VOLUME 5 | ISSUE 3 | **OCTOBER 2014** ISSN 1757-9597

Longitudinal and Life Course Studies: International Journal

Published by

SLLS Society for Longitudinal and Life Course Studies

Inside this issue

Co-sponsors:

VU university medical center

ISER INSTITUTE FOR SOCIAL & ECONOMIC RESEARCH

SRC SURVEY RESEARCH CENTER UNIVERSITY OF MICHIGAN

LLCS EDITORIAL BOARD

EDITORIAL COMMITTEE

Executive Editor

John Bynner, *Institute of Education, UK*

Health Sciences

Section Editor **- Scott Montgomery,** *Örebro University Hospital and Örebro University, Sweden*

Associate Editors **- David Blane**, *Imperial College Medical School, UK*

- **Rachel Knowles**, *Institute of Child Health, University College London, UK*

Social and Economic Sciences

Section Editor **- Richard Layte,** *Trinity College, Dublin, Ireland*

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 **- Jeylan Mortimer,** *University of Minnesota, USA*

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**, *Queen's University Belfast & University College Dublin*; **Richard Wiggins**, *Institute of Education, UK*; **Dieter Wolke**, *University of Warwick, UK*

SUBSCRIPTIONS

For immediate access to current issue and preceding 2014 Issues (Volume 5):

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

PAPERS

LLCS 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

Editorial

John Bynner

The 16th issue of the Journal brings news on three fronts: editorial team, annual conference and current and future issues.

Editorial team

Welcome on board to three new Section Editors confirmed at the last Editorial Committee meeting in July. Scott Montgomery from Örebro University Hospital and Örebro University, Sweden, has taken over from Michael Wadsworth as Section Editor for Health Sciences. Richard Layte from Trinity College Dublin has taken over from Robert Erikson as Section Editor for Social and Economic Sciences. Jeylan Mortimer from the University of Minnesota has taken over from Barbara Maughan as Section Editor for Development and Behavioural Sciences. This is also the opportunity to thank the three departing Section Editors for their great contribution to getting the Journal launched and helping it achieve the successful position it holds today; through them we have been extremely fortunate in benefiting from the advice of the excellent reviewers who also play a crucial part in maintaining our high standards.

Annual conference

The Society for Longitudinal and Life Course Studies annual conference in Lausanne, 9-11 October, had as its theme this year "Longitudinal Research and Social Policy". Six of the conference symposia were devoted to different features of, and issues arising from, the research policy interface. The emphasis was on looking for common ground and best practice examples across the world of policy implications, impact and brokerage of longitudinal research and how communication and other difficulties are best resolved. The prominence given to policy gives good prospects of expanding the Journal's scope for the first time to include papers in this area.

But the conference programme was by no means restricted to policy-oriented papers. The conference has never attracted so many papers across the full range of SLLS interests, with four strands running through the programme, including two keynote talks, 200 papers and 41 posters presented. All this is grist to the mill for the Journal, as those starting their lives as SLLS conference presentations form a key source of our papers. On the other hand, the conference is by no means the only route to publication in LLCS. Whenever you hear of a colleague working on a paper in the general area of longitudinal and life course studies, do urge them to consider LLCS as their first choice for publication.

Current issue

This issue contains four papers reflecting the range of LLCS coverage. The first paper based on data from the UK Millennium Cohort Study is devoted to children's socio-emotional wellbeing under different structures experienced since birth. The second paper moves to the other end of the life course, using data from the UK 'National Survey of Health and Development' (1946 Cohort) extending to the over-60s, to investigate factors in social integration in the later years. A four-fold typology is produced combining marital status with level of integration.

The next two papers shift the emphasis again - this time to methodology and procedure. In a fascinating case study, the means of reconstructing, (through administrative and health record data linkage) a sample from a Scottish longitudinal survey starting in 1936 and another starting in 1947, is reported; there are many lessons for those undertaking comparable challenges in the UK and other countries. The final paper moves again, this time to Australia and the different longitudinal vehicle of the household panel study. Capitalising on a switch in data collection procedure, the focus is on the important issue of whether 'dependent' as opposed to 'independent' interviewing produces the lowest reporting of occupational change between surveys. It does!

Security and progress

The Journal as always seeks high quality papers on an ever-widening scale from an ever-expanding range of countries. Our aim is to move now from three to four issues annually. And that of course costs money. The greatest help our readers and authors can give to the Journal is to join the Society individually and to encourage their institutions to take up corporate membership. Library subscriptions are a major source of income for us. So the final plea is that you urge your university library to join the forty others that subscribe to the Journal, if they have not already done so, and invest in a set of hardback copies as back-up to the online version. Your Journal needs you!

SLLS Annual Conference 2015

Dublin, Ireland

SAVE THE DATE! The sixth annual conference of the

Society for Longitudinal and Life Course Studies

will be held in

Dublin, 18th - 21st October 2015

Socio-economic inequalities in profiles of social integration across adulthood: evidence from a British birth cohort study

Kate Harvey
Dr Graciela Muniz Terrera
Professor Diana Kuh
Dr Mai Stafford
MRC Unit for Lifelong Health and Ageing at University College London
k.harvey2@nhs.net

(Received January 2014 Revised May 2014) http://dx.doi.org/10.14301/llcs.v5i3.284

Abstract

Social integration in older age is a key quality of life component and is associated with reduced mortality and morbidity risk. There are socio-economic differences in social integration, but the influence of different indicators of socio-economic position on long-term change in social integration at older ages is not known. This study aimed to identify profiles of social integration across adulthood and explore the impact of various socio-economic indicators at different ages. Data were drawn from the MRC National Survey of Health and Development. A latent class analysis used measures of contact with friends and family, participation in group activities and marital status at ages 36, 43 and 60-64 to identify profiles of change in social integration for men and women. One-step analyses related profiles to father's occupation-based socio-economic position, own educational attainment and head of household occupation-based socio-economic position. Four profiles of social integration were identified for men: high and maintained, married; medium and maintained, married; declining, married; and declining, unmarried. Higher head of household occupation and educational attainment were associated with greater likelihood of maintained integration. Four profiles of social integration were also identified for women: high and maintained, married; high and maintained, unmarried; declining group participation, unmarried; and declining group participation, married. Higher socio-economic position on all indicators was associated with greater likelihood of maintained integration. Lower socio-economic groups are more likely to experience declining social integration by early old age. Support to promote social integration may be particularly important for those with lower occupational grade or education.

Keywords: United Kingdom; social integration; social engagement; socio-economic position; gender; cohort; longitudinal; National Survey of Health and Development

Background

Social integration

Social integration has been defined as attachment to society through ties to family and friends and formal links to community institutions (Berkman, Glass, Brissette, & Seeman, 2000; Fothergill et al., 2011). As such, it is a structural measure of the quantity of social networks as opposed to a functional measure of the subjective experience of social support or loneliness, and is commonly assessed through the construction of indices comprising contact with friends and family, participation in clubs or associations and marital

status (Berkman, & Syme, 1979; Berkman et al., 2004; Stringhini et al., 2012; Vonneilich et al., 2011).

Studies suggest that social integration changes across the life course, with a small decrease in network size at older ages and an increasing emphasis on contact with family members as opposed to non-family members (Kalmijn, 2012). One explanation for this pattern of change is 'socio-emotional selectivity theory' (Carstensen, Isacowitz, & Charles, 1999) whereby older individuals prioritise social relationships that provide immediate emotional payoff over those that confer future advantage. Studies have suggested, however, that there is heterogeneity in change over time, with both increasing and decreasing profiles of social integration (Bassuk, Glass, & Berkman, 1999; van Tilburg, 1998) and some variation by gender (Gallant & Dorn, 2001).

Most longitudinal analyses are short, focus on later life (Cerhan & Wallace, 1997) and, where they explore change in social integration, focus on specific life events such as marriage, parenthood, divorce or widowhood (Kalmijn, 2012). An exception is the Americans' Changing Lives survey, that has been used to explore comprehensively assessed within-person longitudinal profiles of social integration in over 60 year olds across longer time periods. Women were found to have consistently higher levels of social integration than men over an eight year period (Thomas, 2011a). Two further studies using prospective data over 17 years in this population describe five different profiles of change in social integration and clearly demonstrate the heterogeneity of change in social integration in older age (Thomas, 2011b; 2012).

Patterns of social integration may vary by socio-economic position

In middle and older age, lower levels of social contact (Krause & Borawski-Clark, 1995; Van Groenou & Van Tilburg, 2003; Weyers et al., 2008) and leisure time activity (Scherger, Nazroo, & Higgs, 2010; Shaw, Krause, Gallant, & McGeever, 2010) have been found in individuals with lower income or educational attainment. There are mixed findings, however, with other studies finding higher levels of social contact in the lower socio-economic groups (Kubzansky, Berkman, Glass, & Seeman, 1998) and no work examining change across long periods of time.

