Growing Up in Scotland: Data Workshop
December 2011

Working with Longitudinal Data
'An introduction'

Roxanne Connelly
University of Stirling

These slides were prepared by
Professor Vernon Gayle

www.longitudinal.stir.ac.uk

Structure of this Presentation

- Introduction
- Longitudinal Data
- Temporal Data Collection
- Data Collection Modes
- Longitudinal Data Structures
- Benefits of Longitudinal Data
- Longitudinal Models
- Conclusions
- Where Next?

Introduction

The focus of this talk is the analysis of longitudinal social survey data

- Cross-sectional data
  - Respondents surveyed at only one time point

- Longitudinal data
  - Repeated contacts
  - Respondents surveyed at multiple time points

Introduction

- For many social research projects cross-sectional data will be sufficient

- Most social research projects can be improved by the analysis of longitudinal data

- Some research questions require longitudinal data

Introduction

- Some research questions require longitudinal data

  – Flows into and out of childhood poverty

  – The effects of family migration on the woman's subsequent employment activities

  – Numerous policy intervention examples

  – Numerous examples relating to 'individual' development

Introduction

Longitudinal research is the lifeblood of the study of individual development. It has been pointed out many times that the most important questions concerning individual development can be answered only be applying a longitudinal design whereby the same individuals are followed through time.

(Bergman & Magnusson 1990)

Longitudinal Data

Much of the time we are only interested in how some social phenomena affects a later outcome

Primary school experiences → Standard Grade results (S4)

We are using longitudinal data but standard cross-sectional techniques are still suitable

Analytically this is fairly trivial

Longitudinal Data

Repeated outcome measures (one per contact)

ID	Year	Age	Employment
1	1991	16	Student
1	1992	17	Student
1	1993	18	Student
1	1994	19	Unemployed
1	1995	20	Employed (ft)
1	1996	21	Employed (ft)
1	1997	22	Employed (ft)
1	1998	23	Maternity Leave
1	1999	24	Family Care
1	2001	25	Employed (pt)

Analysis of repeated measures (i.e. multiple outcomes) is more complex

Specialist techniques are required!

Longitudinal Social Science Datasets

Two main forms of micro social longitudinal datasets

1 Panel Dataset

– Repeated contacts data collection

– Sociologist Paul Lazarsfeld opinion research in 1930s

– Common example is the Household Panel Study

Longitudinal Social Science Datasets

- Cohort Study
  - Repeated contacts data collection
  - Principally concerned with charting the development of a particular 'group' from a certain point in time
  - (simply a specific form of panel design in my view)

  - A birth cohort of babies born in a particular year
  - A youth cohort, a group of pupils who completed compulsory education in the same year
  - A group of newly qualified doctor

Temporal Data Collection

Prospective ↑

Retrospective ↓
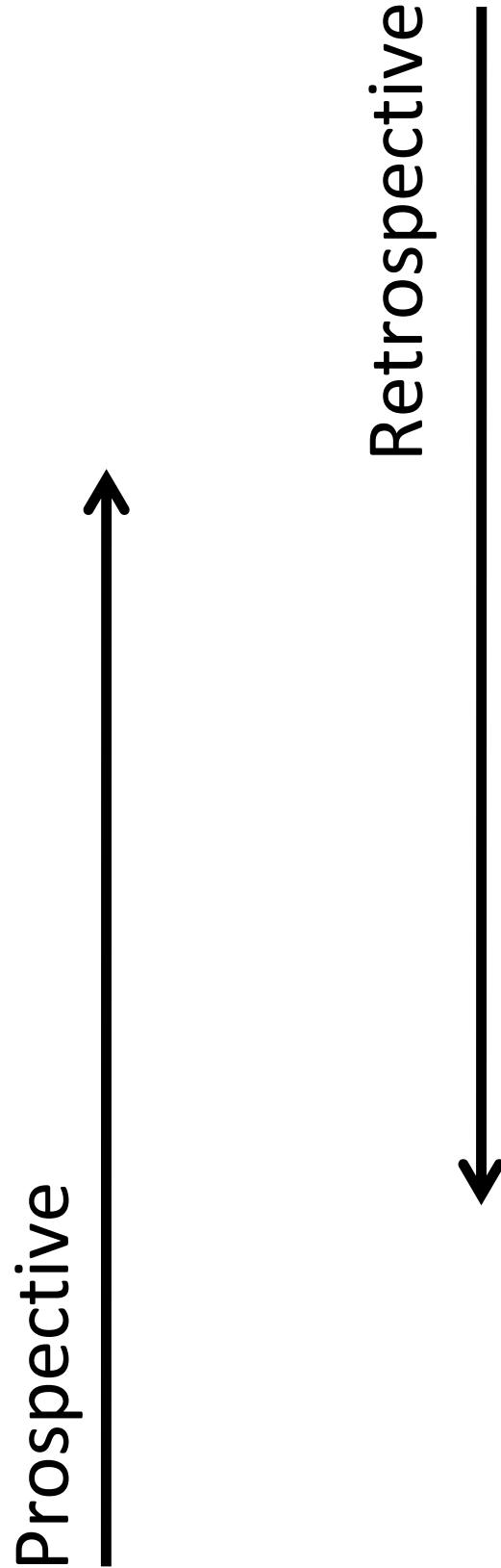
Most studies collect a mixture of both

