VOLUME 4 | ISSUE 2 | **MAY 2013** ISSN 1757-9597

Longitudinal and Life Course Studies: International Journal

Published by

SLLS Society for Longitudinal and Life Course Studies

A feast of varied content!

- *Editorial with good news about finances and access plus an introduction to the topic of social-biological links*
- *Research paper on unemployment sequences and well-being*
- *Methodological paper on the role of respondent characteristics in effective longitudinal tracking*
- *Study profile documenting a new longitudinal resource for studying childhood obesity*
- *Discussion paper on the interaction between social exposures and biological processes in socio-biological transitions as a basis for understanding health*
- *Tutorial devoted to two contrasting approaches to analysing repeated measures of binary outcomes and their implications for modelling panel data*

Longview Promoting Longitudinal and Life Course Studies

Co-sponsors:

VU university medical center

ISER INSTITUTE FOR SOCIAL & ECONOMIC RESEARCH

LLCS EDITORIAL BOARD

EDITORIAL COMMITTEE

Executive Editor

John Bynner, *Institute of Education, UK*

Health Sciences

***Section Editor* - Michael Wadsworth**, *MRC Unit for Lifelong Health and Ageing London, UK;* **Carol Dezateux,** *Institute of Child Health, University College London, UK (Child Health & Development)*

Associate Editors - **David Blane**, *Imperial College Medical School, UK*
- **Rachel Knowles**, *Institute of Child Health, University College London, UK*
- **Scott Montgomery**, *Karolinska University Hospital, Sweden*

Social and Economic Sciences

***Section Editor* - Robert Erikson**, *University of Stockholm, Sweden*

Associate Editors - **Paul Gregg**, *University of Bath, UK*
- **John Hobcraft**, *University of York, UK*
- **Jan Smit,** VU Medical Centre, Netherlands

Statistical Sciences and Methodology

***Section Editor* - Harvey Goldstein**, *University of Bristol, UK*

***Associate Editor* - Sheila Bird,** *MRC Biostatistics Unit, University of Cambridge*

Development and Behavioural Sciences

***Section Editor* - Barbara Maughan**, *Institute of Psychiatry, Kings College, UK*

***Associate Editors* - Lars Bergman**, *University of Stockholm, Sweden*
- **Amanda Sacker**, *University College London, UK*

BOARD MEMBERS

Mel Bartley, *University College London, UK;* **Hans-Peter Blossfeld**, *European University Institute*, **Paul Boyle**, *University of St Andrews*; **Nick Buck**, *University of Essex*, UK; **Richard Burkhauser**, *Cornell University, USA;* **Jane Costello**, *Duke University, USA;* **Tim Croudace**, *University of Cambridge, UK;* **George Davey-Smith**, *University of Bristol, UK*; **Lorraine Dearden**, *Institute for Fiscal Studies, UK*; **Ian Deary**, *University of Edinburgh, UK*; **Glen Elder**, *University of North Carolina, USA;* **Peter Elias**, *University of Warwick, UK*; **Leon Feinstein**, *Cabinet Office, UK*; **Andy Furlong**, *University of Glasgow, UK*; **Frank Furstenberg**, *University of Pennsylvania, USA*; **John Gray**, *University of Cambridge, UK*; **Rebecca Hardy**, *University College London, UK*; **Walter Heinz**, *University of Bremen, Germany;* **Felicia Huppert**, *Addenbrooke's Hospital, Cambridge*, **Marjo-Ritta Jarvelin**, *Imperial College London, UK*; **Heather Joshi,** *Institute of Education, UK*; **Kathleen Kiernan**, *University of York, UK;* **Harvey Krahn**, *University of Alberta*; **Di Kuh**, *University College London, UK*; **Dean Lillard**, *Cornell University, USA*; **Steve Machin**, *London School of Economics*, *UK*; **Robert Michael**, *University of Chicago, USA*; **Sir Michael Marmot**, University College, London; **Jeylan Mortimer**, *University of Minnesota, USA*; **Brian Nolan**, *University College Dublin, Ireland*; **Lindsay Paterson**, *University of Edinburgh, UK*; **Ian Plewis**, *University of Manchester, UK*; **Chris Power**, *Institute of Child Health, UK*; **David Raffe**, *University of Edinburgh, UK*; **Steve Reder**, *University of Portland, USA*; **Marcus Richards**, *Medical Research Council, UK*; **Ingrid Schoon**, *Institute of Education, UK*; **John Schulenberg**, *University of Michigan, USA*; **Jackie Scott**, *University of Cambridge, UK*; **Rainer Silbereisen**, *University of Jena, Germany;* **Heike Solga**, *Social Science Research Centre, Germany;* **Håkan Stattin,** *Orebro University, Sweden*; **Fiona Steele**, *University of Bristol, UK*; **Alice Sullivan**, *Institute of Education, UK*; **Kathy Sylva**, *University of Oxford, UK*; **Gert Wagner**, *German Institute for Economic Research, Germany*; **Chris Whelan**, *Economic & Social Research Institute, Ireland*; **Richard Wiggins**, *Institute of Education, UK*; **Dieter Wolke**, *University of Warwick, UK*

SUBSCRIPTIONS

For immediate access to current issue and succeeding 2013 Issues (Volume 4) and all three Issues published in 2012 (Volume 3):

***Individual rate:* £20 – discounted rate £55 for 3 years** (Euro and US Dollar equivalents)

***Library* rate: £200** (Euro and US Dollar equivalents) to register any number of library readers.

***Society for Longitudinal and Life Course Studies (SLLS) Membership:* £65, *Student* £26, *Corporate* £300** (Euro and Dollar equivalents).

SLLS members have automatic free access to the journal among many other benefits. For further information about SLLS and details of how to join including the application form go to http://www.slls.org.uk/pages/membership.shtml

All issues are accessible free of charge 12 months after publication.

Print on demand

An attractive A5 printed version for each Issue of the journal, including all back numbers, is available at a price of £10 (Euro and US dollar equivalents) plus packaging and postage. If you want to take advantage of this opportunity of having hard copies to order, go to http://www.slls.org.uk/pages/publications.shtml where you will find a link to our online bookshop. Depending on distance, you will receive the order within two weeks of submitting it.

Please note

The reselling of personal registration subscriptions and hard copies of the journal is prohibited. Personal subscription and orders must be purchased with a debit or credit card - proof of identity may be requested.

SLLS disclaimer

The journal's editorial committee makes every effort to ensure the accuracy of all information (journal content). However the Society makes no representations or warranties whatsoever as to the accuracy, completeness or suitability for any purpose of the content and disclaims all such representations and warranties, whether express or implied to the maximum extent permitted by law. SLLS grants authorisation for individuals to photocopy copyright material for private research use only.

CONTENTS

EDITORIAL

86 - 87 **Editorial**
John Bynner

PAPERS

88 – 104 **Labour force sequences, unemployment spells and their effect on subjective well-being set points**
Cara L Booker, Amanda Sacker

105 – 118 **The role of respondent characteristics in tracking on longitudinal surveys: evidence from the UK Millennium Cohort Study**
Lisa Calderwood

STUDY PROFILE

119 – 135 **Design and characteristics of a new birth cohort, to study the early origins and ethnic variation of childhood obesity: the BiB1000 study**
Maria Bryant, Gill Santorelli, Lesley Fairley, Jane West, Debbie A Lawlor, Raj Bhopal, Emily Petherick, Pinki Sahota, Andrew Hill, Noel Cameron, Neil Small, John Wright, the Born in Bradford Childhood Obesity Scientific Group

COMMENT AND DEBATE

136 – 146 **Social-biological transitions: how does the social become biological?**
David Blane, Michelle Kelly-Irving, Angelo d'Errico, Melanie Bartley, Scott Montgomery

TUTORIAL

147 - 165 **Subject specific and population average models for binary longitudinal data: a tutorial**
Camille Szmaragd, Paul Clarke, Fiona Steele

This journal can be accessed online at: www.llcsjournal.org

Published by
The Society for Longitudinal and Life Course Studies
Institute of Education, 20 Bedford Way, London, WC1H 0AL
info@slls.org.uk http://www.slls.org.uk

SLLS Society for Longitudinal and Life Course Studies

AUTHOR GUIDELINES SUMMARY

Submission of Papers

All papers, written in the English language, should be submitted via the *LLCS* website as a Microsoft Word file. If there is a good reason why this is not possible, authors are requested to contact the Journal Manager **before** submitting the paper. All subsequent processes involving the author are carried out electronically via the website.

Preparation of Texts

Length. The paper should normally be **approximately 5000 words**, with longer papers also accepted up to a maximum of 7000 words. This word count excludes tables, figures and bibliography.

Font and Line Spacing. Please use Calibri (or similar sans serif) font, 12pt, with 1.5 line spacing in all main text, single space in figure or table captions.

Page Layout. All text should be justified to the left hand margin (all margins of at least 2.5cm) with no indents at the start of paragraphs and a line space separating paragraphs. All headings and sub-headings used within the body of the paper should also be left justified. Please do **NOT** use automatic text formatting.

Weblinks. To help our readers to look up cited references or other information available on the web, authors should ensure that all such references are activated.

Ensuring an Anonymous (Blind) Review. Please submit papers with a full detailed title page. Once a paper has been submitted to *LLCS* via the website, it will be 'anonymised' by the removal of all author name(s) and institution names on the title page, and any identifying electronic document properties will also be removed. Authors do not need to remove their name(s) from the main text or references but any reference to their work or themselves should be expressed in the third person.

Abstract. The abstract (no subheads or paragraphs) should be no more than 250 words (not part of the main word count).

Keywords. These should be included just below the author list (minimum 3 maximum 10).

Abbreviations. Words to be abbreviated should be spelt out in full the first time they appear in the text with the abbreviations in brackets. Thereafter the abbreviation should be used.

References. Please use the Harvard Style and refer to examples in the full Guidelines.

Authors not complying with these reference guidelines will be asked to make all the necessary alterations themselves, if their paper is accepted for publication.

Notes. As a general rule, supplementary notes should be avoided, but if thought to be essential, they should *not* appear on the page as 'Footnotes', but instead, be included as Endnotes.

Supplementary material. Supplementary material may be uploaded with the submission, and if the paper is published, will be visible to the reader via a link on the RHS of the screen.

Submissions containing graphs, tables, illustrations or mathematics. All graphs, tables and illustrations should be embedded in the submitted text, and have clear, self-explanatory titles and captions. Mathematical expressions should be created in Word 2003 (or a compatible package) with equation editor 3.0, unless the author has good reason to use other software, in which case please contact the Journal Manager. All biological measures should be reported in SI units, as appropriate, followed, in the text, by traditional units in parentheses.

Author Citation*.* If the paper is accepted for publication, a version of the paper with all authors cited in full on the title page will be used. Only individuals who have contributed substantially to the production of the paper should be included.

Copy Editing. All accepted manuscripts are subject to copy editing, with reference back to author with suggested edits and queries for response. Particular care should be taken by the author to check references.

Proofs. The corresponding author will be asked to view a layout proof of the article on the website and respond with final amendments within 3 days.

(full Author Guidelines at: http://www.llcsjournal.org/index.php/llcs/about/submissions#authorGuidelines**)**

Open Journal System

The ***LLCS*** journal is produced using the Open Journal System, which is part of the Public Knowledge Project. OJS is open source software made freely available to journals worldwide, for the purpose of making open access publishing a viable option for more journals. Already, more than 8,000 journals around the world use this software.

Copyright Notice

Authors who publish with Longitudinal and Life Course Studies agree to the following terms:

- Authors retain copyright and grant the Journal right of first publication with the work, simultaneously licensed under a Creative Commons Attribution License that allows others to share the work with an acknowledgement of the work's authorship and initial publication in this journal.
- Following first publication in this Journal, Authors are able to enter into separate, additional contractual arrangements for the non-exclusive distribution of the journal's published version of the work (e.g., post it to an institutional repository or publish it in a book), with an acknowledgement of its initial publication in this journal, provided always that no charge is made for its use.
- Authors are permitted and encouraged to post their work online (e.g. in institutional repositories or on their own website) prior to and 'as approved for publication', as this can lead to productive exchanges and additional citations, as well as complying with the UK Research Councils' 'Green Model' for open access research publication.

Editorial

John Bynner
Executive Editor

Sustainability

Much good news to report! With the membership of the Society for Longitudinal and Life Course Studies (SLLS) continually expanding and a good response to our invitation to members' University libraries to subscribe to the Journal, the prospects for the Journal's finances are looking good. It is particularly pleasing that 10 major libraries in different countries and varying in size responded immediately to our invitation to subscribe to the Journal, ensuring that all their users (staff, students and visitors) gain immediate access to it. We expect to boost the number further with the follow-up letter that has just gone out to those still undecided, offering enhancements in the ways that subscriptions can be paid. The letter also draws attention to other facilities of much interest to libraries, such as the newly established SLLS virtual bookshop, through which all Journal issues can be printed on demand as backup to the online version.

Editorial matters

Other items discussed at the recent editorial committee meeting included expanding the number of editors engaged in organising reviews for the Journal. This will broaden the range of expertise available to handle the interdisciplinary papers which we hope will be an increasing feature of the Journal.

Open Access

The other major item of much concern to all engaged in learned societies online journal publishing has been addressed in past editorials - *Open Access.* The UK Research Councils (UK-RC) have recently published a policy statement setting out their requirement of open access publication for all papers reporting research funded by them. Their position is particularly important because it sounds as if research councils in other countries, if not already adopting the same policy, are likely to follow suit.

The UK-RC favoured initially the 'Gold Model' of open access, in which part of the budgets through which they fund research has to be made available, via the universities receiving the funding, to pay the Article Processing Charge (APC) to the publisher who produces any paper based on it. The alternative approach, now accepted at least *pro tem,* is the 'Green Model' in which papers 'accepted for publication' are made available in manuscript form via repositories, where any reader can access them. Providing such a paper is deposited within six months of publication for STEM disciplines, and up to twelve months, or even two years for Social Science and Humanities disciplines, then the open access requirement is met.

From early on in its history LLCS has deposited all papers, once published, in the British Library, where they can be accessed through the Reading Room and such deposits will be made available to all of the other six UK libraries that serve as 'Legal Deposit' repositories. Complementing this provision, all LLCS authors will be asked in future to make available their 'approved for publication' manuscripts, by depositing them in their own University repositories, thus extending access online through that route.

UK-RC acknowledge that it will take perhaps five years for this change in the requirements of online publishing to be fully in place. In our own case as supporters of open access in principle for research papers, while recognising the need for financial sustainability of the journals that publish them, we are pleased to be able to work within the new policy framework.

This Issue

This week's issue has a varied range of content reflecting most of the modes of publication that the Journal offers. We start with two papers, the first on the relationship between unemployment sequences and well-being. The next is methodological reporting on the optimum means of tracking individuals over time, in longitudinal surveys where intervals between follow-ups are extended, such as in the adult stages of birth cohort studies. The paper addresses the issue of office and field tracking and such factors as respondent characteristics in determining the most successful method for tracking them. The next paper is a Study Profile based on the *Born in Bradford* study, in which a special sub-study dataset was developed for investigating childhood obesity. This is followed by a discussion paper introduced below by members of the Population Health Sciences Section editorial team. The final item is a methodological tutorial, devoted to two different approaches to the analysis of longitudinal repeated measures of binary outcomes in panel studies.

Discussion Paper

In this issue, the paper on s*ocial-biological transitions* by David Blane and colleagues, strongly affirms the need for longitudinal and life course data for research into the mechanisms by which social exposures influence and interact with biological processes including genetic effects. Although ideas about such influences have been around for a long time, the opportunities to test some of them have only relatively recently begun to be possible, because of developments in biological measures and their integration with longitudinal data.

This new paper presents a review of the elements involved in social influences on biological processes, and puts forward guiding principles for thinking about them and the processes that are involved. The models proposed have relevance for studies of individuals over many years of life, and there are already some longitudinal studies in Britain and elsewhere which are undertaking this kind of work. Other studies are now preparing to collect additional biological data for these purposes. Such new research will, as this paper explains, give new insight into how social experience becomes the biological reality of physical and mental ill-health, and provides an essential key to understanding the persistent socio-economic differences in health. This approach could also provide a framework for public health measures to improve health. These proposals also have relevance for comparisons of societies as they change over time. There is already compelling evidence that societies undergoing social and economic transition experience change in the patterns of prevailing illness as the transition occurs. Schooling, Lau, Tin and Leung (2010) showed how in non-western societies, as living conditions of early life improve, and exposure to infection and poor diet is reduced, consequent improvements in growth are associated with changed vulnerability in the long-run to such disease as ischaemic heart disease, diabetes and some cancers.

Clearly it is important now for those who undertake longitudinal studies to consider the implications of these ideas both for individual studies, and for comparative research involving many studies. We would like to encourage discussion in *Longitudinal and Life Course Studies* on this topic, and also at the annual meeting of the Society for Longitudinal and Life Course Research in Amsterdam this September:

http://www.longstudies.longviewuk.com/pages/conference.shtml Please send your contributions to us at the journal or contact your local SLLS representative listed in the global representatives list on:
http://www.longstudies.longviewuk.com/pages/membership.shtmln

Schooling, C.M., Lau, E.W.L., Tin, K.Y.K., & Leung, G.M. (2010). Social disparities and cause-specific mortality during economic development. *Social Science and Medicine 70*, 1550-1557.

SLLS Society for Longitudinal and Life Course Studies

SLLS International Conference

Growing Up and Growing Old:
Health Transitions Throughout the Lifecourse

Amsterdam, The Netherlands
23-25 September 2013

http://www.longstudies.longviewuk.com/pages/conference.shtml

Labour force sequences, unemployment spells and their effect on subjective well-being set points

Cara L. Booker
Institute for Social & Economic Research, University of Essex
cbooker@essex.ac.uk
Amanda Sacker
Department of Epidemiology and Public Health, University College, London

(Received January 2013 Revised May 2013)

Abstract

Drawing upon recent psychological literature, we examine the effect of employment statuses pre- and post-unemployment on levels of subjective well-being (SWB), and the return to pre-unemployment levels, i.e. set points. Data came from the British Household Panel Survey. SWB was measured using the GHQ-12 and a question on life satisfaction; Employment status was self-reported. Multilevel, jointed, piecewise, growth curve regression models were used to explore associations by gender, specifically whether different labour force sequences produced different growth curves and rates of adaptation. Overall, there was a tendency for men and women to return to well-being set points for both outcomes. However, findings showed differences by labour force sequence and SWB measure. Women who experienced unemployment between spells of employment returned to their SWB set point at a faster rate of return for GHQ than for life satisfaction, while for men, the rates of return were similar to each other. Women who were employed prior to unemployment and then became economically inactive showed a return to their GHQ set point, but there was no return to their life satisfaction set point. Economically inactive participants pre-unemployment, who then gained employment, also showed a return to their well-being set point. After economic inactivity and then unemployment, only men experienced a significant increase in life satisfaction upon return to economic inactivity. The findings showed that following unemployment, return to subjective well-being set point was quicker for people who became employed than for people who became economically inactive. There were also differences in the return to SWB set point by type of economic inactivity upon exiting unemployment.

Keywords: labour force status; unemployment; subjective well-being; set points

Introduction

The subjective well-being (SWB) of the population is one of the many measures that are used to compare and rank countries. The Gallup World Poll has collected well-being measures as well as other socio-demographic and political characteristics from over 120 countries (Deaton, 2008; OECD, 2010). This information has been used in reports and news stories that declare the "happiest country in the world" and in research which examines why there are country differences and what factors are most associated with better SWB (Deaton, 2008; OECD, 2010, 2011).

Subjective well-being is composed of three distinct domains (1) Emotional responses including positive and negative affect, (2) Domain specific satisfaction and (3) Overall life satisfaction (Diener, 2000; Diener, Suh, Lucas, & Smith, 1999). These domains have been widely examined with respect to various risk factors and moderators. However, researchers tend to focus on only one domain at a time (i.e. either life satisfaction or negative affect). Psychometric analysis of measures of these domains have shown them to be distinct constructs that are fairly independent and in some cases inversely correlated (i.e. positive and negative affect) (Diener & Emmons, 1984; Diener, Suh, Lucas, & Smith, 1999).

The hedonic treadmill, a theory that all individuals have a neutral level of happiness and that that they return to this equilibrium level, or "set point" after different life events was first postulated by Brickman and Campbell (1971). More recently, empirical and theoretical developments have contributed to the debate of the true stability of SWB, e.g. whether major or minor life events can have permanent effects on SWB, and whether the SWB domains have differing set points (Diener, Lucas, & Scollon, 2006; Headey, 2006, 2008a, 2008b; Headey & Wearing, 1992). While early research utilised cross-sectional data, current research has focused on using longitudinal cohort and panel data to assess the stability of SWB and the permanence of SWB set points (Clark, Diener, Georgellis, & Lucas, 2008; Clark & Georgellis, 2010; Clark, Georgellis, & Sanfey, 2001; Frederick & Loewenstein, 1999; Lucas, 2005, 2007a, 2007b; Lucas, Clark, Georgellis, & Diener, 2003). A revision of the hedonic treadmill theory, by Diener, Lucas and Scollon (2006) incorporated the findings from empirical evidence leading to changes in five assumptions of the theory. The first revision is that set points are not neutral as stated by the hedonic treadmill: more often than not, people are happy and positive (Diener, Lucas, & Scollon, 2006). Secondly, there is variation in set points between individuals. Thus some people may have higher levels of well-being than others, rather than all individuals having the same, neutral, set point (Diener, Lucas, & Scollon, 2006). Thirdly they proposed that there are multiple set points corresponding to the different dimensions of SWB and the direction of long-term trends can differ dependent on the dimension (Diener, Lucas, & Scollon, 2006). Fourthly, they postulate that individuals' set points can be altered. This is a major change from the hedonic treadmill which says that individuals cannot modify their overall levels of happiness. However Diener, Lucas and Scollon (2006) provide both cross-sectional and longitudinal data that show this to be untrue. Finally, they state that adaptation to life events varies across individuals (Diener, Lucas, & Scollon, 2006). It is the last two points that are the main focus of this study.

If we take it as a given that the subjective well-being set points of individuals can be changed, do certain life events such as unemployment cause these changes or do people adapt to these events and return to their set point? There have been many studies that have examined set points and whether there are any changes to them with respect to different life events such as marriage (Diener, Lucas, & Scollon, 2006; Lucas, 2007a; Lucas, Clark, Georgellis, & Diener, 2003; Luhmann & Eid, 2009), divorce (Diener, Lucas, & Scollon, 2006; Lucas, 2005, 2007a; Lucas, Clark, Georgellis, & Diener, 2003; Luhmann & Eid, 2009), unemployment (Diener, Lucas, & Scollon, 2006; Lucas, Clarke, Georgellis, & Diener, 2004; Luhmann & Eid, 2009) and disability (Lucas, 2007b; Lucas, Clark, Georgellis, & Diener, 2003). With respect to unemployment, people do appear to reset their set point. That is, SWB does not return to pre-unemployment levels, i.e. measures of SWB post-unemployment are consistently lower than the measurements of SWB prior to unemployment.

However there are a few issues that previous studies have not addressed which may affect whether there is a return to a SWB set point. First, pre- and post-unemployment labour force status has not been examined. A recent study showed that there were differential effects on GHQ scores during unemployment by pre-unemployment status (Booker & Sacker, 2011). They found that participants who had multiple spells of unemployment following economic inactivity experienced increased distress with each spell. This was in opposition to participants who experienced unemployment following employment whose distress decreased with each unemployment spell (Booker & Sacker, 2011). These effects suggest that the labour force status both pre- and post-unemployment may have an effect on whether individuals will return to their SWB set point. Examining the effects of economic inactivity on SWB may also serve to explore the effects of hidden unemployment (Beatty, Fothergill, & Macmillan, 2000), i.e. those who report being out of the labour force rather than report being unemployed, or those who do not meet the International Labour Organization measures of unemployment.

Additionally, the type of economic inactivity that one is involved with may further impact the return to SWB. The rate of return for someone who is retired may be different than the rate of return for someone who is taking care of their family or long-term sick. Exploration of this topic has not previously been examined and may have implications for future policy.

The third topic is that studies to date have only looked at life satisfaction as their measure of SWB. As Diener and colleagues (2006) showed, there are different set points for different domains of SWB and individuals may adapt differently for each. We test this theory by comparing two measures of SWB, the 12-item General Health Questionnaire (GHQ-12) (D. Goldberg & Williams, 1991) and overall life satisfaction.

Finally, there is some ambiguity in the literature regarding time given to return to the SWB set point. Research conducted by Lucas et al. (2004) looked at short-term adaptation, i.e. a return to set point 1-2 years after a life event, and found no return to set point. Other studies have looked at longer periods of adaptation with mixed findings (Clark & Georgellis, 2010; Lucas, 2007a; Lucas, Clark, Georgellis, & Diener, 2003). The current study makes no assumptions about whether there are improvements to well-being following unemployment nor the length of time needed to return to SWB set point. We address this and other topics by asking:

1. Do individuals return to their well-being set point after exiting unemployment? If so, how long does it take to return to that set point?
2. Does the labour force sequence experienced prior to and post unemployment influence whether individuals return to their set point, and if so, does the time needed to return differ by sequence?
3. Does the type of economic inactivity experienced prior to or post unemployment influence the return to SWB set point for individuals, and if so, how do rates of return differ by economic inactivity state?

Methods

Participants

The data used for this study come from the British Household Panel Survey (BHPS), a nationally representative longitudinal population survey, which began in 1991 with over 10,000 individuals in about 5,500 households (Taylor, with Brice, Buck, & Prentice-Lane, 2009). A detailed description of the sampling procedure and survey methods is provided by Taylor et al. (2009). Eighteen years of data, 1991-2008, were included in this study. Only participants who had experienced at least one spell of unemployment were included in the analyses, reducing the sample to 1,491 persons with 19,505 person-year observations.

Measures

Labour force status

Self-reported labour force status was obtained from annual surveys. The data allowed for in-depth investigation of the different employment states. Participants who reported that they were employed or self-employed were categorised as employed; all others (i.e. retired, maternity leave, family care, long-term illness, full-time education and other) were categorized as economically inactive. The International Labour Organisation definition of unemployment was used to identify unemployed participants, these were people who were out of work but were actively looking for work within the past month.

Subjective well-being

The 12-item General Health Questionnaire (GHQ-12) was used as the measure of psychological well-being. The GHQ-12 has been validated to screen for minor psychiatric morbidity, specifically distress and anxiety (D. Goldberg & Williams, 1988; D. P. Goldberg et al., 1997). A continuous scoring methodology was used where the scores ranged from 0-36. The cut-off for determination of a case with this scoring method is 12 (D. P. Goldberg et al., 1997). For the analysis, scores were reversed so that higher scores indicate better well being, and a case is determined by a score of 24 or less.

In order to account for natural time trends in GHQ scores, annual individual scores were centred around the year's grand mean (i.e. everyone in the study regardless of whether they experienced unemployment or not). These scores were then standardized to the overall grand mean (i.e. the mean GHQ score for everyone over the 18 waves of data).

One question was used as the measure of overall life satisfaction. This question was scored on a 7-point scale ranging from a low of "Completely dissatisfied" to a high of "Completely satisfied."

Control variables

Pre-study unemployment, household income, limiting long-term illness, age and gender, were included as control variables. Pre-study unemployment was a dichotomous indicator of whether individuals had experienced any bouts of unemployment prior to beginning their participation in the BHPS. Annual household income was adjusted for inflation, equivalised for household composition using the OECD modified equivalence scale, log transformed and then standardised to the study mean.

Limiting long-term illness (LLTI) status was a dichotomous variable determined from two questions. The first question asked "Do you have any of the health problems or disabilities listed on this card?" If the participant answered yes to this question then they are asked "Does your health in any way limit your daily activities compared to most people of your age?" In two waves, these questions were not asked and alternatives were chosen. Those questions asked "During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health? Have you…been limited in the kind of work or other activities?" and "During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems. Have you…cut down on the amount of time you spend on work or other activities?" Participants who experienced any limiting impairment/disability were categorized as having a LLTI, while all others were categorised as not having a LLTI.

To control for possible cohort effects, age was calculated as the participant's age in 1991 and standardized to have a mean of 0 and a standard deviation of 1. A squared term of standardised age accounted for differential and non-linear relationships between subjective well-being and age. Men were designated as the reference category for gender.

Analytic Scheme

Three-piece jointed growth curve models were developed to analyse the effects of unemployment on subjective well-being. The three pieces correspond to the period before unemployment, Stage 1; the period of unemployment, Stage2; and the years following unemployment, Stage 3. The overall model did not distinguish between labour force statuses in Stages 1 or 3. Four separate versions of this model were then run which selected samples based on differing labour force status sequences prior to and after unemployment. The four labour force sequences are:

1) Employed – Unemployed – Employed (EUE),
2) Employed – Unemployed – Inactive (EUI),
3) Inactive – Unemployed – Employed (IUE), and
4) Inactive – Unemployed – Inactive (IUI).

These overall and sequence-specific models were analysed using multilevel, jointed, piecewise, growth curve regression models (Bollen & Curran, 2006; Chou, Yang, Pentz, & Hser, 2004), with annual measurements (Level 1) nested within individuals (Level 2), with SAS/STAT software Version 9.1. (SAS Institute Inc., 2003). Details of models can be found in Table S1 in the supporting on-line information. A benefit of using multilevel regression is the ability to model data that is unbalanced and has unequally spaced measurement occasions as is common in panel studies. Only the first unemployment spell was examined; if a participant experienced a second unemployment spell all data were censored beginning with the first year of the second unemployment spell. Only consecutive years of employment status are included in the analysis. For example, if a participant was economically inactive for the first three years of their involvement in the study, then became employed for two years and then became unemployed, the three years of economic inactivity are censored, only the two years of employment immediately before the years of unemployment are included in the analysis. Similarly, if a participant is employed, becomes unemployed, returns to employment and then becomes economically inactive three years following their re-employment all years of economic inactivity following re-employment are censored. Continuously employed or economically inactive participants were censored if a change in employment status or attrition occurred.

The two growth curve parameters specified for each of the three periods include a linear and quadratic effect of time. The model is shown below:

$$Y_{ij} = t_{1ij}b_{1j} + t_{1ij}^{2}b_{1j} + t_{2ij}b_{2j} + t_{2ij}^{2}b_{2j} + t_{3ij}b_{3j} + t_{3ij}^{2}b_{3j} + \Sigma b_{nij}X_{n} + \varepsilon_{ij} + \mu_{i}$$

where Y is the outcome of either GHQ-12 score or overall life satisfaction score, t represents the repeated time measures in each stage, t^2 is a quadratic term for the time variable t, ε_{ij} is the random effect for individual i at measurement occasion j, and μ_i is the random intercept for individual i. The parameters $\Sigma b_{nij}X_n$ represent the covariates included in the model: age, age squared, gender and pre-study unemployment. Variable t_1 is equal to 0 in the first wave of stage one and then increases by one for all waves of stage 1 and then is constant at the last value of stage 1 for the remaining 2 stages. Variable t_2 is equal to 0 during stage 1 and is equal to 1 at the first wave of stage 2 and increases by one for every wave in that stage. Similar to t_1, t_2 remains constant throughout stage 3. The coding for t_3 is similar to that for t_2. However t_3 continues to increase until the most recent wave of data collection or until the participant has dropped out, changed employment status or experienced a second unemployment spell, at which point they are censored from the analysis.

As a comparison to those who experienced unemployment, growth curve trends for participants who were continuously employed or continuously economically inactive were also modelled. There were 5,257 continuously employed participants with 42,655 person-years, and 4,260 continuously economically inactive participants with 27,248 person-year observations. In these models, a single stage was modelled together with the control variables previously described.

Where sample sizes allowed, growth curve regression models were estimated for specific economic inactivity states for the overall and EUI, IUE and IUI labour force sequences. Retirement, family care, full-time student and long-term sick states were modelled. In the case of IUI, the pre-unemployment economic inactivity was not broken down by economic inactivity status, only post-unemployment economic inactivity.

Results

Descriptive Statistics

Table 1 provides the baseline descriptive statistics for the men and women in this sample. The overall sample was equally divided by gender. Overall, 8% of men and 3% of women reported having been unemployed prior to enrolment in the study. The average age was 31.39 (SD = 14.01) for men and 31.16 (SD = 12.8) for women. Breakdown by labour force sequence showed that a greater proportion of men who experienced EUE and EUI reported pre-study unemployment than those who were economically inactive in Stage 1 (p-value = 0.05), no significant differences were observed among women. EUI men had a higher mean age than EUE and IUE men, while IUE participants were the youngest for both men and women (p-value <0.0001). There were no significant differences in GHQ-12 scores for men who experienced different labour force sequences and only marginal differences between IUI and IUE women (p-value = 0.08). There were no significant differences in life satisfaction or annual household income between participants with different labour force status sequences.

Overall Trajectories

The gender stratified parameter estimates are given in Table 2. A significant coefficient for the linear term indicates an increase or decrease in the SWB measure score, while a significant quadratic coefficient indicates that the rate of change was accelerating or decelerating. Overall, the acceleration/deceleration was small compared to the linear slopes.

The figures are based on a hypothetical exemplar respondent's employment history consisting of four years prior to unemployment (periods 1 to 4), two years of unemployment (periods 5 and 6), and the immediate four years after unemployment (periods 7 to 10). The SWB set point is defined as the mean GHQ-12 or life satisfaction score pre-unemployment, periods 1 to 4. In order to illustrate the impact of different labour force sequences on SWB set points, we estimate the time taken to return to this exemplar respondent's mean pre-unemployment SWB score after exiting unemployment.