Inconsistencies in the relationship between socio-economic position and social integration may arise from differences in the type of group or network measured. Those with lower educational attainment, income or occupational social class are suggested to have more frequent contact with family, but smaller friendship networks and less frequent participation in groups (Granovetter, 1973; Park & Roberts, 2003; Thomas, 2012). Findings from the British Household Panel Study (in men only) suggest that the types of organisational affiliations also vary with socio-economic position in older age, with higher socio-economic groups being affiliated with civic and religious groups, and lower socio-economic groups being affiliated with social clubs (Perren, Arber & Davidson, 2003). Socio-economic inequalities in social integration may depend on age and gender, with smaller inequalities for women (Marmot, Bosma, Hemingway, Brunner, & Stansfeld, 1997; Van Groenou & Van Tilburg, 2003) and at younger ages (Stringhini et al., 2012).

The English Longitudinal Study of Ageing has contributed to descriptions of change in social network ties amongst the over 50 years old population, with observations of large variations in individual level change in social integration over eight years (Jivraj, Nazroo, & Barnes, 2012). When examining the socio-economic factors associated with a move into social detachment, authors found poorer, lower-educated, unhealthy and married individuals to be more likely to become socially detached (with detachment defined as not being a member of a group, not going to cultural event and not having at least weekly contact with a friend or family member).

Most studies have focused on socio-economic indicators in adulthood. but a small number have considered parental education or occupational socio-economic position as a possible determinant of their offspring's social networks in adulthood, although the association with other measures of socio-economic position is found to be stronger (Van Groenou & Van Tilburg, 2003; Hatch & Wadsworth, 2008). Several plausible conceptual pathways have been suggested to explain these relationships, with suggestions that socio-economic position at a number of different points in life may independently affect patterns of adult social integration.

Childhood socio-economic position is related to a number of characteristics of early life living situation, including housing and neighbourhood characteristics (Van de Mheen, Stronks, Van den Bos & Mackenbach, 1997) family attitudes, beliefs and behaviours (Van Groenou & Van Tilburg, 2003; Singh-Manoux & Marmot, 2005) and risk of exposure to early life and family stresses (Lehman, Taylor, Kiefe, & Seeman, 2005). These may affect an individual's physical and psychosocial development, socialisation and network formation later in life (Repetti, Taylor, & Seeman, 2002) and may result in decreased interpersonal capacity and networks (Gallo, Espinosa de los Monteros, & Shivpuri, 2009; Matthews, Gallo, & Taylor, 2010). Education may directly relate to adult social networks by means of the availability of cognitive resources or skills to develop personal relationships (Van Groenou & Van Tilburg, 2003). Education may also act as a mediator, explaining the effect of a range of lifetime exposures including anxious or sad adolescent affect on adult social integration (Hatch & Wadsworth,2008) .Occupational measures of socio-economic position may relate to social networks via available material resources, prestige or social norms that facilitate social integration. It has been suggested that lower socio-economic position may lead to a restricted breadth of relationships (lower relational radius) in order to prevent recognition of social disadvantage or feelings of shame (Weyers et al., 2008), with withdrawal from extended social networks but a stronger reliance on family, kinship and neighbourhood. There may also be a direct effect of poverty on social integration due to lack of financial resources to enable participation.

Whilst there is considerable interest within the epidemiological literature on the possible impact of lack of social integration on morbidity and mortality risk (Fratiglioni, Paillard-borg, & Winblad, 2004; Gorman & Sivaganesan, 2007; Holt-Lunstad, Smith, & Layton, 2010; Seeman et al., 2011; Zunzunegui et al., 2004), it is also likely that the onset of poor health or functional limitation results in declining social integration, and this pathway might also explain why those in lower socio-economic positions are less socially integrated. Participation in the labour market involves a degree of social contact but a clear distinction is drawn in the literature between social engagement or integration, and social interactions that are paid for or enforced (Prohaska, Anderson, & Binstock, 2012). In line with this body of literature, we elected to focus on social ties and interpersonal aspects of social integration outside of the workplace in the current study, and paid employment was not included as an indicator of social integration. Participation in the labour market has strong and well-established associations with childhood circumstances and education and we were concerned that socio-economic pathways alone should not drive our analyses.

In summary, several studies have considered socio-economic inequalities in social integration cross-sectionally or with short-term follow-up. One recent study examined socio-economic differences in social detachment (Jivraj et al., 2012). Studies have not, however, investigated the relative contribution of indicators of socio-economic position, at different points in the life course, on patterns of change in social integration over mid to late adulthood.

This study examined longitudinal patterns of social integration between the ages of 36 and 60-64, i.e. over a 24 to 28 year period. It aimed to identify profiles of changing social integration and to explore their association with indicators of socio-economic position at different ages, in a nationally representative cohort of men and women.

Data

Participants: the 1946 British birth cohort

The MRC National Survey of Health and Development (NSHD), also known as the 1946 British birth cohort, is a population based birth cohort, originally consisting of a socially stratified sample of 5362 singleton births in one week in March 1946. Data have been collected at 23 points since birth. The main data collections in adulthood were undertaken when study members were 26, 36, 43, 53 and 60-64 years of age (Wadsworth, Kuh, Richards, & Hardy, 2006).The cohort generally has high rates of marriage and high marital stability over time.

Measures

Social integration

Consistent with commonly-used definitions of social integration, this study included indicators of marital status, group participation and contact with friends and family. Questions assessing frequency of participation in different clubs or associations

(group participation), frequency of informal social contact (contact with friends and family), and marital status were included in NSHD at ages 36, 43 and 60-64 years (see Appendix 1) though with slightly different question wording. Frequency of group participation was summarised from items about doing, helping to run or being involved in a number of social, secular and religious group activities. A binary variable (participating in at least one activity weekly or more often, participating less frequently or not at all) was derived for each sweep to allow comparison across the different time points. Frequency of contact with friends or family was summarised from items asking about how often participants met, visited or were visited by friends, parents, partner's parents or other relatives. A binary variable (seeing friends or family three times a month or more, less often or never) was derived at each sweep to maintain consistency across the different time points. It was not possible to quantify how many social contacts were seen at the reported frequencies. Marital status at each sweep was summarised as a binary variable (married, not married). Rates of cohabitation outside of marriage are low in this cohort (about 3.5% of respondents at age 36).

Socio-economic position

Childhood socio-economic conditions were captured by father's occupation at age 4 (or ages 11 or 15 if this was missing, n=54). Occupations were coded according to the UK Registrar General's Social Class scheme: professional; managerial and technical; skilled (non-manual); skilled (manual); partly skilled; and unskilled (Galobardes, Shaw, Lawlor, Lynch, & Davey Smith, 2006). Educational attainment was measured as highest recorded qualification at age 26, categorised as none, below secondary, O-level or equivalent (usually attained at age 16), and A-level (High School) or above (usually attained at age 18 or older). Early-adult socio-economic position was captured as head of household occupation at age 36 (defined at this time as male occupation or own if there was no male in the household). Again, this was coded according to the UK Registrar General's Social Class scheme.

Statistical methods

Bivariate associations between each measure of socio-economic position and frequent group participation, frequent contact with friends and family and marital status were examined separately for men and women. The latent class analysis used data on three observed categorical indicators of social integration (group participation, contact with friends or family and marital status) at three time points, to identify unobserved profiles of social integration from ages 36 to 60-64. Individuals were grouped according to the similarity of their responses to identify 'classes' of people within the study population. A priori, the number of classes (or profiles of social integration) was unknown so models with an increasing number of classes were fitted to the data, and model selection was conducted via comparison of Bayesian Information Criterion (BIC) values as recommended by Nylund, Asparouhov, and Muthén (2007). To assess the quality of the classification we considered the model's entropy. The entropy is a function that takes values between O and 1, with high values indicating high certainty of classification. Latent class analysis and mixture modelling was performed in MPlus (version 5). All other analyses were performed in STATA version 12.

The association of childhood, educational and early adult measures of socio-economic position with classification of the different adult profiles of social integration was explored in the latent class model. This one step approach was chosen in preference to a two stage approach (using likely class membership as an outcome in multinomial regression) as it has been shown to be more sensitive to detection of effect size and is preferable where the number of covariates is small enough to make computation feasible (Clark & Muthén, 2009; Feingold, Tiberio & Capaldi, 2014). Models use full information maximum likelihood (FIML) for missing data when deriving social integration profiles, including all individuals who have data on one or more observed variables at one or more time point, and we restricted the analytical sample to those with complete data for covariates (socio-economic position variables). Missing data are assumed to be missing at random (Little & Rubin, 2002).

Analytic sample

Of the 2,815 men in the original cohort, 958 (34%) provided no social integration data because of death or other loss to follow-up, and a further 370 men had incomplete socio-economic data. Thus 1,487 men were included in these analyses (998

provided social integration data on three occasions, 382 on two occasions and 107 on one occasion). Of the 2,547 cohort women, 719 did not provide social integration data, 386 did not provide socio-economic data, and 1,442 were included in these analysis. Of these, 1,063 (277, 102) provided social integration data on three (two, one) occasions.

Men included in the model were more likely to have O level or above education (57% compared to 47%, p<0.001) and more likely to be married (82% were married compared to 71%, p<0.001) than those not included in analyses. Women included in the model were more likely to be educated to O level or above (53% compared to 46%, p=0.02) and have frequent friend or family contact (80% compared to 72%, p=0.01), but had slightly lower head of household occupation (only 9% were in households headed by a professional compared to 15% of those not included, p=0.05). Married people, women with frequent group participation at age 36 and women with frequent contact with friends and family at age 60-64 were over-represented among those with complete socio-economic data (p<0.05). Taken together, this indicates that our analytical sample tended to have more advantaged socio-economic circumstances and greater social integration than those not included.

