristics of those lost to follow-up and the implications for bias should be reported in papers.
- The need for transparency in reporting attrition was recognised but many expressed anxiety about publication bias, educating editors being seen as a necessary corollary.
- The group agreed that it would be useful to have guidelines for reporting attrition.

Methodology

The afternoon session comprised four presentations devoted to methodology for analysing missing data and dropout in cohort studies. The cohort studies discussed included the Adult Day Health Care study, the Complementary Comfort Care study, the Avon Longitudinal Study of Parents and Children (ALSPAC), the Antiretroviral Therapy in Lower Income Countries Collaboration, the National Child Development Study (NCDS), and the MRC Cognitive Function and Aging Study. The MRC has substantial input into the last four of these studies.

In her keynote presentation, Professor Paula Diehr from the University of Washington focussed on the treatment of deaths in cohort studies, a particular problem among aging cohorts where death may account for a high proportion of missing values. Various statistical approaches to dealing with missing data in cohort studies were discussed by the four speakers including individual-specific regression based imputation, multiple imputation (based on multivariate normal models or chained equations), inverse probability weighting, and Heckman selection models. All presenters emphasised the importance of sensitivity analysis, at least through the presentation of the complete case analysis alongside the missing data analysis. Moreover, there was a general consensus that minimising missing data at the collection stage was better than subsequent adjustment.

The discussion at the end of the session covered three themes: a discussion of the relative advantages and disadvantages of some of the common methods of handling missing data; recommendations and advice regarding the reporting of results following the analysis of missing data; and a "wish list" covering the most pressing needs of researchers at the workshop. This discussion is summarised below.

1. Methods for the analysis of missing data

- Inverse probability weighting: the key question was whether the probability of response could be predicted from the available data. If so then inverse probability weighting provides a simple solution. However, this method is suitable when data are missing for the outcomes only i.e. when data are complete for other covariates of interest. Moreover, the weighting scheme should be adjusted to allow for the fact that the weights used are estimated (the weights themselves are not observed).
- The Heckman-type models jointly model the outcome of interest and the probability of inclusion in the sample or the probability of missing data. The disadvantage of such models is the requirement for independent instruments which are related to the probability of inclusion in the sample but which are not related to the outcome.

- Multiple imputation methods appear promising but still have limitations. The need to build imputation models adds to model complexity. A chained equation approach may be dependent on the ordering of variables whilst an approach based on the multivariate normal distribution has difficulties coping with categorical variables.

2. Recommendations and advice

- The magnitude of bias introduced by missing data is unknown and will vary from one study to another (and indeed may vary from one analysis to another).
- It was agreed that it was important to undertake sensitivity analysis. At the most basic level this could be a comparison of complete case analysis with missing data analysis; more complex sensitivity analysis could include trying different methods for dealing with missing data or changing the underlying assumptions or imputation models.
- A need for guidelines on the reporting of missing data analyses was recognised. The simplest guidelines should cover the need to report complete case analysis alongside the missing data analysis. This is necessary even in cases where complete case analysis are clearly going to be inadequate e.g. for studies that include informative dropout. The presentation of complete case analysis alongside the missing data analysis should overcome the concerns of space-conscious journals.

3. Wish list

- There was some discussion surrounding the possibility of creating complete multiple imputed datasets that could be stored and accessed for subsequent analyses, with the idea that the imputation need only be done once. However, given the need for an imputation model to take into account factors and interactions of importance in the analysis it seems unlikely that all potential analyses could be considered in advance. Moreover, more complex imputation models tend to cause technical problems with the imputation algorithms.
- There is a recognised need for more widespread training in methods for the analysis of missing data for epidemiologists and other researchers involved in cohort studies.

Programme

Alastair H Leyland

(MRC Social and Public Health
Sciences Unit)

Introduction

Cohort maintenance

Pat Crawford

(UC Berkeley School of Public
Health)

Prevention of attrition: keys to
successful cohort maintenance

Gita Mishra

(MRC National Survey of Health and
Development)

Loss and representativeness in a 53
year follow up of a national birth
cohort

Nirupa Dattani

(Office for National Statistics)

Cohort management in Secondary
User Services (SUS)

Louisa Blackwell

(Office for National Statistics)

LS attrition: best practice and lessons
learned

Cohort maintenance summary

Methodology

Paula Diehr

(School of Public Health and
Community Medicine, University of
Washington)

Death and missing data in longitudinal
studies: quality of life at the end of
life

Jonathan Sterne

(MRC Health Services Research
Collaboration)

Can methods that deal with missing
data reduce bias or improve precision
in analysis of longitudinal data?