Table 1. Baseline Descriptive Statistics, by Gender *

	Overall (n=1491)		Employed - Unemployed - Re-employed (EUE) (n=674)		Employed - Unemployed - Inactive (EUI) (n=246)		Inactive - Unemployed - Employed (IUE) (n=300)		Inactive - Unemployed - Inactive (IUI) (n=271)		*p-value* [+]
Men											
Pre-Study Unemployment, % Yes	8	--	9	--	12	--	5	--	3	--	0.05
Limiting Long-term Iliness, % Yes	7	--	4	--	3	--	7	--	22	--	<0.0001
Age, mean (95% CI)	31.39	(30.39, 32.40)	32.06	(30.83, 33.29)	40.61	(38.12, 43.09)	20.53	(18.51, 22.55)	36.27	(33.69, 38.85)	<0.0001
GHQ-12 Score, mean (95% CI)	25.84	(25.47, 26.20)	26.05	(25.56, 26.54)	25.41	(24.40, 26.43)	26.09	(25.29, 26.90)	24.85	(23.80, 25.90)	0.16
Life Satisfaction Score, mean (95% CI)	5.30	(5.12, 5.48)	5.16	(4.91, 5.41)	5.60	(4.98, 6.22)	5.30	(4.96, 5.64)	5.63	(5.14, 6.11)	0.29
Annual Household Income (95% CI)	0.04	(0.04, 0.05)	0.05	(0.04, 0.06)	0.04	(0.03, 0.06)	0.04	(0.02, 0.05)	0.03	(0.01, 0.05)	0.51
Women											
Pre-Study Unemployment, % Yes	3	--	3	--	2	--	1	--	1	--	0.87
Limiting Long-term Iliness, % Yes	13	--	8	--	11	--	10	--	23	--	<0.0001
Age, mean (95% CI)	31.16	(30.24, 32.08)	33.02	(31.58, 34.47)	34.69	(32.73, 36.64)	22.63	(20.69, 24.56)	32.60	(30.83, 34.37)	<0.0001
GHQ-12 Score, mean (95% CI)	24.82	(23.34, 24.26)	24.05	(23.29, 24.81)	24.03	(23.02, 25.05)	24.47	(23.46, 25.48)	22.65	(21.72, 23.59)	0.04
Life Satisfaction Score, mean (95% CI)	5.15	(4.88, 5.21)	5.06	(4.74, 5.38)	5.30	(4.90, 5.71)	4.97	(4.66, 5.29)	4.96	(4.63, 5.28)	0.56
Annual Household Income (95% CI)	0.05	(0.05, 0.06)	0.05	(0.04, 0.06)	0.06	(0.05, 0.08)	0.04	(0.02, 0.06)	0.06	(0.05, 0.08)	0.20

*95% CI = 95% Confidence Interval; Annual Household Income log transformed and grand centred

[+] p-value for Pre-Study Unemployment and Limiting Long-term Illness based on X^2; p-value for Age, GHQ-12, Life Satisfaction and Annual Household Income based on least squares mean comparisons

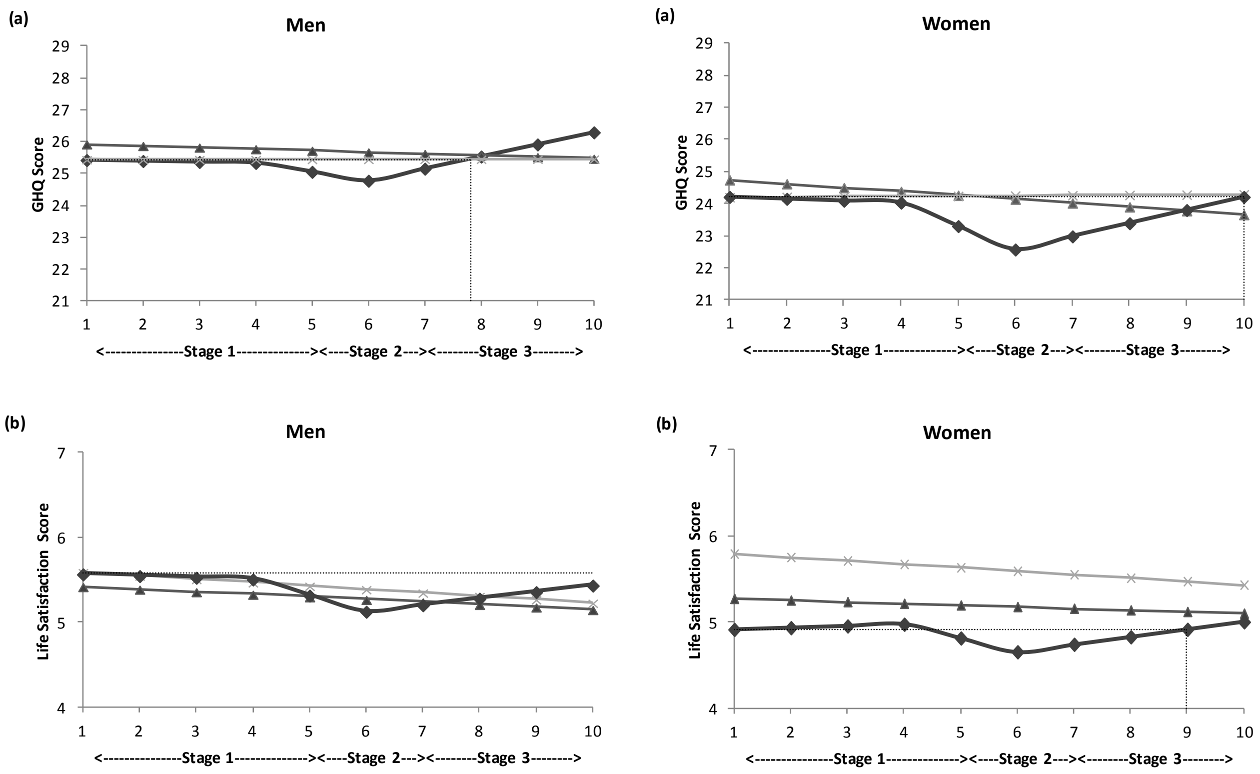
Table 2. Overall Growth Models, by subjective well-being outcome and gender

	GHQ				Life Satisfaction			
	Men		Women		Men		Women	
	Regression Coefficient	SE	Regression Coefficient	SE	Regression Coefficient	SE	Regression Coefficient	SE
Intercept	25.43****	0.27	24.21****	0.33	5.57****	0.20	4.91****	0.24
Pre-Unemployment								
Linear	-0.03	0.09	-0.06	0.10	-0.02	0.03	0.02	0.03
Quadratic	0.01	0.01	-0.00	0.01	0.00	0.00	-0.00	0.00
Unemployment Spell								
Linear	-0.28*	0.14	-0.99****	0.23	-0.19***	0.05	-0.21**	0.07
Quadratic	0.04	0.02	0.26**	0.08	0.02	0.01	0.05*	0.02
Post-Unemployment								
Linear	0.40****	0.07	0.43****	0.09	0.08****	0.02	0.09****	0.03
Quadratic	-0.02***	0.00	-0.02***	0.01	-0.00*	0.00	-0.00*	0.00
Age	-0.58****	0.14	-0.24	0.17	-0.02	0.04	-0.02	0.05
Age Squared	0.60****	0.11	0.29*	0.14	0.12**	0.04	0.09*	0.04
Pre-Study Unemployment	-0.13	0.42	-0.06	0.82	-0.22	0.15	-0.29	0.27
Limiting Long-term Illness	-2.18****	0.20	-3.14****	0.22	-0.48****	0.06	-0.41****	0.06
Annual Household Income	0.47	1.28	-2.97*	1.41	0.02	0.43	-1.49***	0.41

** <0.05; ** <0.01; *** <0.001; **** <0.0001*

Figure 1 shows the overall growth curves for GHQ (a) and life satisfaction scores (b) by gender, based on the parameter estimates in Table 2. There were significant decreases in levels of SWB during unemployment and significant increases in the post-employment period for both men and women. Women experienced decelerating declines in SWB while unemployed and decelerating improvements post-employment, while men showed deceleration during post-unemployment only. These changes in well-being indicate that during the years of unemployment there was a decrease in GHQ scores, which levelled off over time for women, and during the period following unemployment there was improvement, albeit with deceleration, back towards the pre-unemployment GHQ score. Calculation of the increase in GHQ score post-unemployment provided an estimate of 1.81 years for men and 4 years for women, for our exemplar respondents to return to their pre-unemployment GHQ score.

Figure 1. Overall growth curve models (-♦-) for GHQ (a) and life satisfaction (b) by gender. **Graph shows trend lines for continuously employed (-Δ-) and continuously economically inactive (-×-) participants. The horizontal (--) is the set point line and the vertical (--) is where the curve crosses the set point line. The numbers along the x-axis are years.**



The growth curve of life satisfaction is similar to that for GHQ (Figure 1b). Life satisfaction reduced throughout the unemployment spell, for women there was a decreasing rate of change, and then a gradual increase again after exiting unemployment. The return to pre-unemployment life satisfaction took slightly longer than for GHQ, 5.71 years for our exemplar male respondent, but was shorter for our female exemplar at 2.99 years.

Figure 1 also includes comparison trajectories for participants who were continuously employed or economically inactive. (See Table S1a in the supporting on-line Appendix A for parameter estimates). For men, the GHQ trajectories were slightly better than for those who experienced unemployment with a rate of decrease across the years for those continuously employed, but a marginal increase across the years for the continuously economically inactive. Continuously employed women also had better GHQ scores than continuously economically inactive or those who experienced unemployment. Continuously employed women also experienced declines in their SWB with their GHQ scores dipping below that of the continuously economically inactive after 6 years. Life satisfaction scores for men and women who were continuously economically inactive were slightly higher (better SWB) compared to the continuously employed. Life satisfaction decreased more quickly for those who were continuously economically inactive than for those who were continuously employed across the years. Levels of life satisfaction between the continuously employed and continuously economically inactive were more similar for men than for women.

Trajectories by Labour Force Sequence

While the overall curves show that there were significant non-linear trajectories for both SWB outcomes, we are interested in potential differential trajectories dependent on the labour force sequence experienced, specifically with respect to return to set point. Table 3 provides the gender specific parameter estimates for the different labour force sequences for each outcome.

Both men and women who experienced a bout of unemployment in-between employment spells, EUE, had decreases in their SWB levels during unemployment, which then increased again on re-employment. These decreases in SWB during unemployment and subsequent increase upon re-employment gradually slowed down. Our male exemplar returned to GHQ set point after 5.19 years, while the return took 5.02 years for our female exemplar (Figures 2a and 3a).

In contrast to EUE, participants who experienced EUI had improvements in their GHQ score during economic inactivity only. Our male exemplar did not have a return to GHQ set point, while the return to GHQ set point was much shorter for our exemplar woman at 2.87 years (Figures 2b and 3b).

Table 3. Growth models by labour force sequence, by subjective well-being outcome and gender⁺

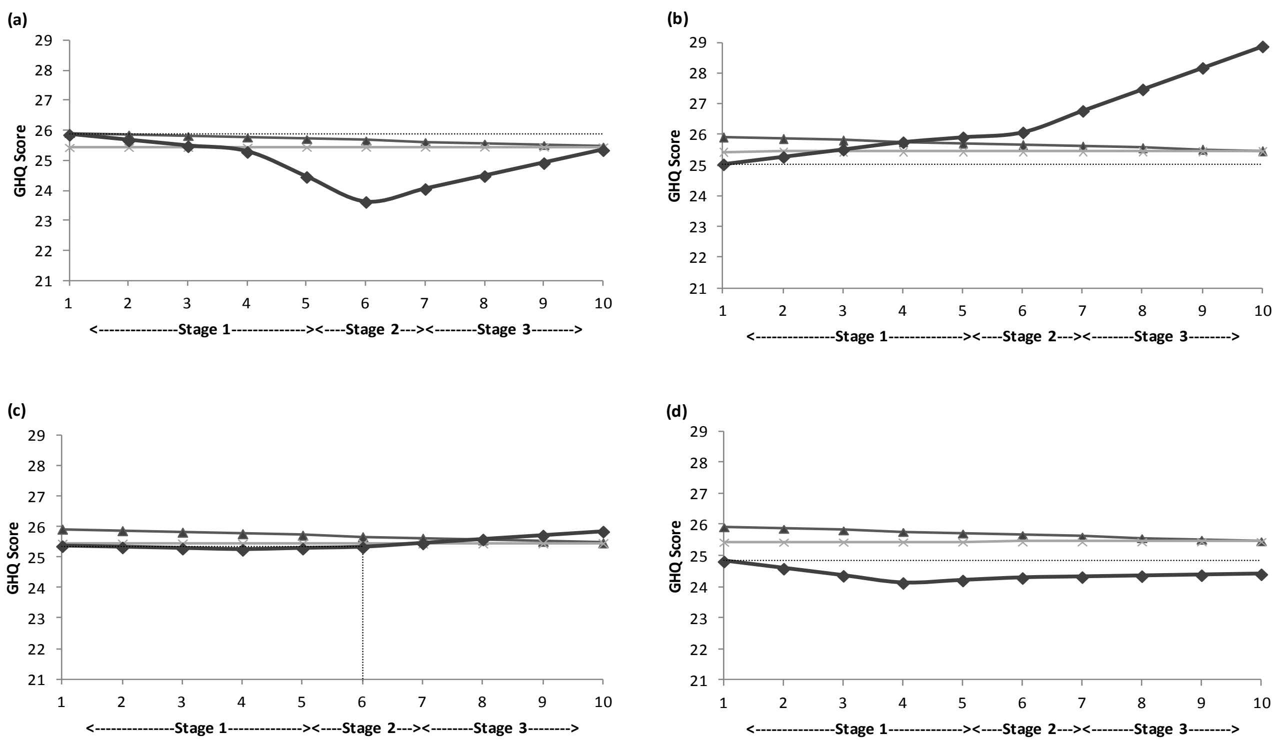
	GHQ				Life Satisfaction			
	Men		Women		Men		Women	
	Regression Coefficient	SE	Regression Coefficient	SE	Regression Coefficient	SE	Regression Coefficient	SE
EUE								
Intercept	25.86	0.34	24.16****	0.50	5.81****	0.24	4.51****	0.33
Employed (E)								
Linear	-0.19	0.11	-0.20	0.14	-0.03	0.03	-0.01	0.04
Quadratic	0.01	0.01	0.01	0.01	0.00	0.00	-0.00	0.00
Unemployed (U)								
Linear	-1.07****	0.22	-1.49****	0.36	-0.36****	0.08	-0.23*	0.10
Quadratic	0.24****	0.06	0.41***	0.11	0.06**	0.02	0.04	0.03
Re-Employed (E)								
Linear	0.46****	0.09	0.55***	0.14	0.14****	0.02	0.07	0.04
Quadratic	-0.03****	0.01	-0.03**	0.01	-0.01****	0.00	-0.00	0.00
EUI								
Intercept	25.03****	0.78	24.24****	0.69	5.61****	0.66	4.23****	0.55
Employed (E)								
Linear	0.24	0.25	-0.06	0.22	-0.06	0.08	-0.08	0.07
Quadratic	-0.00	0.02	-0.01	0.02	0.00	0.01	0.00	0.01
Unemployed (U)								
Linear	0.16	0.34	-0.57	0.48	-0.14	0.13	-0.22	0.17
Quadratic	-0.01	0.04	0.11	0.15	0.02	0.01	0.09	0.05
Inactive (I)								
Linear	0.70***	0.20	0.46*	0.19	0.03	0.06	0.02	0.06
Quadratic	-0.03	0.01	-0.01	0.01	0.00	0.00	0.00	0.00
IUE								
Intercept	25.36****	0.67	25.48****	0.84	5.14****	0.49	6.59****	0.52
Inactive (I)								
Linear	-0.04	0.33	0.12	0.23	-0.01	0.10	-0.07	0.07
Quadratic	-0.02	0.06	0.01	0.02	-0.01	0.02	0.01	0.01
(Table 3 cont'd)								
Unemployed (U)								
Linear	0.04	0.42	-0.75	0.59	-0.25	0.15	-0.09	0.17
Quadratic	0.04	0.11	0.13	0.23	0.07	0.04	0.01	0.06
Employed (E)								
Linear	0.13	0.17	0.61**	0.20	0.05	0.05	0.19***	0.05
Quadratic	-0.00	0.01	-0.05**	0.02	0.00	0.00	-0.01**	0.00

(Table 3 cont'd)

IUI								
Intercept	24.82****	1.00	22.83****	0.79	5.94****	0.77	4.85****	0.57
Inactive (I)								
Linear	-0.23	0.42	0.39	0.25	-0.15	0.14	0.09	0.07
Quadratic	0.02	0.05	-0.01	0.03	0.01	0.02	-0.01	0.01
Unemployed (U)								
Linear	0.08	0.38	-0.98	0.57	-0.23	0.14	-0.37*	0.19
Quadratic	-0.03	0.05	0.37	0.24	0.02	0.02	0.06	0.07
Return to Inactivity (I)								
Linear	0.03	0.23	0.15	0.19	-0.15*	0.07	0.05	0.05
Quadratic	-0.00	0.02	-0.01	0.02	0.01	0.00	-0.00	0.00

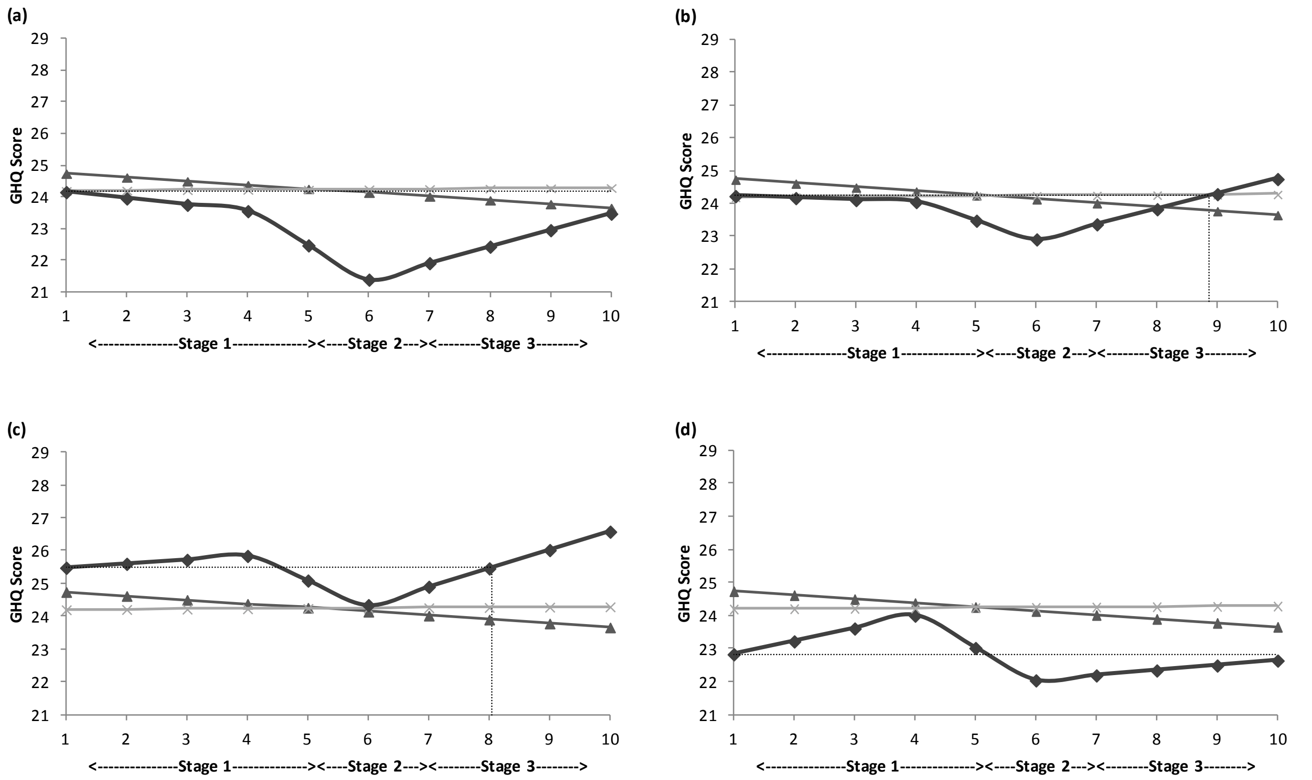
*[*] <0.05; ** <0.01; *** <0.001; **** <0.0001*
[+] Controlled for age, age squared, gender, limiting long-term illness, household income and pre-study unemployment; EUE = Employed - Unemployed - Employed; EUI = Employed - Unemployed - Inactive; IUE = Inactive - Unemployed - Employed; IUI = Inactive - Unemployed – Inactive

Figure 2. GHQ by labour force sequence growth curve models for men (-♦-): EUE (a), EUI (b), IUE (c) and IUI (d). Graph shows trend lines for continuously employed (-Δ-) and continuously economically inactive (-x-) participants. The horizontal (--) is the set point line and the vertical (--) is where the curve crosses the set point line, where applicable. The numbers along the x-axis are years.



The GHQ score of participants who experienced IUE non-significantly changed during economic inactivity and unemployment, and then significantly increased with some deceleration on employment for women only (Figure 3c). The return to the GHQ set point for our exemplar woman was between that for EUE and EUI participants at 2.04 years (Figure 3c). There were no significant changes in GHQ scores, at any stage, for participants who experienced IUI (Figures 2d and 3d).

Figure 3. GHQ by labour force sequence growth curve models for women (-♦-): EUE (a), EUI (b), IUE (c) and IUI (d). Graph shows trend lines for continuously employed (-Δ-) and continuously economically inactive (-×-) participants. The horizontal (--) is the set point line and the vertical (--) is where the curve crosses the set point line, where applicable. The numbers along the x-axis are years.



Three labour force sequences showed significant increases in life satisfaction post-unemployment with gender differences, Table 3. Men who experienced EUE had deterioration during unemployment and an increase in life satisfaction when re-employed, with slowing down in the rates of deterioration or increase during both periods. Women experienced a decrease in life satisfaction while unemployed only. The return to life satisfaction set point took 5.07 years for men and 7 years for women (Figures 4a and 5a).

There were no significant increases or decreases in life satisfaction at any stage for EUI participants.

Upon employment, the life satisfaction of women IUE participants increased steadily, there were no significant changes to life satisfaction for men. The return to life satisfaction set point was 2.17 years for our female IUE exemplar respondent (Figure 5b).

Female IUI participants showed significant decline in life satisfaction during unemployment, but no significant increase on return to inactivity was observed (Figure 5d). Male IUI respondents however, did show a significant decrease in life satisfaction upon return to inactivity (Figure 4d).

Figure 4. Life Satisfaction by labour force sequence growth curve models for men (-♦-): EUE (a), EUI (b), IUE (c) and IUI (d). Graph shows trend lines for continuously employed (-Δ-) and continuously economically inactive (-×-) participants. The horizontal (--) is the set point line and the vertical (--) is where the curve crosses the set point line, where applicable. The numbers along the x-axis are years.

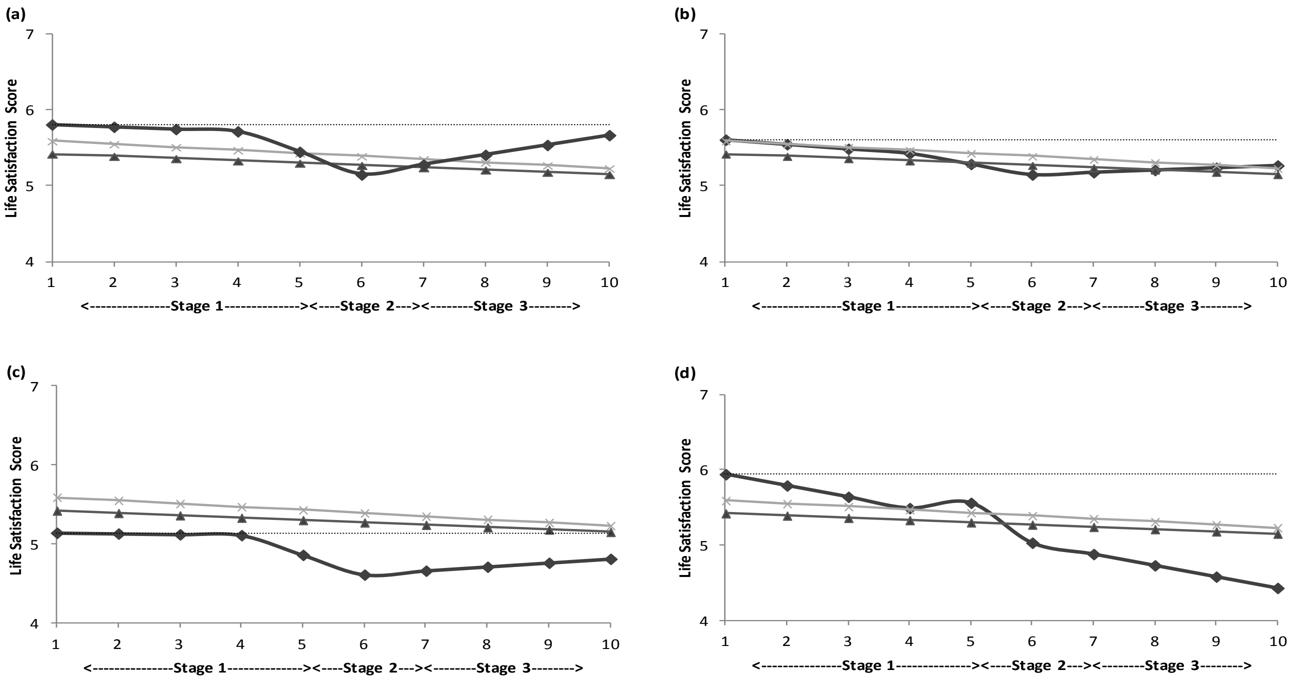
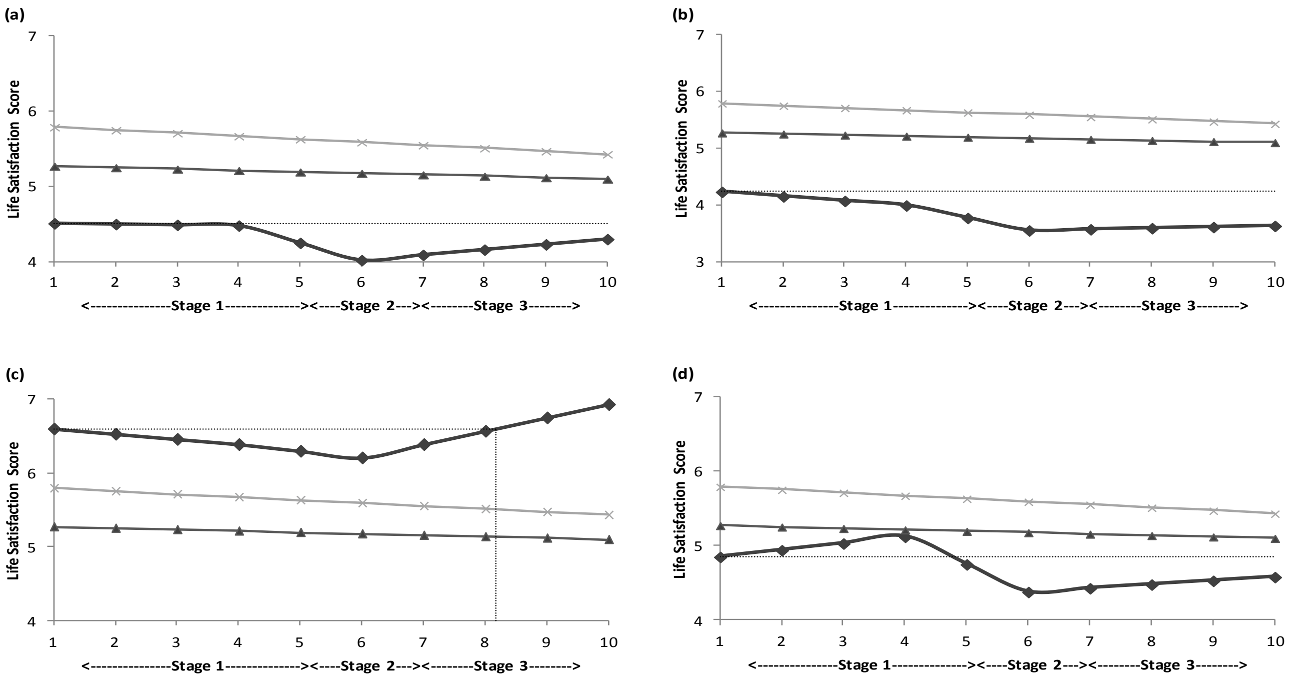


Figure 5. Life Satisfaction by labour force sequence growth curve models for women (-♦-): EUE (a), EUI (b), IUE (c) and IUI (d). Graph shows trend lines for continuously employed (-Δ-) and continuously economically inactive (-×-) participants. The horizontal (--) is the set point line and the vertical (--) is where the curve crosses the set point line, where applicable. The numbers along the x-axis are years.



In addition to SWB trajectory differences by labour force sequence, we were able to explore SWB trajectories for selected economically inactive states, specifically retirement, family care, full-time student and long-term sick. In the overall models, only long-term sick respondents showed a significant increase in GHQ score post-unemployment (See Table S2a in the supporting on-line Appendix B for parameter estimates). Participants who retired showed a significant decrease and those who became full-time students showed significant increases in life satisfaction upon exit from unemployment (See Table S2b in the supporting on-line Appendix B for parameter estimates). These findings did not change significantly by labour force sequence; EUI long-term sick were the only participants who experienced an increase in GHQ score upon their change from unemployment to long-term sick (See Table S3a in the supporting on-line Appendix C for parameter estimates). Both EUI and IUI participants experienced a decrease in life satisfaction upon retirement, i.e. post-unemployment, (See Tables S3a and S3b in the supporting on-line Appendix C for parameter estimates). Participants who were involved in family care showed a decrease in life satisfaction before becoming unemployed (IUE) (See Table S3c in the supporting on-line Appendix C for parameter estimates).

Discussion

Main findings

In the non-labour force sequence specific model, there was a return to GHQ and life satisfaction set points after exiting unemployment for both men and women. However, investigation of different labour force sequences showed variation in return to SWB set point. Men who entered economic inactivity after unemployment were much less likely to return to either their GHQ or life satisfaction set points, while men who exited to employment returned to both SWB set points. Conversely, women who exited to economic inactivity returned to their GHQ set point at a much faster rate than their life satisfaction set point. Differential changes in levels of distress and life satisfaction were observed upon further examination of types of economic inactivity. This finding provides evidence of the heterogeneity of the economically inactive and suggests that analysis of the effects of economic inactivity be stratified by type.

The revisions offered by Diener and colleagues (2006) are mainly supported by the findings from this study with potential new insights. While there was a return to the SWB set point in the overall model, the rate of return differed by labour force sequence as well as the measure of SWB. In general, both men and women included in this study had high levels of life satisfaction and lower levels of distress. The pre-unemployment distress levels of men and women were similar to those of the continuously employed or economically inactive. This pattern held for male life satisfaction, however women who experienced unemployment had lower life satisfaction pre-unemployment compared to continuously employed and economically inactive women. These findings support the idea that people are generally happy and positive as postulated by the first revision (Diener, Lucas, & Scollon, 2006) rather than neutral per the hedonic treadmill (Brickman & Campbell, 1971).

The second revision states that there are differences in the set points of individuals. This revision is supported by the finding of different levels of distress and life satisfaction at stage 1 for the different labour force sequences. While these differences may be small, larger differences can be seen between the set points of the different labour force sequences and those who were continuously employed or economically inactive. Men and women also appear to have different SWB set points. Similar differences for other life events, including marriage, divorce and disability were found by Lucas (2007a).

The finding that there were improvements in GHQ scores on exiting the labour market but no corresponding increase in life satisfaction for both men and women who experienced EUI supports the third revision. This revision states that there are multiple set points and that changes to these set points over time may not correspond due to the different dimensions of SWB (Diener, Lucas, & Scollon, 2006). Further exploration of the economic inactivity state post-unemployment showed that participants who became long-term sick had a decrease in their distress levels with no corresponding change to their life satisfaction. Similar findings are seen in the IUE family care and IUI retired participants where there were significant reductions in life satisfaction while GHQ levels did not significantly change. The third revision suggests that returns to set points may occur for certain SWB domains but not for others, a finding observed with these data.

This study shows that overall there is a return to GHQ and life satisfaction set points following unemployment, however the time to return is affected by the labour force sequence experienced, with some sequences showing no return. This finding offers inconclusive support to the fourth hedonic treadmill revision, which states that set points can be altered. However some considerations need to be made which may increase support for this revision. One caveat is that the time to return to set point can vary, dependent on the length of unemployment. The figures in this paper provide a picture when a 2-year unemployment spell is experienced. In cases where SWB worsened during spells of unemployment, the length of unemployment spell becomes important. Taking into account any levelling-off effects, the longer one is unemployed, the longer it may take to return to set point. Experiencing multiple or repeated life events may also permanently alter the set point. Recent analysis of BHPS participants showed that those who experienced multiple unemployment spells following economic inactivity had worse reactions with each unemployment spell (Booker & Sacker, 2011), making it less likely that these participants return to their SWB set point. The return to SWB set point may also be affected by national policies regarding unemployment and other social welfare benefits; these findings may not be universal phenomena. Additionally, for people who become employed after unemployment, the type of job may influence whether and how quickly they return to their SWB set point. If the post-unemployment job is of lower status or income or for fewer hours, then the return may be slower than if a person were to get a job appropriate for their qualifications and needs (Fineman, 1987).

Finally, Diener and colleagues offered a fifth revision to the hedonic treadmill, which states that adaptation varies across individuals. While this study did not specifically examine individual variations in adaptation, it did examine group differences in adaptation. Our findings showed faster rates of return to the GHQ set point for those participants who experienced IUE as compared to those who experienced EUE or EUI, and particularly slow adaptation was seen for IUI participants. Similar life satisfaction adaptation patterns differences were observed for women, while the time of return for IUE men was in between the time for EUE and EUI men. There were also differences between the two measures of SWB, and in some cases where there was a return to set point for one measure, there was not for the other measure. Additionally, there were differences in return to SWB set point by gender and by economic inactivity state, e.g. retirement compared with family care. One possible explanation for the lack of return to GHQ set point for men but not women is that their contribution to the total household income is larger. Therefore when men experience unemployment or enter economic inactivity, their increase in distress becomes so great that they are not able to return to their set point.

Comparisons with previous studies

Other studies (Lucas, 2007a; Lucas, Clarke, Georgellis, & Diener, 2004; Luhmann & Eid, 2009) have provided evidence of alterations to set point. Clark and colleagues (Clark, Diener, Georgellis, & Lucas, 2008; Clark & Georgellis, 2010) observed a slight return to set point among both British and German women who experienced unemployment, while there was no return for men in either country. In addition, full return to life satisfaction set points after experiencing other life events, such as marriage, divorce, birth of a child, etc., were observed for both men and women in these same studies (Clark, Diener, Georgellis, & Lucas, 2008; Clark & Georgellis, 2010). These differences in reactions to different life events raise the question of the generalisability of these findings to life events other than unemployment, and more research could explore whether there are differential changes to SWB set points for different life events. The effects of the combinations of these events should also be investigated. The cumulative or possibly synergistic effects of experiencing multiple life events within weeks or months of each other may greatly impact one's subjective well-being and whether a return to pre-event(s) set point is obtained. The sequence and anticipation of these events may also be important to the subjective well-being response.

Limitations

There are some limitations to this study. First, the method in which the SWB and labour force status were measured may result in different rates of return to SWB set point. Labour force status was self-reported as the current status at the time of the interview. This means that a person could have experienced different labour force states throughout the year, which might impact on subjective well-being. The GHQ requires assessment of recent behaviours and emotions, while other SWB measures such as symptom checklists do not require comparisons to past behaviours. The return to set point may be different for a checklist than for the GHQ, however symptoms were not measured in the BHPS.