Results

Frequent group participation, frequent contact with friends and family and marital status show variations by age and by gender, as shown in table 1.

Table 1. Social integration by age and gender

	Males			Females		
	Age 36	**Age 43**	**Age 60-64**	**Age 36**	**Age 43**	**Age 60-64**
	% (n)	*% (n)*	*% (n)*	*% (n)*	*% (n)*	*% (n)*
Group participation						
at least weekly	74 *(1098)*	37 *(493)*	52 *(497)*	59 *(846)*	37 *(478)*	49 *(502)*
less than weekly or not at all	26 *(389)*	63 *(844)*	48 *(596)*	41 *(596)*	63 *(825)*	51 *(522)*
Contact with friends and family						
at least weekly	*69 (1031)*	*80 (1062)*	*67 (583)*	*71 (1022)*	*77 (1003)*	*80 (742)*
less than weekly or not at all	*31 (455)*	*20 (273)*	*33 (282)*	*29 (416)*	*23 (295)*	*20 (188)*
Married	*84 (1252)*	*82 (1101)*	*82 (776)*	*86 (1245)*	*81 (1062)*	*75 (752)*
Not married	*16 (235)*	*18 (238)*	*18 (166)*	*14 (197)*	*19 (242)*	*25 (248)*

Frequent group participation was most prevalent at age 36 for both men and women, with 74% and 59% participating in activities once a week or more respectively. Frequent participation fell at age 43 (only 37% of both genders reported participating once a week or more) but increased again at age 60-64 (52% of men and 49% of women participated at least once a week). A different pattern was seen for frequent contact with friends and family, with the highest rates of weekly contact being at age 43 for men and 60-64 for women. There were gender differences in social integration although these varied at different ages. Men had more frequent group participation than women at age 36, but no difference in group participation was found at other ages. At age 60-64, a higher proportion of women had at least weekly contact with friends or family than men (80% compared to 67%, p<0.001). Men were more likely than women to be married at age 60-64 (82% compared to 75%, p<0.001).

Levels of educational attainment and age 36 socio-economic position were lower for women than for men (Table 2). At age 36, frequent group participation was not associated with socio-economic indicators. Frequent contact with friends and family at age 36 was more prevalent among men and women in lower socio-economic positions on all three indicators. Being married at age 36 was more prevalent among men with higher education and higher early adult socio-economic position (occupation at age 36).

Table 2: Socio-economic position and social integration at age 36 and 60-64

	Total frequency in analytical sample (%)	Frequent group participation at age 36 (%)	Frequent contact with friends and family at age 36 (%)	Married at age 36 (%)	Frequent group participation at age 60-64 (%)	Frequent contact with friends and family at age 60-64 (%)	Married at age 60-64 (%)
MALES							
Childhood socio-economic position (Father's occupation)							
I	*5.6*	73.5	60.2	77.1	56.1	66.7	74.2
II	*17.3*	67.7	64.1	86.8	55.8	73.2	88.8
IINM	*18.2*	76.3	66.8	86.0	62.5	70.3	83.5
IIIM	*30.3*	75.6	75.1	86.7	44.7	66.1	80.6
IV	*21.9*	72.6	68.6	80.6	44.9	60.2	80.3
V	*6.8*	79.2	74.3	79.2	57.6	73.5	84.5
		p=0.1	*p=0.006*	*p=0.2*	p=0.02	p=0.1	p=0.5
Education							
A-Level or above	*41.6*	74.6	61.7	86.3	62.2	66.3	84.4
O-level or equivalent	*15.5*	73.9	68.6	85.7	57.6	69.6	87.1
Below secondary	*6.0*	74.2	74.2	83.2	50.0	64.0	76.8
No qualifications	*37.0*	72.9	77.6	81.5	64.6	68.7	78.3
		p=0.5	*p<0.001*	*p=0.02*	p<0.001	p=0.06	p=0.02
Head of household socio-economic position (occupation) at age 36							
I	*12.2*	70.7	61.9	85.6	66.2	63.8	82.7
II	*32.7*	73.7	61.9	88.7	59.7	65.7	84.8
IINM	*10.1*	74.7	64.7	84.7	57.3	76.6	87.1
IIIM	*32.2*	75.0	76.4	82.3	37.5	68.8	82.8
IV	*11.2*	71.7	81.3	76.5	42.3	64.0	69.5
V	*1.7*	88.0	84.0	72.0	35.3	71.4	70.6
		p=0.3	*p<0.001*	*p<0.001*	p<0.001	p=0.5	p=0.01
FEMALES							
Childhood socio-economic position (Father's occupation)							
I	*6.1*	67.1	58.0	86.4	75.3	83.3	81.4
II	*17.0*	36.3	73.0	83.7	59.3	77.0	76.2
IINM	*18.6*	55.2	63.8	87.7	57.4	75.9	75.6
IIIM	*30.5*	59.8	73.1	88.0	40.6	82.8	75.6
IV	*21.2*	53.1	74.7	84.3	36.1	78.7	72.7
V	*6.7*	61.5	77.8	88.5	41.4	87.2	69.1
		p=0.06	*p=0.002*	*p=0.7*	p<0.001	p=0.4	p=0.1
Education							
A-Level or above	*27.3*	58.9	59.2	83.3	69.6	78.1	76.7
O-level or equivalent	*25.7*	60.0	72.4	88.4	51.4	79.4	76.5
Below secondary	*9.2*	67.4	81.8	82.6	40.0	86.4	71.8
No qualifications	*37.9*	55.5	76.1	88.1	28.8	79.9	73.5
		p=0.3	*p<0.001*	*p=0.1*	p<0.001	p=0.4	p=0.3
Head of household socio-economic position (occupation) at age 36							
I	*9.2*	58.3	57.6	94.7	64.7	73.5	58.0
II	*29.9*	60.6	68.2	84.0	41.5	81.4	73.1
IINM	*15.9*	57.6	68.2	75.6	44.0	76.5	74.4
IIIM	*28.1*	60.0	76.5	93.1	39.9	83.4	78.9
IV	*14.2*	56.5	75.7	85.5	36.7	76.5	68.2
V	*2.6*	42.1	84.2	81.6	18.2	86.4	66.8
		p=0.2	*p<0.001*	*p=0.7*	p<0.001	p=0.4	p=0.1

Notes. % shown for the total analytic sample are the proportion with different socio-economic position. %s shown for the other subgroups are the proportions of that social class reporting weekly or more social contact or group participation. P-values are from test for linear trend across socio-economic groups.

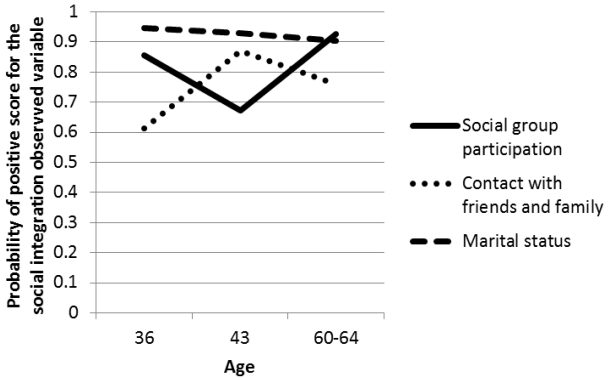
By age 60-64, test for linear trend indicated a general pattern of lower group participation in the more socio-economically disadvantaged groups. However, there was indication of non-linearity by father's occupation and educational attainment among men, with those in the middle socio-economic positions having the lowest frequent group participation. Frequent contact with friends and family at age 60-64 was not socio-economically patterned. Marriage at age 60-64 was less prevalent among more socio-economically disadvantaged men.

Longitudinal profiles of social integration - men

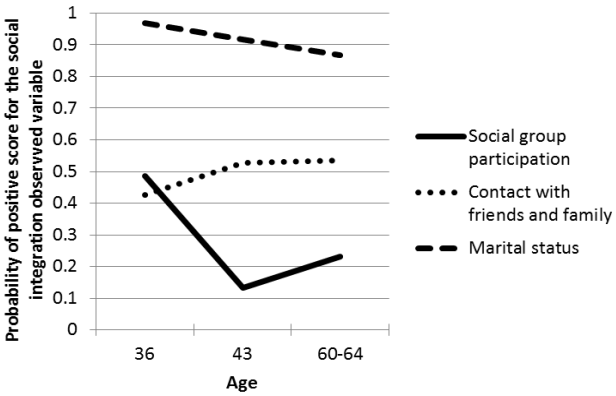
A model with four latent class profiles of social integration between ages 36 and 60-64 provided the best fit to the data in men (see Appendix 2 for fit statistics). The profiles of observed variables within each class are shown in figure 1. For men, the entropy was 0.569 which, whilst acceptable, demonstrates some uncertainty in class classification (0.6 is considered good, 0.5 poor)(Asparouhov & Muth, 2014).

