Longitudinal studies in ageing research

Dr Faiza Tabassum

Department of Epidemiology & Public Health

University College London (UCL)

f.tabassum@ucl.ac.uk

Third FLARE Summer School, San Lorenzo Del Escorial, Madrid, Spain, 5-8 Sep 2011

Organisation of the talk

What are longitudinal studies?

Issues in analysing longitudinal data

Commonly used models to analyse longitudinal data

Missing values in longitudinal data

Why to have longitudinal studies on ageing

English longitudinal study of ageing (ELSA)

Conclusions

Bibliography

What are longitudinal studies?

Longitudinal data consist of outcome measurements repeatedly taken on each experimental unit over time.

Multiple or repeated measurements of the same variables are made for each individual in the study over a period of time

In longitudinal studies, we can study:

i. changes over time within individuals (ageing effects)
ii. differences among people in their baseline levels (cohort effects)

What are longitudinal studies?

- Repeated observations are likely to be correlated , so assumption of independence is violated
- Special statistical methods are needed to incorporate the covariance structure.
- The limitation of OLS in such situations is that it ignores the fact that the observations on the same subject are dependent or correlated.

Types of Longitudinal Research

There are three major types of longitudinal studies:

- **Panel Study:** Involves sampling a cross-section of individuals.
- **Retrospective Study:** Involves looking to the past by looking at historical information such as medical records.
- **Cohort Study:** Involves selecting a group based on a specific event such as birth, geographic location or historical experience.

  the results that are obtained from long-term cohort studies are of substantially superior quality to retrospective/cross-sectional studies, and cohort studies are considered the gold standard in observational epidemiology.

Issues in analysing longitudinal data

- Longitudinal data may either be continuous or categorical or a mixture of both
- Longitudinal data trajectories may be highly complicated, and there may be large variations between individuals
- There are often missing data or dropouts;
- Some variables may be measured with errors;
- In some studies the number of variables may be large while the sample sizes may be small.
- In longitudinal data analysis, new statistical methods are required to address one or more of the above problems since standard methods are not directly applicable.

Analysing longitudinal data

First step: exploratory analysis; graphs, descriptive tables, smooth curves

The observations within-subjects are correlated, so special statistical techniques are required for analyses

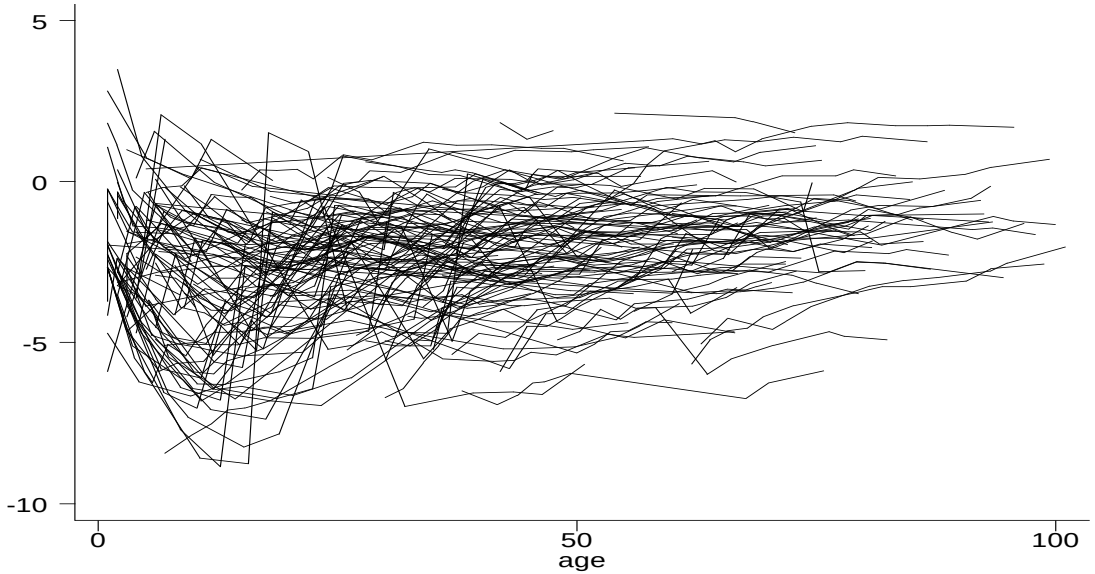
Moreover measures on the same subject close in time tend to be more highly correlated than measures far apart in time