Secondly, this study did not examine any potential moderators that could help to explain whether there are certain characteristics (e.g. coping skills, personal and household income and savings, potential earning capability, and personality) which allow people to return to their set point, either at all or faster than others. While other studies have shown differences in return to SWB set point between high and low earners (Georgellis, Gregoriou, Healy, & Tsitsianis, 2008) and introverts and extraverts (Clark & Georgellis, 2010) these studies have observed unemployment spells only and did not examine pre-unemployment labour force status. The findings from this study show differences in return to set point, and future studies could usefully address potential moderators of these effects.

Thirdly, we did not take into account the wider socio-economic context of the UK between 1991 and 2008. It is possible that unemployment which happened during a time of relative economic stability and low rates of unemployment, would be experienced differently from those spells occurring during a time of economic decline such as the post-2008 'Great Recession'.

Finally, we did not examine what happens to people who experience more than one unemployment spell. The first, and in some cases only, unemployment spell experienced while enrolled in the BHPS was considered in this study. People who experience multiple unemployment spells may not return to the SWB level that was observed prior to the first unemployment spell, suggesting a permanent change in their set point. Nevertheless the variable indicating pre-study unemployment was not significant in all models for both outcomes, although there may still be some unknown variables that may affect the return to SWB set point. We do not know the time since pre-study unemployment spells, the labour force status before or after pre-study unemployment, or if there were any other life events which may have had an effect on an individual's SWB levels. Future studies should look at the effects of multiple exposures to different types of life events on subjective well-being set points.

Conclusions

This study has shown that the return to subjective well-being set point for men and women who experience a bout of unemployment, depends on the pre- and post-unemployment labour force sequence experienced and the SWB outcome examined. Men who became employed following a period of unemployment were more likely to return to their pre-unemployment SWB set points, while men who entered economic inactivity did not. Conversely, while women were likely to return to their SWB set points, the return to GHQ set point was quicker than their return to life satisfaction set point. Examination of the specific types of economic inactivity produced interesting patterns. Participants who retired also experienced a reduction in life satisfaction while participants who became economically inactive due to long-term health problems had a reduction in psychological distress. Current changes of and debates over governmental policies worldwide, to increase retirement age and decrease the number of economically inactive persons may have the unintended consequence of adversely impacting on levels of subjective well-being in this population group. Finally, time to return to SWB set point may vary with the duration of unemployment.

These findings suggest that an increased understanding of the effects of different types of economic inactivity on subjective well-being is indicated. Researchers and policy makers should address the psychological impacts and potential further health risks due to entering economic inactivity, particularly among men.

Acknowledgments

The support of both the Economic and Social Research Council (ESRC) and the University of Essex is gratefully acknowledged. Data from the British Household Panel Survey (BHPS) were supplied by the ESRC Data Archive. This research was supported by the ESRC Research Centre on Micro-Social Change (MISOC), ESRC References Number RES-535-25-0090 and the ESRC International Centre for Lifecourse Studies in Society and Health, ESRC References Number RES-596-28-0001.

References

Beatty, C., Fothergill, S., & Macmillan, R. (2000). A theory of employment, unemployment and sickness. *Regional Studies, 34*, 617-630.

Bollen, K. A., & Curran, P. J. (2006). *Latent Curve Models: A Structural Equation Perspective.* Hoboken, NJ: John Wiley & Sons, Inc.

Booker, C. L., & Sacker, A. (2012). Psychological well-being and reactions to multiple unemployment events: Adaptation or sensitisation? *Journal of Epidemiology and Community Health, 66,* 832-838.

Brickman, P., & Campbell, D. T. (1971). Hedonic Relativism and Planning the Good Society. In M. H. Appley (Ed.), *Adaptation Level Theory: A Symposium.* (pp. 287-302). New York: Academic Press.

Chou, C.-P., Yang, D., Pentz, M. A., & Hser, Y.-I. (2004). Piecewise growth curve modeling approach for longitudinal prevention study. *Computational Statistics & Data Analysis, 46*, 213-225.

Clark, A. E., Diener, E., Georgellis, Y., & Lucas, R. E. (2008). Lags and leeds in life satisfaction: A test of the baseline hypothesis. *Economic Journal, 118*, F222-F243.

Clark, A. E., & Georgellis, Y. (2010). Back to baseline in Britain: Adaptation in the BHPS. *Working Paper No. 2010-02*. Paris: Paris School of Economics.

Clark, A. E., Georgellis, Y., & Sanfey, P. (2001). The psychological impact of past unemployment. *Economica, 68*, 221-241.

Deaton, A. (2008). Income, health and well-being around the world: Evidence from the Gallop World Poll. *Journal of Economic Perspectives, 22*, 53-72.

Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist, 55*, 34-43.

Diener, E., & Emmons, R. A. (1984). The independence of positive and negative affect. *Journal of Personality & Social Psychology, 47*, 1105-1117.

Diener, E., Lucas, R. E., & Scollon, C. N. (2006). Beyond the hedonic treadmill: Revising the Adaptation Theory of Well-being. *American Psychologist, 61*, 305-314.

Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin, 125*, 276-302.

Fineman, S. (1987). Back to employment: Wounds and wisdom. In D. Fryer & P. Ullah (Eds.), *Unemployed People* (pp. 268-284). Milton Keynes: Open University Press.

Frederick, S., & Loewenstein, G. (1999). Hedonic adaptation. In D. Kahneman, E. Diener & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology.* (pp. 302-329). New York: Russell Sage Foundation.

Georgellis, Y., Gregoriou, A., Healy, J., & Tsitsianis, N. (2008). Unemployment and life satisfaction: A non-linear adaptation process. *International Journal of Manpower, 29*, 668-680.

Goldberg, D., & Williams, P. (1988). *A User's Guide to the General Health Questionnaire.* Windsor: NFER-Nelson.

Goldberg, D., & Williams, P. (1991). *A User's Guide to the General Health Questionnaire.* London: NFER-Nelson.

Goldberg, D. P., Gater, R., Sartorius, N., Ustun, T. B., Piccinelli, M., Gureje, O.,& Rutter, C. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychological Medicine, 27*, 191-197.

Headey, B. (2006). Subjective well-being: Revisions to dynamic equilibrium theory using national panel data and panel regression methods. *Social Indicators Research, 79*, 369-403.

Headey, B. (2008a). Life goals matter to happiness: A revision of set-point theory. *Social Indicators Research, 86*, 213-231.

Headey, B. (2008b). The set-point theory of well-being: Negative results and consequent revisions. *Social Indicators Research, 85*, 389-403.

Headey, B., & Wearing, A. (1992). *Understanding Happiness: A Theory of Subjective Well-being.* Melbourne: Longman Cheshire.

Lucas, R. E. (2005). Time does not heal all wounds: A longitudinal study of reaction and adaptation to divorce. *Psychological Science, 16*, 945-950.

Lucas, R. E. (2007a). Adaptation and the set-point model of subjective well-being: Does happiness change after major life events? *Current Directions in Psychological Science, 16*, 75-79.
Lucas, R. E. (2007b). Long-term disability is associated with lasting changes in subjective well-being: Evidence from two nationally representative longitudinal studies. *Journal of Personality & Social Psychology, 92*, 717-730.
Lucas, R. E., Clark, A. E., Georgellis, Y., & Diener, E. (2003). Reexamining adaptation and the Set Point Model of happiness: Reactions to changes in marital status. *Journal of Personality & Social Psychology, 84*, 527-539.
Lucas, R. E., Clarke, A. E., Georgellis, Y., & Diener, E. (2004). Unemployment alters the set point for life satisfaction. *Psychological Science, 15*, 8-13.
Luhmann, M., & Eid, M. (2009). Does it really feel the same? Changes in life satisfaction following repeated life events. *Journal of Personality & Social Psychology, 97*, 363-381.
OECD. (2010). *OECD Factbook 2010: Economic, Environmental and Social Statistics*: OECD Publishing.
OECD. (2011). OECD Better Life Initiative. Retrieved June 06, 2011., from http://www.oecdbetterlifeindex.org/
SAS Institute Inc. (2003). SAS/STAT Software (Version 9.2). Cary, NC.
Taylor, M. F., with Brice, J., Buck, N., & Prentice-Lane, E. (Eds.). (2009). *British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices.* Colchester: University of Essex.

Longitudinal and Life Course Studies 2013 Volume 4 issue 2 Pp 105 - 118 **ISSN 1757-9597**

The role of respondent characteristics in tracking on longitudinal surveys: evidence from the UK Millennium Cohort Study

Lisa Calderwood
Centre for Longitudinal Studies, Institute of Education, University of London
l.calderwood@ioe.ac.uk

(Received December 2012 Revised May 2013)

Abstract

Longitudinal surveys typically devote considerable resources to tracking procedures designed to minimise attrition through failure to locate sample members who move. Although these tracking procedures are often very successful, there is relatively little methodological evidence about the relative success, and cost-effectiveness, of different tracking procedures (Couper & Ofstedal, 2009). This paper extends the existing literature by exploring the effectiveness of office tracking and field tracking separately, and by examining the role of respondent characteristics as a determinant of tracking success rates. These issues are explored using the Millennium Cohort Study, a large-scale birth cohort study in the UK. The existing research on tracking procedures has been based on household panel surveys, but in the context of a birth cohort study with relatively high mobility rates among the study population and longer intervals between waves, the effectiveness of office tracking procedures is particularly important. Our main finding, that respondent characteristics are related to overall tracking success rate, implies that survey practitioners should consider ways of improving their tracking procedures for certain groups of respondents who are the least likely to be located through existing methods.

Keywords: tracking; attrition; mobility; non-response; Millennium Cohort Study

1. Introduction

As longitudinal surveys aim to follow sample members over time, they employ a range of procedures designed to minimise sample loss due to failure to locate those who move. From a scientific perspective, this is crucially important. As residential mobility is driven by social processes such as relationship and employment changes, failure to locate mobile sample members can lead to biased estimates of change in these and other important domains of substantive interest to data users.

Many longitudinal surveys have developed highly successful procedures to minimise sample loss through failure to locate. For example, the Panel Survey of Income Dynamics (PSID) and the Health and Retirement Study (HRS) successfully located 97-98 per cent of sample members who moved between the 2003-2005 and 2002-2004 waves of these US panel studies, and the German Socio-Economic Panel (GSOEP) and the British Household Panel Survey (BHPS) have tracking rates of 96 per cent and 94 per cent respectively (Couper & Ofstedal, 2009). These authors also show that in PSID and HRS, around 90 per cent of sample members who move and were located, took part in the next wave of data collection. Similarly, research on the Millennium Cohort Study (MCS) in the UK has shown that conditional on location, families who moved after wave 1 were as likely as non-mobile families to take part in wave 2 (Plewis, Ketende, Joshi, & Hughes, 2008). This combination of high rates of tracking and a high likelihood that located sample members will be interviewed, means that the resources involved in tracking on longitudinal surveys are generally viewed as 'money well spent' by survey practitioners.

However, as Couper and Ofstedal (2009) point out, there is very little methodological evidence on the relative success, and cost-effectiveness, of different tracking procedures. They argue that survey design features, such as the interval between waves and the tracking procedures used, are a major determinant of tracking success and as these are under the control of survey practitioners, research should focus on optimising their design and evaluating their cost-effectiveness. This has led to increasing interest in improving the effectiveness of tracking procedures and in particular, on the optimal design of between-wave mailings, with randomised experiments being carried out on these mailings on the BHPS (Fumagalli, Laurie, & Lynn, 2010), PSID (McGonagle, Couper, & Schoeni, 2011) and the MCS (Calderwood, 2012). Although this research has undoubtedly improved knowledge in relation to between-wave mailings, there has been little attention in the literature to other tracking procedures.

This paper extends the existing literature by exploring the effectiveness of tracking carried out remotely in the office prior to the start of data collection and tracking carried out by interviewers in the field during data collection, and investigating how respondent characteristics are related to tracking success. By examining both office and field tracking, we provide indicative evidence on cost-effectiveness; as office tracking is remote, it is less expensive than field tracking, and therefore increasing the proportion of movers that are located using office-based methods should lead to an improvement in cost-effectiveness. For longitudinal surveys with high mobility rates, long intervals between waves and without the resources to carry out field tracking, improving the effectiveness of office tracking will be particularly important. This paper also addresses the role of respondent characteristics as a determinant of tracking success rates. We examine how the office tracking rate, field tracking rate and overall tracking rate are related to a range of respondent characteristics which we hypothesise may be related to tracking success. If certain types of sample members are more difficult to locate than others, or more likely to be located through different tracking methods, this may have implications for survey practice and the design of tracking procedures. For example, it may be that tracking procedures should be tailored for different types of respondents. To our knowledge the relationship between the characteristics of sample members and tracking success has not been explored directly before.

We examine these different tracking success rates using data from the Millennium Cohort Study, a large-scale birth cohort study in the UK. We explore mobility over a two-year period between wave 2 (at age 3) and wave 3 (at age 5), for families who took part at wave 2. As this is a study of families with young children, the between-wave mobility rate is relatively high. Around one in five (21%) of the wave 2 co-operating families had moved in this two-year period. The interval between waves on cohort studies is not fixed, rather it varies with the age of the sample member, so unlike panel surveys, which tend to have relatively short, fixed intervals between waves, cohort studies can have much longer intervals between waves. In this context, increasing the proportion of movers located using office tracking procedures is particularly important.

The next section reviews the existing evidence in this area and develops our hypotheses, the third section describes the data and methods that we use, the fourth and fifth sections present and discuss our results and the final section concludes with some implications for survey practice and recommendations for further research.

2. Background

The extensive range of tracking procedures that can be employed by longitudinal surveys is well-documented and there are several examples in the literature reporting the procedures used on different surveys. The earliest examples of published research on this topic are from the US in the 1960s and 1970s, published by the American Association of Public Opinion Research. Eckland (1968) presents tracking rates from several US longitudinal surveys carried out in the 1960s and reviews the tracking procedures used on these studies (which include postal and telephone services and directories, other public records, and employing local co-ordinators to search at 'grass-roots' level). Crider, Willits, & Bealer, (1971) and McAllister, Goe, & Butler, (1973) extend this literature by reviewing the tracking procedures used on a particular large-scale longitudinal survey and emphasising the need to collect extensive contact information, including full names and date of birth of study participants and the details of contact persons i.e. friends and/or relatives of sample members who may know where they are if they moved.

More recently, Laurie, Smith, & Scott, (1999) reviewed the tracking procedures used on the British Household Panel Survey. They make a distinction between prospective and retrospective tracking procedures. *Prospective tracking* aims to prevent loss of contact in the event of a change of address by ensuring that multiple alternative methods of contacting sample members are collected and kept as up-to-date as possible. This includes collecting email addresses and multiple phone numbers from sample members as well as the contact details of one or more contact persons. Most surveys encourage sample members to get in touch with changes to their contact information by providing toll-free telephone numbers, email addresses and websites. 'Keep-in-touch' mailings are also often sent between waves of data collection to prompt sample members to confirm or update their contact details. These mailings can also lead to the discovery of a move, if they are returned to sender by the current occupiers. *Retrospective tracking* involves trying to find sample members who are known to have moved. This includes attempting to contact the new occupiers and neighbours of the sample member's last known address and the contact persons given by sample members. This can be done by post, telephone, email or face-to-face. Retrospective tracking can also include seeking new addresses and other contact information in public records such as electoral roles, phone and postal directories as well as administrative data sources.

Couper and Ofstedal (2009) extend this classification of tracking methods by making a further distinction between office tracking and field tracking. Office tracking is often prospective e.g. sending out between-wave mailings and processing returns, receiving updates to contact information from sample members. It can also be retrospective, involving active attempts to locate sample members who are known to have moved through e.g. post office returns from between-wave mailings. Sometimes office-based tracking can involve automated processing of large numbers of cases at the same time e.g. matching to administrative records, which is a more efficient use of resources than case-by-case review. However, often office-based tracking, for example sending emails, letters, making phone calls and searching directories, does require staff to review cases individually. This can be resource intensive, particularly for large-scale surveys. As this is done remotely, usually from a centralised location, it is less expensive than field tracking, which involves interviewers making personal visits to the last known addresses of sample members, their neighbours and contact persons. This incurs additional direct costs of travel in addition to the labour costs associated with making these tracking attempts. Field tracking is usually only carried out on longitudinal surveys which use face-to-face data collection. It is relatively uncommon for surveys which use only remote methods of data collection to use field-based tracking, though this is not unheard of, particularly for local area surveys. However, face-to-face surveys typically only carry out field tracking during data collection waves, rather than between waves.

The difference in the relative costs of field tracking compared with office tracking implies that face-to-face longitudinal surveys should aim to locate as many movers as possible using office-based methods prior to the start of data collection for a wave, in order that resources are only used on more expensive interviewer tracking in the field for sample members who cannot be located through office tracking. However, office tracking can only be carried out prior to data collection for known movers, and it is not usually possible to identify all movers prior to the start of fieldwork i.e. some sample members may not be identified as movers until the interviewer visits the issued address. In addition, very recent moves, which take place during the data collection period, cannot, by their nature, be identified before the start of fieldwork. For this reason, office and field tracking are often carried out iteratively during the data collection period. Movers identified by interviewers during fieldwork, but for whom a new address cannot be found through field tracking, are returned for office tracking during the data collection period, and subsequently re-issued to interviewers if a new address is found.

Overall tracking success will depend largely on the range of tracking procedures adopted by the study and the amount of resources devoted to tracking. Many large-scale longitudinal surveys employ a wide range of tracking procedures, both prospectively and retrospectively and in the office as well as in the field, and for this reason have very high rates of tracking success. For example, the Panel Survey of Income Dynamics (PSID) and the Health and Retirement Study (HRS) successfully located 97-98 per cent of sample members who moved between the 2002-2004 and 2003-2005 waves of these US panel studies, and the German Socio-Economic Panel (GSOEP) and the British Household Panel Survey (BHPS) have tracking rates of 96 per cent and 94 per cent respectively (Couper and Ofstedal, 2009).

However, as well as being a function of the amount of tracking effort made by the study, tracking success is also related to the behaviours of the study members themselves. Tracking is sometimes characterised as something that is 'done to' sample members, but it is important to remember that they can influence how easy or difficult it is for the study to locate them if they move. An obvious example of this is by notifying the study team when they move. Other behaviours likely to be associated with tracking success include leaving a forwarding address with neighbours/new occupiers, responding to keep-in-touch exercises and providing multiple sources of contact information. It is plausible to hypothesise that different types of respondents, living in different circumstances, may be more or less likely to exhibit these behaviours, and hence that respondent characteristics may be related to tracking success. In general, very little attention has been given in the literature to how the characteristics of movers are related to the likelihood of successful tracking; to our knowledge the relationship between the characteristics of sample members and tracking success has not been explored directly before.

Couper and Ofstedal (2009) discuss the role of respondent characteristics in tracking success and argue that individuals with large family and social networks will be easier to locate then individual who are socially isolated. However, due to the high overall tracking rate in the HRS and PSID, they do not examine how individual characteristics are related to tracking success. Call (1990) explores how individual characteristics are related to the number of contact persons given by respondents on the National Survey of Families and Households in the US. They find that younger and older respondents, ethnic minorities and single persons provided fewer contact persons. Although they acknowledge that this may be because such individuals have fewer potential contacts to provide, rather than as a result of unwillingness, it is nevertheless likely that respondents who provided fewer contact persons will be less easy to locate if they move. The literature on between-wave mailings also shows that younger sample members (McGonagle et al., 2011), lower-educated sample members and those who speak languages other than English at home (Calderwood, 2012) are less likely to return these mailings, which also makes successful tracking less likely. We hypothesise that a range of individual, family and housing characteristics will be associated with tracking success.

In relation to individual characteristics, we expect age, ethnic group and education to be related to tracking success. Specifically, it is hypothesised that younger, non-white and lower-educated sample members are less likely to be successfully tracked. Previous research has established that younger sample members, those from non-white ethnic groups and lower-educated sample members tend to have lower response rates on many longitudinal surveys, and we would expect these characteristics to also be negatively associated with tracking success. For some minority groups, poor English language skills may also make tracking more difficult, particularly tracking through members of the extended family e.g. grandparents.

In relation to family characteristics, we expect family type and family employment situation to be associated with tracking success. We also hypothesise that lone parents will be less likely to be successfully tracked than couple families. In part, this is because lone parent families tend to have lower response rates in general than couple families and we expect this to also be related to tracking success. Additionally, for lone parent families, it is likely that less contact information will be available to use for tracking i.e. we only have contact details for one respondent and one contact person, whereas in couple families, we collect contact details for two respondents and, for many families, two contact persons i.e. one for each parent. We expect that sample members in paid employment will be more likely to be tracked than those who are not in work. This is primarily because employment tends to be positively associated with taking part overall.

In relation to housing characteristics, we hypothesise that tenure and accommodation type will be associated with tracking success. Specifically, we expect that sample members living in rented accommodation and those living in flats will be less likely to be successfully tracked than families living in owner-occupied accommodation and in houses. Living in rented accommodation is associated with less stability in terms of residential moves, which in turn is likely to be associated with lower rates of tracking success, as renters are, in general, less likely to develop social ties with their neighbours and less likely to leave forwarding addresses for new occupiers. New tenants of rented properties may also be less likely to return mailings for previous occupants to their sender, meaning that moves may remain undiscovered for longer. This is also true of the new occupants of flats, rather than houses, particularly flats which are part of multiple-occupancy blocks. Flats can also be more difficult for interviewers to gain access to than houses, which makes it more difficult to speak to new occupants and neighbours, and is another reason to expect that tracking is less successful for sample members living in flats.

We will examine how these characteristics are associated with office tracking success, field tracking success and overall tracking success. We hypothesise that the same individual, family and housing characteristics that are associated with successful tracking overall will be associated with successful office and field tracking, and that the direction of the relationships between these characteristics will be the same for both office and field tracking. Moreover, we would expect that individual characteristics will be more strongly associated with office tracking success than field and overall tracking success, because office tracking is more dependent on the proactive behaviour of respondents than field tracking.

3. Data and Methods

The Millennium Cohort Study (MCS) is a longitudinal birth cohort study following the lives of over 19,000 children in the UK who were born in 2000 and 2001. The sample was drawn from records of recipients of a universal benefit for children, and was initially geographically clustered by electoral ward with an over-representation of areas with high proportions of Black or Asian families, disadvantaged areas and areas in the three smaller UK countries. There have been five waves of the study so far, when the cohort member was aged 9 months, then at 3, 5, 7 and 11 years of age. At all waves, interviews were conducted with both resident parents, and from the second wave onwards, data has been collected directly from the cohort member. The study has also collected data from siblings and teachers as well as consents to link to administrative data for cohort member, parents and siblings. More information about the design of the study can be found in Plewis (2007).

The MCS employs both office and field tracking. The study provides a Freephone number, email address and a website through which cohort families can inform the study’s cohort maintenance team if they change their address or contact details. Contact details for study members are updated annually between-waves through the mailing of a reply-slip which is pre-printed with all of the families’ contact details i.e. address, names, phone numbers, email address and contact person details. Undelivered mail, usually indicating that the family has moved, is returned to the study by the post office, which triggers retrospective office-based tracking by the cohort maintenance team. Multiple attempts are made to contact sample members, their nominated contact person and the current occupiers of the address previously occupied by sample members through telephone, mail, email and text messaging. Specialist tracing software which combines publicly available Post Office, electoral and phone records is also used routinely in the office for individual searches. As the fieldwork for the study is carried out face-to-face, interviewers also attempt to track families who have moved by making personal visits to the last known addresses of cohort members, neighbours and, if local, their contact persons, in addition to attempting contact through phone and mail. During the fieldwork period, movers who cannot be located by interviewers are returned for office tracking.

This paper examines tracking between wave 2, carried out at age 3 in 2003-4 and wave 3, which took place at age 5 in 2006. We restrict the sample to the families who took part in wave 2 and use respondent characteristics measured at this survey in our analysis. As wave 2 was the first follow-up wave, and only families who were interviewed at baseline (wave 1), were followed up, almost all of the families in our sample had taken part in both wave 1 and wave 2[i] i.e. there was little variation in terms of their participation history. In total, 15,590 families took part in wave 2, and 3,278 (21%) of them had moved by wave 3. For a very small number of families (1%) which were not issued to the field at wave 3 it is not possible to know with certainty whether or not they moved. For all other cases, it is possible to know with a very high degree of certainty whether or not they moved because, even if they didn't participate in the survey, an interviewer visited their address and established whether or not they were still resident. This mobility rate is lower than the proportion of families who moved (over a slightly longer period) between wave 1 and wave 2 of MCS, which was 38 per cent (Plewis et al., 2008). It is very similar to the mobility rates observed in a two-year period in PSID, around 21-22 per cent (McGonagle et al., 2011) and comparable to rates observed in BHPS, around 10 per cent each year (Laurie et al., 1999).

We define the office tracking rate to be the percentage of all movers who were located using office tracking methods alone prior to the start of data collection, the field tracking rate as the percentage of movers *not located by office tracking* who were located by interviewers in the field, and the overall tracking rate as the percentage of all movers who were located by either office or field tracking. Under these definitions, office tracking necessarily takes place prior to the start of fieldwork and therefore prior to field tracking; so they are sequential processes. Importantly, the field tracking rate, defined in this way, is a conditional success rate i.e. we choose to analyse the outcome of field tracking only for those who were not found by office tracking. This means that the office and field tracking rates are not directly comparable. It should also be noted that some of the cases defined as being located using field tracking may have also been tracked in the office during fieldwork, and so may have been tracked through a combination of field and office tracking. It is unclear whether these cases could have been located using office tracking alone, and we therefore include them in the field tracking rate. Additionally, as it is not possible to identify all movers prior to the start of fieldwork, office tracking was not attempted for all movers prior to the start of fieldwork i.e. some of the 'movers not located through office tracking' did not actually receive office tracking (prior to the start of fieldwork).

Figure 1 gives the office and field tracking rates and Figure 2 gives the overall tracking rate.

Figure 1. Office and field tracking rates between wave 2 and wave 3 of MCS

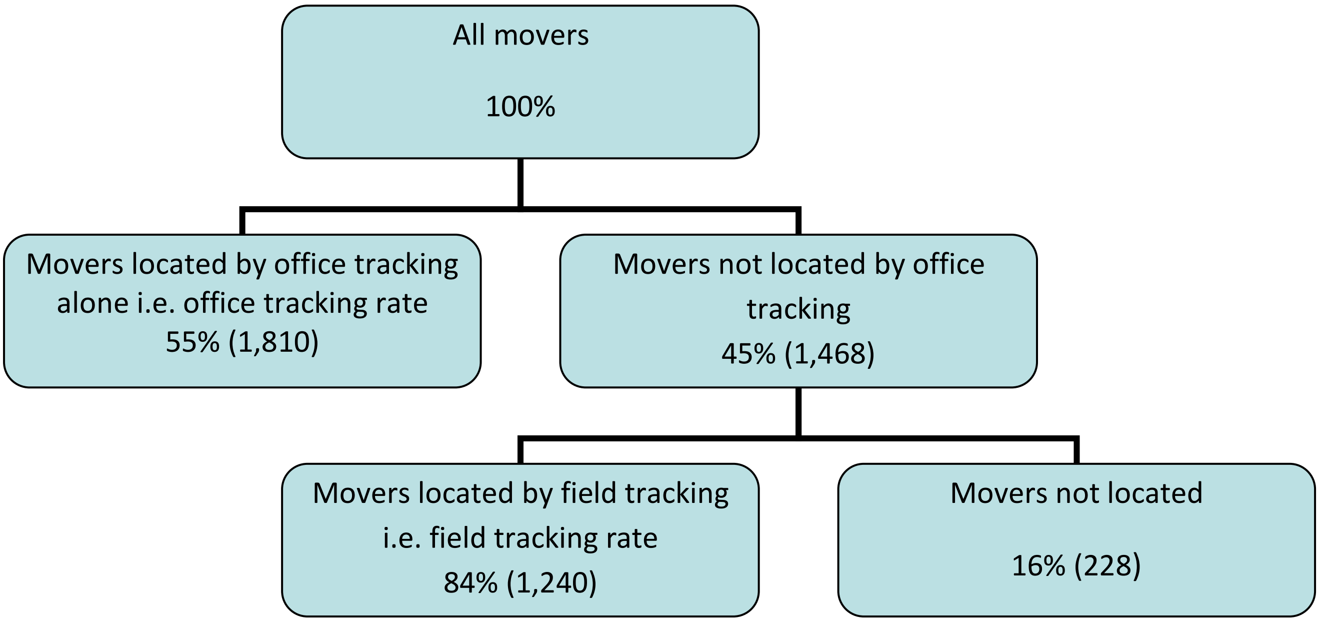


Figure 2. Overall tracking rates between wave 2 and wave 3 of MCS

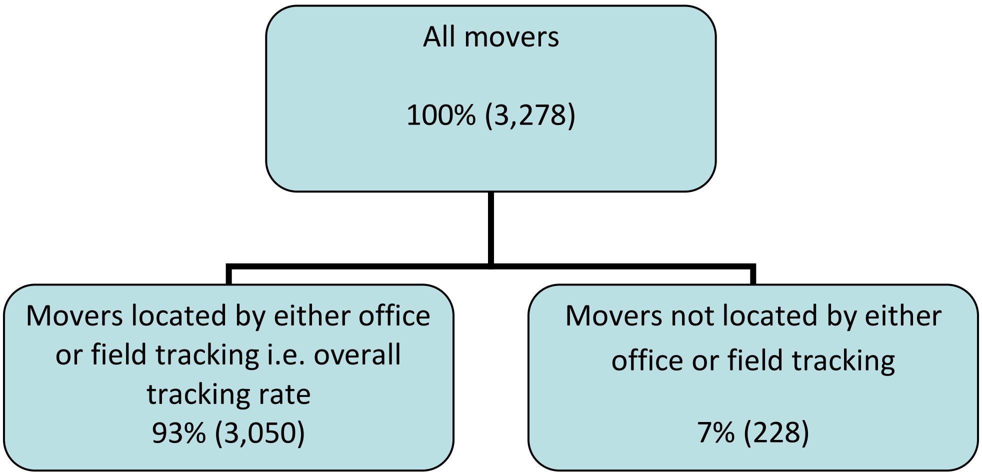


Figure 1 shows that 55 per cent of all movers were successfully located using office tracking alone prior to the start of fieldwork. This is the office tracking rate. Figure 1 also shows that, of the movers not located prior to the start of fieldwork using office tracking, 84 per cent were successfully located during fieldwork using field tracking. This is the conditional field tracking rate. Figure 2 shows that 93 per cent of all movers were located by either office tracking prior to the start of fieldwork or by field tracking. This is the overall tracking rate.

The overall level of tracking success is comparable with other major longitudinal surveys. It is encouraging that the majority of movers are located prior to the start of data collection, but it is difficult to know how this proportion compares with other surveys as there is limited published research in this area. One exception is the paper by Laurie et al. (1999) which reports that around half of movers on BHPS are found using office tracking prior to the start of fieldwork.

Our analysis will test the hypothesis that sample members who are younger, non-white, lower-educated, lone parents, not working, living in flats or living in rented accommodation will all be less likely to be located by comparing how these characteristics, measured at wave 2, are associated with the office, field and overall tracking rate. As noted earlier, the MCS involves interviews with up to two resident parents. The individual characteristics used, i.e. age, ethnic group and education, are those of the main respondent, who is almost always the child's natural mother. Age and ethnic group are self-reported by the main respondent. The education measure used is based on qualifications. The main respondent's highest academic or vocational qualifications are mapped to an equivalent level on a standard scale which is used in the UK for National Vocational Qualifications (NVQs). The highest level, Level 5, is equivalent to a postgraduate degree and the lowest level, Level 1, is equivalent to a General Certificate of Secondary Education (GCSE), usually taken at age 16, in at least one subject, awarded at the lower range of grades i.e. D-G. Family type is derived from household composition, and household employment status uses both household composition and the working status of main and partner respondents. Most partner respondents are the natural or step-father of the child. Tenure and accommodation type are reported by the main respondent on behalf of the household.

4. Results

Table 1 gives the office tracking rate prior to the start of fieldwork for all movers, the field tracking rate for movers not located by office tracking prior to the start of fieldwork, and the overall tracking rate for all movers, by the selected individual, family and housing characteristics of sample members at wave 2.

Table 1 shows that ethnic group and education had a statistically significant relationship with the office tracking rate, all of the characteristics, except for age, had a statistically significant relationship with the field tracking rate, and all of the characteristics, including age, had a statistically significant relationship with the overall tracking rate. The observed relationships between the tracking rates and the characteristics chosen were in the hypothesised direction.

Looking firstly at the office tracking rate, there is a strong relationship with both ethnic group and education. For example, almost 59 per cent of white movers were located using office tracking prior to the start of fieldwork compared with around 44 per cent of Black or Black British movers. Similarly, over 60 per cent of movers with relatively high education levels (Level 4 or 5 qualifications) were located using office tracking compared with less than half, around 48/49 per cent, of those with the lowest level of education (Level 1) or no qualifications. There is also a clear age gradient in office tracking success, with older respondents more likely to be located using this method, though this relationship is not statistically significant. The office tracking rate was not related to family and housing characteristics. Although overall these results do not support our hypotheses that all of the chosen characteristics would be related to the office tracking rate, it is interesting that the only statistically significant relationships observed are with individual characteristics i.e. ethnic group and education. This supports our secondary hypothesis that individual characteristics would be most strongly related to office tracking, as this method is more dependent on the proactive behaviour of respondents than field tracking. In addition, prior research on MCS has shown that ethnic group and education are both significantly related to responding to the 'keeping-in-touch' mailing, which is one of the main office tracking methods used (Calderwood, 2012).