Figure 1. Male profiles of social integration

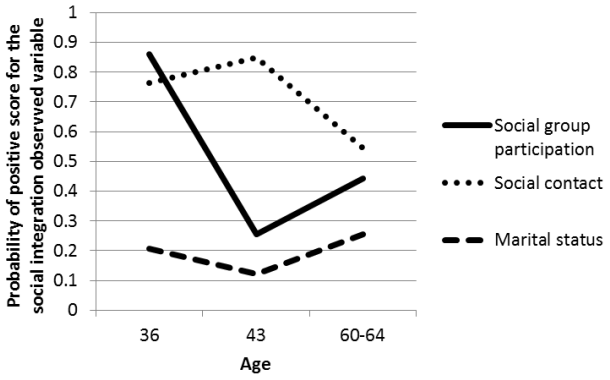
"High and maintained, married men" (36%)



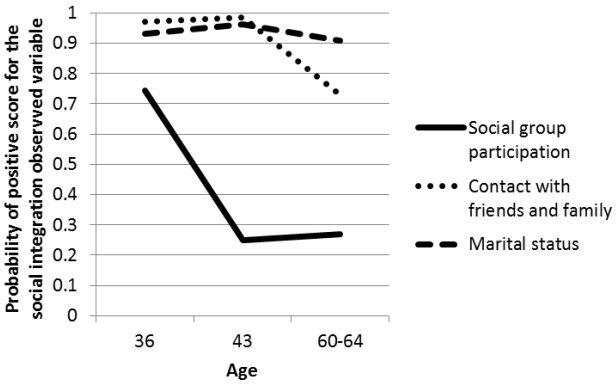
"Medium and maintained, married men" (20%)



"Declining, unmarried men" (14%)



"Declining, married men" (30%) *



*Notes. *denotes reference class in latent class model.*

The profile of declining married social integration was taken as the reference in the latent class regression analyses, as it was large in size and had a comparable group among women (see next section). Each step increase in educational attainment was associated with an odds ratio of 1.64 (that is, a 64% higher likelihood) of following the profile, with high and maintained married social integration compared to the profile of decline. Higher head of household occupation at age 36 was also associated with higher odds of being "high and maintained, married" and "medium and maintained, married", though only the latter attained statistical significance at the 5% level. No relationship was found between childhood socio-economic position and men's social integration.

Table 3. Socio-economic position and profile of social integration for men

	Father's occupation odds ratio (95% CI)	Education at age 26 odds ratio (95% CI)	Head of household occupation at age 36 odds ratio (95% CI)
High and maintained, married men (36%)	0.92 (0.75 to 1.13)	1.64 (1.33 to 2.03)*	1.62 (0.89 to 2.96)
Medium and maintained, married men (20%)	1.01 (0.76 to 1.34)	1.07 (0.67 to 1.71)	1.71 (1.28 to 2.26)*
Declining, unmarried (14%)	0.97 (0.79 to 1.2)	1.19 (0.95 to 1.50)	1.09 (0.71 to 1.66)
Declining, married (30%)	1	1	1

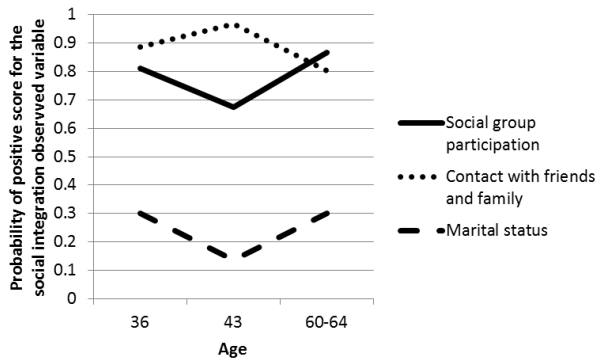
*Note. * represents p ≤ 0.05*

Longitudinal profiles of social integration - women

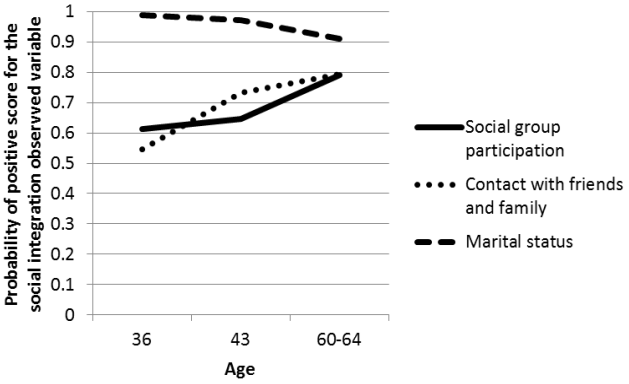
The model fit criteria in the latent class analysis for women identified that the model that best fitted the data had five classes (see Appendix 2) but the difference between the values of BIC was small and, as recommended (Muthen & Muthen, 2007), did not justify the reduction in parsimony and entropy, so a four class model was chosen; i.e. four profiles of social integration were identified between ages 36 and 60-64 (Figure 2). Class classification was good with entropy of 0.720.

Figure 2. Female profiles of social integration:

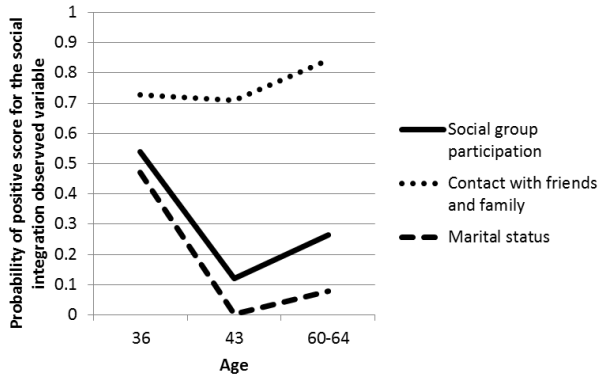
"High and maintained, unmarried women" (8%)



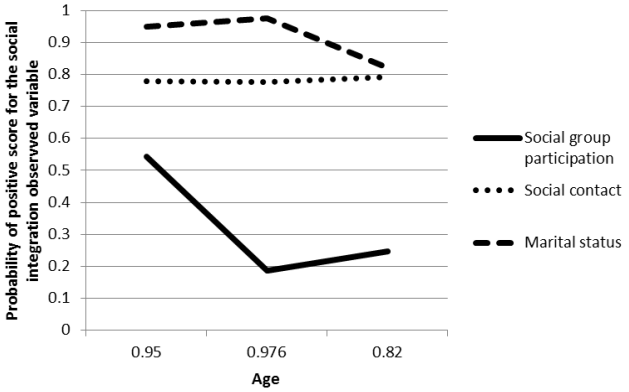
"High and maintained, married women" (29%)



"Declining group participation, unmarried" (10%)



"Declining group participation, married" (53%) *



*Note. *denotes reference class in latent class model*

The "declining group participation, married" was used as the reference category. Women with higher education were more likely to experience high and maintained social integration between ages 36 and 60-64 (Table 4). Each step increase in educational attainment was associated with an odds ratio of 2.89 (that is 189% higher likelihood) of being in the "high and maintained, unmarried" profile and 2.60 (95% confidence interval 1.53 to 4.40) of being in the "high and maintained, married" profile compared to being in the "declining group participation, married" profile, independently of other socio-economic indicators. Socio-economic position at age 36 was also related to the likelihood of following these two profiles. Higher childhood socio-economic position was associated with greater likelihood of being in the "high and maintained, unmarried" profile of social integration compared to the "declining group participation, married" profile.

Table 4. Socio-economic position and profile of social integration for women

	Father's occupation odds ratio (95% CI)	**Education at age 26 odds ratio (95% CI)**	**Head of household occupation at age 36 odds ratio (95% CI)**
High and maintained, unmarried (8%)	1.32 (1.02 to 1.71)*	2.89 (1.62 to 5.18)**	1.31 (1.03 to 1.67)*
High and maintained, married (29%)	1.13 (0.91 to 1.40)	2.60 (1.53 to 4.40)**	1.62 (1.21 to 2.16)*
Declining group participation, unmarried (10%)	0.90 (0.71 to 1.13)	1.01 (0.78 to 1.31)	1.06 (0.81 to 1.38)
Declining group participation, married (53%)	1	1	1

*Notes. * demotes p ≤ 0.05, ** denotes p ≤ 0.001*

Discussion

Within this cohort, multiple profiles of change in social integration between ages 36 and 60-64 were identified, indicating that people do not all experience the same changes in social integration as they age. This confirms recent work undertaken in a US sample of over 60s (Thomas, 2012). For both men and women, group participation changed more with age than either marital status or contact with friends of family. This is consistent with previous research that has identified informal social networks to be more stable over time than civic or leisure activities (Jivraj et al., 2012).

When comparing the four profiles identified for men, higher educational attainment was associated with a higher odds of having high and maintained social integration, compared to experiencing a profile of decline. Higher head of household occupation was also associated with a higher odds of following a profile of "medium and maintained, married men"). No relationship was identified between men's early childhood socio-economic position and the profiles of social integration that followed, suggesting that the socio-economic variables that may affect men's patterns of change in social integration over these ages involve exposures during adolescence or adulthood.