Need to specify an appropriate correlation structure both within- and between- subjects. The choice depends on the design of the survey (equally spaced or not), drop-outs and how far apart the observations are collected

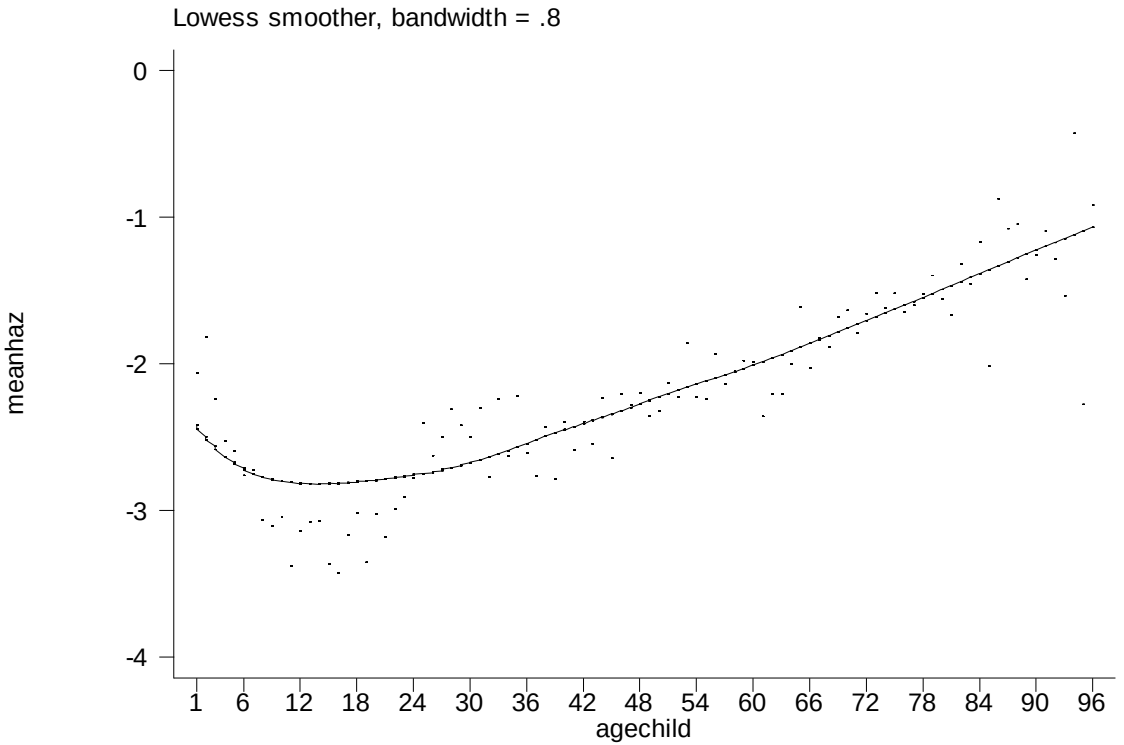
Methods such as ; growth curve models, multilevel models, GEE, path analysis can be used

Height-for-age against age (n=1500) over 12 waves

(Source: Pakistan Panel Data)



Lowess curve of height-for-age



Analysing longitudinal data: including time as a covariate

Time can be the only covariate in the model, eg. to model the growth rates of BMI trajectories across waves as time is considered as the fundamental predictor (Willett et al. 1998)

The effect of time describes the shape of the underlying developmental trajectory. It also gives insight on whether the growth is linear or not.

Including the effects of time in the model can lead to more accurate summaries of complex development and facilitate the testing of interesting hypothesis about the effects of substantive variables.

Commonly used models for longitudinal data

- **Mixed effects models:** in these models random effects are introduced to incorporate the between individual variation and the within-individual correlation in longitudinal data (AKA growth curve, multilevel models)

- **Multilevel models:** natural nesting present where measurement occasions are nested within individuals, so
- **Level 1:** measurements occasions (time)
- **Level 2**: subjects

- **Generalized estimating equations (GEE)** models: in these models the mean structure and the correlation structure are modelled separately without distributional assumptions for the data

- **Transitional models**, in these models the within-individual correlation is modelled via Markov structures

- **Nonparametric models** and semiparametric models: in these models the mean structures are modelled semiparametrically or nonparametrically or the distributional assumptions are assumed to be nonparametric, so these models are more flexible than parametric longitudinal models.