Table 1. Office, field and overall tracking rates for movers by wave 2 characteristics

Wave 2 characteristics	Office tracking rate (% of all movers)	Field tracking rate (% of movers not located by office)	Overall tracking rate (% of all movers)	Sample size (all movers)	Sample size (movers not located by office)
Individual characteristics (main respondent)					
Age					
16-24	51.4	81.8	91.2	698	351
25-29	55.2	84.0	92.8	768	358
30-34	58.4	90.4	96.0	967	416
35-39	61.0	88.8	95.6	602	237
40+	59.6	89.9	95.9	207	89
F-statistic	*2.19*	*2.33*	*3.63*		
p-value	*>0.05*	*>0.05*	*<0.01*		
Ethnic group					
White	58.7	90.2	96.1	2652	1153
Mixed	48.1	76.0	87.5	40	23
Indian	51.4	75.2	87.9	63	30
Pakistani and Bangladeshi	54.9	72.7	87.7	180	84
Black or Black British	43.9	67.4	81.7	98	48
Other	30.4	61.5	73.2	55	33
F-statistic	*4.07*	*8.67*	*15.35*		
p-value	*<0.01*	*<0.001*	*<0.001*		
Educational qualifications					
No qualifications	48.3	77.3	88.3	498	246
Overseas qualifications only	46.1	79.3	89.0	109	55
Level 1 (lowest)	49.8	80.0	89.9	288	150
Level 2	58.1	90.4	96.0	936	414
Level 3	57.9	85.2	93.7	461	206
Level 4	61.3	92.2	97.0	848	345
Level 5 (highest)	62.5	91.7	96.9	120	42
F-statistic	*3.39*	*4.84*	*8.00*		
p-value	*<0.01*	*<0.001*	*<0.001*		
Family characteristics					
Family Type					
Married or cohabiting natural parents	56.9	90.0	95.7	2311	1021
Lone natural mother	58.1	79.1	91.3	789	358
Other family type	51.9	72.6	86.8	178	89
F-statistic	*0.86*	*13.72*	*14.73*		
p-value	*>0.4*	*<0.001*	*<0.001*		
Household employment status					
Main (and/or partner) in work	57.2	89.9	95.7	2456	1105
Main (and partner) not in work	56.2	75.3	89.2	780	345
F-statistic	*0.13*	*33.48*	*32.80*		
p-value	*<0.7*	*<0.001*	*<0.001*		
Housing characteristics					
Tenure					
Own	58.8	92.5	96.9	1562	680
Rent	54.4	80.3	91.0	1462	681
Other	59.3	87.6	94.9	211	88
F-statistic	*1.65*	*16.11*	*19.65*		
p-value	*>0.1*	*<0.001*	*<0.001*		

(Table 1 cont'd)

Accommodation type						
	House	57.5	89.6	95.6	2607	1167
	Flat	55.4	74.7	88.6	628	282
F-statistic		*0.61*	*22.64*	*23.72*		
p-value		*>0.4*	*<0.001*	*<0.001*		
Sample size		3278	1468	3278		

Notes. Design-based F tests (degrees of freedom omitted) were used to test the null hypothesis of no relationship between each of the characteristics and the tracking rates. The analysis was carried out using the svy commands in STATA to adjust for the sample design.

Now turning to the field tracking rate, which, as noted earlier, is conditional on not being located using office tracking alone prior to the start of fieldwork. As hypothesised, this shows a statistically significant relationship with all of the chosen characteristics, with the exception of age. The higher value of the F-statistic for ethnic group and education, compared with the comparable F-statistic for the office tracking rate, shows that the relationship between these characteristics is even stronger for field tracking compared with office tracking. This is particularly notable given the smaller sample size for field tracking, though, as noted earlier, the field tracking rate is conditional on not being located through office tracking. As with the office tracking rate, there is a clear age gradient in the field tracking rate, though it is not statistically significant. In relation to family characteristics, as hypothesised, couple families had higher field tracking rates than lone parent families (90 per cent compared with 79 per cent) and families with at least one parent employed had higher field tracking rates than those without parental employment (90 per cent compared with 75 per cent). Housing characteristics also showed the hypothesised relationships with the field tracking rate. Movers living in rented accommodation were more difficult to locate in the field than those living in owner-occupied accommodation (80 per cent compared with 93 per cent) and those living in flats were more difficult to locate in the field than those in houses (75 per cent compared with 90 per cent).

In relation to the overall tracking rate, as hypothesised, all of the chosen characteristics were statistically significantly associated with the overall tracking rate, including age, which was not significant for either the office or the field tracking rate when examined separately. Overall, these results provide strong evidence that respondent characteristics are important determinants of tracking success and provide support for our hypotheses regarding the relationship between tracking and age, ethnic group, education, family type, employment status, housing tenure and accommodation type.

In order to further investigate these relationships, we carried out multiple logistic regression analysis for each of the three tracking rates, in order to ascertain whether these relationships remain statistically significant when controlling for other characteristics i.e. the rest of our chosen characteristics. Table 2 gives the odds ratios associated with the office tracking rate, conditional field tracking rate and overall tracking rate from the logistic regression models, which included all of our chosen characteristics.

Table 2. Odds ratios of being located through office tracking, being located through field tracking and being located through either office or field tracking from logistic regression models, by wave 2 characteristics

Wave 2 characteristics		Office tracking (all movers) Odds ratios (95% CI)	Field tracking (movers not located by office) Odds ratios (95% CI)	Overall tracking (all movers) Odds ratios (95% CI)
Individual characteristics (main respondent)				
Age				
	16-24	1	1	1
	25-29	1.24 (0.96,1.61)	0.94 (0.57,1.54)	1.02 (0.64, 1.63)
	30-34	1.38 (1.03,1.85)	1.12 (0.59,2.13)	1.28 (0.68,2.40)
	35-39	1.69 (1.23,2.34)	1.27 (0.55,2.93)	1.79 (0.86,3.72)
	40+	1.77 (1.13,2.77)	1.17 (0.44,3.12)	1.61 (0.59,4.35)
p-value		*<0.05*	*>0.9*	*>0.4*
Ethnic group	White	1	1	1
	Mixed	0.68 (0.32,1.44)	0.60 (0.18,2.06)	0.47 (0.16,1.34)
	Indian	0.75 (0.39,1.44)	0.21 (0.88,0.51)	0.22 (0.09,0.54)
	Pakistani and Bangladeshi	0.99 (0.67,1.47)	0.23 (0.09,0.57)	0.31 (0.14,0.67)
	Black or Black British	0.53 (0.31,0.89)	0.25 (0.10,0.63)	0.23 (0.12,0.45)
	Other	0.29 (0.15,0.56)	0.11 (0.04,0.31)	0.08 (0.03,0.20)
p-value		*<0.001*	*<0.001*	*<0.001*
Educational qualifications				
	No qualifications	1	1	1
	Overseas qualifications only	0.92 (0.52,1.62)	1.31 (0.44,3.95)	1.27 (0.48,3.37)
	Level 1 (lowest)	0.99 (0.67,1.46)	0.82 (0.43,1.57)	0.73 (0.42,1.26)
	Level 2	1.34 (1.01,1.78)	1.51 (0.81,2.81)	1.70 (0.89,3.24)
	Level 3	1.32 (0.95,1.83)	0.78 (0.45,1.35)	0.89 (0.54,1.47)
	Level 4	1.42 (1.01,1.99)	1.67 (0.85,3.29)	1.64 (0.88,3.06)
	Level 5 (highest)	1.36 (0.74,2.52)	1.19 (0.29,4.93)	1.37 (0.36,5.14)
p-value		*>0.1*	*>0.9*	*>0.1*
Family characteristics				
Family Type				
	Married or cohabiting natural parents	1	1	1
	Lone natural mother	1.14 (0.88,1.47)	0.83 (0.47,1.48)	1.01 (0.57, 1.79)
	Other family type	1.00 (0.68,1.48)	0.39 (0.18,0.86)	0.39 (0.19,0.79)
p-value		*>0.5*	*>0.05*	*<0.05*
Household employment status				
	Main (and/or partner) in work	1	1	1
	Main (and partner) not in work	1.10 (0.85,1.43)	0.46 (0.30,0.70)	0.52 (0.34,0.81)
p-value		*>0.4*	*<0.001*	*<0.001*
Housing characteristics				
Tenure	*Own*	*1*	*1*	*1*
	Rent	0.94 (0.76,1.17)	0.69 (0.41,1.16)	0.71 (0.44,1,16)
	Other	1.12 (0.71,1.75)	1.42 (0.53,3.80)	1.45 (0.59,3.58)
p-value		*>0.6*	*>0.2*	*>0.1*
Accommodation type				
	House	1	1	1
	Flat	1.03 (0.81,1.32)	0.53 (0.32,0.88)	0.59 (0.36,0.94)
p-value		*>0.7*	*<0.05*	*<0.05*
-2log likelihood		-2058.87	-489.29	-648.97
Sample size		3,040	1,346	3,040

Notes. Design-based F tests (degrees of freedom omitted) were used to test the null hypothesis of no relationship between each of the characteristics and the tracking rates. The analysis was carried out using the svy commands in STATA to adjust for the sample design. Note that the -2log likelihood values are not available using the svy commands and were obtained by re-running the models without the svy commands.

In relation to office tracking, Table 2 shows that, controlling for other characteristics, age and ethnic group are significantly related to the office tracking rate. Compared with the results from Table 1, this shows that, controlling for other factors, education is not significantly related to office tracking, whereas age is now shown to have a significant relationship with office tracking. The direction of the relationship with age and ethnic group is as expected and as we previously observed in Table 1 i.e. older respondents and white movers are more likely to be located than younger respondents and non-white movers. None of the other characteristics are significantly associated with respondent tracking.

In relation to conditional field tracking, Table 2 shows that, after controlling for other characteristics, many fewer characteristics are significantly related to the field tracking rate. In Table 1, all of the chosen characteristics except age were significantly related to field tracking, whereas, controlling for other characteristics, only ethnic group, employment status and accommodation type remain significant. The direction of the relationship between these characteristics and field tracking is as expected and as observed in Table 1 i.e. non-white, non-employed and flat-dwellers are less likely to be tracked in the field. Education, family type and tenure are no longer significantly related to the field tracking rate, controlling for other characteristics. The relationship between age and field tracking remains non-significant, controlling for other characteristics.

Similarly for overall tracking, Table 2 shows that, after controlling for other characteristics, many fewer characteristics are significantly related to the overall tracking rate. Ethnic group, family type, employment status and accommodation type are the only characteristics which remain significantly related to the overall tracking rate, controlling for other characteristics. The direction of the relationship is as expected and as observed in Table 1 i.e. non-white, other family types, non-employed, flat-dwellers are less likely to be tracked overall. Age, education and tenure are no longer significantly related to overall tracking, controlling for other characteristics.

Overall, the results from the multiple logistic regressions provide further evidence that respondent characteristics are important determinants of tracking success and show that, controlling for other characteristics, ethnic group, family type, employment status and accommodation type are strongly related to overall tracking success. However, it should be noted that the -2log likelihood values of the models show that their overall goodness of fit is relatively low, indicating that other factors are also important.

5. Discussion

Overall, these results show that several respondent characteristics show statistically significant relationships with office tracking success, field tracking success and overall tracking success between wave 2 and wave 3 of the UK Millennium Cohort Study.

Controlling for other characteristics using multiple logistic regression showed that ethnic group was related to office tracking, conditional field tracking and overall tracking, age was related to office tracking but not field tracking or overall tracking, employment status and accommodation type were related to both field tracking and overall tracking but not office tracking, and family type was related to overall tracking but not office or field tracking.

Reflecting on our hypotheses in the light of these findings, it is perhaps unsurprising that ethnic group was strongly associated with tracking success, though for surveys like MCS which incorporate over-sampling of minority ethnic groups, this is particularly worrying. Further research is needed to establish the mechanisms through which ethnic group impacts negatively on tracking success, but it seems likely that our hypothesised mechanisms i.e. language barriers, affecting both office and field tracking, are part of the explanation.

It was interesting that age was the only other characteristic which was significantly related to office tracking success, and that it was not related to field or overall tracking success. This gives some support to our hypothesis that individual characteristics would be more strongly related to office tracking than field and overall tracking, as office tracking is more reliant on proactive behaviours by sample members e.g. responding to keeping in touch mailings.

Family type showed an unusual relationship with tracking success; it was not significantly related to either office or field tracking but it was related to overall tracking success, controlling for other characteristics. We hypothesised that the mechanism between tracking success and family type was the additional contact information provided by partners in couple families. However, there was no difference between tracking success rates for lone mother families and couple families once other characteristics were controlled for. Rather, the reason why family type was significantly related to overall tracking was due to a much lower overall tracking rate among 'other family types'. It is unclear why the tracking rate should be much lower for this group and it is surprising that tracking success rates were no different for couple parents and lone mothers, as this implies that the additional contact information collected from partner respondents is of limited value in relation to tracking. However, further research is needed before concluding this. For example, it may be that the quantity and quality of information collected from partners at wave 2 was limited.

It was surprising that employment status was significantly associated with both field and overall tracking success, controlling for other characteristics, as we did not have a clear hypothesis about the mechanism through which this was likely to have an impact on tracking success. This may be reflection of a lower level of commitment to the study among workless families. However, further research is needed to better understand the process through which employment status is related to tracking success.

It was unsurprising that housing characteristics i.e. accommodation type and tenure, both showed the strong relationships with both field and overall tracking success in the bivariate analysis, as the hypothesised mechanisms through which these characteristics were expected to influence tracking were clear and direct. However, only accommodation type, and not tenure, remained significantly associated with field and overall tracking once other characteristics were controlled for. This implies that compositional differences between owners and renters explains the relationship between tenure and field and overall tracking observed in the bivariate analysis. Conversely, the fact that accommodation type remained significantly associated with both field and overall tracking, controlling for other characteristics, shows that this relationship cannot be explained by compositional differences between the types of people living in flats and those living in houses.

6. Conclusions

Overall, this paper has clearly shown that respondent characteristics are related to the successful office, field and overall tracking. In doing so, this paper makes a significant contribution to the survey research literature in this area. Although there are examples in the literature which show how individual characteristics are associated with returns to between-wave mailings and collection of contact information, to our knowledge, the relationship between the characteristics of sample members and tracking success has not been explored directly before. As noted earlier, sample members are sometimes viewed as passive in discussions of tracking procedures, though it is clear that by their actions e.g. notifying the study of a change of address or inaction e.g. failure to leave a forwarding address when they move, they can influence how likely it is that they can be located when they move. Further research is needed to examine this issue in more detail, and in particular, to directly test the hypothesised mechanisms through which these characteristics are related to tracking success. However, in terms of implications for survey practice arising from this paper, given the difficulty associated with tracking families living in flats and families from non-white ethnic groups, it is clear that longitudinal surveys should consider implementing additional and/or tailored tracking methods for these groups, including the collection of additional contact information for those living in flats and use of translated tracking materials and office tracking staff/interviewers who speak minority languages for non-white ethnic groups.

This paper also explored the effectiveness of both office and field tracking, and compared how individual, family and housing characteristics were associated with both office and field tracking. Our aim was to provide evidence which would help other longitudinal surveys to increase the proportion of movers which are found using office tracking methods, which are less expensive than field tracking, and thereby improve the cost-effectiveness of their tracking procedures.

Overall, we found that over half of all movers were located by office tracking prior to the start of fieldwork. We thereby demonstrated that it is possible to locate a relatively high proportion of movers using office tracking, prior to the start of fieldwork. Overall, only two of the individual characteristics i.e. age and ethnic group, were significantly related to office tracking success, providing some support for our hypothesis that individual characteristics would be more strongly associated with office tracking than field and overall tracking, because office tracking is more dependent on the proactive behaviour of respondents. The fact that the other characteristics were not significantly related to office tracking success implies that office tracking prior to fieldwork is not differentially effective for these different types of sample members. In some ways, this is a reassuring finding as it shows that, with the exception of younger and non-white respondents, office tracking procedures are not systematically failing to locate certain types of sample members. Attempting to improve office tracking procedures to make them more effective for younger and non-white respondents e.g. by using tailored and/or translated materials during keeping-in-touch mailings, would seem to be worth exploring.

However, we have also clearly shown the importance of field tracking to achieving high overall tracking rates. In addition, our analysis demonstrated that overall tracking success was significantly related, in the bivariate analysis, to all of the respondent characteristics. This implies that improving the effectiveness of tracking, both in the office and in the field, for these 'hard-to-locate' groups should be the primary aim of further research and improvements to survey practice.

Acknowledgements

The UK Millennium Cohort Study (MCS) is funded by the Economic and Social Research Council (ESRC) and a consortium of UK government departments, and run by the Centre for Longitudinal Studies (CLS). This research was carried out as part the Resource Centre funding for CLS from the ESRC (Ref: RES-579-47-001). I would like to thank Professor John Micklewright for helpful comments on drafts of this manuscript.

References

Calderwood, L. (2012). Improving between-wave mailings on longitudinal surveys: a randomised experiment on the UK Millennium Cohort Study. *CLS Working Paper,* 2012/4. London: Centre for Longitudinal Studies.

Call, V.R.A. (1990). Respondent Cooperation and Requests for Contacts in Longitudinal Research. *National Survey of Families and Households Working Paper No.35.* Madison: University of Wisconsin-Madison.

Crider, D.M., Willits, F.K., & Bealer, R.B. (1971). Tracking Respondents in Longitudinal Surveys. *Public Opinion Quarterly, 35*, 613-620.

Couper, M.P., & Ofstedal, M.B. (2009). Keeping in Contact with Mobile Sample Members. In P. Lynn (ed.) *Methodology of Longitudinal Surveys* (pp. 183-203). Chichester: John Wiley & Sons.

Eckland, B.K. (1968). Retrieving Mobile Cases in Longitudinal Surveys. *Public Opinion Quarterly, 32*, 51-64.

Fumagalli, L., Laurie, H., & Lynn, P. (2010). Experiments with Methods to Reduce Attrition in Longitudinal Surveys. *ISER Working Paper Series*, No. 2010-04.

Laurie, H., Smith, R., & Scott, L. (1999). Strategies for reducing nonresponse in a longitudinal panel survey. *Journal of Official Statistics, 15*, 269-282.

McAllister, R.J., Goe, S.T., & Butler, E.W. (1973). Tracking Respondents in Longitudinal Surveys: Some preliminary considerations. *Public Opinion Quarterly, 37*, 413-416.

McGonagle, K.A., Couper, M.P., & Schoeni, R.F. (2011). Keeping Track of Panel Members: An Experimental Test of a Between-Wave Contact Strategy. *Journal of Official Statistics*, *27*, 319-338.

Plewis, I. (ed.) (2007). *The Millennium Cohort Study: Technical Report on Sampling (4th. ed.).* London: Institute of Education, University of London.

Plewis, I., Ketende, S.C., Joshi, H., & Hughes, G. (2008). The Contribution of Residential Mobility to Sample Loss in a British Birth Cohort Study: Evidence from the first two waves of the Millennium Cohort Study. *Journal of Official Statistics*, *24*, 364-385.

Endnote

[i] Families interviewed at wave 2 included a small number (692) of 'new' families, who were first recruited to the study at wave 2 and who had not been approached at wave 1.

Longitudinal and Life Course Studies 2013 Volume 4 Issue 2 Pp 119 - 135 **ISSN 1757-9597**

STUDY PROFILE

Design and characteristics of a new birth cohort, to study the early origins and ethnic variation of childhood obesity: the BiB1000 study

M Bryant[1,2], G Santorelli[2], L Fairley[2], J West[2], DA Lawlor[3], R Bhopal[4], E Petherick[2], P Sahota[5], A Hill[6], N Cameron[7], N Small[8], J Wright[2], Born in Bradford Childhood Obesity Scientific Group*

[1]Clinical Trials Research Unit, University of Leeds
[2] Bradford Institute of Health Research, Bradford Teaching Hospitals NHS Trust
[3] MRC Centre for Causal Analyses in Translational Epidemiology, University of Bristol
[4] Centre for Population Health Sciences, University of Edinburgh
[5]Nutrition and Dietetics, Leeds Metropolitan University
[6]Leeds Institute of Health Sciences, University of Leeds
[7] Centre for Global Health and Human Development, Loughborough University
[8]School of Health Studies, University of Bradford

***** m.j.bryant@leeds.ac.uk

******Authors plus**: Amanda Farrin, Helen Ball, Carolyn Summerbell, Sally Barber, Diane Farrar and Rosie McEachan**

(Received October 2012 Revised January 2013)

Abstract

Epidemiological evidence indicates that early life factors are important for obesity development but there are gaps in knowledge regarding the impact of exposures during pregnancy and early life, especially in South Asian children. There is a corresponding lack of evidence to guide development of culturally-appropriate, obesity prevention programmes. This paper describes the methodology and characteristics of participants in Born in Bradford 1000 (BiB1000), a nested cohort of the Born in Bradford prospective birth cohort. BiB1000 aims to enable a deep and extensive understanding of the predictors and influences of health-related behaviours to develop a culturally-specific obesity prevention intervention. 1,735 mothers agreed to take part in detailed assessments focused on risk factors of obesity. Of these, 1,707 had singleton births. Data were collected from the families during pregnancy, at birth and when the infant was aged 6, 12, 18, 24 and 36 months. Approximately half of the mothers (n=933) are of South Asian ethnicity; of which, just under half were born in the UK. Prevalence of obesity in BiB1000 is similar to the full BiB cohort and to UK national averages. In addition to pre-specified hypothesised targets for obesity prevention, (e.g. parental feeding styles, diet and activity), BiB1000 is exploring qualitative determinants of behaviours and other exposures with a lesser evidence base (e.g. food environments, sleep, parenting practices). These data will enable a rich understanding of the behaviours and their determinants in order to inform the development of a culturally-relevant, childhood obesity prevention intervention.

Keywords: birth cohort, ethnicity, childhood, obesity, prevention, South Asian

Background

Born in Bradford (BiB) is a multi-ethnic, birth cohort study (based in Bradford, a northern city in the UK) aiming to examine environmental, psychological and genetic factors that impact on maternal and child health and wellbeing (Raynor and Born in Bradford Collaborative Group, 2008; Wright et al., 2012). BiB1000 is a sub-sample of this cohort which is specifically examining the determinants of childhood obesity, in order to aid development of a tailored prevention intervention, by recruiting women during pregnancy and following them up until the infant is aged 3 years.

There is a disappointing lack of evidence for effective childhood obesity prevention interventions (Summerbell et al., 2005). Data indicate that origins of obesity begin in early childhood and that children of South Asian origin are at particular risk of excess adiposity (Saxena, Ambler, Cole, & Majeed, 2004). There are gaps in our knowledge about the impact of important exposures in pregnancy and early life. Maternal gestational weight gain and gestational glucose metabolism, together with greater birth weight and rapid postnatal growth, are all associated with later obesity, though it is unclear if these are driven by causal mechanisms as currently hypothesized (Reilly et al., 2005). Infant weight gain is consistently associated with subsequent risk of childhood, adolescent and adult obesity and this risk is particularly high for infants with very rapid weight gain (>1.33 SDS, or two centile band crossings) (Baird et al., 2005). Parental obesity is associated with childhood obesity (Whitaker, Wright, Pepe, Seidel, & Dietz, 1997) and although the mechanisms are unclear, they are likely to be influenced by an interaction between genes and the environment. Further, the prevalence of obesity in children born to less educated parents is greater than more educated counterparts (Wang & Beydoun, 2007).

However, it is the social, behavioural and environmental influences on childhood obesity that offer most potential for modification of obesity. A review of systematic reviews of early determinants of obesity identified the following factors associated with an increased risk of childhood obesity: maternal smoking, short sleep duration, less than 30 minutes of daily physical activity, consumption of sugar-sweetened drinks, screen-viewing, and parental feeding practices (Monasta, 2010). Importantly, however, the evidence of causality is not clear, especially in multi-ethnic populations.

BiB1000 aims to enable a deep and extensive understanding of the predictors and influences of health-related behaviours in a multi-ethnic sample, in order to develop a feasible and appropriate culturally-specific childhood obesity prevention intervention. The aim of this paper is to profile the BiB1000 cohort by providing an overview of the BiB1000 study methodology and describing maternal and infant demographic, obstetric and anthropometric characteristics.

Methods and Procedures

Study recruitment

The BiB1000 population were not selectively sampled. Instead, all mothers recruited to the full Born in Bradford study between August 2008 and March 2009, who had completed the baseline questionnaire, were approached to take part in BiB1000 during their routine 26-28 week glucose tolerance test. A sample size of 1080 was calculated based on the statistical ability to detect a difference in infant growth of 0.67 z-scores (one centile band) in weight at age over 1 year, and allowed for a 5% annual attrition. However, once recruitment had begun (and was highly successful), the team decided to oversample the population by up to 70% to optimise the amount of data that were available across all assessments. Ethical approval was obtained from the Bradford Research Ethics Committee and all participants provided written informed consent prior to inclusion in the research.

Data collection

Trained bilingual study administrators collected information from mothers in participants' homes, hospital-based clinics and in local Children's Centres. Anthropometric measurements were taken and structured questionnaires were self-completed. Routinely collected data were extracted from the maternity IT system (eClipse) and the Child Health system in Bradford and Airedale Primary Care Trust. A summary of all measures taken at each assessment is shown in Table 1. In addition, all data collection forms are freely available to download at http://www.borninbradford.nhs.uk/research_documents.htm.

Table 1. BiB1000 data collection measures (with references for each where appropriate)

	Booking (10-14 wks)	**Baseline** (26-28 wks)	**Birth**[2]	**6 mths**	**12 mths**	**18 mths**	**2 yrs**	**3 yrs**
ANTHROPOMETRY								
Mother height[1]		●						
Mother weight	●	●		●	●	●	●	●
Mother arm circumference		●						
Mother triceps		●						
Infant weight			●	●	●	●	●	●
Infant length				●	●	●	●	●
Infant head circumference			●	●	●	●	●	●
Infant abdominal circumference			●	●	●	●	●	●
Infant mid-arm circumference			●					
Infant skin folds (triceps, sub- scapular, thigh)			● (triceps and sub-scapular)	●	●	●	●	●
MATERNAL CHARACTERISTICS & OUTCOMES								
Age of menarche		●						
Previous births (stillbirths/deaths included)		Hospital records						
Obstetric History[3]	Extracted by hand from medical notes							
Glucose tolerance, lipids, insulin, Vitamin D		●						

	Booking (10-14 wks)	**Baseline** (26-28 wks)	**Birth**[2]	**6 mths**	**12 mths**	**18 mths**	**2 yrs**	**3 yrs**
DEMOGRAPHICS								
Residence type (National Centre for Social Research and University College London. Department of Epidemiology and Public Health, 2010)		●						
Household structure (National Centre for Social Research and University College London. Department of Epidemiology and Public Health, 2010)				●	●		●	
Measures of poverty (Willitts, 2006)[4]		●						
Educational status (Office for National Statistics, 2005)[5]		●						
Employment status ((National Centre for Social Research and University College London, 2003) Module J & NS-SEC[6])		●		●	●		●	
Marital status and cohabitation (Office for National Statistics, 2005)		●		●	●		●	
Ethnic origin (Office for National Statistics, 2005)		●						
Family tree (family relationships)		●						
Migration history (Office for National Statistics, 2005)		●						
Childcare arrangements (Australian Institute of Family Studies, 2007)				●	●	●	●	
Food Security (Bickel, Nord, Price, Hamilton, & Cook, 2000)					●			
HEALTH RELATED BEHAVIOURS								
General health (National Centre for Social Research and University College London, 2003)				●	●	●	●	●
Alcohol use (National Centre for Social Research and University College London. Department of Epidemiology and Public Health, 2010)		●						

	Booking (10-14 wks)	**Baseline** (26-28 wks)	**Birth**[2]	**6 mths**	**12 mths**	**18 mths**	**2 yrs**	**3 yrs**
Smoking behaviour(National Centre for Social Research and University College London, 2003)		●		●			●	
Consumption of caffeinated drinks/water/bread products*		●						
Use of supplements/vitamins*		●						
Breast Feeding (Australian Institute of Family Studies, 2003) (Section B)				●	●	●	●	●
FFQ Parent (SFFFQ[7]: Cade et al., unpublished)		●		●		●		●
Eating habits*				●				
Parent physical activity(Australian Institute of Health and Welfare (AIHW), 2003)		●		●		●		●
Infant Diet (Marriott, et al., 2008; Sheehy et al., 2008)					●	●		●
Mother screen (Wareham et al., 2003)				●	●	●	●	●
Child eating with others*							●	
Infant screen time (Wareham et al., 2003)				●	●	●	●	●
Sleep questions*				●	●	●	●	●
Sleep diary*						●		●
Child physical activity*							●	●
Caregivers feeding styles (Hughes, Power, Orlet Fisher, Mueller, & Nicklas, 2005)					●		●	
Parenting Practices (PPQ)(Australian Institute of Family Studies, 2003) (Section F)				●			●	
PSYCHOLOGICAL WELL-BEING								
General Health (GHQ28) (Boardman, 1987)		●		●		●		
Growth perception (Stunkard, Sorensen, & Schulsinger, 1983)				●			●	

	Booking (10-14 wks)	**Baseline** (26-28 wks)	**Birth**[2]	**6 mths**	**12 mths**	**18 mths**	**2 yrs**	**3 yrs**
Maternal mental health (Kessler, Andrews, & Colpe, 2002)					●		●	
Depression screening (National Institute of health and Clinical Excellence (NICE), 2007)					●			
ENVIRONMENTAL INDICES								
Food outlet mapping		●						
Home food availability inventory (researcher conducted)*						●		●
Foods in the home checklist*						●		●
OTHER								
Childhood Illness (Golding, Pembrey, Jones, & ALSPAC Study Team, 2001) (My Daughter Questionnaire)				●		●	●	●
Infant Characteristics (ICQ)(Bates, Freeland, and Lounsbury, 1979)				●				

Notes. [1]*Booking BMI calculated using measured height at baseline and measured weight at booking;* [2] *Includes measures taken at birth (i.e. infant weight) and prior to discharge by paediatrician;* [3] *Includes gestational diabetes, parity, diabetes, pre-eclampsia, hypertension, delivery, adverse outcomes;* [4] *Includes material deprivation, subjective poverty, benefits received, financial coping;* [5]*Equivalent Pakistan qualifications provided by the School of Lifelong Learning and Development (University of Bradford) based on standard classifications used regarding HEFCE admissions/credits procedures* [6]*NS-SEC National Statistics Socio Economic Classification;* [7] *FFQ=Food Frequency Questionnaire, SFFFQ= Short Form Food Frequency Questionnaire; * Newly developed for BiB1000.*

Anthropometry: Infant weight was measured at birth by midwives, and infant head circumference, mid upper arm circumference, abdominal circumference, sub-scapular skin fold and triceps skin fold were taken within the first 24 hours following delivery by paediatricians and midwives who were trained in measurement techniques according to written guidelines. Circumference measurements were taken using Lasso-o tapes (Harlow Printing Ltd South Shields, UK). Head circumference was measured at the most anterior part of the head (frontal eminence) and the most posterior part of the head (maximal head circumference). Abdominal circumference was taken at the umbilicus.

Postnatal measures of infant weight, length, head circumference and abdominal circumference were collected as part of routine practice by health visiting teams. These were supplemented by additional skinfold measures taken by specially trained BiB1000 study health workers at 6 months (actual range 4.9 to 9.4 months), 12 months (actual range 10.7-18.3 months), 18 months (actual range 15.2-22.9), 2 years (actual range 23.4-28.5 months) and 3 years (35.4 - 40.6), obtained using Tanner/Whitehouse Calipers (Holtain Ltd, UK) on the left side of the body. Infant weight was measured using Seca baby scales (Harlow Healthcare Ltd, London, UK). Length was measured using a standard issue neonatometer (Harlow Health Care, London, UK). Maternal weight was taken at all assessments using Seca 2in1 scales (Harlow Healthcare Ltd, London, UK). Maternal height was measured at baseline, and booking BMI (Body Mass Index (BMI kg/m^2)) was derived from antenatal booking weight (at ~12 weeks pregnancy) and baseline height. Reliability testing of the growth data measurements by BiB health workers indicated good quality control for inter- and intra-observer technical error of measurements (r=0.96-1.00) (Johnsonet al., 2009).

Pregnancy and birth outcomes*:* Mothers' date of birth, parity, date of delivery, mode of delivery, birth weight, infant gender and gestational age were collected from eClipse. Medical and obstetric data were extracted by hand from medical records. Gestational diabetes was diagnosed by glucose tolerance test based on WHO (WHO/NCD/NCS/99.2) thresholds for impaired glucose tolerance or impaired fasting glucose, (i.e. fasting plasma glucose ≥6.0mmol/l and/or 2-hr post-challenge glucose ≥ 7.8mmol/l) at 26 weeks.

Demographics*:* The majority of demographic data were obtained using structured self-reported questionnaires (Table 1). Items were generated and modified from the Millennium Cohort Study (National Centre for Social Research and University College London, 2003), Growing Up in Australia (Australian Institute of Family Studies, 2003), the 2001 Census (Office for National Statistics, 2005) and the European Prospective Investigation into Cancer and Nutrition (EPIC) (Riboli et al., 2002) questionnaires. Ethnicity and migration history was ascertained using items on the age that participants moved to the UK, plus questions on the country, cultural background and town of birth, (and name of Biraderi or other caste system) of each parent and grandparent.

Behavioural measures*:* Behavioural measures were collected as part of the structured administered questionnaires, including; feeding style and practices, parental and infant diet, mental health, parental and infant activity, sleep patterns, home food availability, parenting practices and other health behaviours (e.g. smoking, alcohol consumption). Validated questionnaires were used where available, although appropriate ethnic modifications were made and tested using expertise within Bradford. Assessors were trained to collect data in a sensitive manner, including ensuring privacy were necessary and using female assessors of South Asian origin for interviewing other South Asian women.

All questionnaires were translated into Urdu (the national language of Pakistan) and Mirpuri, as the Mirpuri population are the single largest sub-group of the Pakistani populations in Bradford. Transliteration involved translation, back-translation and several rounds of piloting by bilingual and monolingual groups in collaboration with local experts in Bradford (Bradford Talking Media). Since Mirpuri does not have a written form, transliterations were made available for administration by bilingual study administrators.

Qualitative and objective data collection: Qualitative and objective methodologies were employed in sub-samples to explore the lifestyles, behaviours and environments in this multi-ethnic population, including (a) food outlet mapping (b) home food availability inventories (c) a mealtime observation study of maternal feeding styles and (d) interviews with families.

(a) The association between food outlet location, deprivation, weight status and ethnicity was analysed using individual level data using geographic information systems (GIS) methodology. The study area included five inner city wards in Bradford Metropolitan District Council (BMDC) with a range of ethnic population mix (1.2%-63.8% South Asian). A radius of one ward in each direction was included to minimise edge effects in the analysis. Food outlet details were obtained from the BMDC's list of food outlets and the Bradford Yellow Pages (index of local businesses). Data were validated by physical 'ground truthing' a sample of the study area in a random selection of output areas (OAs) to ensure both that a food outlet existed where the list expected one, and whether there were any additional food outlets over and above what was expected.

(b) Home food availability inventories were conducted in an opportunistic sample of 100 participants during 18 month assessments. Researchers measured *all* foods from *all* storage areas within the categories of fruit, vegetables, snack foods and sugar-sweetened beverages. These categories were chosen because; (1) they are often the target of obesity interventions; (2) there is some evidence that their intake is related to obesity in children (Nicklas, Yang, Baranowski, Zakeri, & Berenson, 2003); (3) and/or early literature indicates a relationship between availability in the home and either diet or obesity (Byrd-Bredbenner & Maurer Abbot, 2009). Data collection method, staff training and quality assurance were conducted using a standardised protocol using well-established methodologies from previous research (Bryant et al., 2008).