For women, indicators of higher socio-economic position from childhood and adulthood were associated with greater likelihood of maintaining social integration between ages 36 and 60-64. Multiple pathways may be in operation to link socio-economic position and social integration in later life, with early life experiences having a lasting impact on women's social integration in later life. Our analyses suggest that childhood socio-economic position was somewhat less strongly associated with social integration profiles than was age 36 socio-economic position, though note that at least some of any causal effect of childhood socio-economic position would be expected to operate through education and adult occupational status which were controlled for here. This is consistent with Van Groenou & Van Tilburg (2003) who found that network size in older age was associated with the adult household's, but not father's, socio-economic position. The strength of association between socio-economic position, especially

education, and patterns of social integration for women is novel, as previous research suggests that relationships between socio-economic position and social integration may be weak for women (Marmot, Bosma, Hemingway, Brunner, & Stansfeld, 1997; Van Groenou & Van Tilburg, 2003).

Higher wealth and education at baseline has previously been suggested to be protective against declining social integration or social detachment (Jivraj et al., 2012) but this is the first study to find multiple measures of socio-economic position across childhood and early adulthood to be independently associated with changes in social integration across adulthood. Considering the findings within a social capital framework (Putnam, 2000), group participation may be viewed as a type of bridging social capital, related to access to resources, whereas social contact with friends and family can be viewed as bonding social capital, providing contact between similar individuals but accessing more restricted forms of support. The emergence of socio-economic differences in social integration by early old age may represent differences in the availability of resources, and potentially be important for health and wellbeing within this age group.

Strengths and limitations

This is the first study to identify an independent relationship between multiple measures of socio-economic position and the change in social integration experienced during mid to late adulthood. The use of a data-driven approach such as latent class analysis allows consideration of patterns of change in the components of social integration, patterns that would be masked by the use of a summary index. This is also the first study to take a longitudinal approach to the measurement of both socio-economic position and social integration, considering socio-economic characteristics at multiple points through to early adulthood and social integration for almost thirty years of follow up.

The study does, however, have a number of limitations. There were differences in item wording and response categories (see Appendix 1) which may have affected responses, in particular regarding the lower group participation at age 43, where items were more specific than in other sweeps. Dichotomising the observed variables was necessary for data harmonization but this lost some detail and it was not possible to explore patterns of change in different types of relationship (e.g. kin and non-kin) or different groups or activities (e.g. leisure and civic). The analysis treated change in each social integration indicator independently and did not explore the impact of life events such as widowhood or divorce on social contact and group participation. The 'not married' group includes never married, widowed and divorced, and these potentially have different profiles of other social integration indicators including informal contact with friends and family, and social group participation. The full information maximum likelihood method used here, in common with multiple imputation approaches, assumed the data to be missing at random. Our analytical sample appeared to over-represent more socio-economically advantaged and socially integrated study members, but we are not able to know whether associations between socio-economic exposures and integration are similar amongst excluded individuals. This is a data-driven approach to summarising multiple variables, and emerging latent classes will depend on the analytic sample. Additionally, the models for men did not have high entropy, which means that there may be uncertainty associated with class allocation. It should be acknowledged that, although data covers a 30-year time period, there are only 3 observations of each indicator of social integration.

This study aimed to consider the structural dimensions of social relationships as the data on functional measures were not available over such a long time period. Considering patterns of change in functional social support would provide a more detailed picture of the changes in social relationships over time and it would be useful to see this developed further in future research. As already discussed, and in line with much of the past literature on this topic, we have not included work contacts within our definition of social integration. Different changes in work role and work-related social contacts across the socio-economic gradient may be related to the changes in non-work-related social integration observed here. Declining health or the onset of limitations may be another key pathway that we have not explored.

Conclusions

The study highlights the higher risk of declining social integration amongst those with lower socio-economic position. Having positive social relationships is a key component of quality of life as defined by older people (Bowling & Gabriel, 2007) and lack of social integration is associated with an increased likelihood of loneliness (Wenger, Daviesj, Shahtahmasebi, & Scott, 1996), morbidity, mortality and institutionalisation (Nicholson, 2012; Steptoe, Shankar, Demakakos, & Wardle, 2013). With social integration increasingly viewed as a resource for healthy ageing, this study supports a growing body of literature highlighting the importance of fostering social integration in later life (Bowling, Holt-lunstad, Scharf, & Walker, 2012; Gorman & Sivaganesan, 2007; Social Care Institute for Excellence, 2011) and provides evidence of the potential socio-economic inequalities in social integration in later life.

References

Asparouhov, T., & Muth, B. (2014). Auxiliary Variables in Mixture Modeling : 3-Step Approaches Using Mplus. *Structural Equation Modelling: a Multidisciplinary Journal, 21*, 1-13.

Bassuk, S., Glass, T., & Berkman, L. (1999). Social disengagement and incident cognitive decline in community-dwelling elderly persons. *Annals of Internal Medicine, 131*(3), 165–173. http://dx.doi.org/10.7326/0003-4819-131-3-199908030-00002

Berkman, L. F., Glass, T., Brissette, I., & Seeman, T. E. (2000). From social integration to health: Durkheim in the new millennium. *Social Science & Medicine, 51*(6), 843–57. http://dx.doi.org/10.1016/S0277-9536(00)00065-4

Berkman, L. F. & Syme, S. L. (1979). Social networks, host resistance, and mortality: a nine year follow-up study of Alameda County residents. *American Journal of Epidemiology, 109*(2), 186–204.

Berkman, L. F. Melchior, M. Chastang, J. F., Niedhammer, I., Leclerc, A., Goldberg, M.. (2004). Social integration and mortality: a prospective study of French employees of electricity of France-gas of France: The GAZEL cohort. *American Journal of Epidemiology, 159*(2), 167–174. http://dx.doi.org/10.1093/aje/kwh020

Bowling, A., & Gabriel, Z. (2007). Lay theories of quality of life in older age. *Ageing and Society, 27*(06), 827–848. http://dx.doi.org/10.1017/S0144686X07006423

Bowling, A., Holt-lunstad, J., Scharf, T., & Walker, A. (2012). *Loneliness: the state we 're in*. A report of evidence compiled for the Campaign to End Loneliness. http://campaigntoendloneliness.org/toolkit/download/loneliness-the-state-were-in/

Carstensen, L., Isacowitz, D., & Charles, S. (1999). Taking time seriously: a theory of socioemotional selectivity. *American Psychologist, 54*(3), 165–181. http://dx.doi.org/10.1037/0003-066X.54.3.165

Cerhan, J. R., & Wallace, R. B. (1997). Change in social ties and subsequent mortality in rural elders. *Epidemiology, 8*(5), 475–481. http://dx.doi.org/10.1097/00001648-199709000-00001

Clark, S. L., & Muthén, B. (2009). *Relating latent class analysis results to variables not included in the analysis*. Retrieved from www.statmodel.com/downloads/relatinglca.pdf

Feingold, A. Tiberio, S. Capaldi, D. (2014). New approaches to examining associations with latent categorical variables: applications to substance use and aggression. *Psychology of Addictive Behaviours, 28*(1), 257-67. http://dx.doi.org/10.1037/a0031487

Fothergill, K. E., Ensminger, M. E., Robertson, J., Green, K. M., Thorpe, R. J., & Juon, H.-S. (2011). Effects of social integration on health: A prospective study of community engagement among African-American women. *Social Science & Medicine, 72*(2), 291–8. http://dx.doi.org/10.1016/j.socscimed.2010.10.024

Fratiglioni, L., Paillard-borg, S., & Winblad, B. (2004). An active and socially integrated lifestyle in late life might protect against dementia. *The Lancet, Neurology, 4*(Box 6401), 343–353. http://dx.doi.org/10.1016/S1474-4422(04)00767-7

Gallant, M. P., & Dorn, G. P. (2001). Gender and race differences in the predictors of daily health practices among older adults. *Health Education Research, 16*(1), 21–31. http://dx.doi.org/10.1093/her/16.1.21

Gallo, L. C., Espinosa de los Monteros, K., & Shivpuri, S. (2009). Socioeconomic status and health: what is the role of reserve capacity? *Current Directions in Psychological Science, 18*(5), 269–274. http://dx.doi.org/10.1111/j.1467-8721.2009.01650.x

Galobardes, B., Shaw, M., Lawlor, D. A, Lynch, J. W., & Davey Smith, G. (2006). Indicators of socioeconomic position (part 2). *Journal of Epidemiology and Community Health, 60*(2), 95–101. http://dx.doi.org/10.1136/jech.2004.028092

Gorman, B. K., & Sivaganesan, A. (2007). The role of social support and integration for understanding socioeconomic disparities in self-rated health and hypertension. *Social Science & Medicine, 65*(5), 958–75. http://dx.doi.org/10.1016/j.socscimed.2007.04.017

Granovetter, M. S. The strength of weak ties. *American Journal of Sociology, 78*(6), 1360-1380 http://dx.doi.org/10.1086/225469

Hatch, S. L., & Wadsworth, M. E. (2008). Does adolescent affect impact adult social integration? Evidence from the British 1946 birth cohort. *Sociology, 42*(1), 155–177. http://dx.doi.org/10.1177/0038038507087358

Holt-Lunstad, J., Smith, T. B., & Layton, J. B. (2010). Social relationships and mortality risk: a meta-analytic review. *Public Library of Science Medicine, 7*(7), e1000316.