Longitudinal Data Structures

A simple example of a panel (repeated contacts) dataset

ID	Year	Age	Gender	Employment	Marital Status
1	1991	16	Female	Student	Single
1	1992	17	Female	Student	Single
1	1993	18	Female	Student	Single
1	1994	19	Female	Unemployed	Single
1	1995	20	Female	Employed (ft)	Cohabiting
1	1996	21	Female	Employed (ft)	Cohabiting
1	1997	22	Female	Employed (ft)	Cohabiting
1	1998	23	Female	Maternity Leave	Married
1	1999	24	Female	Family Care	Married
1	2001	25	Female	Employed (pt)	Separated

Longitudinal Data Structures

"Long" format dataset

ID	Year	Age	Gender	Employment	Marital Status
1	1991	16			
1	1992	17			
1	1993	18			
1	1994	19			
1	1995	20			
1	1996	21			
1	1997	22			
1	1998	23			
1	1999	24			
1	2001	25			

Longitudinal Data Structures

Repeated outcome measures (one per contact)

ID	Year	Age	Gender	Employment	Marital Status
1	1991	16	Female	Student	Single
1	1992	17	Female	Student	Single
1	1993	18	Female	Student	Single
1	1994	19	Female	Unemployed	Single
1	1995	20	Female	Employed (ft)	Cohabiting
1	1996	21	Female	Employed (ft)	Cohabiting
1	1997	22	Female	Employed (ft)	Cohabiting
1	1998	23	Female	Maternity Leave	Married
1	1999	24	Female	Family Care	Married
1	2001	25	Female	Employed (pt)	Separated

Longitudinal Data Structures

Time constant explanatory variables

ID	Year	Age	Gender	Employment	Marital Status
1	1991	16	Female		Single
1	1992	17	Female		Single
1	1993	18	Female		Single
1	1994	19	Female		Single
1	1995	20	Female		Cohabiting
1	1996	21	Female		Cohabiting
1	1997	22	Female		Cohabiting
1	1998	23	Female		Married
1	1999	24	Female		Married
1	2001	25	Female		Separated

Longitudinal Data Structures

Time changing explanatory variables

ID	Year	Age	Gender	Employment	Marital Status
1	1991	16	Female	Student	Single
1	1992	17	Female	Student	Single
1	1993	18	Female	Student	Single
1	1994	19	Female	Unemployed	Single
1	1995	20	Female	Employed (ft)	Cohabiting
1	1996	21	Female	Employed (ft)	Cohabiting
1	1997	22	Female	Employed (ft)	Cohabiting
1	1998	23	Female	Maternity Leave	Married
1	1999	24	Female	Family Care	Married
1	2001	25	Female	Employed (pt)	Separated

Longitudinal Data Structures

Time to an event

Time to first childbirth

1991

1998

ID=1

X
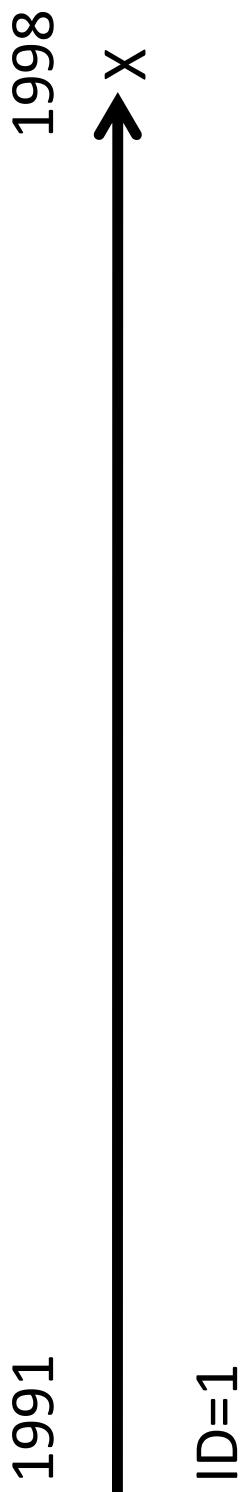
Benefits of Longitudinal Data

- Some research questions require longitudinal data

  – Micro-level change over time

    • Flows into and out of childhood poverty
    • The effects of family migration
    • Policy intervention examples
    • 'Individual' development

Benefits of Longitudinal Data

- Additional methodological benefits

  – Temporal ordering of events (direction of causality)

  – Improved control for omitted explanatory variables (residual heterogeneity)

  – Improved control for the effects of previous states (state dependence)

  – Exploring the effects of both ageing and cohort membership (age-period-cohort effects)

Longitudinal Models

- Two main modelling approaches in social science research

1. Event history analysis, time to an event

   – Also known as duration analysis; survival analysis; failure time analysis; duration economics; hazard modelling

   Generally time is continuous and we model the probability of an event occurring given that it has not already occurred (hazard)

Longitudinal Models

- Two main modelling approaches in social science research

2. Panel data analysis

  - Regression models suitable for repeated observations
  - Time generally conceptualised as being discrete
  - Extension of standard regression models (glm)
  - Closely related to multilevel modelling (glmm)
  - More advanced versions (e.g. dynamic models)

  - Alternative terminology
    - variance components models; hierarchical linear models; cross-sectional time series; random effects modelling

Conclusion

- For many social research projects cross-sectional data will be sufficient

- Most social research projects can be improved by the analysis of longitudinal data

- Some research questions require longitudinal data

Conclusion

- Longitudinal are not a panacea but data facilitate

  – The study of micro-level change social change over time (and also social stability)

  – A better understanding of the temporal ordering of events (direction of causality)

  – Improved control for omitted explanatory variables (residual heterogeneity)

  – Improved control for the effects of previous states (state dependence

  – Exploration of the effects of both ageing and cohort membership (age-period-cohort effects)

Where Next?

- More advanced skills are required
  - Extension from standard modelling techniques
  - Software (I recommend Stata)

    www.longitudinal.stir.ac.uk