- **Bayesian models:** prior information or information from similar studies are incorporated for Bayesian

- **Survival models** for longitudinal data to model events occur in life at various time points

Plus and negatives of the above models

Each of these modelling approaches offers its own advantages and disadvantages.

Mixed effects models allow for individual-specific or subject-specific inference but require distributional assumptions

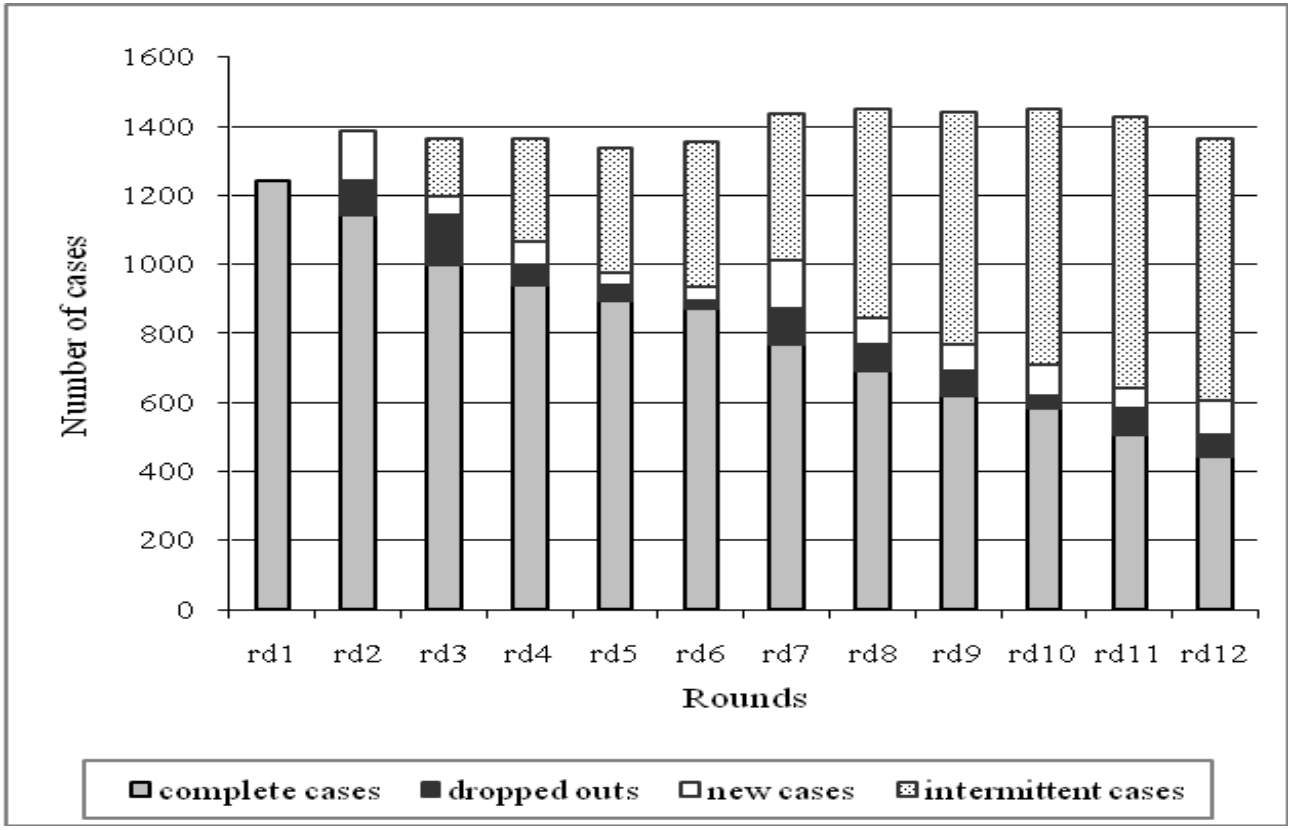
GEE models are robust to distributional assumptions but may be less efficient.