Jivraj, S., Nazroo, J., & Barnes, M. (2012). 3. Change in social detachment in older age in England. In The Dynamics of Ageing: Evidence From the English Longitudinal Study of Ageing 2002-2010 48–97.

Kalmijn, M. (2012). Longitudinal analyses of the effects of age, marriage, and parenthood on social contacts and support. *Advances in Life Course Research, 17*(4), 177–190. http://dx.doi.org/10.1016/j.alcr.2012.08.002

Krause, N., & Borawski-Clark, E. (1995). Social class differences in social support among older adults. *The Gerontologist, 35*(4), 498–508. http://dx.doi.org/10.1093/geront/35.4.498

Kubzansky, L. D., Berkman, L. F., Glass, T. a, & Seeman, T. E. (1998). Is educational attainment associated with shared determinants of health in the elderly? Findings from the MacArthur Studies of Successful Aging. *Psychosomatic Medicine, 60*(5), 578–85. http://dx.doi.org/10.1097/00006842-199809000-00012

Lehman, B. J., Taylor, S. E., Kiefe, C. I., & Seeman, T. E. (2005). Relation of childhood socioeconomic status and family environment to adult metabolic functioning in the CARDIA study. *Psychosomatic Medicine, 67*(6), 846–54. http://dx.doi.org/10.1097/01.psy.0000188443.48405.eb

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*, second edition. Hoboken, New Jersey: Wiley.

Marmot, M. G., Bosma, H., Hemingway, H., Brunner, E., & Stansfeld, S. (1997). Contribution of job control and other risk factors to social variations in coronary heart disease incidence. *Lancet, 350*(9073), 235–9. http://dx.doi.org/10.1016/S0140-6736(97)04244-X

Matthews, K. a, Gallo, L. C., & Taylor, S. E. (2010). Are psychosocial factors mediators of socioeconomic status and health connections? A progress report and blueprint for the future. *Annals of the New York Academy of Sciences, 1186*, 146–73. http://dx.doi.org/10.1111/j.1749-6632.2009.05332.x

Muthen, B., & Muthen, L. (2007). *MPlus User Forum: difference in BIC values*. Retrieved from http://www.statmodel.com/discussion/messages/23/2232.html

Nicholson, N. R. (2012). A review of social isolation: an important but underassessed condition in older adults. *The Journal of Primary Prevention, 33*(2-3), 137–52. http://dx.doi.org/10.1007/s10935-012-0271-2

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(4), 535–569. http://dx.doi.org/10.1080/10705510701575396

Park, A. Roberts, C. (2003). The ties that bind. In British Social Attitudes: The 19th report. London: Sage.

Perren, K., Arber, S., & Davidson, K. (2003). Men's organisational affiliations in later life: the influence of social class and marital status on informal group membership. *Ageing and Society, 23*(1), 69–82. http://dx.doi.org/10.1017/S0144686X02001034

Prohaska, T. R., Anderson, L. A., & Binstock, R. H. (2012). *Public health for an ageing society*. Baltimore: The Johns Hopkins University Press.

Putnam, R. (2000). *Bowling alone: the collapse and revival of American community*. New York: Simon and Schuster.

Repetti, R., Taylor, S., & Seeman, T. (2002). Risky families: family social environments and the mental and physical health of offspring. *Psychological Bulletin, 128*(2), 330–366. http://dx.doi.org/10.1037/0033-2909.128.2.330

Scherger, S., Nazroo, J., & Higgs, P. (2010). Leisure activities and retirement: do structures of inequality change in old age? *Ageing and Society, 31*(01), 146–172. http://dx.doi.org/10.1017/S0144686X10000577

Seeman, T. E., Miller-Martinez, D. M., Stein Merkin, S., Lachman, M. E., Tun, P. A, & Karlamangla, A. S. (2011). Histories of social engagement and adult cognition: midlife in the U.S. study. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 66* Suppl 1, i141–52. http://dx.doi.org/10.1093/geronb/gbq091

Shaw, B. A., Krause, N., Gallant, M., & McGeever, K. (2010). Age Differences and Social Stratification in the Long-Term Trajectories of Leisure-Time Physical Activity. *Journal of Gerontology: Social Sciences, 65B*(6), 756–766. http://dx.doi.org/10.1093/geronb/gbq073

Singh-Manoux, A., & Marmot, M. (2005). Role of socialization in explaining social inequalities in health. *Social Science & Medicine, 60*(9), 2129–33. http://dx.doi.org/10.1016/j.socscimed.2004.08.070

Social Care Institute for Excellence. (2011). *Preventing loneliness and social isolation : interventions and outcomes*. Retrieved from http://www.scie.org.uk/publications/briefings/files/briefing39.pdf

Steptoe, A., Shankar, A., Demakakos, P., & Wardle, J. (2013). Social isolation, loneliness, and all-cause mortality in older men and women. *Proceedings of the National Academy of Sciences of the United States of America, 2013*, 1–5.

Stringhini, S., Berkman, L., Dugravot, A., Ferrie, J. E., Marmot, M., Kivimaki, M., & Singh-Manoux, A. (2012). Socioeconomic status, structural and functional measures of social support, and mortality: The British Whitehall II Cohort Study, 1985-2009. *American Journal of Epidemiology, 175*(12), 1275–83. http://dx.doi.org/10.1093/aje/kwr461

Thomas, P. A. (2011a). Gender, social engagement, and limitations in late life. *Social Science & Medicine, 73*(9), 1428–35. http://dx.doi.org/10.1016/j.socscimed.2011.07.035

Thomas, P. A. (2011b). Trajectories of social engagement and limitations in late life. *Journal of Health and Social Behavior, 52*(4), 430–43. http://dx.doi.org/10.1177/0022146511411922

Thomas, P. A. (2012). Trajectories of social engagement and mortality in late life. *Journal of Aging and Health, 24*(4), 547–68. http://dx.doi.org/10.1177/0898264311432310

Van de Mheen, H., Stronks, K., Van den Bos, J., & Mackenbach, J. (1997). The contribution of childhood environment to the explanation of socio-economic inequalities in health in adult life: a retrospective study. *Social Science & Medicine, 44*(1), 13–24. http://dx.doi.org/10.1016/S0277-9536(96)00090-1

Van Groenou, M. I. B., & Van Tilburg, T. (2003). Network size and support in old age: differentials by socio-economic status in childhood and adulthood. *Ageing and Society, 23*(5), 625–645. http://dx.doi.org/10.1017/S0144686X0300134X

Van Tilburg, T. (1998). Losing and gaining in old age: changes in personal network size and social support in a four-year longitudinal study. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 53*(6), S313–23. http://dx.doi.org/10.1093/geronb/53B.6.S313

Vonneilich, N., Jöckel, K.-H., Erbel, R., Klein, J., Dragano, N., Weyers, S., Moebus, S., Siegrist, J., von dem Knesebeck, O. (2011). Does socioeconomic status affect the association of social relationships and health? A moderator analysis. *International Journal for Equity in Health, 10*(1), 43-51. http://dx.doi.org/10.1186/1475-9276-10-43

Wadsworth, M., Kuh, D., Richards, M., & Hardy, R. (2006). Cohort Profile: The 1946 National Birth Cohort (MRC National Survey of Health and Development). *International Journal of Epidemiology, 35*(1), 49–54. http://dx.doi.org/10.1093/ije/dyi201

Wenger, G. C., Daviesj, R., Shahtahmasebi, S., & Scott, A. (1996). Social isolation and loneliness in old age: review and model refinement. *Ageing and Society, 16*, 333–358. http://dx.doi.org/10.1017/S0144686X00003457

Weyers, S., Dragano, N., Möbus, S., Beck, E.-M., Stang, A., Möhlenkamp, S., Jockel, K. H., Erbel, R., Siegrist, J. (2008). Low socio-economic position is associated with poor social networks and social support: results from the Heinz Nixdorf Recall Study. *International Journal for Equity in Health, 7*(13). http://dx.doi.org/10.1186/1475-9276-7-13

Zunzunegui, M. V, Koné, A, Johri, M., Béland, F., Wolfson, C., & Bergman, H. (2004). Social networks and self-rated health in two French-speaking Canadian community dwelling populations over 65. *Social Science & Medicine, 58*(10), 2069–81. http://dx.doi.org/10.1016/j.socscimed.2003.08.005