(c) Ethnic and weight status differences in meal structures and mealtime interactions were investigated in an observational study when infants were aged 18 and 27 months. A daytime meal of 38 mother-child dyads was video-recorded. Mothers were selected according to their ethnicity (non-South Asian, South Asian) or weight status (healthy weight, obese). Mealtimes were coded using the Mealtime Observation Schedule (Sanders & Le Grice, 1989). The schedule measured positive and negative parent behaviours, and positive and aversive child behaviours.

(d) A purposive sample (by ethnicity, parity, cohabitation with extended family) of 14 mothers (and husbands/partners where possible) were interviewed in their homes when the baby was 4 months old to ascertain key cultural and social differences in feeding practices. Researchers sought to elicit the interviewee's responses in three areas: 1) The interviewee belief in what constitutes a healthy diet for their baby; 2) who in the family makes choices about diet for the baby, (are these choices influenced by advice from professionals such as Health Visitor or Children's Centre staff); and 3) interviewee self-efficacy to make decisions regarding the baby's feeding. Interviews were recorded and subsequently transcribed and translated.

Dealing with missing data

Missingness of variables such as obstetric data was dealt with by 'back-fill' (i.e. review of case notes by a physician and entry if available). For consideration of future analysis, missingness will be dealt with on a case-by-case basis and will employ imputation techniques as appropriate. Variability in missingness methodology is based on the study exposure and outcomes, and the amount of missingness in each dataset (i.e. <2% missing will likely not be imputed).

Results

Of 1,916 eligible women, 1,735 agreed to take part in the study. Of these, 28 mothers gave birth to twins. Descriptive statistics are provided here for singleton births only (n=1,707). 77%, 75%, 74%, 70% and 70% have been followed-up at 6, 12, 18, 24 and 36 month assessments respectively. 47% of participants completed all assessments to date, with 17% formally withdrawn from the research. The greatest impact on attrition rates was the inability to make contact with participants in order to book appointments, despite adoption of multiple efforts to contact participants. For example, calls were placed during evenings and weekends; participant details were searched on primary care electronic records; health visitors were approached to verify contact details; letters were mailed; and opportunistic visits were made to family homes.

Demographic information is shown in Table 2 according to maternal ethnicity. BiB1000 characteristics are similar to that of the full BiB cohort (Wright et al., 2012); with a similar distribution of age, marital status and parity. Demographic differences by ethnicity within BiB1000 were also observed, with White British mothers tending to be younger, educated to a lower level, less likely to be married or cohabiting, and having fewer children than other ethnic groups. Gestational diabetes was greatest in South Asian ethnicities, with 10.5% prevalence in Pakistani women compared to 5.5% in White British. Gestational age, proportion of pre-term infants, mode of birth and the proportion of stillbirths were similar between ethnic groups. Over 35% of White British women reported smoking during pregnancy, compared to 4% and 3% of Pakistani and 'Other' South Asian ethnicities.

Table 3 shows descriptive data for maternal and infant anthropometry by ethnic group. A little over 40% of BiB1000 women were of normal weight (BMI 18.5-24.9) at booking, according to WHO criteria (World Health Organisation, 1998), with 25.6% and 18.2% of the sample overweight (BMI 25-29.9) or obese (BMI ≥30) respectively. By ethnicity, BMI was greatest in White British women compared to other groups. Birth weight of infants of Pakistani and Other South Asian mothers was on average approximately 200 grams lighter than in other ethnicities. These women were also more likely to have a lower birth weight infant (<2.5kg) and less likely to have a macrosomic infant (>4kg) than White British women, though differences in the proportion of women having an infant less than 2.5kg was not statistically significant. Circumference and skinfold thickness also differed between ethnic groups, indicating some ethnic differences in body fat distribution that warrant further investigation. However, it is important to note that up to 20.4% of circumference and skinfold measurements are missing. These measurements, normally taken in hospital at birth, were not always possible due to early discharge or competing demands (e.g. neonatal hearing screening; first examination; midwife checks; photographer etc.).

Table 2. Demographic and obstetric characteristics by maternal ethnicity

		White British (n=652)		Pakistani (n=808)		Other S. Asian[1] (n=125)		Other[2] (n=122)		All (n=1707)		p-value[4]
		N	%	N	%	N	%	N	%	N	%	
Maternal age	Mean (95% CI)	26.4	(25.9, 26.9)	27.5	(27.1, 27.8)	28.5	(27.6, 29.4)	27.2	(26.1, 28.3)	27.1	(26.9, 27.4)	0.0001
	<20 years	85	13.0	25	3.1	2	1.6	12	9.8	124	7.3	<0.0001
	20-24 years	187	26.7	230	28.5	26	20.8	29	23.8	472	27.7	
	25-29 years	190	29.1	295	36.5	43	34.4	41	33.6	569	33.3	
	30-34 years	108	16.6	164	20.3	37	29.6	21	17.2	330	19.3	
	≥35	82	12.6	94	11.6	17	13.6	19	15.6	212	12.4	
Maternal education	No qualifications	133	20.4	208	25.7	18	14.4	16	13.1	375	22.0	<0.0001
	School	242	37.1	261	32.3	28	22.4	25	20.5	556	32.6	
	Further	100	15.3	97	12.0	18	14.4	18	14.8	233	13.7	
	Higher	119	18.3	202	25.0	47	37.6	36	29.5	404	23.7	
	Foreign unknown	48	7.4	23	2.9	10	8.0	22	18.0	103	6.0	
	Missing	10	1.5	17	2.1	4	3.2	5	4.1	36	2.1	
Marital status	Married/cohabiting	473	72.6	767	94.9	120	96.0	95	77.9	1455	85.2	<0.0001
	Single	178	27.3	39	4.8	5	4.0	27	22.1	249	14.6	
	Missing	1	0.2	2	0.3	0		0		3	0.2	
Father's emp't status	Employed, non-manual	319	48.9	277	34.3	59	47.2	58	47.5	713	41.8	<0.0001
	Employed, manual	153	23.5	311	38.5	34	27.2	27	22.1	525	30.8	
	Self-employed	56	8.6	116	14.4	18	16.4	6	4.9	196	11.5	
	Student	15	2.3	12	1.5	4	3.2	12	9.8	43	2.5	
	Unemployed	91	14.0	79	9.8	10	8.0	17	13.9	197	11.5	

		White British (n=652)		**Pakistani (n=808)**		**Other S. Asian[1] (n=125)**		**Other[2] (n=122)**		**All (n=1707)**		**p-value[4]**
		N	%	N	%	N	%	N	%	N	%	
	Unknown	13	2.0	11	1.4	0		1	0.8	25	1.5	
	Missing	5	0.8	2	0.3	0		1	0.8	8	0.5	
Parity	0	303	46.5	244	30.2	48	38.4	58	47.5	653	38.3	<0.0001
	1	191	29.3	234	29.0	40	32.0	37	30.3	502	29.4	
	2	81	12.4	145	18.0	24	19.2	11	9.0	261	15.3	
	≥3	61	9.4	168	20.8	12	9.6	8	6.7	249	14.6	
	Missing	16	2.5	17	2.1	1	0.8	8	6.6	42	2.5	
GAD (weeks)[3]	Mean (95% CI)	39.1	(39.0, 39.3)	39.2	(39.0, 39.3)	39.2	(39.0, 39.4)	39.3	(39.0, 39.6)	39.2	(39.1, 39.2)	0.861
Preterm infants	(<37 weeks)	41	6.3	39	4.8	4	3.2	5	4.1	89	5.2	0.377
Mode of birth	Vaginal birth	507	77.8	646	80.0	97	77.6	88	72.1	1338	78.4	0.358
	Caesarean Section	138	21.2	155	19.2	27	21.6	31	25.4	351	20.6	
	Missing	7	1.1	7	0.9	1	0.8	3	2.5	18	1.1	
Still birth	Yes	0		0		0		0		0		
Pre-eclampsia	Yes	9	1.4	15	1.9	1	0.8	0		36	1.5	0.378
Gestational diabetes	Yes	36	5.5	85	10.5	15	12.0	2	1.6	138	8.1	<0.0001
Maternal smoking during pregnancy	Yes	230	35.3	32	4.0	4	3.2	25	20.5	291	17.1	<0.0001

Notes. [1] Other S. Asian = Indian (n=73); Bangladeshi (n=42); Asian other (n=10); [2] Other = White other (n=38); Black (n=34); Mixed White and Black (n=14); Mixed White and South Asian (n=8); Other (n=28). [3] Gestational age at delivery; [4] Continuous data were compared using ANOVA and categorical data using Chi-Squared tests.

Table 3. Maternal and infant anthropometry[1]

		White British (n=652)		Pakistani (N=808)		Other SA (n=125)		Other (n=122)		All (n=1707)		p-value[2]
		N	%	N	%	N	%	N	%	N	%	
MATERNAL												
Booking BMI	Mean (95% CI)	26.8	(26.3, 27.3)	25.3	(24.9, 25.6)	25.5	(24.4, 26.6)	24.8	(23.9, 25.7)	25.8	(25.6, 26.1)	<0.0001
BMI category[1]	Underweight	18	2.8	45	5.6	5	4.0	6	5.0	74	4.3	0.002
	Normal weight	254	39.0	355	43.9	60	48.0	54	44.0	723	42.4	
	Overweight	161	24.7	218	27.0	27	21.6	31	25.6	437	25.6	
	Obese	151	23.2	124	15.4	22	17.6	14	11.6	311	18.2	
	Missing[2]	68	10.4	66	8.2	11	8.8	17	14.1	162	9.5	
INFANT												
Birth weight (g)	Mean (95% CI)	3318	(3272, 3363)	3129	(3092, 3165)	3094	(3010, 3178)	3280	(3198, 3362)	3209	(3183, 3235)	<0.0001
	Missing	7	1.1	7	0.9	1	0.8	3	2.5	18	1.1	
Low birth weight	<2.5kg	46	7.1	72	8.9	9	7.2	5	4.1	132	7.7	0.2450
	Missing	7	1.1	7	0.9	1	0.8	3	2.5	18	1.1	

		White British (n=652)		Pakistani (N=808)		Other SA (n=125)		Other (n=122)		All (n=1707)		p-value[2]
		N	%	N	%	N	%	N	%	N	%	
Large birth weight	>4kg	70	10.7	35	4.3	2	1.6	8	6.6	115	6.7	<0.0001
	Missing	7	1.1	7	0.9	1	0.8	3	2.5	18	1.1	
Head	Mean (95% CI)	34.40	(34.27, 34.53)	34.07	(33.98, 34.18)	34.12	(33.88, 34.36)	34.46	(34.21, 34.72)	34.23	(34.16, 34.30)	0.0002
circumference (cm)	Missing	23	3.5	28	3.5	5	4.0	6	4.9	62	3.6	
Abdominal	Mean (95% CI)	31.74	(31.55, 31.93)	30.72	(30.54, 30.90)	30.60	(30.15, 31.05)	31.56	(31.13, 32.0)	31.16	(31.04, 31.28)	<0.0001
circumference (cm)	Missing	55	8.4	65	8.0	11	8.8	8	6.6	139	8.1	
Mid-arm	Mean (95% CI)	10.87	(10.78, 10.95)	10.63	(10.56, 10.70)	10.62	(10.44, 10.80)	10.77	(10.61, 10.94)	10.73	(10.68, 10.78)	0.0002
circumference (cm)	Missing	48	7.4	61	7.6	11	8.8	8	6.6	128	7.5	
Subscapular skinfold	Mean (95% CI)	4.84	(4.74, 4.94)	4.67	(4.58, 4.75)	4.61	(4.42, 4.81)	4.78	(4.55, 5.01)	4.73	(4.68, 4.79)	0.0364
(mm)	Missing	133	20.4	130	16.1	18	14.4	21	17.2	302	17.7	
Triceps skinfold (mm)	Mean (95% CI)	5.35	(5.25, 5.45)	5.16	(5.08, 5.24)	4.96	(4.78, 5.14)	5.26	(5.04, 5.48)	5.22	(5.16, 5.28)	0.0018
	Missing	133	20.4	130	16.1	18	14.4	21	17.2	302	17.7	

Notes. [1] Other S. Asian = Indian (n=73); Bangladeshi (n=42); Asian other (n=10); [2]Other = White other (n=38); Black (n=34); Mixed White and Black (n=14); Mixed White and South Asian (n=8); Other (n=28).[3]Gestational age at delivery; [4] Continuous data were compared using ANOVA and categorical data using Chi-Squared tests.

Discussion

This Research Note describes the profile and characteristics of a multi-ethnic nested birth cohort in Bradford, which aims to examine the patterns and aetiology of childhood obesity as a needs assessment in order to develop a tailored obesity prevention intervention. In addition to BiB measurements (Raynor and Born in Bradford Collaborative Group, 2008, Wright et al., 2012) further quantitative and qualitative data were collected on the BiB1000 participants to explore risk factors for, and consequences of obesity. A unique quality of this cohort is its ethnic composition, with approximately 50% being of South Asian origin. This new study will have the advantage of data on food availability and families' feeding styles. Antenatal data are freely available and have the capacity of combining with other similar cohorts to strengthen our understanding of early life and obesity outcomes. A registry of such cohorts can be found at http://www.birthcohorts.net. Postnatal data will also be available for open access upon completion of the BiB1000 study in mid-2014. Anyone wishing to access the BiB1000 data are advised to contact the BiB team via completion of an 'Application proforma' from the website. (http://www.borninbradford.nhs.uk/research_proformas.htm). The website provides full details about how to apply for data and further details for interested collaborators.

Trends shown in BiB1000 background characteristics data indicate ethnic inequalities in health that are consistent with existing literature. For example, infants born to women of Pakistani origin in BiB1000 are significantly lighter than babies born to White women. Previous research shows similar patterns (Leon & Moser, 2012; Margetts, Mohd Yusof, Al Dallal, & Jackson, 2002) and this is evident across generations moving to the UK, indicating a complex, and largely unknown, relationship between environmental influences (Kelly, et al., 2008) and biological mechanisms (Margetts, Mohd Yusof, Al Dallal, and Jackson, 2002). A continuous positive relationship (up to macrosomic weight) has been observed between birth weight and health (short and long term), with lower birth weight babies at greatest risk, including perinatal mortality (Ashworth, 1998). Consistent with existing literature (Ali & Dornhorst, 2011; Dornhorst et al., 1992), BiB1000 women of South Asian origin were more than twice as likely to have gestational diabetes compared to White women, despite a lower prevalence of obesity and a significantly lower number of South Asian women smoking during pregnancy.

The proportion of BiB1000 women categorised as overweight (26%) was similar to national prevalence data for the UK (27%) (Food Standards Agency Office, 2010). Rates of obesity in BiB1000 (18.2%) were lower than national rates (29%), but were comparable with age-specific rates in England (21.3%) in women aged 25-34 years (Health Survey for England 2010, 2012), suggesting that the higher national rates may be partially explained by the inclusion of women aged up to 65 years. Furthermore, South Asian women in BiB1000 had a lower prevalence of obesity compared to White British women. Thus, given the high proportion of South Asian women in this cohort compared to the UK generally, this may also partly explain the lower prevalence of obesity overall. BiB1000 weight status is based on data collected at week 12 of pregnancy, at which point there may be some (though negligible) impact on maternal weight. When comparing prevalence to other pregnant samples (where measurement was taken within the first trimester) obesity prevalence is similar. For example, in a recent national UK sample of 98,511 predominantly White British pregnant women, the prevalence of obesity was 15.6% (Heslehurst, Rankin, Wilkinson, & Summerbell, 2010). Clinically, the classification of obesity within pregnant women is often described as a BMI of 35kg/m^2 or greater, since this is the BMI beyond which referrals are often made. Using this cut-off, national data suggests approximately 5% of women are obese in the first trimester of pregnancy ((CMACE), 2010). By comparison, 7.3% of BiB1000 women had a BMI greater than 35.

Conclusion

While there is evidence supporting individual predictors of childhood obesity (e.g. parental obesity (Whitaker, Wright, Pepe, Seidel, & Dietz, 1997)) little is currently known about the role of early modifiable exposures that are associated with early childhood obesity and how they may interact to influence risk of obesity. This is especially apparent for South Asian populations living in the UK. Existing obesity programmes are generally based on best practice and often fail to consider culturally-specific influences of healthy weight behaviours. BiB1000 offers an opportunity to collect rich longitudinal data to create a picture of a child's environment over time, which will help to understand and explain the influence of this environment on behaviours, and will inform the development of a feasible culturally-specific intervention to prevent childhood obesity.

Acknowledgements

We are grateful to all the families who took part in this study, to the midwives for their help in recruiting them, the paediatricians and health visitors and to the Born in Bradford team which included interviewers, data managers, laboratory staff, clerical workers, research scientists, volunteers and managers.

This article presents independent research commissioned by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research funding scheme (RP-PG-0407-10044). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

References

Ali, S. and Dornhorst, A. (2011). Diabetes in pregnancy: health risks and management. *Postgraduate Medical Journal*. doi: 10.1136/pgmj.2010.109157

Ashworth, A. (1998). Effects of intrauterine growth retardation on mortality and morbidity in infants and young children. *European Journal of Clinical Nutrition, 52*, S34-41.

Australian Institute of Family Studies. (2003). Growing up in Australia: The Longitudinal Study of Australian Children: Wave I: LSAC Discussion Paper No. 2 Retrieved February 2012, from http://www.aifs.gov.au/growingup/pubs/discussion/dp2/index.html

Australian Institute of Family Studies. (2007). Growing up in Australia: The Longitudinal Study of Australian Children: Wave II: LSAC Discussion Paper No. 5 Retrieved February 2012 http://www.growingupinaustralia.gov.au/pubs/discussion/dp5/index.htmlAustralian Institute of Health and Welfare (AIHW). (2003). The Active Australia Survey: A guide and manual for implementation, analysis and reporting. Canberra. Media and Publishing, AIHW.

Baird, J., Fisher, D., Lucas, P., Kleijnen, J., Roberts, H., and Law, C. (2005). Being big or growing fast: systematic review of size and growth in infancy and later obesity. *British Medical Journal, 331*, 929.

Bates, J. E., Freeland, C. A. B., and Lounsbury, M. L. (1979). Measurement of Infant Difficultness. *Child Development, 50*, 794-803.

Bickel, G., Nord, M., Price, C., Hamilton, W., and Cook, J. (2000). Measuring Food Security in the United States: Guide to Measuring Household Food Security. In U. S. D. o. A. (USDA) (Ed.), (Revised Edition of Report Number 3 ed.).

Boardman, A. P. (1987). The General Health Questionnaire and the detection of emotional disorder by General Practitioners. A replicated study. *The British Journal of Psychiatry, 151*, 373-381. doi: 10.1192/bjp.151.3.373

Bryant, M., Ward, D., Hales, D., Vaughn, A., Tabak, R., and Stevens, J. (2008). Reliability and validity of the Healthy Home Survey: A tool to measure factors within homes hypothesized to relate to overweight in children. *International Journal of Behavioral Nutrition and Physical Activity, 5*, 23-33.

Byrd-Bredbenner, C., and Maurer Abbot, J. (2009). Differences in Food Supplies of U.S. Households with and without Overweight Individuals. *Appetite, 52*, 479-484.

CMACE (Centre for Maternal and Child Enquiries) . (2010). *Maternal obesity in the UK: Findings from a national project*. London: CMACE.Dornhorst, A., Paterson, C. M., Nicholls, J. S. D., Wadsworth, J., Chiu, D. C., Elkeles, R. S., and Beard, R. W. (1992). High Prevalence of Gestational Diabetes in Women from Ethnic Minority Groups. *Diabetic Medicine, 9*, 820-825. doi: 10.1111/j.1464-5491.1992.tb01900.x

Food Standards Agency Office, U. (2010). National Diet Nutrition Survey: headline results from year 1(2008/2009).

Golding, J., Pembrey, M., Jones, R., and ALSPAC Study Team. (2001). ALSPAC -

The Avon Longitudinal Study of Parents and Children. I. Study methodology. *Paediatric and Perinatal Epidemiology, 15*, 74-87.

Health Survey for England 2010, T. N. I. C., Lifestyles Statistics. (2012). Statistics on obesity, physical activity and diet.
Heslehurst, N., Rankin, J., Wilkinson, J. R., and Summerbell, C. D. (2010). A nationally representative study of maternal obesity in England, UK: trends in incidence and demographic inequalities in 619 323 births, 1989-2007. *International Journal of Obesity, 34*, 420-428.
Hughes, S. O., Power, T. G., Orlet Fisher, J., Mueller, S., and Nicklas, T. A. (2005). Revisiting a neglected construct: parenting styles in a child-feeding context. *Appetite, 44*, 83-92. doi: 10.1016/j.appet.2004.08.007
Johnson, W., Cameron, N., Dickson, P., Emsley, S., Raynor, P., Seymour, C., and Wright, J. (2009). The reliability of routine anthropometric data collected by health workers: A cross-sectional study. *International Journal of Nursing Studies, 46*, 310-316. doi: DOI: 10.1016/j.ijnurstu.2008.10.003
Kelly, Y., Panico, L., Bartley, M., Marmot, M., Nazroo, J., and Sacker, A. (2008). Why does birthweight vary among ethnic groups in the UK? Findings from the Millennium Cohort Study. *Journal of Public Health 31*, 131-137.
Kessler, R. C., Andrews, G., and Colpe, L. J. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychology Medicine, 32*, 959-976.
Leon, D. A., and Moser, K. A. (2012). Low birth weight persists in South Asian babies born in England and Wales regardless of maternal country of birth. Slow pace of acculturation, physiological constraint or both? Analysis of routine data. *Journal of Epidemiology and Community Health, 66*, 544-551.
Margetts, B. M., Mohd Yusof, S., Al Dallal, Z., and Jackson, A. A. (2002). Persistence of lower birth weight in second generation South Asian babies born in the United Kingdom. *Journal of Epidemiology and Community Health, 56*, 684-687.
Marriott, L. D., Inskip, H. M., Borland, S. E., Godfrey, K. M., Law, C. M., Robinson, S. M., and Southampton Womens Study Group. (2008). What do babies eat? Evaluation of a food frequency questionnaire to assess the diets of infants aged 12 months. *Public Health Nutrition, 12*, 967-972.
Monasta, L. (2010). Early-life determinants of overweight and obesity: a review of systematic reviews. *Obesity Reviews, 11*, 695-708.
National Centre for Social Research and University College London. (2003). Millenium Cohort First Study: CAPI Questionnaire Documentation. London.
National Centre for Social Research and University College London. Department of Epidemiology and Public Health. (2010). Health Survey for England, 2004 (2nd Edition, ed.). Colchester, Essex.
National Institute of Health and Clinical Excellence (NICE). (2007). Antenatal and postnatal mental health Guideline 45.
Nicklas, T. A., Yang, S.-J., Baranowski, T., Zakeri, I., and Berenson, G. (2003). Eating patterns and obesity in children: The Bogalusa Heart Study. *American Journal of Preventive Medicine, 25*, 9-16. doi: Doi: 10.1016/s0749-3797(03)00098-9
Office for National Statistics. (2005). Census 2001: General report for England and Wales (Vol. ISBN 1403987688.).
Raynor, P., and Born in Bradford Collaborative Group. (2008). Born in Bradford, a cohort study of babies born in Bradford, and their parents: Protocol for the recruitment phase. *BMC Public Health, 8*, 327.
Reilly, J. J., Armstrong, J., Dorosty, A. R., Emmett, P. M., Ness, A., Rogers, I., Steer, C., and Sherriff, A. (2005). Early risk factors for obesity in childhood: cohort study. *British Medical Journal, 330*, 1357. doi: doi:1136/bmj.38470.670903.ED
Riboli, E., Hunt, K., Slimani, N., Ferrari, P., Norat, T., Fahey, M., P, Norat, T., Fahey, M., Charrondière, U.R., Hémon, B., Casagrande, C., Vignat, J., Overvad, K., Tjønneland, A., Clavel-Chapelon, F., Thiébaut, A., Wahrendorf, J., Boeing, H., Trichopoulos, D., Trichopoulou, A., Vineis, P., Palli, D., Bueno-De-Mesquita, H.B., Peeters, P.H., Lund, E., Engeset, D., González, C.A., Barricarte, A., Berglund, G., Hallmans, G., Day, N.E., Key, T.J., Kaaks, R., and Saracci, R. (2002). European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutrition, 5*, 1113-1124. doi: doi:10.1079/PHN2002394
Sanders, M. R., and Le Grice, B. (1989). *Mealtime Observation Schedule: An observers manual.* . Unpublished technical manual. Department of Psychiatry, University of Queensland, Herston, Queensland, Australia.
Saxena, S., Ambler, G., Cole, T. J., and Majeed, A. (2004). Ethnic group differences in overweight and obese children and young people in England: cross sectional survey. *Archives of Disease in Childhood, 89*, 30-36.
Sheehy, C., McNeill, G., Masson, L., Craig, L., Macdiarmid, J., Holmes, B., and Nelson, M. (2008). *Survey of sugar intake among children in Scotland.* Commissioned by Food Standards Agency Scotland. Research Project S14029 FSA (Scotland). Scottish Centre for Social Research.
Stunkard, A. J., Sorensen, T. I., and Schulsinger, F. (1983). *Use of the Danish Adoption Register for the Study of Obesity and Thinness*. New York: Raven Press.

Summerbell, C. D., Waters, E., Edmunds, L., Kelly, S., Brown, T., and Campbell, K. J. (2005). *Interventions for preventing obesity in children.* Cochrane Database of Systematic Reviews 10.1002/14651858.CD001871.pub2(3).
Wang, Y., and Beydoun, M. A. (2007). The Obesity Epidemic in the United States—Gender, Age, Socioeconomic, Racial/Ethnic, and Geographic Characteristics: A Systematic Review and Meta-Regression Analysis. *Epidemiologic Reviews, 29*, 6-28. doi: 10.1093/epirev/mxm007
Wareham, N. J., Jakes, R. W., Rennie, K. L., Schuit, J., Mitchell, J., Hennings, S., and Day, N. E. (2003). Validity and repeatability of a simple index derived from the short physical activity questionnaire used in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Public Health Nutrition, 6,* 407-413.
Whitaker, R. C., Wright, J. A., Pepe, M. S., Seidel, K. D., and Dietz, W. H. (1997). Predicting Obesity in Young Adulthood from Childhood and Parental Obesity. *New England Journal of Medicine, 337*, 869-873. doi:10.1056/NEJM199709253371301
Willitts, M. (2006). Measuring child poverty using material deprivation: possible approaches; Department of Work and Pensions Working Paper 28.
World Health Organisation. (1998). Obesity: preventing and managing the global epidemic. Report of a WHO consultation (Vol. WHO/NUT/98.1). Geneva.
Wright, J., Small, N., Raynor, P., Tuffnell, D., Bhopal, R., Cameron, N., and West, J. (2012). Cohort profile: The Born in Bradford multi-ethnic family cohort study. *International Journal of Epidemiology*. doi: 10.1093/ije/dys112

COMMENT AND DEBATE

Social-biological transitions: how does the social become biological?

David Blane
Imperial College London, UK
d.blane@imperial.ac.uk.
Michelle Kelly-Irving
INSERM U1027 Toulouse, France
Angelo d'Errico
University of Torino, Italy
Melanie Bartley
University College London, UK
Scott Montgomery
Örebro University Hospital & Örebro University, Sweden

(Received January 2013 Revised May 2013)

Abstract

The present discussion paper sets forward a model within the life course perspective of how the social becomes biological. The model is intended to provide a framework for thinking about such questions as how does social class get into the molecules, cells and tissues of the body to produce social class differences in life expectancy and cause of death? A categorisation of social exposures and biological processes is suggested; and some principles governing their inter-relations proposed. The paper ends by suggesting two public health applications of this approach.

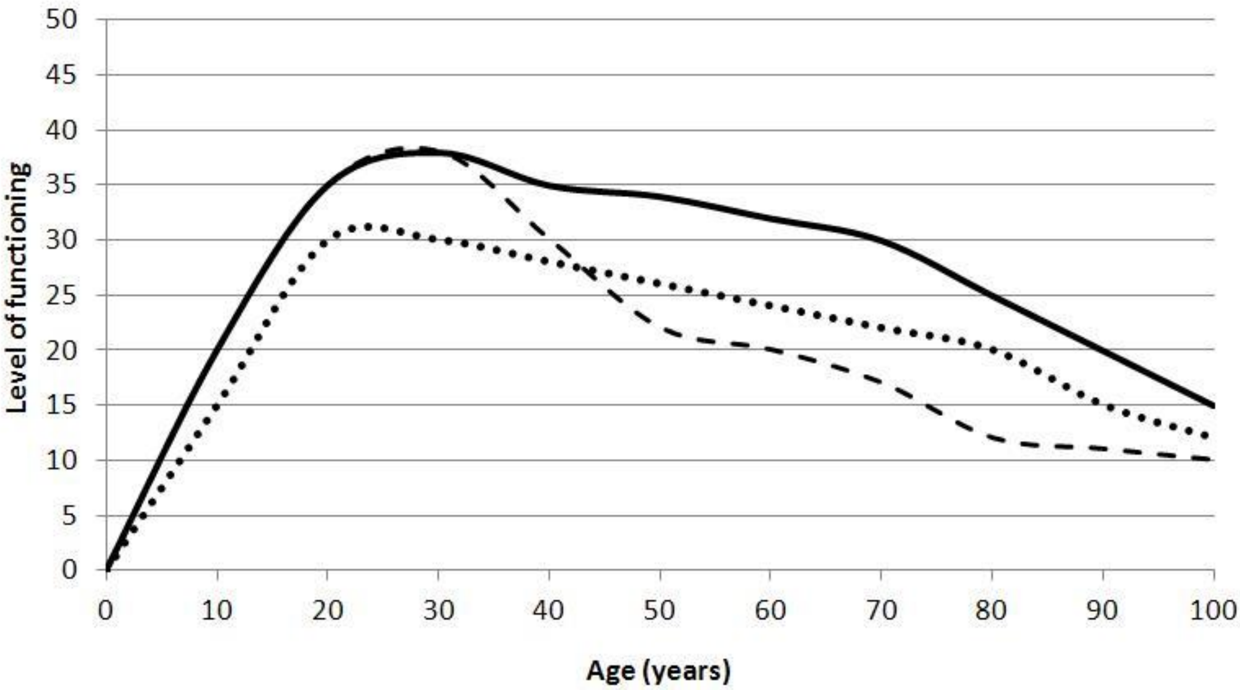
Keywords: Life course; social exposures; biological processes; social-biological transitions; public health.

Introduction

Many published studies report that social circumstances during childhood are associated with later health during adulthood or older ages (among the most distinguished recently are: Johnson & Schoeni, 2011; Maty, Lynch, Raghunathan & Kaplan, 2008; Osler et al., 2003) without specifying and testing the intervening social and biological processes. From the point of view of public health, this omission is serious because potential intervention points may be missed. So, a useful next step is to consider the pathways between the social and the biological and to ask how the social becomes biological, contributing thereby to a growing international literature on the topic (Arbeev et al., 2011; Butterworth, Cherbuin, Sachdev & Anstey, 2012; Eisenberger, Taylor, Gable, Hilmert & Lieberman, 2007; Wolfe, Evans & Seeman, 2012). The present paper aims to suggest an organising framework for the operationalization and testing of pathways from the social to the biological over the life course. The framework is intended to be comprehensive, in order to avoid premature emphasis on any one particular pathway or process; for example, concentrating on psycho-social factors because material hardship is assumed to be rare in rich countries or prioritising health behaviours because these are judged to be realistic and acceptable targets for change.

It is helpful to think about the biological part of social-biological transitions as situated within the context of a generalised version of Strachan and Sheikh's sequential model of life course functioning (Figure 1). The first stage, from conception and early intra-uterine life to late adolescence or early twenties, is a period of growth and development from a single fertilised cell to an adult human. The second stage, starting in early adulthood, is a period of decline from maximum attained growth to loss of function, overt disease and death. Strachan and Sheikh described this model for lung function (Strachan & Sheikh, 2004), but it can be generalised usefully to other aspects of functioning and health. Social exposures during the first stage can influence the proportion of optimum growth attained, for example the stunting of physical height by nutritional or emotion deprivation (Figure 1: dotted line). Social exposures during the second stage can influence the rate at which functioning is lost, for example occupational fumes and dusts damaging bronchiolar surfaces and reducing lung function (Figure 1: dashed line). The social structure often delivers continuity of disadvantage, or advantage, across these two biological stages of the life course to combine sub-optimal growth with accelerated decline; or optimal growth with delayed decline. Such processes, plausibly, contribute to social class differences in longevity, healthy life expectancy and physical fitness at older ages.

Figure 1. Life course growth and decline in functioning: Strachan-Sheikh Model



(Source: Strachan & Sheikh 2004)

The timing of the biological sequence of growth and decline needs to be seen in relation to the standard of living life cycle in which the two dominant social institutions of wage labour and nuclear family interact over the life course among the majority of individuals in industrialised countries to produce two phases of household relative affluence, in early adulthood and late middle age, and three phases of household relative hardship, in childhood, families with dependent children and old age (Falkingham & Hills, 1995; Goldberg, Wheeler & Sydenstriker, 1920; Rowntree, 1902; Townsend, 1979). In Rowntree's words: *A labourer is thus in poverty and therefore underfed a) in childhood – when his constitution is being built up, b) in early middle life - when he should be in his prime, c) in old age* (Rowntree 1902:170). As a result, two of the most important biological stages of life in the Strachan-Sheikh model (child and adolescent growth and development; old age decline and loss of function) coincide with two of the standard of living life cycle's three phases of household relative hardship. Fuel poverty (Webb, Blane & de Vries, 2012) and excess winter mortality (ONS 2012) illustrate the contemporary relevance of the latter to physical health, while the long-term consequences of household stress and social exclusion for child development (Dearing 2008; Gershoff, Aber, Raver & Lennon, 2007) illustrate the former's contemporary relevance to mental health. Social transitions (Bartley, Blane & Montgomery, 1997) also need to be considered in this social-biological context.

JN Morris gave precision to estimates of health-relevant financial hardship, by using the results of 50 years of scientific research world-wide to specify the constituents of a healthy life and to cost these for early 21st century England, what he called: the Minimum Income for Healthy Living – MIHL (Morris, Donkin, Wonderling, Wilkinson & Dowler, 2000; Morris & Deeming 2004; Morris, Wilkinson, Dangour, Deeming & Fletcher, 2007). Morris' estimates are conservative: his young single man is a paragon of near-abstinence, drinking one half-pint of beer per week and using one condom (Morris et al., 2000); while his retired couple do not incur the additional costs which come with the physical or mental disability afflicting some 40 per cent of their peers (Morris et al., 2007). The broad picture is clear. National insurance entitlements, in the form of Jobseeker's Allowance for the young man and state retirement pensions for the elderly couple, amount to only around one-half of their Minimum Income for Healthy Living (MIHL). Statutory alternatives and additions to national insurance entitlements, in the form of the minimum wage for a young man and the means-tested pension premium for the elderly couple, are still several pounds short of their MIHL. Any deviation from near-abstinence at younger ages or the onset of disability at older ages would push the person further below MIHL. And this assault on the person's health occurs every week, plausibly accumulating over time.