Transitional models may be particularly attractive for discrete data

Bayesian models borrow information from previous or similar studies

Nonparametric or semiparametric models are appealing for complex longitudinal data

Missing values in longitudinal data



Tabassum, F (2004) *Modelling Growth Trajectories of Children: A Longitudinal Analysis of Individual and Household Effects on Children's Nutritional Status in Rural Pakistan, Unpublished PhD thesis*

Missing values in longitudinal data

Missing values arise in the analysis of longitudinal data whenever one or more of the sequences of measurements from units within the study are incomplete

The units are incomplete in the sense that intended measurements are not taken, are lost or are otherwise unavailable

Four types of missing data (Siddiqui, et al. 1996):

- **Item-non response:** failure to answer a specific item by a participant;
- **Unit non-response:** some participants may miss entire survey but are present for later surveys
- **Dropouts:** participants may be impossible to follow-up
- Participants enrol in a study and then decide not to participate

Missing values; some terminology

- **Missing at random (MAR)** or **Missing Completely at Random (MCAR)** refers to the condition that a missing observation may depend on the observed components of data but not on the unobserved

- For example, people who are depressed might be less inclined to report their income, and thus reported income will be related to depression

- Depressed people might also have a lower income in general

- If we have a high rate of missing data on income among depressed individuals, the existing mean income might be lower than it would be without missing data.

- However, if, within depressed patients the probability of reported income was unrelated to income level, then the data would be considered MAR, though not MCAR

- Missing at random or missing completely at random condition is satisfied in these cases, and the mixed effects model will utilize the nonmissing measurement data but this is not the case if OLS regression methods are used

- **IMPUTATION:** is a technique which is used to fill up the missing values; most of the statistical software have incorporated commands to do the imputation, however, sometimes, the researchers have to use a 'model-based' technique for imputation.

Why to have longitudinal studies on ageing?

Why to have longitudinal studies on ageing?

Ageing populations are an increasing issue for the Western world as the proportion of people over sixty is growing.

For identifying the physiologic, psychological, socio-economic and other types of factors across the lifespan, affecting onset and progression of disease with advancing age

Scientists are learning what happens as people age and how to sort out changes due to ageing from those due to disease or other causes.

Longitudinal cohorts, longitudinal data sets, would be a valuable resource for facilitating future research on ageing changes across the lifespan

Particularly, to plan efficient health care system and pension policies for the ageing population

English longitudinal study of ageing (ELSA)

ELSA aims to study a sample of people over the age of 50 every two years in order to see how people's health, economic and social circumstances change over time

ELSA has longitudinal information across 4 waves on variables like: financial factors, household composition, social class, quality of life, lifestyle factors. Nurse visits are done on every alternative waves. This is the only longitudinal study on ageing at national level in England

The longitudinal information may help the government plan for an ageing population and longer periods of retirement, and ensure that the UK's healthcare and pension systems will be able to meet everyone's needs.

ELSA design is based on Health and retirement survey (HRS) of USA

Hypothetical examples how the longitudinal information can be used

Longitudinal information can be used in a number of ways, eg.

Modelling BMI across the critical periods; pre- and post-retirement;

Piecewise regression can be used to model changes during critical periods

The **piecewise model** allows separate slopes to be fitted to the observations representing the periods before and after a 'critical period' or 'event'

In this way, the growth trajectories of BMI, before and after the retirement can be calculated by using the mixed effects models

Likewise, looking at wealth/income trajectories both before and after the on-set of disease

Other ways can also be used, eg:

Summing BMI (obese or no) trajectories across the waves then to look at the risk factors associated of BMI with cardiovascular health disease (CHD)

Summing disability or no across waves or to examine the transition in the status of disability across waves and how it is associated with various financial variables or household structure

Movements in one status to another across waves, eg, social class mobility across the lifecourse and its associations on various health outcomes

PATH ANALYSIS APPROACH

Sometimes, we are interested in modelling both the direct and indirect effects simultaneously,

EG. associations of obesity with CHD is a direct effect but

there could be other pathways and indirect effects such as:

Childhood factors (social class, psychological factors)

Adult life factors (own SC, life style factors, psychological factors, etc....)

Recent events (SC, life style, psychological, etc...)

Which of these has stronger effects in explaining the associations b/w obesity and CHD?

Conclusion

Longitudinal research provides valuable, unique and critical insight into problems and issues of interest to social scientists

Longitudinal studies particularly on ageing helps researchers to investigate both current issues and changes over time

These studies also allow policy makers to examine the effects of unexpected health events such as cancer, heart attacks on loss of earning after the onset of disease

Worldwide trend in ageing by comparing various ageing studies around the world can be calculated

Longitudinal analysis is distinctive from cross-sectional analysis as it addresses dependency among measurements taken on the same experimental unit

- Such data are collected to address research questions that are concerned with changes in the mean response or potentially varying mean differences over time, in contrast to cross-sectional data that are concerned with the mean response and mean differences at a single time point.