Appendix 1

Social integration variables

	1982 (age 36)	1989 (age 43)	2006-2011 (age 60-64)
Frequency of social group participation:	Nurse led interviews: *Do you do any of the following in your spare time and on average in the last year have you done this at least daily, weekly, less frequently? (chess bridge or similar, church or religious activities, going to the theatre concerts, going to pub/club, helping to run a club or playgroups or school, local government or trade union or political work, playing a musical instrument with others, voluntary social welfare work).* *Are there any other things you do with other people in your spare time that I haven't yet asked about? What are they and how often?*	Nurse led interviews: *In your spare time, do you help to run or are you currently involved in any of the following activities and do you take part weekly, monthly or less often? (Church activities, playgroup nurseries or school, local government, trade unions, voluntary services, sports clubs, evening classes/adult education, other organisations).*	Postal questionnaire: *In your spare time are you involved in any of the following activities, if yes have you taken part in the last 12 months weekly, monthly, less often or never? (Church-related group or religious activities, job related association, recreational groups, civic political group, other voluntary work, other groups or clubs, other social activities).*
Frequency of contact with friends or family:	Nurse led interviews: *How often do you have friends to spend the evening with you at home, or to spend some time at home with you on the weekend? (Once a week or more often, not more than once a fortnight, not more than once a month, rarely or never, unknown).* *How often do you see your parents and/or your wife's/husband's/partner's parents? (Once a week or more often, not more than once a fortnight, not more than once a month, rarely or never, not relevant no parents, unknown).*	Nurse led interviews: *On average, how often would you say you met friends or relatives socially? (Never, 1-2 times a month, 3-5 times a month, 6-10 times a month, 11-15 times a month, more than 15 times a month).*	Nurse led interview: *Thinking of all your relatives or friends, how often do you regularly visit or are you visited by these people? (Never/almost never, once every few months, about once a month, about once a week, almost daily).*
Marital status	Nurse led interviews: *And so may I just check at the moment you are single and not living with a partner, single i.e. never married and living with a partner, married, widowed and not living with a partner, separated and not living with a partner, divorced and not living with a partner, widowed and living with a partner, separated or divorced and living with a partner?*	Nurse led interviews: *So can I just check, are you currently single never married, married, widowed, separated, divorced?*	Postal questionnaire: *So, are you currently single that is never married, married & living with husband/wife, married & separated from husband/wife, divorced, widowed?*

Appendix 2a

Table 2a. Latent class analysis fit statistics for social integration for men in the MRC NSHD (age 43 to 60-64)

	2 classes	3 classes	4 classes	5 classes
Information criteria[a]				
Schwarz's BIC	12169.5	12031.9	*12023.5*	12064.9
Smallest class percentage	16	14	14	7
Entropy[b]	0.823	0.613	0.569	0.587

Abbreviations: BIC, Bayesian Information Criterion
[a] minimum information criterion is shown in italic type
[b] based on estimated posterior class membership probabilities

Appendix 2b

Table 2b. Latent class analysis fit statistics for social integration for women in the MRC NSHD (age 43 to 60-64)

	2 classes	3 classes	4 classes	5 classes	6 classes
Information criteria[a]					
Schwarz's BIC	12248.0	12077.6	12060.8	*12056.7*	12112.8
Smallest class percentage	18	16	8	7	7
Entropy[b]	0.863	0.657	0.720	0.664	0.633

Abbreviations: BIC, Bayesian Information Criterion
[a] minimum information criterion is shown in italic type
[b] based on estimated posterior class membership probabilities

Family structure and socio-emotional wellbeing in the early years: a life course approach

Anna Pearce
Steven Hope
Hannah Lewis
Catherine Law
anna.pearce@ucl.ac.uk

Department of Health Policy Research Unit in the Health of Children, Young People and Families, and the Population, Policy and Practice Programme, UCL Institute of Child Health, University College London, UK

(Received August 2013 Revised August 2014) http://dx.doi.org/10.14301/llcs.v5i3.264

Abstract

Children living in reconstituted and lone parent families are at greater risk of poorer socio-emotional wellbeing than those in couple families. A life course approach can help us to understand the dynamics of family structure and how they influence child wellbeing, through consideration of sensitive and critical periods, accumulation, and trajectories of stability or change. We do this using data on 10,357 children from the UK Millennium Cohort Study from infancy to middle childhood. Family structure (natural couple, lone parent, reconstituted family) was measured at 9 months, 3, 5 and 7 years. We used a structured life course approach to examine how family structure might influence socio-emotional wellbeing throughout childhood. We also considered the role of early-life selection, and cumulative poverty (number of sweeps spent in income poverty). We found no evidence of sensitive or critical periods for exposure to certain family types. A measure capturing trajectories of family structure stability or change was as predictive of socio-emotional wellbeing at age 7 as a saturated measure representing all permutations of family structure over time. Compared to children living in a natural couple family throughout, all other groups were more likely to experience poor socio-emotional wellbeing, although children who were living in a natural couple family which transitioned to a lone parent family had a lower prevalence ratio (PR 1.80 [95% confidence interval: 1.54, 2.10]) than the other trajectory types, such as lone parent family throughout (PR=2.77 [2.34, 3.29]), or a lone parent family which transitioned to a reconstituted family (2.66 [1.99, 3.56]). Number of sweeps spent in a lone parent or reconstituted family was also as predictive of poor socio-emotional wellbeing as the saturated model, with the elevated risk increasing incrementally with every sweep spent in either of these family types (PR=1.86 [1.52, 2.26]) for 1 sweep, rising to 2.87 [2.46, 3.56] for 4). The association between both family structure measures and socio-emotional wellbeing were, in most cases, substantially attenuated after adjustment for early-life selection factors (such as maternal social class and separation of the mother's parents in childhood) and cumulative poverty throughout childhood. This analysis confirms that policies to provide support to vulnerable families in the early years and to reduce poverty are likely to benefit child wellbeing.

Keywords: family structure, socio-emotional wellbeing, early-life, poverty, life course

Introduction

Significant changes have occurred over recent decades in the family environments of children, including dramatic rises in lone parenthood and the incidence of family reconstitution (Bradshaw 2011; Ferri & Smith, 2003). Cross-sectional evidence has shown that children from lone parent and reconstituted families are at increased risk of poorer socio-emotional health in childhood (Amato & Keith, 1991; Dunn, Deater-Deckard, Pickering, O'Connor, & Golding, 1998; McMunn, Nazroo, Marmot, Boreham, & Goodman, 2001; Pearce, Lewis, & Law, 2013; Wadsworth, Burnell, Taylor, & Butler, 1985), although, due to the fluidity of family structure, it is important to examine these relationships longitudinally.

In life course epidemiology there are several complementary perspectives regarding the ways in which an individual's environment might affect their health and wellbeing over time. These have traditionally been examined across the whole of the life course, although they are also relevant over shorter periods of time, such as throughout childhood. They therefore might be used to enhance our understanding of the relationship between family structure dynamics and child socio-emotional wellbeing. The first perspective is referred to as *accumulation*, whereby the longer spent in adverse circumstances the greater the negative impact on health (Mishra et al., 2009). It is possible that potential negative effects of long-term lone parenthood on child socio-emotional wellbeing may strengthen over time, as exposure to economic hardship accumulates; on the other hand a child's wellbeing may improve with time, for example, as the level of conflict resulting from an initial family breakdown subsides (Amato & Keith, 1991). The second perspective pays more attention to the timing of exposure and hypothesises that there are particular periods during which exposures will have greater (*sensitive period*) or essential/irreversible (*critical period*) effects on health (Mishra et al., 2009). For example there is evidence that younger children find it easier than adolescents to adapt to a natural parent re-partnering (Rodgers & Pryor, 1998). A third perspective proposes that transitions or continuity in an exposure can influence outcomes (Hardy, Mishra & Kuh, 2008). Particular transitions in family circumstances may be especially influential for child wellbeing; for example the formation of step-families might be protective (through buffering against economic hardship), or stressful (due to changing relationship dynamics in the household) for children (Amato & Keith, 1991). A fourth perspective refers to social and health selection, whereby individuals have a greater or lesser likelihood of experiencing particular health and social statuses based on earlier experiences or exposures (Graham & Power, 2004). For example, mothers born into more disadvantaged backgrounds are more likely to experience relationship breakdown and to have poorer mental wellbeing (both of which may in turn influence the socio-emotional health of their child) (Conger, Conger, & Martin, 2010). However, few studies have investigated if any of these potential life course influences can be demonstrated empirically.

Analysis of the UK Millennium Cohort study examined whether transitions between family types (married, cohabiting, lone) between two time points were associated with behaviour at the start of primary school (Kiernan & Mensah, 2010). Compared to children living with married parents at age 9 months and at 5 years, children who lived in other family types (such as stable lone parent families, or where married parents separated and a new family was formed) were more likely to display externalising behavioural problems at age 5 years, before and after adjustment for a range of maternal socio-economic, demographic and health factors. Differences were also seen in internalising behaviour, although these were removed after adjustment for maternal characteristics. Two US studies set out to examine multiple life course processes through which family structure might be related to child socio-emotional wellbeing. One aimed to disentangle the independent effects of continuous exposure to lone parent and reconstituted families, and the number and nature of any transitions experienced, on socio-emotional wellbeing. The authors found evidence of cumulative effects of living in a lone parent or reconstituted family; they also found that behavioural problems increased with the number of transitions experienced, but that the type of transition mattered (a transition into a lone parent family by age 12 was associated with poorer behaviour scores, whereas a transition into a reconstituted family was not) (Magnuson & Berger, 2009). The other study found that time spent in a lone parent family between infancy and age 4 years

was associated with poorer socio-emotional wellbeing (suggesting accumulation), but that this relationship was removed after adjustment for baseline maternal characteristics (suggesting selection). The number of transitions in family structure ('instability') between ages 5 and 14 years were also predictive of behavioural problems over the same period, and remained so after adjustment for maternal characteristics and cumulative lone parenthood in the early years (Fomby & Cherlin, 2007). These findings may indicate that transitions in family structure, and the frequency of change, influence child wellbeing; they might also suggest that adolescence is a sensitive period for exposure to lone parenthood or reconstituted families, whereas early childhood is not.