While Morris' constituents of a healthy life are intended as universal, his costings of MIHL were specific to England. It is not legitimate to generalise them unthinkingly to other countries, because countries vary in their welfare state regime and, specifically, the extent to which existence is de-commodified. "*De-commodification occurs when a service is rendered as a matter of right, and when a person can maintain a livelihood without reliance on the market*" (Esping-Andersen, 1990:21-22). Welfare state regimes and the extent of de-commodification in a particular country evolve over time (Esping-Andersen, 1996; Esping-Andersen, 1999), providing the international context to social-biological transitions.

Attempts to understand social-biological transitions can be made unnecessarily difficult by the use of terms which blur the distinction between the social and the biological. *Material,* for example, has been used to describe both social exposures derived from the spheres of production or consumption and biological processes triggered by the impact of external matter, whether living or inert. Similarly, *psycho-social* has been used to describe both exposure to environmental stressors and biological processes that are implicated in acute and chronic stress responses, including the hypothalamic-pituitary-adrenal (HPA) axis and adrenal cortical and medullary hormones. The present paper tries for clarity by describing separately the different types of social exposure and the different types of biological process; and then suggesting the principles which govern how the two interact.

Social exposures

Structural: Those aspects of current social organisation which inevitably disadvantage a proportion of the population and affect negatively their health; or confer advantage and benefit health. Such aspects include the distribution of income and wealth within a society, the way it organises the production and distribution of the necessities of life, the quality of its housing stock and how it distributes medical care. The organisation of production affects health via exposure to occupational fumes and dusts, physically arduous work (NRC, 2001) and psychosocial work strain (da Costa & Viera, 2010); and the extent to which these are controlled by health & safety protection. The distribution of income and wealth affects health via the MIHL, the proportion of life spent below MIHL and, possibly of particular importance, the proportion of the period from conception to late adolescence and the proportion of life after retirement spent below

MIHL. The quality of the housing stock affects health via its ability to protect inhabitants against the prevailing climate of their region (Blane, Bartley & Mitchell, 2000; Mitchell, Blane & Bartley, 2002), via exposure to indoor pollutants, nitrogen dioxide, carbon monoxide and allergens associated with asthma (Laquatra, Maxwell & Pierce, 2005) and via crowding's influence on the dose and sequence of infections. In addition, it interacts with income to determine fuel poverty risk, while its proximity to industry and main roads determines air pollution exposure, with its risk of sub-optimal lung development during childhood and adolescence (Gauderman et al., 2004) and chronic lung disease at older ages (Schikowski et al., 2005). The social distribution of medical care, through fee for service or geographic mis-match with morbidity, affects health via access to preventive and therapeutic care (van Doorslaer, Koolman & Jones, 2004).

Behavioural: Habits and behaviours which affect health and are, at least to some extent, subject to individual autonomy, choice and decision-making. Health-relevant behaviours are often part of a culture or sub-culture which shapes its members' behaviour and thus patterns of disease, such as the low rates of coronary heart disease associated with Japanese and Mediterranean diets (Iso, 2011, Willet, 2006) and the low rates of tobacco smoking and lung cancer among Hispanic Americans (CDC, 2010). Health-related behaviours include: (a) tobacco smoking and age at its initiation and cessation; (b) level and intensity of physical exercise, including leisure-time; (c) dietary preferences in relation to sugar, salt, fibre, fresh vegetables and fruit, oily fish, saturated and transfats, slow release carbohydrates, nuts and seeds, alcohol; (d) illness behaviour, particularly use of preventive measures like immunisation and an informed response to the onset and natural history of chronic degenerative diseases; (e) use of condoms to prevent transmission of micro-organisms; (f) self-medication with drugs of recreation and addiction.

Inter-personal: Aspects of social interaction which affect emotions and feelings, including life events, social participation, social integration and social support. Health at all ages can be affected. Social support during working life can buffer the effect of work strain on self-efficacy and self-esteem, thereby weakening its association with coronary heart disease and clinical depression (Siegrist, 1996; Wahrendorf, Blane, Bartley, Dragano & Siegrist, 2013). At older ages, social participation confers psychological resilience when faced with adversity (Netuveli, Wiggins, Montgomery, Hildon & Blane, 2008); while major life events increase mortality risk, as when death of spouse raises six-month mortality risk (Kaprio, Koskenvuo & Rita, 1987) and the risk of cardiovascular death is increased in the month following a cancer diagnosis though what are thought to be CNS-mediated mechanisms (Fang et al., 2012). Early childhood is particularly important because brain maturation is fastest and neuroplasticity greatest (Lichtman, 2001), which development sets subsequent levels of endocrine function in the hypothalamic-pituitary-adrenal axis and carcinogenesis (Kelly-Irving, Mabile, Grosclaude, Lang & Delpierre, 2013). The association between slower growth during childhood, as measured by height, signals the importance of this early development, due to its association with poorer cognitive function (Montgomery, Ehlin & Sacker, 2006a) which is relevant to the risk of further structural social exposures (Montgomery, Bartley, Cook & Wadsworth, 1996), as well as later biological outcomes indicated by the association of slower child growth with higher blood pressure in old age (Montgomery, Berney & Blane, 2000).

Biological processes

Material: Living (bacteria, viruses) and inert (asbestos fibres, folic acid) materials which have an impact on the body's structure and immune system. Impact can be beneficial (essential gut flora; folic acid-dependent embryonic neural tube development), harmful but contained (antibodies; scar tissue) or pathological (respiratory tuberculosis; mesothelioma).

CNS-mediated: (aka psycho-social). Social events and circumstances whose physiological effects are mediated via neurological and hormonal pathways from perception and emotions to the central nervous system and thus to peripheral receptors (Brunner & Marmot, 2006). Psychosocial stress alters neuroendocrine hormone levels, angiogenesis, tumour growth and cell migration; and down-regulates cellular immune responses, particularly through glucocorticoid and adrenergic signalling pathways (Herman & Cullinan, 1997; Lupien, McEwan, Gunnar & Heim, 2009).

Epigenetic: The apparent inconsistency between estimates of heritability derived from twin studies and those found in genome-wide association studies (Maher, 2008; Goldstein, 2009) led to an appreciation that genetic material acts within a complex nexus: proximally other surrounding genes and distally the life course material and social environment. Epigenetics refers to the ability of the social environment to trigger or suppress the process by which a gene acts as the template for production of a biologically active protein. For example, DNA methylation status, which with histone acetylation is the most common form of epigenetic modification (Nise, Falaturi & Erren, 2010), is associated with the social environment in early life, when lack of emotional warmth produces hyper-methylation of the glucocorticoid receptor gene in the brain hippocampus of both laboratory rats (Weaver et al., 2004) and human beings (McGowan et al., 2009; Perroud et al., 2011) to increase hypothalamic-pituitary-adrenal axis reactivity (Mazzio & Soliman, 2012). It is also suggested that epigenetic changes may be heritable over one or two generations if the environmental impact occurs when ova and sperm are forming (Pembrey, 2002). In a Swedish example (Bygren, Kaati & Edvinsson, 2001), boys who experienced high food availability around ages 9-12 years, when spermatogenesis increases, were more likely to have children and grandchildren with raised mortality rates, mostly due to type 2 diabetes and cardiovascular disease; while poorer food availability at ages 9-12 years was associated with greater longevity in descendants (Kaati, Bygren & Edvinsson, 2002; Kaati, Bygren, Pembrey & Sjostrom, 2007). Despite such suggestive results, epigenetics is a new area of research, so it is premature to judge the importance of its contribution to health and changes in health.

Principles

Each type of social exposure can work through any or all of the biological processes. A person's occupation (structural exposure) can affect their health through both material (asbestos) and CNS-mediated (effort-reward imbalance) biological processes. Tobacco smoking (behavioural exposure), likewise, can affect a person's health through both inhaled carcinogens (material process) and the low self-esteem of addiction (CNS-mediated process). And death of spouse (inter-personal exposure) can affect health through both poverty (material process) and grief and isolation (CNS-mediated process).

Both social exposures and biological processes can cumulate and interact. Living in a deprived area (structural exposure) is associated with high levels of inhaled fine particulate matter (material process), low social support (CNS-mediated process) and physical inactivity (behavioural exposure) leading to metabolic syndrome and cardiovascular disease risk. Although no longitudinal studies have assessed the combined effect of these factors on cardiovascular disease risk, there is evidence that adverse environmental exposures and psychosocial stressors cluster in deprived areas and disadvantaged groups (Bolte, Tamburlini & Kohlhuber, 2009); and may interact to amplify the effects of air pollution on health (Jerrett et al., 2004). In another example: cognitive performance at age 7 years, as measured by reading and mathematics test scores, reduces linearly with the number of disadvantages (structural, behavioural, inter-personal) experienced since birth (Bartley 2012). The impact of prior disadvantage on biological development is shown by the shorter stature at age seven years of children with poorer cognitive performance (Montgomery et al., 2006a) and this marker of disadvantage is independently associated with adult unemployment risk (Montgomery et al., 1996), underlining that disadvantage and associated exposures accumulate over life.

The relative importance of an exposure or process can vary with life course stage. Of particular interest biologically are childhood and adolescence, when brain plasticity is greatest and optimum growth and development may be at risk, and older ages when biological frailty develops. The effect of environmental stressors at different life stages depends upon the brain areas that are developing or declining at the time of exposure (Lupien et al., 2009). Reduced cognitive development during the early school years, as measured by reading and spatial test scores at age 7 years, is associated with irregular bedtimes, particularly in early childhood at age 3 (Kelly, Kelly & Sacker, 2011). Similarly, being economically inactive due to permanent sickness during early working life (ages 16-29 years) is associated with worse mental health than permanent sickness at other ages (30-49; 50-65 years)(Flint 2012). Structural and inter-personal disadvantage during early life may be more damaging than at other ages because it affects both normal child weight and neurological development and is associated with adult obesity and Type 2 diabetes (Olsson, Hulting & Montgomery, 2008; Osika & Montgomery, 2008a; Stenhammer et al., 2010).

Exposures and processes can be negative or positive. Negative exposures and processes (hazards, risk factors, pathology) traditionally have received the greater attention, but positive good health (optimal growth and development, resilience, longevity) are important also. Adult respiratory health, as measured by high FEV_1 (forced expiratory volume in one second) and no phlegm production, is associated with a good diet, as indexed by high consumption of fresh fruit (Kelly, Sacker & Marmot, 2003). Pre-existing social participation is associated with resilience at older ages, in the sense of mental health returning quickly to its pre-adversity level (Netuveli et al., 2008). Some protective exposures are associated with good health only in the presence of adversity – these are the true resilience factors. Having been breast fed in infancy is not itself related to adolescent mental health but, in those whose parents have divorced, infant breast feeding protects later psychological well-being (Montgomery et al., 2006b). Similarly, a warm childhood relationship with parents, and its consequent secure attachment style, is associated with achievement of a higher occupational grade only in those without elite education (Bartley, Head & Stansfeld, 2007; Stansfeld, Head, Bartley & Fonagy, 2008). Conversely, economic adversity in early old age may eliminate the benefits of earlier advantages, particularly among those who experienced a relatively advantaged childhood: *disappointment paradox* (Osika et al., 2006; Osika & Montgomery 2008b).

Example. These are complex processes by which socially patterned exposures are linked to biological changes which then lead on to health outcomes. More than one pathway may be involved, and probably is in most cases. Health outcomes in turn may influence subsequent social exposures in an iterative cycle. It may be helpful to provide one example that is already seen in the literature. Adverse experiences in childhood (ACE) have now been associated in prospective studies with both problematic social relationships and obesity in adulthood, as well as increased levels of inflammatory or pro-inflammatory substances, such as C-reactive protein and interleukin-6 (Danese, Pariante, Caspi, Taylor & Poulton, 2007; Pollitt et al., 2007; Taylor, Lehman, Kiefe & Seeman, 2006). In very recent work, adverse experiences in childhood appear to be associated with adult obesity (Thomas, Hypponen & Power, 2008), all-cause mortality (Brown et al., 2009) and with cancer incidence (Brown et al., 2010; Kelly-Irving et al., 2013). In biological terms, obesity is associated with increased risk of most female cancers, perhaps because adipose tissue alters the steroid milieu by absorbing and storing steroids (Patterson et al., 2013). Apart from cancer, a pro-inflammatory profile and obesity have additional adverse outcomes, including depression (Soczynska et al., 2011).

Public health applications

A potentially important use of the life course approach and the principles of the social-biological transition, is to pin-point when during life certain processes are most harmful and when others are most helpful, and at what levels (individual; household; neighbourhood; school; social structure) – then supporting or initiating social interventions at those times and levels. In the present economic context, this may mean science demonstrating (again) the relevance to public health and well-being of programmes threatened by budget cuts.

Tobacco smoking reduction. The problem with using high price to deter tobacco smoking is that the remaining smokers are disproportionately poor, so high price can be a second assault on the health of the undeterred, adding deepened poverty to tobacco toxins (Lang, Jusot, Visier, Menvielle & Lombrail, 2012). A life course understanding of nicotine addiction suggests support for other strategies than price alone. The natural history of tobacco smoking involves adolescent social smoking becoming adult addictive smoking followed by repeated quit attempts, triggered by acute chest infections in middle age, leading to successful cessation or chronic obstructive pulmonary disease. Promising alternatives to further increases in the price of tobacco include: de-glamourise tobacco smoking to make it unattractive to adolescents; remove state subsidies to the cost of tobacco (Armed Forces; international travellers) to reduce progression to addictive smoking; prohibit tobacco smoking in workplaces and public places to protect others, and self, from damage; train physicians to see every acute chest infection as a chance to initiate and support quitting. In terms of public health interventions, the difficulties associated with smoking cessation need to be considered, such as the need for several attempts at quitting. The significance of smoking among certain groups, such as pregnant women's risk of long-term damage to offspring health (Toschke, Montgomery, Pfeiffer & von Kries, 2003), ought to also be taken into account as well as the possible side effects of cessation (Lang et al., 2012).

Causal maps. A causal map includes the social exposures and biological processes known to be relevant to a specified health outcome. The causal map of lung cancer might include living in a deprived area (structural exposure) leading to environmental exposure to asbestos (material process) and to polycyclic aromatic hydrocarbons (material process) from tobacco smoking (behavioural exposure) and traffic exhaust fumes (structural exposure). Such maps (Figures 2 & 3) identify the variables to be measured and draw attention to where existing data sets could be strengthened.

Figure 2. Suggested causal map for links between area deprivation and lung cancer

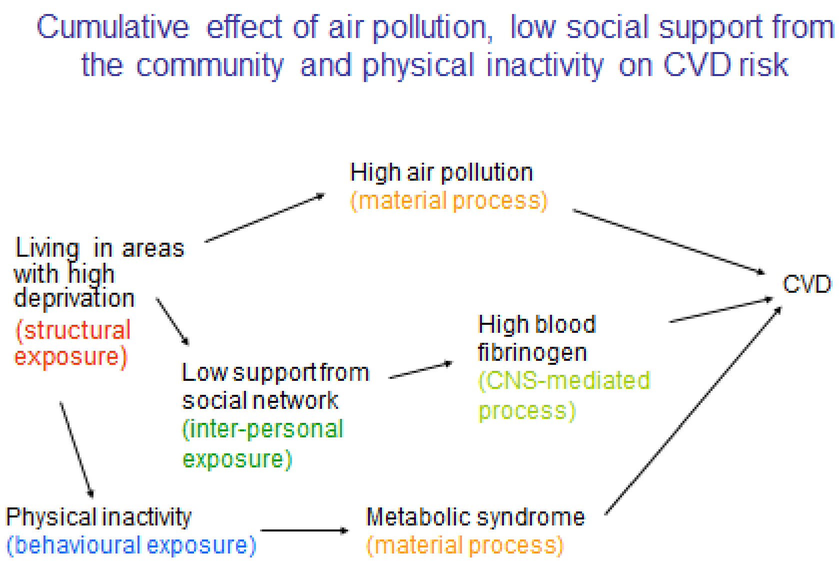
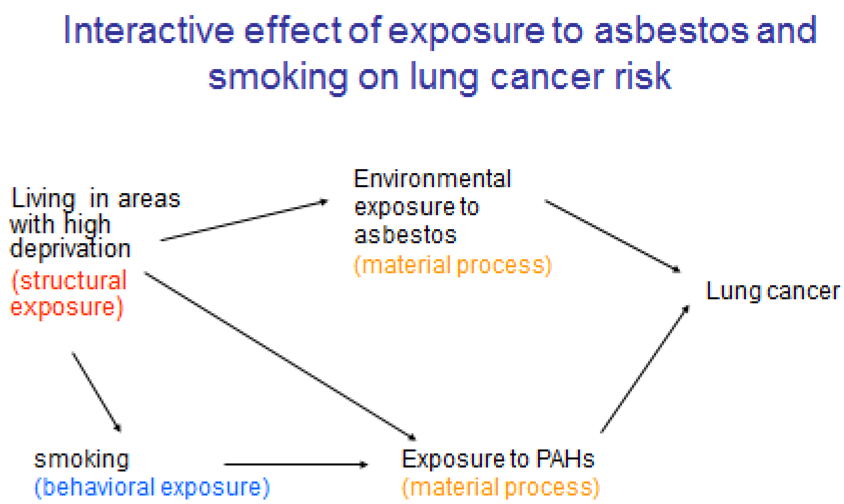


Figure 3. Suggested causal map for area deprivation and cardiovascular disease



Conclusions

The present discussion paper has tried to sketch a comprehensive framework for thinking about social-biological transitions. As well as counter-balancing the excessive exuberance with which epigenetics or neuroscience or health-related behaviours are sometimes treated as *the* main pathway through which the social and biological interact, the framework also draws attention to the scarcity of studies collecting information on the different types of social exposure and the different types of biological process. The title of the framework as social-biological transitions implies that the social drives the biological, which may seem contentious to biologist members of the life course inter-disciplinary research community. On the face of it, the relative stability of biological structures does not fit easily with the rapidity of social change; for example, it is difficult to reconcile the recent rapid increase in life expectancy at middle age with the apparently timeless biology of cell senescence. Undoubtedly, in time, the extra-cellular or epigenetic processes involved will be described, but it is likely that these will be pathways between the social and the biological, not the drivers of population changes in longevity.

In summary. The epidemiological tradition within life course research combines longitudinal data analysis with the social and biological sciences. Such research will become more plausible and relevant to public health when it *specifies* and *tests* the social and biological pathways between social circumstances and later health. This paper has suggested a model for thinking about such pathways in terms of social exposures, biological processes and the principles governing their interaction.

Acknowledgements

This discussion paper is indebted heavily to the participants of the workshop on social-biological transitions held in London 31 May-June 1st 2012 which brought together, in addition to the present authors, members of INSERM U1027 Toulouse (Thierry Lang, Laurence Mabile, Cyrille Delpierre), Turin Longitudinal Study (Giuseppe Costa, Chiara Marinacci, Roberto Leombruni, Silva Stringhini) and ESRC *International Centre for Life Course Studies in Society and Health* - ICLS (Meena Kumari, Yvonne Kelly, Amanda Sacker); an event funded as part of UK Economic and Social Research Council award RES-596-28-0001.

References

Arbeev, K., Akushevich, I., Kulminski, A., Arbeeva, L., Akushevich, L., Culminskaya, I. & Yashin, A (2011). Age trajectories of physiological indices in relation to healthy life course. *Mechanisms of Ageing and Development 132,* 93-102. DOI:10.1016/j.ma.

Bartley, M., Blane, D. & Montgomery, S. (1997). Health and the life course: why safety nets matter. *British Medical Journal 314*, 1194-1196.

Bartley, M., Head, J. & Stansfeld, S. (2007). Is attachment style a source of resilience against health inequalities at work? *Social Science & Medicine 64*, 765-775.

Bartley, M. (Ed.) (2012). *Life Gets Under Your Skin*. (Pp. 4-5) London, University College London.

Blane, D., Bartley, M. & Mitchell, R. (2000). The inverse housing law and respiratory health. *Journal of Epidemiology and Community Health 54*, 745-749.

Bolte, G., Tamburlini, G. & Kohlhuber, M. (2010). Environmental inequalities among children in Europe: evaluation of scientific evidence and policy implications. *European Journal of Public Health 20*, 14-20.

Brown, D., Anda, R., Tiemeier, H., Felitti, V., Edwards, V., Croft, J. & Giles, W. (2009). Adverse childhood experiences and the risk of premature mortality. *American Journal of Preventive Medicine 37*, 389-396.

Brown, D., Anda, R., Felitti, V., Edwards, V., Malarcher, A., Croft, J. & Giles, W. (2010). Adverse childhood experiences are associated with the risk of lung cancer: a prospective cohort study. *BMC Public Health 10*:20.

Brunner, E. & Marmot, M. (2006). Social organization, stress and health. In M. Marmot & R. Wilkinson (Eds). *Social Determinants of Health* 2nd edition (pp. 6-30). Oxford: Oxford University Press.

Butterworth, P., Cherbuin, N., Sachdev, P. & Anstey, K. (2012). The association between financial hardship and amygdala and hippocampal volumes: results from the PATH through life project. *Social Cognitive and Affective Neuroscience 7*, 548-556.

Bygren, L., Kaati, G. & Edvinsson, S. (2001). Longevity determined by paternal ancestors' nutrition during their slow growth period. *Acto Biotheoretica 49*, 53-59.

CDC - Centers for Disease Control and Prevention. (2010). Cigarette use among high school students - United States, 1991-2009. *Morbidity and Mortality Weekly Report (Jul 9) 59*, 797-801.

da Costa, B. & Vieira, E. (2010). Risk factors for work-related musculoskeletal disorders: systematic review of recent longitudinal studies. *American Journal of Industrial Medicine 53*, 285-323.

Danese, A., Pariante, C., Caspi, A., Taylor, A. & Poulton, R. (2007). Childhood maltreatment predicts adult inflammation in a lifecourse study. *Proceedings of the National Academy of Sciences of the United States of America 104*, 1319-24.

Dearing, E. (2008). Psychological costs of growing up poor. *Annals of New York Academy of Science 1136*, 324-332.

Eisenberger, N., Taylor, S., Gable, S., Hilmert, C. & Lieberman, M. (2007). Neural pathways link social support to attenuated neuroendocrine stress responses. *NeuroImage 35*, 1601-1612.

Esping-Andersen, G. (1990). *The Three Worlds of Welfare Capitalism.* Cambridge: Polity Press.

Esping-Andersen, G. (1996). *Welfare States in Transition: national adaptations in global economies.* London: Sage Publications.

Esping-Andersen, G. (1999). *Social Foundations of Industrial Economies.* Oxford: Oxford University Press.

Falkingham, J. & Hills, J. (Eds). (1995). *The Dynamics of Welfare: The welfare state and the life cycle.* Hemel Hempstead: Prentice Hall.

Fang, F., Fall, K., Mittleman, M., Sparén, P., Ye, W., Adami, H. & Valdimarsdóttir, U. (2012). Suicide and cardiovascular death after a cancer diagnosis. *New England Journal of Medicine 366*, 1310-1318.

Flint, E. (2012). Unemployment and Mental Health in British Household Panel Study. University of London Library (Doctoral Thesis).

Gauderman, W., Avol, E., Gilliland, F., Vora, H., Thomas, D., Berhane, K., McConnell, R., Kuenzli, N., Lurmann, F., Rappaport, E., Margolis, H., Bates, D. & Peters, J. (2004). The effect of air pollution on lung development from 10 to 18 years of age. *New England Journal of Medicine 351*, 1057-1067.

Gershoff, E., Aber, .L, Raver, C. & Lennon, C. (2007). Income is not enough: incorporating material hardship into models of income associations with parenting with child development. *Child Development 78*,70-95.

Goldberg, J., Wheeler, G. & Sydenstriker, E. (1920). A study of the relation of family income and other economic factors to pellagra incidence in seven cotton-mill villages of South Carolina in 1916. *Public Health Reports 35*,2673-2714.

Goldstein, D. (2009). Common genetic variation and human traits. *New England Journal of Medicine 360*, 1696-1698.

Herman, J. & Cullinan, W. (1997). Neurocircuitry of stress: central control of the hypothalamo-pituitary-adrenocortical axis. *Trends in Neurosciences 20*, 78-84.

Iso, H. (2011). Lifestyle and cardiovascular disease in Japan. *Journal of Atherosclerosis and Thrombosis 18*, 83-88.

Jerrett, M., Burnett, R., Brook, J., Kanaroglou, P., Giovis, C., Finkelstein, N. & Hutchison, B. (2004). Do socioeconomic characteristics modify the short term association between air pollution and mortality? Evidence from a zonal time series in Hamilton, *Canadian Journal of Epidemiology and Community Health 58*, 31-40.

Johnson, R. & Schoeni, R. (2011). Early-life origins of adult disease: national longitudinal population-based study of the United States. *American Journal of Public Health 101*, 2317-2324.

Kaprio, J., Koskenvuo, M. & Rita, H. (1987). Mortality after bereavement: prospective study of 95,647 widowed persons. *American Journal of Public Health 77*, 283-287.

Kaati, G., Bygren, L. & Edvinsson S. (2002). Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period. *European Journal of Human Genetics 10*, 682-688.

Kaati, G., Bygren, L., Pembrey, M. & Sjöström, M. (2007). Transgenerational response to nutrition, early life circumstances and longevity. *European Journal of Human Genetics 15*, 784-790.

Kelly, Y., Kelly, J. & Sacker, A. (2011). Time for bed? The relationship between bedtimes and socio-emotional and cognitive development in 7 year old children. Findings from the UK Millenium Cohort Study. *Journal of Epidemiology and Community Health 65 (Suppl 2)*, A39-A40.

Kelly, Y., Sacker, A. & Marmot, M. (2003). Nutrition and respiratory health in adults: findings from Health Survey for Scotland. *European Respiratory Journal 21*, 664-671.

Kelly-Irving, M., Mabile, L., Grosclaude, P., Lang, T., & Delpierre, C. (2013). The embodiment of adverse childhood experiences and cancer development: potential biological mechanisms and pathways across the life course. *International Journal of Public Health 58,* 3-11.

Lang, T., Jusot, F., Visier, L., Menvielle, G. & Lombrail, P. (2012). Réduire la consommation de tabac: comment prendre en compte les inégalités sociales de santé? *Actualité et Dossier en Santé Publique 81*, 44-46.

Laquatra, J., Maxwell, L. & Pierce, M. (2005). Indoor air pollutants: limited-resource households and child care facilities. *Journal of Environmental Health 67*, 39-43.

Litchman, J. (2001). Developmental neurobiology overview: synapses, circuits and plasticity. In D. Bailey, T. Bruer, F. Symons & J. Lichtman J. (Eds). *Critical Thinking about Critical Periods*. (Pp. 27-44) Baltimore: Paul H. Brookes.

Lupien, S., McEwen, B., Gunnar, M. & Heim, C. (2009). Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nature Reviews Neuroscience 10*, 434-445.

Maher, B. (2008). Personal genomes: the case of missing heritability. *Nature 456*, 18-21.

Maty, S., Lynch, J., Raghunathan, T. & Kaplan, G. (2008). Childhood socioeconomic position, gender, adult body mass index, and incidence of type 2 diabetes mellitus over 34 years in the Alameda County Study. *American Journal of Public Health 98*, 1486-94.

Mazzio, E. & Soliman, K. (2012). Basic concepts of epigenetics: impact of environmental signals on gene expression. *Epigenetics 7*, 119-130.

McGowan. P, Sasaki, A., D'Alessio, A., Dymov, S., Labonté, B., Szyf, M., Turecki, G. & Meaney, M. (2009). Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nature Neuroscience 12*, 342-348.

Mitchell, R., Blane, D. & Bartley, M. (2002). Elevated risk of high blood pressure: climate and the inverse housing law. *International Journal of Epidemiology 31*, 831-838.

Montgomery, S., Bartley, M., Cook, D. & Wadsworth, M. (1996). Health and social precursors of unemployment in young men in Great Britain. *Journal of Epidemiology and Community Health 50*, 415-422.

Montgomery, S., Berney, L. & Blane, D. (2000). Pre-pubertal stature and blood pressure in early old age. *Archives of Disease in Childhood 82*, 358-363.

Montgomery, S., Ehlin, A. & Sacker, A. (2006a). Pre-pubertal growth and cognitive function. *Archives of Disease in Childhood 91*, 61-62.

Montgomery, S., Ehlin, A. & Sacker, A. (2006b), Breast feeding and resilience against psychosocial stress. *Archives of Disease in Childhood 91*, 990-994.

Morris, J.N., Donkin, A., Wonderling, D., Wilkinson, P. & Dowler, E. (2000). A minimum income for healthy living. *Journal of Epidemiology and Community Health 54*, 885-889.

Morris, J.N. & Deeming, C. (2004) Minimum Income for Healthy Living (MIHL): next thrust in UK social policy? *Policy & Politics 32*, 441-454.

Morris, J.N., Wilkinson, P., Dangour, A., Deeming, C. & Fletcher, A. (2007). Defining a minimum income for healthy living (MIHL): older age, England. *International Journal of Epidemiology 36*, 1300-1307.

NRC - National Research Council, Institute of Medicine. (2001). *Musculoskeletal disorders and the workplace: low back and upper extremities.* Washington DC: National Academy Press.

Netuveli, G., Wiggins, R., Montgomery, S., Hildon, Z. & Blane, D. (2008). Mental health and resilience at older ages: bouncing back after adversity in the British Household Panel Survey. *Journal of Epidemiology and Community Health 62*, 987-991.

Nise, M., Falaturi, P. & Erren, T. (2010). Epigenetics: origins and implications for cancer epidemiology. *Medical Hypotheses 74*, 377-382.

Olsson, G., Hulting, A. & Montgomery, S. (2008). Cognitive function in children and subsequent type 2 diabetes mellitus. *Diabetes Care 31*, 514-516.

ONS – Office for National Statistics. (2012). Excess winter mortality in England and Wales 2010-2011 (Final). Newport: ONS Mortality Analysis Team.

Osika, W., Ehlin, A. & Montgomery, S. (2006). Does height modify the risk of angina associated with economic adversity? *Economics and Human Biology 4*, 398-411.

Osika, W. & Montgomery, S. (2008a). Physical control and coordination in childhood and adult obesity among members of a longitudinal birth cohort. *British Medical Journal 337*, a699. DOI: 10.1136/bmj.a699.

Osika, W. & Montgomery, S. (2008b). Economic disadvantage modifies the association of height with low mood in the US, 2004: the disappointment paradox. *Economics and Human Biology 6*, 95-107.

Osler, M., Andersen, A., Due, P., Lund, R., Damsgaard, M. & Holstein, B. (2003). Socioeconomic position in early life, birth weight, childhood cognitive function, and adult mortality: a longitudinal study of Danish men born in 1953. *Journal of Epidemiology and Community Health 57*, 681-6.

Patterson, R., Rock, C., Kerr, J., Natarajan, L., Marshall, S., Pakiz, B. & Cadmus-Bertram, L. (2013). Metabolism and breast cancer risk: frontiers in research and practice. *Journal of the Academy of Nutrition and Dietetics 113*, 288-96.

Pembrey, M. (2002). Time to take epigenetic inheritance seriously. *European Journal of Human Genetics 10*, 669-671.

Perroud, N., Paoloni-Giacobino, A., Prada, P., Olié, E., Salzmann, A., Nicastro, R., Guillaume, S., Mouthon, D., Stouder, C., Dieben, K., Huguelet, P., Courtet, P. & Malafosse, A. (2011). Increased methylation of glucocorticoid receptor gene (NR3C1) in adults with a history of childhood maltreatment: a link with the severity and type of trauma. *Translational Psychiatry 1*, e59.

Pollitt, R., Kaufman, J., Rose, K., Diez-Roux, A., Zeng, D. & Heiss, G. (2007). Early-life and adult socioeconomic status and inflammatory risk markers in adulthood. *European Journal of Epidemiology 22*, 55-66.
Rowntree, B.S. (1902). *Poverty: A study of town life.* London: Nelson.
Schikowski, T., Sugiri, D., Ranft, U., Gehring, U., Heinrich, J., Wichmann, H. & Krämer, U. (2005). Long-term air pollution exposure and living close to busy roads are associated with COPD in women. *Respiratory Research 6*, 152.
Siegrist, J. (1996). Adverse health effects of high effort-low reward conditions. *Journal of Occupational Health Psychology 1*, 27-41.
Soczynska, J., Kennedy, S., Woldeyohannes, H., Liauw, S., Alsuwaidan, M., Yim, C. & McIntyre, R. (2011). Mood disorders and obesity: understanding inflammation as a pathophysiological nexus. *Neuromolecular Medicine 13*, 93-116.
Strachan, D. & Sheikh, A. (2004). A life course approach to respiratory and allergic diseases. In D. Kuh & Y. Ben-Shlomo. (Eds). *A Life Course Approach to Chronic Disease Epidemiology,* 2nd edition (pp. 240-259). Oxford: Oxford University Press.
Stansfeld, S., Head, J., Bartley, M. & Fonagy, P. (2008). Social position, early deprivation and the development of attachment. *Social Psychiatry and Psychiatric Epidemiology 43*, 516-526.
Taylor, S., Lehman, B., Kiefe, C. & Seeman, T. (2006). Relationship of early life stress and psychological functioning to adult C-reactive protein in the coronary artery risk development in young adults study. *Biological Psychiatry 60*, 819-24.
Thomas, C., Hypponen, E. & Power, C. (2008). Obesity and type 2 diabetes risk in mid-adult life: the role of childhood adversity. *Pediatrics 121*, e1240-1249.
Toschke, A., Montgomery, S., Pfeiffer, U. & von Kries, R. (2003). Early intrauterine exposure to tobacco: inhaled products and obesity. *American Journal of Epidemiology 158*, 1068-1074.
Townsend, P. (1979). *Poverty in The United Kingdom: a survey of household resources and standards of living*. London: Penguin.
van Doorslaer, E., Koolman, X. & Jones, A. (2004). Explaining income-related inequalities in doctor utilisation in Europe. *Health Economics 13*, 629-647.
Wahrendorf, M., Blane, D., Bartley, M., Dragano, N. & Siegrist, J. (2013). Working conditions in mid-life and mental health in older ages: results from SHARELIFE. *Advances in Life Course Research* 18, 93-102.
Webb, E., Blane, D. & de Vries, R. (2012). Housing and respiratory health at older ages. *Journal of Epidemiology and Community Health.* DOI: 10.1136/jech-2012-201458.
Weaver, I., Cervoni, N., Champagne, F., D'Alessio, A., Sharma, S., Seckl, J., Dymov, S., Szyf, M. & Meaney, M. (2004). Epigenetic programming by maternal behavior. *Nature Neuroscience 7*, 847-854.
Willett, W. (2006). The Mediterranean diet: science and practice. *Public Health Nutrition 9*, 105-110.
Wolfe, B., Evans, W. & Seeman, T. (Eds). (2012). *The Biological Consequences of Socioeconomic Inequalities.* New York: Russell Sage Foundation.