Ian Plewis

(Bedford Group for Lifecourse and
Statistical Studies, Institute of
Education)

Incorporating information about non-
response into analyses of NCDS data

Fiona Matthews

(MRC Biostatistics Unit)

Attrition and its effects – examples
from analysis of the MRC cognitive
function and aging study

Methodology summary

Keynote speakers

Pat Crawford is on the faculty of both the Department of Nutritional Sciences and Toxicology and the School of Public Health at the University of California, Berkeley. She is also a Cooperative Extension Nutrition Specialist and was instrumental in founding the Center for Weight and Health of which she is now co-director. Nationally recognized as an authority on childhood obesity, Dr. Crawford has served as principal investigator on numerous large-scale projects including the longitudinal NHLBI Growth and Health Study, a 10-year study of the development of obesity and cardiovascular risk factors in African-American and White girls ages 9-19 years, and the 5-state FitWIC child obesity intervention initiative, the first effort to prevent obesity in our nation's largest supplemental food program for mothers and children. A widely published author and consultant, Dr. Crawford authored the American Dietetic Association's position paper on childhood obesity and the American Public Health Association's resolution on the prevention of pediatric overweight. She co-authored the recently published reference book *Obesity: Dietary and Developmental Influences*. Dr. Crawford is currently conducting epidemiologic studies on the prevention of overweight as well as intervention and evaluation studies in the field of pediatric obesity, with particular emphasis on the areas of school and community intervention.

Paula Diehr is professor of Biostatistics and Health Services. She has been involved in mental health services, different insurance and provider plans, health status, diagnostic rules for headache, cough and ankle trauma, health promotion evaluation, health services for older adults, people without health insurance, utilization and cost of health services, survey methods, community-based analyses, and years of healthy life or quality of life. She is generally interested in health outcomes. She has published recently in the area of obesity, survival and years of healthy life for older adults. She is fellow of the American Statistical Association, the Association for Health Services Research, and the American Academy for the Advancement of Science.

Participants

Unit/Study

SPHSU

UC Berkeley
Washington
1946 cohort
ALSPAC

ONS LS

NCDS cohort

MRC Biostats

PHSRN Network Manager
Wellcome
UCL

Imperial
NSHD
QUB

HNR

Relachs

Habits
LSHTM
SLS
ERC

British Heart Health

NESS

Epidemiology
CTSU
Midspan
CC75C

Contact

Alastair Leyland
Seeromanie Harding
Mel Whitrow
Pat Crawford
Paula Diehr
Gita Mishra
Jonathan Sterne
Kate Thomas
Lynn Molloy
Louisa Blackwell
Nirupa Dattani
Ian Plewis
Sosthenes Ketende
Fiona Matthews
Ian White
Angela Wood
Peter Craig
Jacob Sweiry
Chris Power
Leah Li
Claudia Thomas
Gopalakrishnan Netuveli
Diana Kuh
Dermot O'Reilly
Michael Rosato
Petra Lahmann
Adrian Mander
Stephen Stansfeld
Yasmin Khati
Susan Carnell
Kristina Stanfield
Lin Hattersley
Shirley Simmonds
Jason Poole
Richard Morris
Lucy Lennon
Helena Romaniuk
Guido Maggio
Nita Forouhi
Jane Armitage
Carole Hart
Blossom Stephan

E-mail

a.leyland@msoc.mrc.gla.ac.uk
seeromanie@msoc.mrc.gla.ac.uk
Melissa@msoc.mrc.gla.ac.uk
crawford@socrates.berkeley.edu
pdiehr@u.washington.edu
g.mishra@nshd.mrc.ac.uk
Jonathan.Sterne@bristol.ac.uk
cdxkt@bristol.ac.uk
Lynn.Molloy@bristol.ac.uk
Louisa.Blackwell@ons.gsi.gov.uk
Nirupa.Dattani@ons.gsi.gov.uk
i.plewis@ioe.ac.uk
s.ketende@ioe.ac.uk
fiona.matthews@mrc-bsu.cam.ac.uk
lan.White@mrc-bsu.cam.ac.uk
Angela.Wood@mrc-bsu.cam.ac.uk
peter.craig@scotland.gsi.gov.uk
j.sweiry@wellcome.ac.uk
c.power@ich.ucl.ac.uk
g.netuveli@imperial.ac.uk
d.kuh@nshd.mrc.ac.uk
d.oreilly@qub.ac.uk
m.rosato@qub.ac.uk
kate.homan@mrc-hnr.cam.ac.uk
s.a.stansfeld@qmul.ac.uk
s.carnell@ucl.ac.uk
Kristina.stanfield@lshtm.ac.uk
lin_hattersley@hotmail.com
sjs@mrc.soton.ac.uk
jp2@mrc.soton.ac.uk
r.morris@pcps.ucl.ac.uk
lucy.lennon@pcps.ucl.ac.uk
h.romaniuk@bbk.ac.uk
g.maggio@bbk.ac.uk
nf250@medschl.cam.ac.uk
jane.armitage@ctsu.ox.ac.uk
C.L.Hart@udcf.gla.ac.uk
bcms2@cam.ac.uk