- In particular, missing data, dropouts, and measurement errors are very common in longitudinal studies, and many of these issues need to be addressed simultaneously in order to draw reliable conclusions from the data.

- Moreover, longitudinal trajectories of observed data are often very complex. Therefore, statistical analyses of complex longitudinal data can be very challenging, and much research remains to be done.

Resources on longitudinal data analysis on web

- http://www.ats.ucla.edu/stat/examples/alda.htm
- http://www.ats.ucla.edu/stat/seminars/
- http://www.bristol.ac.uk/cmm/links/
- http://www.restore.ac.uk/resources/longitudinal.html
- http://tigger.uic.edu/~hedeker/long.html

Short courses on longitudinal data

- http://www.s3ri.soton.ac.uk/cass/programme.php
- http://www.essex.ac.uk/summerschool/
- Upcoming Online Workshop: Analyzing Repeated Measures Data: The GLM and Mixed Model Approaches http://theanalysisinstitute.com/workshops/Repeated-Measures
- MIXED MODELS ANALYSIS OF MEDICAL DATA USING SAS (Edinburgh, 12-14 October 2011) http://www.lifelong.ed.ac.uk/mixedmodels/
- STATA net-courses http://www.stata.com/netcourse/
- http://www.restore.ac.uk/Longitudinal/workshop_materials.html

Bibliography

- Albert, P. S. (1999) Longitudinal data analysis (repeated measures) in clinical trials. *Statistics in Medicine*, 18: 1707-1732.

•Brown, H. and Prescott, R. (1999) *Applied Mixed Models in Medicine*. John Wiley & Sons Ltd, England.

•Cnaan, A., Laird, N. M. and Slasor, P. (1997) Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in Medicine*, 16: 2349-2380.

•Diggle, P. J., Liang, K-Y. and Zeger, S. L. (1994) *Analysis of Longitudinal Data*. Oxford: Clarendon Press.

Fitzmaurice, G., M. Davidian, G. Molenberghs, and G. Verbeke, Longitudinal Data Analysis, Boc Raton, Florida: Chapman & Hall/CRC, 2008.

•Goldstein, H. and Brown, W. (2002) Multilevel modelling of medical data. *Statistics in Medicine*, 21(2): 3291-3315.

•Laird, N. M., and Ware, J. H. (1982) Random-effects models for longitudinal data. *Biometrics*, 38: 963-974.

•Little, R. J. A. and Rubin, D. B. (1987) *Statistical analysis with missing data*. New York: Wiley.

•Newsom, J et al. 2011 (eds) Longitudinal Data Analysis: A Practical Guide for Researchers in Aging, Health, and Social Sciences. Routledge Academic

•Siddiqui, A., J. Nicolson et al. (1996) Factors affecting attrition in a longitudinal smoking prevention study. Preventative Medicine, 25: 554-60.

•Singer, J. D. and Willett, J. B. (2003) *Applied longitudinal data analysis*. Oxford University Press.

•Snijders, T. A. B. and Bosker, R. J. (1999) *Multilevel Analysis*. London: Sage Publications.

- Verbeke, G. and Molenberghs, G. (2000) *Linear mixed models for longitudinal data*. Springer-Verlag.
- Willett, J. B., Singer, J. D. and Martin, N. C. (1998). The design and analysis of longitudinal studies of development and psychopathology in context: Statistical models and methodological recommendations. *Development and Psychopathology*. 10: 395-426.
- **STUDIES ON COMPARING VARIOUS COUNTRIES**
- Tabassum,F., Verropoulou,G., Tsimbosa,C., Gjonca,E., Breeze,E. (2009). Socio-economic inequalities in physical functioning: a comparative study of English and Greek elderly men. ***Ageing and Society*** 29(07), 1123-1140.
- Banks J., M. Marmot, Z. Oldfield, and J.P. Smith. 2006. "Disease and Disadvantage in the United States and in England." *JAMA 295(17):2037-45.*
- European Union 2007. Decision 1350/2007/EC of the European Parliament and of the Council of 23 October 2007 Establishing a Second Programme of Community Action in the Field of Health (2008-13)**.** Official Journal of the European Union, 20 November, L 301/3