Longitudinal and Life Course Studies 2013 Volume 4 Issue 2 Pp 147 – 165 ISSN 1757-9597

Subject specific and population average models for binary longitudinal data: a tutorial

Camille Szmaragd, Paul Clarke, Fiona Steele
University of Bristol, UK
Camille.Szmaragd@bristol.ac.uk

(Received March 2013 Revised May 2013)

Summary

Using data from the British Household Panel Survey, we illustrate how longitudinal repeated measures of binary outcomes are analysed using population average and subject specific logistic regression models. We show how the autocorrelation found in longitudinal data is accounted for by both approaches, and why, in contrast to linear models for continuous outcomes, the parameters of population average and subject specific models for binary outcomes are different. To illustrate these points, we fit different models to our data set using both approaches, and compare and contrast the results obtained. Finally, we use our example to provide some guidance on how to choose between the two approaches.

Keywords: autocorrelation, British Household Panel Survey, hierarchical models, logistic regression, marginal models, mixed effects models, multilevel models, random effects models, repeated measures.

1 Introduction

In this tutorial, we consider the analysis of repeated measures longitudinal data and how to choose the most appropriate method of statistical analysis. We restrict our attention to data from panel surveys like the British Household Panel Survey (BHPS) (ISER 2010) in which the survey waves take place at regular intervals, and at each wave all sample members are surveyed at approximately the same point in time.

The analysis of longitudinal data typically involves questions about the relationship between an outcome variable and its predictors, in much the same way as when only cross-sectional data are available. The difference is that we have outcomes, and possibly predictor variables, from each wave to include in the analysis. Longitudinal models allow these measures to be incorporated correctly, and for us to fully exploit the information contained by the data.

Our tutorial is aimed at quantitative social scientists familiar with linear and logistic regression models but less familiar with modelling longitudinal data. It is principally about two basic types of longitudinal methods called 'population average' (PA) and 'subject specific' (SS) models. In the case of linear models for continuous outcomes, the two approaches are very similar, but differences emerge when the two are used to analyse binary outcomes. To illustrate our tutorial, we analyse the relationship between mental health and employment status using 18 waves of BHPS data. As we will discuss, the choice depends partly on the type of research question to be answered, and whether this question is answered by estimating the effects of time-invariant or time-varying predictor variables.

While we use *Stata* (StataCorp, 2011) to fit PA and SS models to our example data, it is important to note that we do not attempt to present a step-by-step guide on how this is done. A *Stata* '`.do`' file containing the commands used to carry out the analyses presented here is provided as supplementary material, but we have tried as far as possible to make this article independent of the software package. Instead, we focus on highlighting the distinction between the two approaches, and discuss the reasons why an analyst might report results from using one rather than the other.

The article is arranged as follows: The data example is introduced in Section 2, followed in Section 3, by a discussion of the potential and limitations of longitudinal data for answering complex research questions. In Section 4, we formally introduce and compare PA and SS models, and in Section 5, we show graphically why the coefficients of both models are different. In Section 6, we use both approaches to analyse the BHPS data, and finally, in Section 7, we discuss how our analysis could be extended, and give pointers for further reading on this subject.

2 Data Example

The main question in our illustrative analysis concerns the association between employment status and mental health. In this section, we focus on describing the features of a particular longitudinal data set, and leave the description of how we model these data until the analysis of mental health and employment status is formally introduced in Section 6. The reader should note that Steele, French and Bartley (2013) carry out a comprehensive analysis of the same data set using more advanced longitudinal modelling techniques, but for illustrative purposes we present a simplified analysis.

The data are extracted from waves 1 to 18 of the British Household Panel Survey (BHPS) in which each wave took place annually following the first wave in 1991 (ISER 2010). Our sample comprises 9,192 men aged 16-64. In an ideal world, the data would be 'balanced' in the sense that we observe the General Health Questionnaire (GHQ) score and employment status 18 times for each individual subject. However, our data set is 'unbalanced' because some subjects appear in it for the first time after the first wave in 1991, and some appear for the last time before the final wave in 2008. To take two examples, this may be because some subjects in sample households do not join the panel until after their 16th birthdays, and/or drop out (temporarily or permanently) of the BHPS.

The mental health outcome is based on the GHQ, which in its raw form measures anxiety and depression on a scale from 0 to 36. Instead of using the raw GHQ scores, however, we follow others and construct a binary variable of GHQ 'caseness', where a GHQ case is defined as a subject whose GHQ score is greater than 12 (Goldberg *et al.* 1997), and refer to this using the variable `ghq_case`. Throughout the paper, we take GHQ caseness to be synonymous with poor mental health.

In terms of predictor variables, we limit ourselves to employment status (`empl`), which is represented by a 3-category measure: employed (E), unemployed (U) and inactive/outside labour force (I). We also include variables for the age of each subject.

Depending on the software package or particular routine being used to fit longitudinal models, the data are stored using one of two formats: wide or long.

In the wide format, the information on a particular subject is contained in one row of the data set. Table 1 displays a subset of our data using the wide format (specifically, waves 1, 2, 6 and 12 out of 18) for ten subjects. Time-varying characteristics like mental health are represented using a separate variable for its values at each wave, so that `ghq_case` is represented by `ghq_w1`, `ghq_w2`, etc.; similarly, employment status and age are represented by the variables `empl_w1`, `empl_w2`, etc. and `age_w1`, `age_w2`, etc., respectively. The wide format is mainly used for analysing longitudinal data using structural equation models, where the repeated measures of GHQ are treated as multivariate responses (e.g. Bollen and Curran 2005).

Table 1. Subset of the BHPS data set represented in wide format. The categories for the employment variable(s) are E - employed, U - unemployed, I – inactive. The three time-varying variables, `ghq_case` (outcome), `empl` and `age` (predictors) are observed at the different time points as indicated by `w1`, `w2`, `w6` and `w12`

pid	ghq_w1	ghq_w2	ghq_w6	ghq_w12	empl_w1	empl_w2	empl_w6	empl_w12	age_w1	age_w2	age_w6	age_w12
1	0	0			U	U			28	29		
2	0	0			U	U			26	27		
3	0	0	0		E	E	E		57	58	62	
4	0	1			E	E			36	37		
5				0				E				21
6			1				E				46	
7	0	0	0		E	E	E		30	31	35	
8	0	0			E	E			23	24		
9			0				I				64	
10	1	0	0		U	U	I		30	31	35	

Table 2 displays subjects 1-5 from table 1 but using the alternative long format. Each row now corresponds to the data on a subject at a particular measurement occasion, which reduces the number of columns/variables at the expense of increasing the number of rows in the data set. For instance, `ghq_case` is now represented using one variable in conjunction with the subject and wave identifiers `pid` and `wave`, respectively.

We also define `occasion` to indicate the measurement occasion for each subject. There are two features of `occasion` which should be noted: first, occasion 1 does not correspond to wave 1 for everyone in the sample because some subjects do not appear until after the first wave; and second, the first record on some subjects is not occasion 1 because of missing data on the analysis variables at earlier occasions.

Note that we have defined two employment status variables here: `empl` is the subject's employment status at that occasion; and `empl1` is employment status at occasion 1. Both `age1` and `empl1` are subject-level variables whose values are fixed across occasions for the same subject.

Table 2. Subset of the BHPS data set represented in long format

pid	wave	occasion	ghq_case	empl	age	empl1	age1
1	1	1	0	Unemployed	28	Unemployed	28
1	2	2	0	Unemployed	29	Unemployed	28
2	1	1	0	Unemployed	26	Unemployed	26
2	2	2	0	Unemployed	27	Unemployed	26
2	3	3	0	Unemployed	28	Unemployed	26
3	1	1	0	Employed	57	Employed	57
3	2	2	0	Employed	58	Employed	57
3	3	3	0	Employed	59	Employed	57
3	6	6	0	Employed	62	Employed	57
4	1	1	0	Employed	36	Employed	36
4	2	2	1	Employed	37	Employed	36
5	12	5	0	Employed	21	Employed	16

The unbalanced nature of our data can be seen in table 1 and table 2: `ghq_case` and employment status are not observed for every subject at every occasion. While this is obvious from the wide format data in table 1, where the missing values are shown by blanks, it is indicated only by breaks in the `occasion` sequence and so less apparent for the long format data in table 2.

3 Why Use Longitudinal Data?

Now that we have seen what a longitudinal data set looks like and discussed some of its features, we recap on the advantages of longitudinal over cross-sectional data.

The first advantage is that longitudinal data allow us to establish the temporal ordering of events. Suppose that we measure mental health and employment status in two waves of a longitudinal panel survey. The mental health of each participant is measured using the GHQ-case variable introduced in Section 2, and the employment status of each participant is classified as either employed (E), unemployed (U) or inactive (I). We can denote employment status at the first and second waves by the categorical variables x_1 and x_2, respectively, and GHQ-case at the first and second waves by y_1 and y_2, respectively. The longitudinal design allows us to establish with certainty that x_1 and y_1 are measured before x_2 and y_2. The time ordering means that neither x_2 nor y_2 can have caused x_1 or y_1, and longitudinal models with 'reverse causation' – in which, say, x_2 predicts y_1 – can be excluded from consideration.

But this does not mean that longitudinal data automatically gives us the answer to causal questions such as "If I change someone's first-wave employment status x_1 then what will happen to his second-wave mental health y_2?" To answer such questions definitively, the data need to come from a longitudinal experimental design, or be adjusted appropriately for confounding bias. For a simple example of a longitudinal experimental design, consider an experiment where we measure the mental health of each participant at wave one, y_1, and then randomise each subject to one of the three employment status groups to obtain x_1. If the subjects all keep their randomised employment status during the follow-up period until y_2 is measured, then the differences in the proportions of GHQ cases between the three employment status groups are 'causal estimates' of employment effect. While an experiment like this would clearly be difficult to implement, the main message is simply this: longitudinal data can help rule out models with reverse causation, but do not guarantee that causal relationships can be estimated.

The second advantage of longitudinal data is that the repeated measurements can be used to improve the precision of our estimates. The argument we set out here comes from Zeger and Liang (1992), who illustrated their point using a simple experiment much like the one just described. So keep in mind the hypothetical experiment where x_1 is randomised employment status, but suppose that mental health at each wave, y_1 and y_2, is measured using the raw GHQ score rather than the GHQ-case indicator. Zeger and Liang showed that the causal effect of employment status can be estimated using the *difference score* (i.e. the difference $y_2 - y_1$), and calculating the difference between the mean difference scores in each of the three employment categories. Not only is this estimate unbiased, it is more precise (i.e. has smaller standard errors) than simply taking the differences between the means of y_2. The improved precision comes about because measures on the same individual, even at different points in time, are typically positively associated; the general term used to describe positive associations between measurements on the same individual is 'autocorrelation'. More generally, however, difference scores cannot be used when we have more than two measurements, which is why we need formal longitudinal data methods to allow for autocorrelation and to improve parameter estimation.

The third advantage of longitudinal data is that within-subject changes, or growth, over time can be explicitly modelled along with the outcome's relationship with time-varying predictors.

4 Population Average and Subject Specific Models

In this section, we will provide a brief review of both population average (PA) and subject specific (SS) models. To help introduce some of the fundamental differences between PA and SS modelling, we introduce each type of model for the more familiar linear case, before moving onto non-linear logistic models.

4.1 PA linear models

In general, we specify longitudinal models for repeated measurements taken on each subject at different points in time, so all longitudinal models have a time dimension. For our example, we could define time by the calendar year in which the wave took place, the wave number, the subject-specific measurement occasion, or the subject's age at a particular wave. In section 6, we discuss the choice of time for our application, but for now we talk simply in terms of time and time-points.

So for subject i and time t, it might be tempting to use the standard linear model

$$y_{it} = \beta_0 + \beta_1 x_{it} + e_{it},$$

where y_{it} is the outcome variable (which would be the raw GHQ score in our illustrative example) and x_{it} represents the predictor variable(s). It is typically assumed that the residual e_{it} is normally distributed with variance σ^2.

There is one model equation defined for each subject at each time-point. The reason we cannot simply fit this model to the longitudinal data is because it assumes that *all* the residuals are independent of each other, but we know that all residuals on the same subject $e_{i1}, e_{i2}, \dots$ are not independent because of autocorrelation.

PA models can be specified and estimated so as to account for autocorrelation. A linear PA model comprises two parts. First is the mean of y_{it} given the covariates

$$E(y_{it}|x_{it}) = \beta_0^{PA} + \beta_1^{PA} x_{it},$$

where $E(y_{it}|x_{it})$ denotes the mean outcome among those subjects with predictor variables x_{it}. Notice that no assumption about the distribution of the residuals has been made.

It turns out that we can estimate the parameters of the PA model provided we specify something about the residual distribution. In fact, it turns out that we only need to specify the (auto)correlation between the residuals $e_{i1}, e_{i2}, \dots$ The autocorrelation structure is specified simply through the choice of 'working correlation matrix' which constitutes the second part of the PA model. We discuss specification of the working correlation matrix in section 4.3.

4.2 PA logistic models

For binary outcomes, one would generally choose a logistic or probit model. We focus on the former because its parameters are conveniently interpreted as log-odds ratios.

As with the PA linear model above, a PA logistic model has two components. The first component can be written as

$$\text{logit}\,\Pr(y_{it} = 1|x_{it}) = \beta_0^{PA} + \beta_1^{PA} x_{it},$$

where $\text{logit}(p) = \log(p/1-p)$ is the usual logit 'link' function for any probability p between 0 and 1. In the context of our example, this means that the log-odds of being a GHQ case is linearly related to employment status and the other predictor variables.

This is identical to the expression for the standard logistic model apart for the t subscript, but we cannot fit the standard logistic model to longitudinal data. There does not appear to be a residual specified in the model above, but there is a hidden residual, and in the standard model these are all assumed to be independent.

The reason we cannot see the residual is because it is hidden from us, but we specified it implicitly when we chose to use the logit link. To show where it is, we note that the logistic model can be represented using latent variables, where there is a continuously distributed outcome variable y_{it}^* hidden from us for which we observed only whether its value is positive (i.e. $y_{it} = 1$) or negative ($y_{it} = 0$). It is further assumed that the hidden outcome variable follows a linear model which depends on the same predictor variable(s) as above *and* a hidden residual e_{it}^* that is logistically distributed. The (standard) logistic distribution in question is a symmetric continuous distribution with a mean of 0 and a variance of 3.29 (the probit model, on the other hand, is based on the assumption that the hidden residuals are normally distributed with mean 0 and variance 1). The 3.29 value emerges again when we discuss the difference between the PA and SS coefficients in section 5.

Hence, we complete the specification of the PA logistic model by specifying the working correlation matrix for the hidden residuals. However, we cannot estimate these residuals as we can for linear models because they are hidden, and we must assume that they all have equal variance (i.e. are homoskedastic).

4.3 Fitting PA models

Liang and Zeger (1986) proposed Generalized Estimating Equations (GEE) as an extension of standard regression estimation procedures to allow for autocorrelation. In the overwhelming majority of applications, PA models are fitted using GEE, and so we focus on GEE estimation. It is the popularity of GEE estimation that has resulted in PA and GEE becoming synonymous, and the presentation of PA models in terms of these estimating equations. We choose not to do this, however, because GEE is not a model and, we feel, doing so introduces unnecessary algebra that makes PA models seem more complex than is actually the case.

In short, GEE is a two-stage method in which the autocorrelation structure is treated as a nuisance to be adjusted for. Stage 1 of GEE involves estimating the 'working correlation matrix', the structure of which the user must specify prior to estimation; to specify this matrix correctly, the user must declare the `occasion` variable. Stage 2 of GEE uses the estimated working correlation matrix to adjust the estimates of the logistic model parameters and standard errors for autocorrelation.

For PA linear models, the structure of the residual autocorrelation can be estimated from the data, and used to choose the working correlation matrix. However, as we have discussed, there is no way of doing this for PA logistic models because the residuals are hidden from us (while we can estimate the autocorrelation between the binary outcomes, this is not generally the same as that for the residuals). Instead, we can fit the model using GEE with different working correlation matrices, and use an appropriate goodness-of-fit criterion to establish which is best.

The four main types of working correlation matrix structure are:

- Independence: The residuals are mutually independent (equivalent to a standard logistic regression model). Without adjustment, the standard errors obtained using this method will be too small unless there is no or very little autocorrelation.
- Exchangeable: Every pair of residuals on a subject has the same correlation so that, for example, the residuals at occasions 1 and 2 have the same correlation as the residuals at occasions 3 and 5, and so on; this is also known as 'compound symmetry'.
- Autoregressive: The correlation decreases exponentially as the time between measurements increases, so that if ρ is the correlation between residuals one occasion apart, then ρ^2 is the correlation between pairs of residuals two occasions apart, and so on, getting smaller and smaller as the gap increases.
- Unstructured: The correlation between a particular pair of residuals is different to that for all other pairs. So the residuals at occasions 1 and 2 have correlation ρ_{12}, which is distinct from the correlation between residuals at occasions 2 and 3 ρ_{23}, and so on.

We herein refer to a PA model fitted using GEE with an exchangeable working correlation matrix as the 'exchangeable' PA model, with the 'independent' PA, 'autoregressive' PA, and 'unstructured' PA models similarly defined.

The exchangeable and autoregressive working correlation matrices both involve one parameter, ρ, whereas the unstructured matrix (as its name suggests) makes no assumptions about structure but involves $(T-1)T/2$ parameters to represent the autocorrelation between the residuals at T occasions. For instance, in our example, there are up to 18 measurement occasions and so the unstructured working correlation matrix has $18 \times 17/2 = 153$ parameters. In practice, the unstructured working correlation matrix should always be used for examples involving few measurement occasions, but when there are many occasions it is often inestimable (i.e. the fitting routine is unable to estimate it and will output only an error), although this is not the case in our illustrative example.

One way to choose between different matrices is to use the quasi-likelihood information criterion (qIC) (Pan 2001). The qIC comprises an overall measure of goodness-of-fit and a penalisation for model complexity (i.e. the number of parameters in the working correlation matrix and predictor variables in the model). Hence, for two models with the same goodness-of-fit, the qIC will indicate that the model with the fewest parameters is to be preferred. We discuss how to use qIC further on in Section 6.

4.4 SS linear models

Subject specific (SS) models handle autocorrelation by including a unique 'effect' for each individual subject, which is separate from the occasion-specific residual. In contrast to PA models, there are many different ways of specifying the individual-level effects in SS models, and many different ways of estimating the resulting models.

An example of a SS linear model is a two-level random intercepts model with the individual subject at level two and the occasion at level one. It can be written as

$$y_{it} = \beta_{0i}^{SS} + \beta_1^{SS} x_{it} + e_{it},$$

where $\beta_{0i}^{SS} = \beta_0^{SS} + u_{0i}$ is the random intercept comprising the fixed intercept ${\beta_0}^{SS}$ and the individual-level residual, or 'random effect', u_{0i}. The classical assumption is that the individual-level residuals are normally distributed with mean zero and homoskedastic variance. The occasion-level residuals $e_{i1}, e_{i2}, \ldots, e_{i18}$ are taken to be independent of each other (and of u_{0i}) just as for standard regression models.

The residuals can be thought of as representing the combined effect on mental health of the variables omitted from the model. Hence, a random effect accounts for autocorrelation by explaining the omitted *time-invariant* variables for an individual subject. This is the key difference between SS and PA models: the PA model makes no explicit assumptions about any random effects or between-subject differences. However, the random intercepts model implies the *same* exchangeable autocorrelation structure as the exchangeable PA model.

Random intercept models can be extended through the addition of 'random slopes'. To construct a random slopes version of the model above, we would specify $\beta_{0i}^{SS} = \beta_0^{SS} + u_{0i}$ (the random intercept) and the random slope $\beta_{1i}^{SS} = \beta_1^{SS} + u_{1i}$, where u_{0i} is the same random effect as before, and u_{1i} is another normally distributed random effect, which can be correlated with u_{0i}. This model allows the effect of x_{it} on mental health to vary between subjects. In practice for longitudinal data, it is common to have a covariate corresponding to the time-point to model 'growth' in the outcome over the study; a random slope for time thus allows each subject to follow a different growth trajectory or trend.

In our example in Section 6, where we use age to index time, we consider a simple linear growth relationship between age and mental health, where the log-odds of poor mental health can increase or decrease linearly with age, at the same rate (random intercept) or different rates (random slope) for each individual. More complex relationships between the outcome and time can be modelled, for example, by including quadratic or higher order polynomial terms (e.g. by including age^2, age^3, etc. in our example). Each additional parameter can be made subject specific by allowing it to vary randomly across individuals in a random slopes model. However, each added parameter (fixed or random) complicates the model and may make its conclusions more difficult to interpret. As a rule, we therefore recommend that readers consult manuals and worked examples of such models before developing complex random slopes models.

Random slopes models allow more complex autocorrelation structures than the simple exchangeable autocorrelation allowed by random intercepts models. The correlation between pairs of residuals is allowed to depend on the time (as we have chosen to index it in the model) between measurement occasions in a complex fashion, but there is no direct correspondence between the autocorrelation structure implied by a random slopes model and the autoregressive or unstructured autocorrelation structures we can specify directly for PA models.

4.5 SS logistic models

The logistic random intercepts model is likewise written

$$\text{logit}\Pr(y_{it} = 1|x_{it}, u_{0i}) = \beta_{0i}^{SS} + \beta_1^{SS} x_{it},$$

where again $\beta_{0i}^{SS} = \beta_0^{SS} + u_{0i}$ is the random intercept, and the occasion-level residual is implicit in the choice of the logit link. The individual-level residual is normally distributed as before and plays the same role as for linear models, namely, accounting for autocorrelation due to the omitted characteristics of the individual subject common to all time points.

The two variance components (for occasion and individual) are on two different scales and cannot be compared directly. However, a useful measure is the intra-class correlation (ICC), which is the proportion of the total residual variation that is due to differences between individuals. The ICC is output by most routines for fitting random intercepts models, but it cannot be estimated from PA models.

The use of a non-linear 'link' function between the probability and the linear predictor on the right hand side leads to a difference in interpretation between the β parameters of the SS models and PA models. We discuss this distinction in more detail in Section 5, and again in the context of our data example in Section 6.

4.6 Fitting SS models

Different estimation approaches can be used for random effects (or mixed effects) models and the various statistical packages offer one or more different methods which can lead to differences in the estimates and standard errors produced. Estimation of the parameters involves maximising the likelihood function associated with the model, and for non-linear models like the logistic this can be computationally intensive for large sample sizes and models with multiple random effects. For our example, we will be using the default maximum likelihood estimation method available with the `xtmelogit` procedure in *Stata*, that is, a Gauss-Hermite adaptive quadrature approximation with 7 integration points.

5 Why is there a difference between SS and PA coefficients?

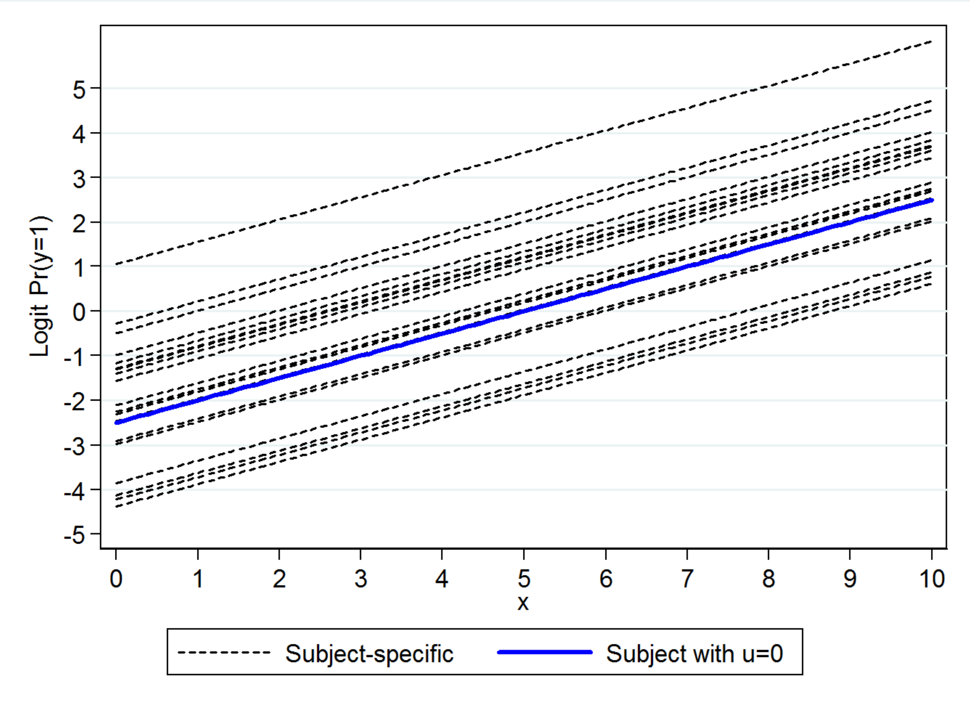
We noted in Section 4 that the coefficients in SS and PA logistic models have different interpretations. Furthermore, as we will see in the analysis presented in Section 6, the SS coefficients will generally be larger than the PA ones. In this section, we use a simulated data set to explain why this is the case for random intercepts SS and PA exchangeable models, which are consistent in this context.

To begin, just consider one time point *t* in a longitudinal study. For each subject *i* we have a binary variable y_{it} and a continuous predictor x_{it} at that time point. The predictor variable is uniformly distributed and so is equally likely to take any value between 0 and 10. Now suppose that the relationship between the predictor and the binary outcome at a given time follows the SS random intercepts model

$$\text{logit Pr}(y_{it} = 1|x_{it}) = -2.5 + 0.5x_{it} + u_{0i},$$

where the random effect u_{0i} is normally distributed with mean 0 and variance 2. This model tells us how the log-odds of a positive response for any subject varies with x_{it}, which we will denote as x.

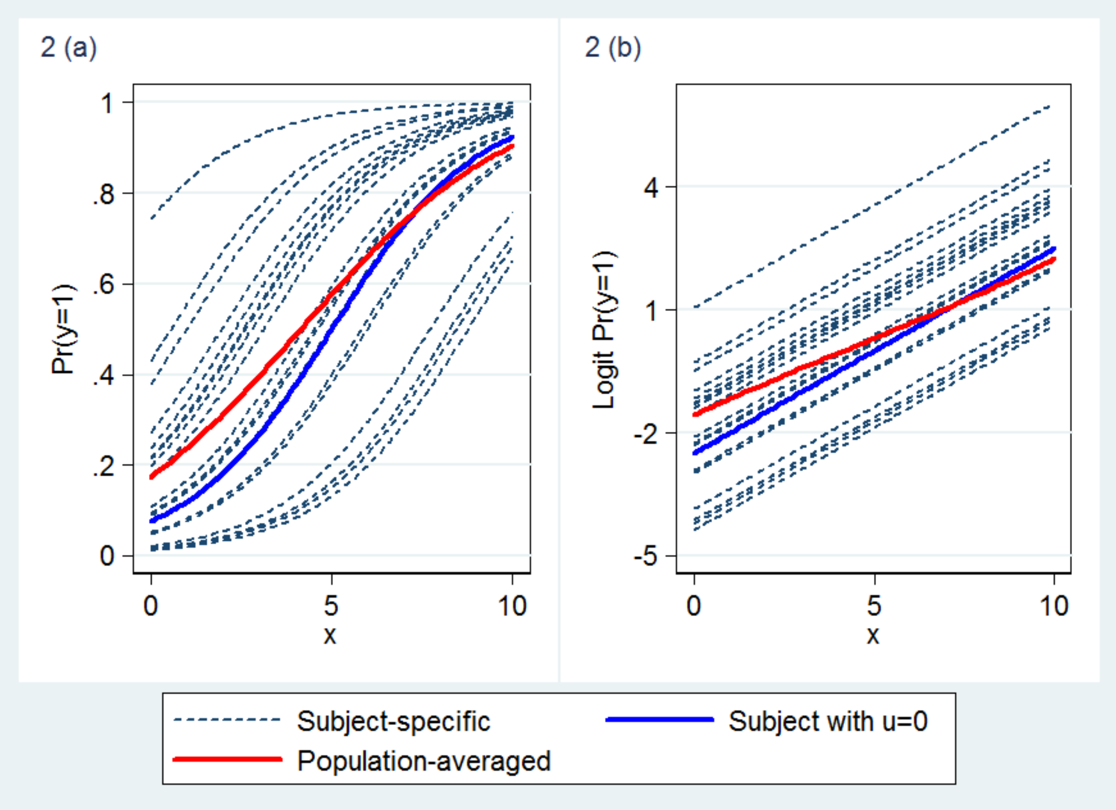
Figure 1. Subject specific effects of $x_{it} = x$ on the log-odds that $y_{it} = 1$



From figure 1, we can see how the SS log-odds increase with x for 20 randomly selected subjects. The relationship is linear, and the blue line shows the log-odds for the mean subject ($u_{0i} = 0$), which happens to equal the log-odds for the median subject because u_{0i} is normally distributed. As this is a random intercepts model, the other lines are parallel to that for the mean/median subject (the slopes can also vary for different subjects under a random slopes model).

However, the situation changes when we consider the relationship between the *probability* of a positive response and x because the relationship is now non-linear. Figure 2a shows the relationship between the positive response probability and x for the same 20 subjects as displayed in figure 1. We can now see that the change in x is smaller for subjects with large positive values of $u_{0i} > 0$ (i.e. the curves near to the top of the figure) than it is for the others. This is because there is a ceiling for probabilities: no matter how much x increases, the probabilities cannot exceed 1 but only get closer to it; this ceiling constrains the effect of x. There is the opposite floor effect for subjects with large negative values ($u_{0i} < 0$).

Figure 2. Subject specific and population average effects of x on (a) the probability that $y_{it} = y = 1$ and (b) the log-odds that $y_{it} = 1$. The mean probability at $x = 5$ and $u_{0i} = 0$ is 0.5



Further displayed in figure 2a is the red PA curve, which is obtained by taking the mean of the SS probabilities at each value of x. The SS curve for an individual at the mean of the random effects distribution ($u_{0i} = 0$) is also shown in blue. Why are the two curves different? It is because of the non-linearity of the logistic function that the probability for the mean subject will not equal the mean probability. Instead, the probability at $u_{0i} = 0$ now corresponds to the *median* probability because, unlike the mean, the median is unaffected by the logistic transformation.

From figure 2a, it can be seen that for values of x between 0 and 7, the PA or mean curve lies above the median curve. This is because the greater part of the spread of the SS curves at these values lies above the blue line, which indicates a positive skew and the median probability exceeding the mean probability. For values of x greater than 7, however, the greater part of the response-probability spread lies below the median, which pushes the mean below the median.

Figure 2b displays the effects of x on the log-odds scale (this plot is identical to figure 1 apart from the addition of the PA line which was obtained by applying the logistic transformation to the PA probabilities in figure 2a). We can see that the coefficient of x from the PA model (β_1^{PA}) is the slope of the PA line in figure 2b, while the coefficient from the SS model (β_1^{SS}) is the slope of the SS line for an individual with $u_{i0} = 0$, and we can see that $\beta_1^{SS} > \beta_1^{PA}$ for most of the time.

The relationship between the PA and SS effects observed in figure 2b holds more generally. It can be shown that the random intercepts SS coefficients are related to the PA coefficients by

$$\beta^{PA} \approx \sqrt{\frac{3.29}{3.29 + \sigma_u^2}}\, \beta^{SS}$$

where σ_u^2 is the between-subject (or random effect) variance from the SS model; and 3.29 is the variance of a standard logistic distribution, which is the within-subject variance under the logistic model. The quantity under the square root is the proportion of the variance that is unexplained by x for the SS model relative to the PA model. It serves to 'scale down' the SS coefficient to obtain the PA coefficient, and implies that the SS and PA coefficients are equal if there is no between-subject variation. The greater the between-subject variation, the greater the SS coefficient is compared to the PA one. This relationship can be extended to more complex random effect structures such as random slopes (Zeger, Liang & Albert, 1988).

A consequence of this is that one should not report the SS effect simply because it is larger than the PA effect. It does not mean that the SS model shows a 'stronger' effect than the PA model; it simply means that there is between-subject variation: the two measures are equivalent but different. However, we should note that the relationship above is only an approximation, and it is possible for β^{SS} to be close to, or even less than, β^{PA} even when the random effect variance is large. Finally, although the SS and PA coefficients can be very different, the ratio of a parameter estimate to its standard error will in general be similar for the two models; thus, significance tests will be unaffected by the choice of model.

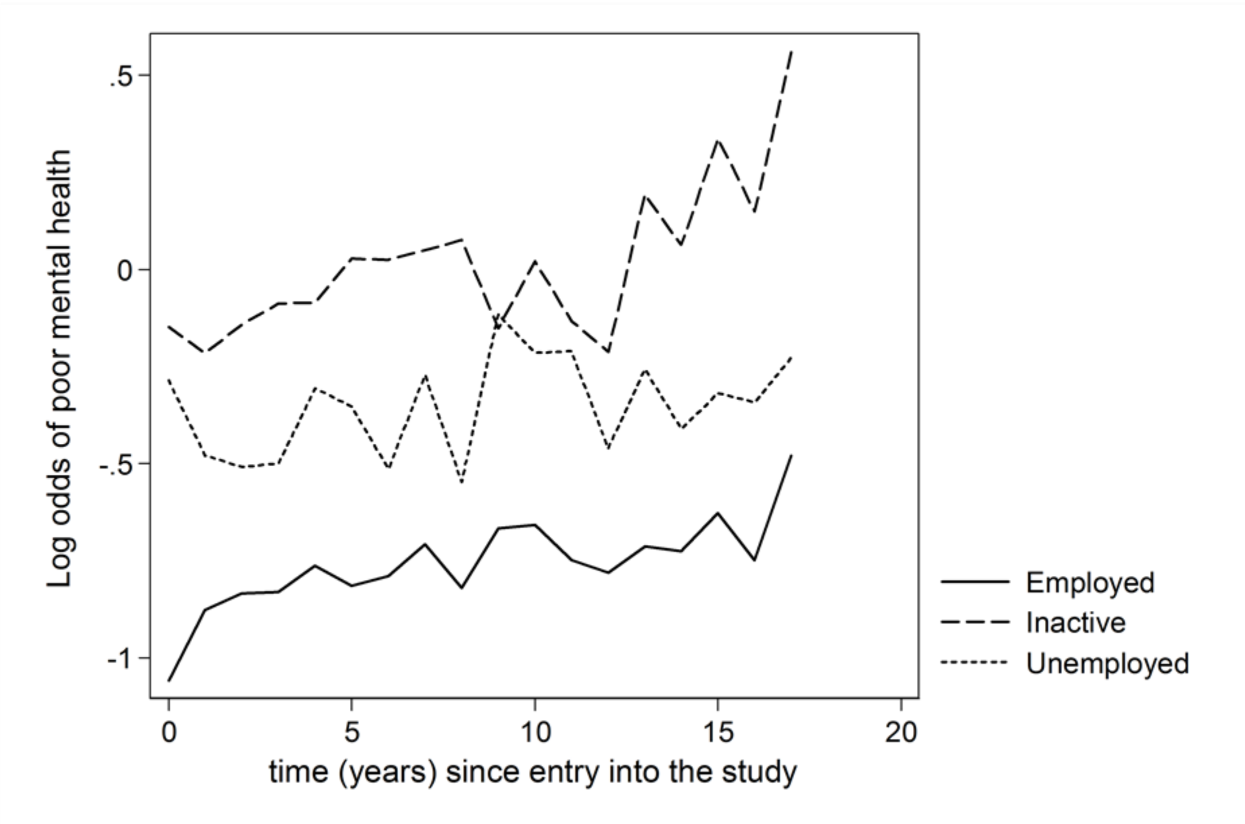
6 Data Analysis

6.1 Preliminaries

We use the BHPS data introduced in Section 2 to analyse the relationship between mental health and employment status. A suitable longitudinal analysis of these data is important because the relationship is a dynamic one that varies from wave to wave, and a simpler analysis would risk losing important information about variation or change in mental health over time. We conduct a simple analysis here to illustrate the most important points regarding the comparison of PA and SS logistic models, but refer the reader elsewhere for further analyses of the same data (Steele et al., 2013).

Before carrying out any modelling, we first look at the raw data. In the figures below, we use `occasion` as a subject specific measure of time for each subject (`occasion` being annually spaced), and in figure 3 we can see how `ghq_case` varies over time for each category of `empl1`. We plot the log-odds of poor mental health (i.e. `ghq_case = 1`) because it is this that is being modelled by both the PA and SS logistic regression models.

Figure 3. Relationship between the log-odds of GHQ "caseness" and the time since entry into the study by employment status at occasion 1



It can be seen from figure 3 that, when looking only at the log-odds of `ghq_case` at the final occasion, and comparing it with the log-odds at the first, that there is an increase in both the employed and inactive groups, but there is less pronounced change among the unemployed. We can also see that the log-odds of `ghq_case` varies for all categories throughout the study period, and this variation cannot be captured without using all the available outcome measures. It is worth emphasising at this point that figure 3 represents the population average log-odds of `ghq_case` in each employment category, which can be very different to its subject specific equivalent.

Figure 4. Relationship between the log-odds of GHQ 'caseness' and the time since entry into the study by time-varying employment status

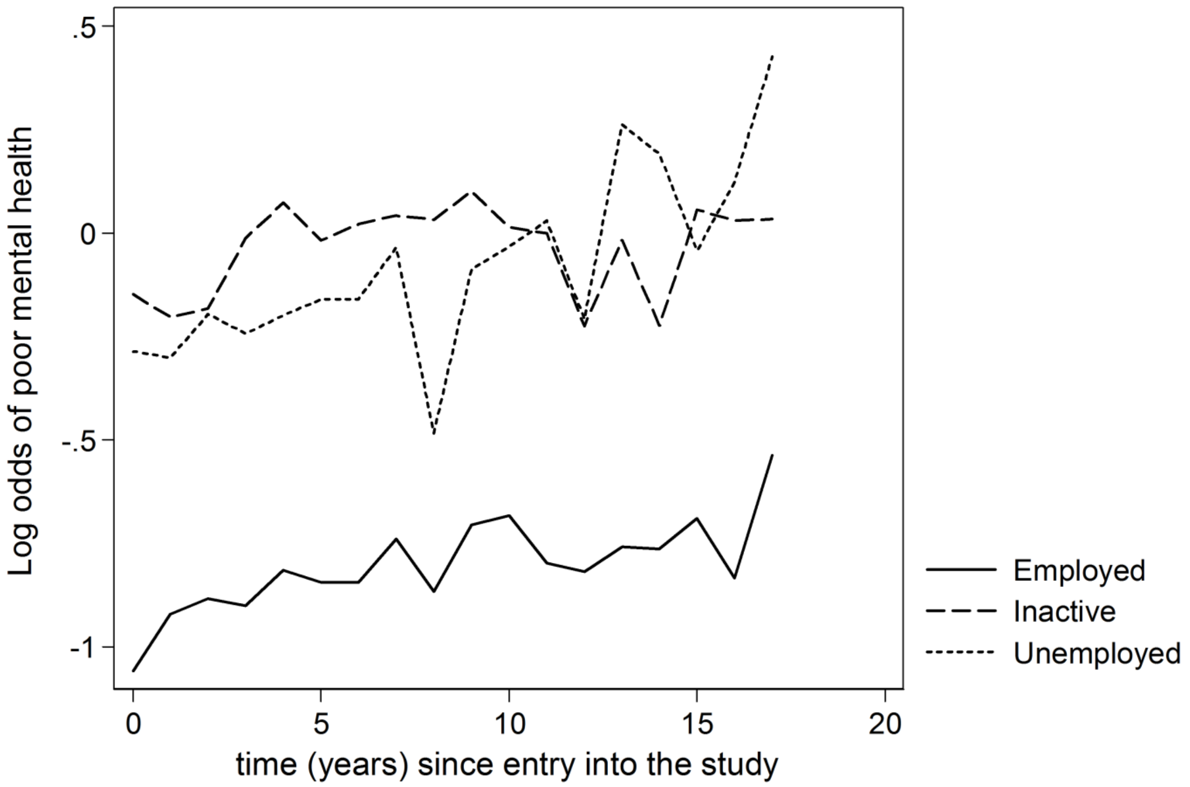


Figure 4 shows the relationship between the log-odds of being a GHQ case/having poor mental health and the time-varying employment variable (`empl`). Comparing figures 3 and 4 highlights that the log-odds of poor mental health among the unemployed increased during the survey. As we shall see later, the difference between the two figures can partly be explained by the change in the employment distribution over the observation period, which confirms the importance of treating employment status as a time-varying predictor. However, we first analyse these data using `empl1` as our explanatory variable, and present estimates of the effect of employment status on mental health using PA and SS models.

6.2 Simple models with no time-varying predictors

To begin, we fit a simple model in which the mental health outcome varies over time but the employment status predictor is that from occasion 1. Using a symbolic 'pseudo code' notation, this model can be written as

```
Log-odds (ghq_case = 1) =
intercept + unemployed_1 +
inactive_1 + agec_1
```

We use this symbolic notation throughout to represent the linear predictor (sometimes called the 'fixed part') common to both PA and SS models. On the left hand side, the outcome `ghq_case` is measured at each occasion and so varies over time. On the right hand side, however, `intercept` represents the constant term (its coefficient is the intercept β_0); `unemployed_1` and `inactive_1` are dummy variables for whether subjects were, respectively, unemployed and inactive on first entry to BHPS; and `agec_1` represents the subject's mean-centred age (that is, the difference between the subject's age and the sample mean of age) at the first occasion. Hence, in contrast to the left hand side, none of the variables in the linear predictor change over time.

We adjust for age at occasion 1 but note that, in substantive (rather than illustrative) applications, we may wish to adjust for a wider range of variables (for example, to adjust for confounding bias). We use mean-centred age so that β_0 can be interpreted as the log-odds of GHQ caseness for an employed subject of mean age. Centring is important for any continuous predictor variable for which we wish to add a random slope (see section 6.3 for an example).

We now fit the symbolic model introduced above using both the PA and SS approaches. These analyses were conducted using *Stata* and the following functions: `logit` for the simple logistic; `xtlogit` with options `pa` and `robust` to fit PA models using GEE; and `xtmelogit` to fit random effects models using the `mle` fitting option. The code we used for this data analysis is provided in the supplementary material and the results are presented in table 3.

The `robust` option uses a 'sandwich estimator' that takes into account that the working correlation matrix may be incorrect. All GEE routines will have a `robust` option (or its equivalent) and this should always be used.

Table 3. Results from fitting the basic model without temporal trend. The models fitted are the simple logistic regression, the PA with independent correlation matrix (Ind), the SS with random intercepts (RI), the PA with exchangeable (Exch), first-order autoregressive (AR1) and unstructured (Uns) correlation matrix. The table shows, for each model parameter, the parameter estimate and (robust) standard error. For the SS_RI model, we also provide the Intra Class Coefficient (ICC) which is a measure of within-individual autocorrelation. For each model, a model diagnostic is provided using either the Log Likelihood (LogLik) for the simple logistic and SS models, or the qIC (measure of model fit similar to the AIC penalising the quasi-likelihood to reflect the complexity of the model; Cui, 2007) for the PA models

	Simple Logistic	PA_Ind	SS_RI	PA_Exch	PA_AR1	PA_Uns
Intercept	-0.774	-0.774	-1.207	-0.800	-0.782	-0.813
	(0.009)	(0.020)	(0.027)	(0.018)	(0.021)	(0.018)
inactive_1	0.684	0.684	0.982	0.644	0.617	0.657
	(0.029)	(0.058)	(0.076)	(0.053)	(0.061)	(0.053)
unemployed_1	0.450	0.450	0.696	0.462	0.455	0.487
	(0.025)	(0.050)	(0.070)	(0.046)	(0.055)	(0.045)
agec_1	0.006	0.006	0.007	0.005	0.005	0.006
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Intercept variance			2.964			
			(0.083)			
ICC			0.474			
			(0.007)			
qIC/LogLik	-45394.7	90820.8	-38793	90829.1	90829.6*	90840.5

**1,713 individuals are omitted from estimation due to unbalanced unequal observations (that is individuals with observations on non-consecutive time points).*

The first column in table 3 contains the results obtained from fitting a simple logistic model that does not account for autocorrelation. The second column (`PA_Ind`) contains the results from fitting the independence PA model using GEE, namely, the independence PA. If we do not use the `robust` option for the GEE, we would expect both sets of estimates to be exactly the same. However, as we have discussed, GEE should always be estimated using `robust`, and so we can see that the estimated standard errors are larger for the PA independence model, because the autocorrelation in the data has been allowed for.

Next we consider the random intercepts (RI) model (`SS_RI`) and the exchangeable PA model (`PA_Exch`). As noted previously, both of these models require that the autocorrelation structure is exchangeable, and both have larger standard errors than the simple logistic model because autocorrelation is accounted for.

Perhaps the most salient feature of table 3 is that the RI model estimates all have larger absolute values than the exchangeable PA model estimates. The estimated PA exchangeable logistic model for person i at occasion t is

$$\begin{aligned}\text{logit Pr}(\text{ghq\_case}_{it} = 1|x_{it}) \\ &= -0.800 + 0.644 \times \text{inactive\_1}_i \\ &+ 0.462 \times \text{unemployed\_1}_i \\ &+ 0.005 \times \text{agec\_1}_i,\end{aligned}$$

and we can interpret the estimates of the two employment status dummies in the usual manner for logistic models. The odds of being a GHQ case for employed people of mean age at occasion 1 are $\exp(-0.800) = 0.44$; the odds *ratio* of being a GHQ case for the unemployed compared to the employed, conditional on age at occasion 1, is $\exp(0.462) = 1.6$, which means that the unemployed are 60 percent more likely to be GHQ cases than the employed; the odds ratio for the inactive relative to the employed is similarly obtained.

If we now look at the estimated random intercepts SS model

$$\begin{aligned}\text{logit Pr}(\text{ghq\_case}_{it} = 1|x_{it}, u_{0i}) \\ &= -1.207 + 0.982 \times \text{inactive\_1}_i \\ &+ 0.696 \times \text{unemployed\_1}_i \\ &+ 0.007 \times \text{agec\_1}_i + u_{0i},\end{aligned}$$

where u_{0i} is normally distributed with mean 0 and variance 2.9. The presence of the random effect (the u_{0i} term) means that each individual subject has his own regression equation. Using this model, for subjects with the same value of u_{0i}, the SS odds of being a GHQ case are $\exp(-1.244) = 0.30$ among subjects employed at the start of the study, and the SS odds ratio of being a GHQ case for unemployed subjects compared to employed ones is $\exp(0.696) = 2.0$.

As was discussed in section 5, the odds ratios obtained using RI models are usually larger than those from the exchangeable PA model (Neuhaus, Kalbfleisch & Hauck, 1991), but this is because both are different measures of the same association. It is important to remember that SS model results should not be reported here just because the odds ratio is larger: it does not mean that the SS model has estimated a 'stronger' effect, it just means that the two coefficients are different measures of association (even though both odds ratios equal 1 if there is no association).

The two approaches are complementary and the most relevant depends on the focus of a particular analysis. In our example, the effect of employment status can legitimately be reported as either a population average or a subject specific effect. If employment status were randomised and fixed throughout the study, as in our hypothetical example in Section 3, then the PA estimate would be akin to a 'causal odds ratio', summarising the effect of employment status across the experimental population rather than for any particular subject. On the other hand, the SS estimate pertains to the effect of employment status on any given subject, that is, what will happen to that subject if he changed only his employment status. (Of course, it is important to remember that estimates based on non-experimental data will only be causal if confounding bias has been adjusted for.)

In fact, every SS model has a corresponding PA model (Lee & Nelder, 2004). For SS logistic models (including random intercepts and random slopes), there is a rule for converting the SS parameters to have a PA interpretation (see section 5). This rule is an approximation based on the assumption that the random effects are normally distributed[1]. Using this rule, which is simply a refinement of the equation displayed in section 5, we can see that there is very little difference between these two sets of estimates. The 'marginalised' estimates of employment at occasion 1 compared to the estimates for the exchangeable PA model are shown in table 4. As we would expect, the marginalised estimates are not exactly equal, partly because the formula is an approximation, and partly because of model differences (e.g. the PA model does not assume a normally distributed random effect).

Table 4. Population average estimates from the exchangeable PA model as per table 3 and derived from the random intercepts logistic model

	RI	PA_Exch
Employed_1	-0.848	-0.800
Inactive_1	0.690	0.644
Unemployed_1	0.489	0.462

Unfortunately, there is no rule for converting PA parameters to SS ones because a PA model can correspond to many different SS models. In some quarters, this is perceived to be a strength of the PA approach, because no assumptions appear to have been made about the distribution of the residuals and random effects. However, a counter-argument to this is that these assumptions are hidden from us and cannot be inspected, whereas the assumptions of SS models are clear and can be relaxed if more advanced SS modelling approaches are used.

While the random intercepts and exchangeable PA models allow for the same type of autocorrelation structure, it is always advisable to explore alternative autocorrelation structure assumptions to improve the estimation accuracy further and ensure the estimates lie as close as possible to the truth. The `robust` option inflates the standard errors, but the estimates may be far too large if the choice of working correlation matrix is poor. For PA models, this can be done by fitting PA models using GEE with two, more complex, autocorrelation structures: the autoregressive and unstructured PA models. These results are displayed in the final two columns of table 3.

We can see from table 3 that the absolute values of the unstructured PA (`PA_Uns`) model estimates are larger than those for the autoregressive PA (`PA_AR1`) model, with the exchangeable PA model estimates lying somewhere between the two. We can use the qIC to choose between the three PA, where the smallest qIC indicates the 'best' fitting model in terms of the balance between goodness-of-fit and simplicity. The code used to calculate the qIC is given in the supplementary material.

The first point to note is that the qIC of the autoregressive PA model is not directly comparable to that for the other two. This reflects that GEE estimation using the autoregressive correlation structure, as implemented in *Stata*, requires that all the individuals in the data set are observed for consecutive occasions and as such does not handle gaps between occasions (users of other software must check if the situation is the same for them). From the data summary in table 3, we can see that only 57,592 (out of 72,173) observations on 7,479 (out of 9,192) individuals have been used. To be comparable, all three PA models should be fitted to the same sample so that the qICs can be compared. We take this approach in the analyses to follow (all of the models are fitted to the reduced sample of 7,479 observations), but simply exclude the autoregressive PA model for consideration here.

Now we must decide which of the independence, exchangeable PA or unstructured PA models to choose. Looking at the estimated working correlation matrix for the exchangeable and unstructured PA (correlation matrix for the independence PA model is the identity matrix), the correlations between different pairs of occasions appear to be unequal, with some pairs having larger correlations than others (the matrix is not shown, but we show how to obtain it using *Stata* in the supplementary material). This would seem to favour the unstructured PA estimates, but the qIC for the unstructured PA model is larger than for the exchangeable one, most likely because the unstructured working correlation matrix has $18 \times 17/2 = 153$ parameters. Similarly the qIC for exchangeable PA is larger than for the independence one. We would not, however, recommend choosing the independence PA model; for a poor choice of correlation matrix, the robust estimates of the standard errors will be overinflated and larger than necessary. Regarding the choice between the exchangeable or the unstructured PA models, either could be used because the parameter estimates for these two models are fairly close to one another.

6.3 Models that allow for time-varying age

How can we relax the exchangeable autocorrelation structure implied by the random intercepts SS model? The approach we follow is to extend the linear predictor of the simple model to explain changes in mental health over time. The rationale for this is that the exchangeable (or even independence) autocorrelation structure becomes more plausible as we explain more systematic variation through the linear predictor.

To do this, we can extend the model to allow a subject's log-odds of GHQ caseness to vary as they age, and therefore capture some of the trend in log-odds we saw in figure 3. By doing this, we assume that changes in mental health over time are driven primarily by a subject's age; the impact of when subjects were born ('cohort effects') and the calendar year when mental health was measured ('period effects') is thus taken to be less important. Using the symbolic notation, this model is written as

```
Log-odds (ghq_case = 1) =
intercept + unemployed_1 +
inactive_1 + agec
```

recalling that we use mean-centred age `agec`, which increases with time and replaces the first-occasion age variable used in section 6.2.

The results obtained from fitting this model using different PA and SS approaches are presented in table 5. Recall that, in contrast to the results in the previous table, we fit the different PA and SS models to the 7,479 individuals in order to make the other results comparable to those obtained using the autoregressive PA model.

Table 5. Results from fitting the model with an age trend. The models fitted are the SS with random intercepts (RI) and random slopes (RS), the PA with exchangeable (Exch), first-order autoregressive (AR1) and unstructured (Uns) correlation matrix. The table shows, for each model parameter, the parameter estimate and (robust) standard error. For the SS models, we also provide the ICC. For each model, a model diagnostic is provided using either the Log Likelihood (LogLik) for the SS models or the qIC for the PA models. These models are fitted to the reduced dataset of 7,479 individuals

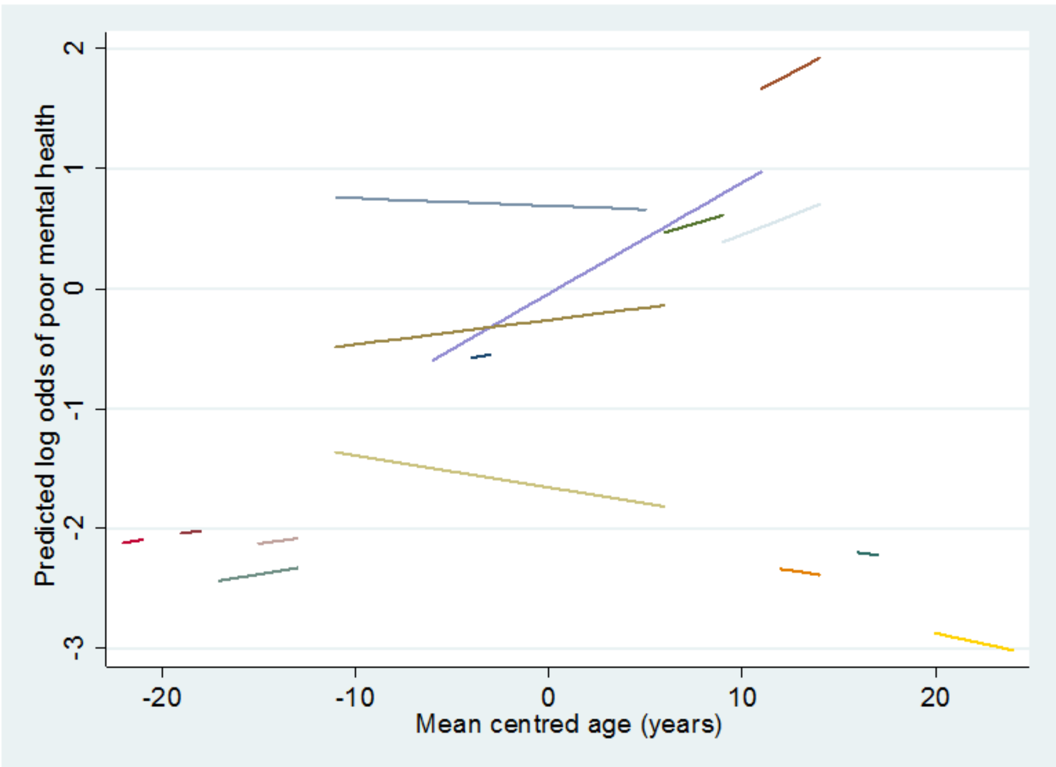
	PA_Exch	PA_AR1	PA_Uns	SS_RI	SS_RS
intercept	-0.819	-0.808	-0.833	-1.238	-1.189
	(0.019)	(0.020)	(0.019)	(0.029)	(0.030)
inactive_1	0.570	0.610	0.590	0.872	0.896
	(0.057)	(0.060)	(0.057)	(0.083)	(0.093)
unemployed_1	0.481	0.470	0.501	0.730	0.741
	(0.053)	(0.055)	(0.052)	(0.081)	(0.080)
agec	0.008	0.007	0.008	0.012	0.010
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Intercept				3.052	2.659
variance				(0.096)	(0.103)
Slope variance					0.006
					(0.0005)
Intercept/slope					0.053
covariance					(0.005)
ICC				0.481	0.447
				(0.008)	(0.010)
qIC/LogLik	90764.5	90755.1	90759.5	-30918.3	-30738.6

Focussing on the SS_RI model (penultimate column in table 5), we note that the parameter for agec in the SS model describes a within-subject trend in mental health with age. More generally, SS models are the more natural choice if within-subject characteristics of growth are the focus of the analysis. The corresponding effect in the PA models is a measure of change in the population average and less useful for describing subjects' growth.

We can extend the random intercepts model by allowing a random slope for the effect of agec. It is sensible to consider this model because, as we can see from figure 5, there is substantial between-subject variation in the time trend. The variance of the random slope measures between-subject variation in age trend, and allows for a more complex autocorrelation structure than the random intercepts model. The use of mean-centred age is particularly important here because the random intercept variance can be interpreted as the between-individual variance at the mean age (rather than age zero), and the intercept-slope covariance is between an individual subject's log-odds of GHQ caseness at the mean age and his rate of change. Centring continuous predictor variables with random slopes can also stabilise the fitting of these models.

The results from fitting the random slope model are displayed in the final column of table 5 (SS_RS). The likelihood ratio test of including the random slope is obtained by taking the difference between the log-likelihoods for the random slopes and intercepts models and multiplying it by -2 to get 359, which, compared to a chi-square on 2 degrees of freedom, reveals substantial evidence to support its inclusion. The random slopes model allows a different effect of agec for each individual, which implies that two different individuals will experience different rates of change in the odds of suffering from poor mental health with age (even if they are in the same employment category). If we were to convert the SS parameter to a PA one using the same method as used for table 4, we would find that the relationship would be quadratic rather than linear (due to the presence of both random intercept and random slope variance, as well as the covariance between the two random terms), which further emphasises that PA models say little about within-subject change.

Figure 5. Sample of individual predicted trajectories based on the random slopes model



6.4 Models that allow for time-varying employment status

Finally, we complete the picture by including employment status as a time-varying variable to utilise fully the information contained in the data. Many of the sample subjects changed employment status at least once during the study with the proportion of men in each employment category varying between waves.

To model time-varying employment status, we include `empl` as a predictor rather than `empl1`. We assume that this prevents reverse causation because `empl` is asked retrospectively, and the subject's current employment status will have been determined at a point prior to the current occasion. Conversely, the subject's mental health varies from day to day, and is measured on the day of the study.

Using our symbolic notation, we can write this model as

```
Log-odds (ghq_case = 1) = intercept
   + unemployed + inactive + agec
```

The variable `agec` is included on the basis of our previous investigations. While `ghq_case` is measured at the time of the survey, employment status is actually determined in the between-time point interval, and so – for the purposes of this application – we take it to be temporally antecedent to the mental health measure. However, we discuss in Section 7 the use of 'lag' variables (namely, including employment status from time points prior to the current one) as a further guard against reverse causation.

As previously, we fitted five different models to the same dataset: exchangeable PA (`PA_Exch`), autoregressive PA (`PA_AR1`), unstructured PA (`PA_Unst`), random intercepts (`SS_RI`), and random slopes (`SS_RS`).

Table 6. Results from fitting the model with temporal trend and time-varying employment. The models fitted are the PA with exchangeable (Exch), first-order autoregressive (AR1) and unstructured (Uns) correlation matrix, the SS with random intercepts (RI) and random slopes (RS), and the PA with AR1 correlation matrix and SS with RS for the model with interactions. The table shows, for each model parameter, the parameter estimate and (robust) standard error. For the SS models, we also provide the ICC. For each model, a model diagnostic is provided using either the Log Likelihood (LogLik) for the SS models or the qIC for the PA models. These models are fitted to the reduced dataset of 7,479 individuals

	PA_Exch	PA_AR1	PA_Unst	SS_RI	SS_RS
intercept	-0.813	-0.821	-0.829	-1.229	-1.220
	(0.020)	(0.021)	(0.020)	(0.029)	(0.032)
agec	-0.0005	-0.0004	-0.0008	-0.0005	-0.0001
	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0005)
inactive	0.464	0.609	0.503	0.688	0.676
	(0.039)	(0.040)	(0.037)	(0.048)	(0.049)
unemployed	0.576	0.647	0.611	0.844	0.843
	(0.041)	(0.041)	(0.040)	(0.054)	(0.054)
ICC				0.478	0.489
				(0.008)	(0.010)
Variance (intercept)				3.017	3.146
				(0.095)	(0.130)
Variance (age)					0.0002
					(0.00004)
Covariance					-0.013
(intercept/wave)					(0.002)
Model Diagnostic (LL or qIC)	90475	90347	90445	-30850	-30800

All the estimates from the various PA models are similar to each other apart for the inactive category of employment, the coefficient of which seems to be influenced by the working correlation structure (although the qIC indicates that the autoregressive PA seems to be the best model). For the SS models, the random slopes model is a better fit to the data than the random intercepts model, providing support towards growth-type models with individual-specific effects.

7 Discussion

In this tutorial, we have considered the differences between using population average (PA) and subject specific (SS) models for the analysis of longitudinal data.

In short, PA models are more appropriate for estimating the average effects of predictors (in our case, employment status) on outcomes. The parameters of PA models are most relevant to measuring the effect of time-invariant predictors; an example of this is in experimental settings (or observational settings where it can be justified that confounding bias has been adjusted for) for estimating the 'average' effects of predictor variables which correspond to a 'treatment' or 'exposure' of interest. However, PA models make no (explicit) assumptions about the distribution of the random effect, and so cannot be used to estimate between-subject variation or subject-level residuals. The nature of GEE estimation for PA models means we cannot use proper goodness-of-fit statistics based on likelihoods, and so must rely on ad-hoc tools like qIC.

The SS models we consider allow a different model for each subject through the use of random effects. In experimental/confounding-adjusted settings, the parameters of these models correspond to the effect of the treatment/exposure on each subject. If the target of the analysis is growth, or more general within-subject change, then SS models are more appropriate than PA models (because, in a nutshell, changes in averages are not the same as average changes for non-linear models). Random slopes can be used to increase the complexity of the SS models (although these models can be difficult to fit) at the expense of modelling assumptions like normality of the random effects. An advantage of SS models is that PA effects can be estimated using marginalisation. For logistic models with normal random effects, one can always use the formulae discussed in sections 5 and 6. Conversely, it is impossible to obtain estimates of the SS parameters from a PA model because there are many SS models that correspond to the same PA model (Lee and Nelder, 2004).

In our application, we handled the missing data problem by excluding any subject-occasion contributions with missing values from the data set. GEEs require that the data are Missing Completely At Random (MCAR) such that the missing values arose in a manner completely independent of the variables in the analysis[2]. On the other hand, SS models require only that the data are Missing at Random (MAR) such that the missing values arose in a manner that depends only on the variables we happen to observe. More generally, weighted GEE estimation can be performed to allow MAR data, and multiple imputation methods can be used for either approach to 'fill in' incomplete data sets under the MAR assumption (Carpenter & Kenward, 2013).

One of the powerful features of longitudinal data is that models with reverse causation can be avoided. In our application, we argued that using employment status to predict mental health, where both were measured at the same occasion, precluded reverse causation, but this argument may be unconvincing to some. To protect against this, one may use 'lagged' employment status from the previous occasion as a predictor instead; another example, used by Steele et al. (2013), is to use the between-occasion change in employment status as the predictor.

The power of longitudinal data to deliver 'causal', or 'policy-relevant', conclusions is limited unless the data come from a randomised experiment, and involves adjusting for confounding bias, just as analyses of cross-sectional data do. Recall that causal effects are not simply associations, but concern the change in a subject's outcome (mental health) if we intervene and change his employment status (e.g. from unemployed to employed), while holding everything else about the subject fixed. The problems of adjusting for confounding bias in longitudinal analyses like these are myriad, and necessitate the use of advanced approaches like simultaneous equation modelling (e.g. Steele et al., 2013), econometric panel data models (e.g. Baltagi, 2008) and marginal structural models (Robins, Greenland & Hu, 1999). Readers new to this subject should be aware that these are advanced techniques beyond the scope of this paper, and should generally avoid using the language of causality when describing any results obtained using the models described in this tutorial.

Acknowledgements

We thank George Leckie for providing *Stata* code to generate the figures presented in section 5.

Supplementary Material

Stata .do file corresponding to the analyses is available online as a Supplementary File.

The data used in the analyses are available on request from the authors. However, as we use BHPS data, anyone making such a request must also provide evidence that they are registered with the ESRC Data Archive.

References

Baltagi, B.H. (2008). *Econometric Analysis of Panel Data* (4th edn.). Chichester: Wiley.

Bollen, K.A., & Curran, P.J. (2005). *Latent Curve Models: A Structural Equation Perspective.* New Jersey: Wiley.

Carpenter, J., & Kenward, M. (2013). *Multiple Imputation and its Application.* London: Wiley.

Cui, J. (2007). QIC Program and Model Selection in GEE Analyses. *The Stata Journal 7*, 209-220.

Goldberg, D.P., Gater, R., Sartorius, N., Ustun, T.B., Piccinelli, M., Gureje, O., & Rutter, C. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychological Medicine, 27*, 191-197

ISER. (2010). *British Household Panel Survey: Waves 1-18, 1991-2009* (7th edn.). University of Essex, Institute for Social and Economic Research [original data producer(s)], Colchester, Essex: UK Data Archive [distributor].

Lee, Y., & Nelder, J.A. (2004). Conditional and Marginal Models: Another View. *Statistical Science 19*, 219-238.

Liang, K-Y., & Zeger, S.L. (1986). Longitudinal Data Analysis using Generalized Linear Models. *Biometrika 73*, 13-22.

Neuhaus, J.M., Kalbfleisch, J.D., & Hauck, W.W. (1991). A Comparison of Cluster-Specific and Population-Averaged Approaches for Analysing Correlated Binary Data*. International Statistical Review 59*, 25-35.

Pan, W. (2001). Akaike's Information Criterion in Generalized Estimating Equations. *Biometrics 57***,** 120-125.

Robins, J.M., Greenland, S., & Hu, F-C. (1999). Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. *Journal of the American Statistical Association 94*, 687-700.

StataCorp. (2011). *Stata Statistical Software: Release 12*. College Station, TX: StataCorp LP.

Steele, F., French, R., & Bartley, M. (2013). Adjusting for selection bias in longitudinal analyses using simultaneous equations modelling: The relationship between employment transitions and mental health*. Epidemiology* (in press).

Zeger, S.L., & Liang, K-Y. (1986). Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics 42*, 121-130.

Zeger, S.L., Liang, K-Y., & Albert, P.S. (1988). Models for Longitudinal Data: A Generalized Estimating Equation Approach*. Biometrics 44***,** 1049-1060.

Zeger, S.L. & Liang, K-Y. (1992). An overview of methods for the analysis of longitudinal data. *Statistics in Medicine 11,* 1825-1839.

Endnotes

[1] For a logistic model, the SS parameters can be marginalised by using the Zeger, Liang & Albert (1988) approximation:

$$\beta_i^{PA} = \beta^{SS}/\sqrt{1 + 0.346 \times \nu_i},$$

where β^{SS}represents the vector of SS parameter estimates, β_i^{PA} the corresponding vector of PA parameter estimates for observation *i*, and ν_i represents the variance of the random part of the linear predictor for observation *i*, which can be different for each individual when random slopes are fitted.

[2] GEE can be used for Missing At Random (MAR) data but the working correlation matrix cannot be consistently estimated using only the observed data, hence the estimates are consistent but can be very inefficient.

CONTENTS

EDITORIAL

86 - 87 **Editorial**
John Bynner

PAPERS

88 – 104 **Labour force sequences, unemployment spells and their effect on subjective well-being set points**
Cara L Booker, Amanda Sacker

105 – 118 **The role of respondent characteristics in tracking on longitudinal surveys: evidence from the UK Millennium Cohort Study**
Lisa Calderwood

STUDY PROFILE

119 – 135 **Design and characteristics of a new birth cohort, to study the early origins and ethnic variation of childhood obesity: the BiB1000 study**
Maria Bryant, Gill Santorelli, Lesley Fairley, Jane West, Debbie A Lawlor, Raj Bhopal, Emily Petherick, Pinki Sahota, Andrew Hill, Noel Cameron, Neil Small, John Wright, the Born in Bradford Childhood Obesity Scientific Group

COMMENT AND DEBATE

136 – 146 **Social-biological transitions: how does the social become biological?**
David Blane, Michelle Kelly-Irving, Angelo d'Errico, Melanie Bartley, Scott Montgomery

TUTORIAL

147 - 165 **Subject specific and population average models for binary longitudinal data: a tutorial**
Camille Szmaragd, Paul Clarke, Fiona Steele

This journal can be accessed online at: www.llcsjournal.org

Published by
The Society for Longitudinal and Life Course Studies
Institute of Education, 20 Bedford Way, London, WC1H 0AL
info@slls.org.uk http://www.slls.org.uk

SLLS Society for Longitudinal and Life Course Studies