It can be hard to disentangle which life course processes are truly occurring, due to the overlap between them (Hallqvist, Lynch, Bartley, Lang & Blane 2004). Mishra and colleagues have proposed a structured approach for considering how exposure via the various life course processes might influence an outcome of interest (Mishra et al., 2009). Firstly a model is estimated, containing a saturated exposure (comprising all permutations of status over time) to predict the outcome. Next, a number of alternative models are estimated, containing measures representing the different life course processes. Each life course model is compared (in terms of model fit using likelihood ratio (LR) tests) with the model containing the saturated measure. Any life course model which offers a model fit as good as the saturated model is then considered further. Since its proposal, the structured life course approach has been applied in several studies exploring the influence of social circumstances on health across the life course (Birnie, et al 2011; Cooper, Mishra & Kuh, 2011; Gustafsson, Persson & Hammarstrom, 2011; Murray et al., 2011; West et al., 2012 ; Wills, Hardy, Black & Kuh, 2010) and throughout childhood (Evans & Kohli, 1997; Giles et al., 2011; Lin, Leung, Hui, Lam and Schooling, 2011).

In this paper we aimed to investigate the relationships between family structure and socio-emotional wellbeing through early to middle childhood using a structured life course approach. In addition, we examined the potential explanatory roles of early-life selection and cumulative poverty.

Methods

We examined data from the Millennium Cohort Study (MCS), which follows children born in the UK in 2000-2002 and registered for the then universal Child Benefit. Data were downloaded from the UK Data Archive, University of Essex, in April 2014. A disproportionally stratified clustered sampling design was used to over-represent children living in Wales, Scotland and Northern Ireland, disadvantaged areas and areas with high proportions of ethnic minority groups (Plewis, 2004). Surveys were carried out by trained interviewers in the home with the main respondent (usually the mother). The first study contact with the cohort child was around age 9 months (MCS1), when information was collected on 18,818 infants (72% of those approached). We analysed data for a further three sweeps when the children were aged 3 (MCS2), 5 (MCS3) and 7 (MCS4) years. More information on the MCS is reported elsewhere (Connelly & Platt, 2014; Hansen, 2014).

Socio-emotional wellbeing

Socio-emotional wellbeing was assessed at 7 years using the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997), completed by the main respondent. We used the total difficulties score, which is made up of 20 items referring to four components: peer problems, conduct disorders, hyperactivity and emotional problems. We classified children, using recommended cut-offs (Goodman, Rowe, & Gan, 2010; Goodman, 1997), as having 'normal' or 'borderline-abnormal' scores. These cut-offs can be used in clinical or research settings to identify children at risk of psychiatric disorders. We also repeated our final models using linear regression and continuous SDQ scores for total difficulties.

Eighteen percent (2,149) of children had missing scores, mainly due to missing entries on 1-2 items on one or more components. We used a recommended method (Goodman et al., 2010) for imputing missing cases through rescaling the average (where data on just one or two items in any one component were missing). This reduced missing SDQ data to 3% (349).

Longitudinal measures of family structure

Family structure was classified at each sweep as: *natural couple families* consisting of two "natural" (wording used in the survey, meaning biological)

parents, *reconstituted families* (one natural parent and one adoptive/foster or step parent), *lone parent families* (one natural father or mother). The 0.8% (94) of children living in rare family structures (e.g. grandparent families) were excluded due to low numbers.

We considered a number of longitudinal family structure measures, each representing the life course processes outlined below and in Figure 1. These were compared to a saturated measure of family structure, which consisted of all permutations of family structure across the four sweeps. This provided 81 possible permutations (3^4), of which 55 were observed in the MCS (see supplementary data, Table S1).

Critical periods

A critical period was taken to occur when exposure at only a particular time point was of importance, and exposure at any other time was inconsequential. We explored ages 9 months, 3, 5 and 7 years as potential critical periods, whereby only family structure for a particular sweep was included in the model, and status at all other sweeps was held at 0. See Figure 1 for additional descriptions and notation.

Sensitive periods

Sensitive periods were said to occur if the effect of exposure at one point in time was larger than the effects of exposure at other times. We examined sensitive periods at MCS1, 2, 3 and 4 by entering family structure at each of the four sweeps into a single model. Where the risk of borderline-abnormal SDQ associated with exposure at any one sweep appeared to be larger compared to the others, we tested this using the *lincom* command in Stata.

Cumulative exposure

Cumulative measures consisted of time spent in a given state, and were therefore well suited to binary measures of exposure; although when dealing with interval variables (or equally spaced ordinal categorical variables) the value from each time point can be summed over time (Cooper et al., 2011). The family structure variable was made up of three nominal categories (natural couple, lone parent, reconstituted family) and it was not possible to assign quantitative values to these groups. We therefore created binary variables through comparing one group to the other two combined. We considered two cumulative measures. The first combined lone parents and reconstituted families (baseline natural couple families), because the socio-emotional wellbeing scores of children from lone parent and reconstituted families were similar at age 7 (89% of children in natural couple families had 'normal' SDQ scores, compared to 75% in reconstituted families and 78% lone parent families) (Pearce et al., 2013). This measure referred to the number of sweeps spent either in a lone parent *or* reconstituted family (0-4).

For the second measure we combined *all* couples (natural and reconstituted) to create a baseline group, compared to lone parent families. We examined this measure because, as postulated by Amato and Keith (1991), it may be that the absence of two parental figures is detrimental to child socio-emotional wellbeing, or the resultant economic hardship comes from living in a lone parent as opposed to a couple family. Indeed the prevalence of income poverty at age 7 years was lower for natural couples (16%) and reconstituted (35%) families compared to lone parent families (60%). This cumulative measure therefore referred to the number of sweeps (0-4) spent living in a lone parent family.

There are two approaches to testing for accumulation (Cooper et al., 2011). With 'strict' accumulation, the level of risk would increase linearly with every sweep exposed, whereas for 'relaxed' accumulation the increase in risk does not have to be linear. Because intervals between MCS sweeps were relatively short and not uniform, we took a "relaxed" approach to measuring accumulation.

Family trajectories

This measure was designed to capture stability or change in family structure. As pointed out by Mishra and colleagues, a 'social mobility' model capturing exposures over more than three time points cannot easily be parameterised (Mishra et al., 2009). They suggest that, in these situations, simpler models ought to be identified. We constructed a measure made up of groups specified *a priori* (*Natural couple throughout*, *Lone family throughout*, *Natural couple to lone*, *Lone to reconstituted family*, *Natural couple to reconstituted family)*. An additional group (*Lone to natural couple*) was included because it was relatively common in the sample (2.6%, N=319); and the smallest groups were combined into the *'Other'* category (comprising children who had experienced two or more changes in family structure, or who were living in a reconstituted family at age 9 months (Table S1)). This measure did not take into account timing of any change.

Poverty

Household income (including benefits) was reported in bands by the main respondent at each sweep. Income poverty was defined as a household income of below 60% of the contemporary national median, before housing costs, and using a modified OECD equivalence scale (Bradshaw & Holmes, 2010). This measure of poverty was chosen because it was collected at, and comparable across, all sweeps. It is also widely employed by researchers and used to monitor national targets and compare international trends. The risk of poor socio-emotional wellbeing has been shown to increase with time spent in poverty (Kiernan & Mensah, 2009; McLeod & Shanahan, 1996; Najman et al., 2010). Therefore we examined the number of sweeps spent in poverty between 9 months and 7 years ranging from 0 (not living in poverty at any sweep) to 4 (living in poverty at all sweeps). We confirmed that it was a suitable longitudinal measure for predicting socio-emotional wellbeing by comparing it to a saturated measure capturing all poverty combinations across the 4 sweeps (LR test p=0.62). The association between poverty and socio-emotional wellbeing is shown in Table 1. Missing income data were multiply-imputed by the data owners (Hansen, 2014).

Potential early-life selection factors

We examined a number of early-life selection factors that were likely to have occurred prior to pregnancy and might have influenced the likelihood of mothers moving into certain family types (Ferri & Smith, 2003; Pearce et al., 2013): whether the mother's parents had permanently separated before she was aged 18 years, her age at MCS child's birth, ethnicity (collapsed into white and non-white due to small numbers), social class (National Statistics Socio-economic Classification [NS-SEC]) and highest educational qualification. All of these selection variables, except ethnicity, were associated with family structure and socio-emotional health, and altered the association between them, and were therefore retained in adjusted models. Social class, education and poverty all capture aspects of social disadvantage. We checked for multi-collinearity using the Variance Inflation Factor (VIF). All VIF values were <5, indicating a low chance of multi-collinearity (Menard, 1995).

In the majority of cases of family breakdown recorded in the MCS the child remained with the mother; therefore we examined early life characteristics of the mother. However we repeated our final models using a measure of household social class (representing the highest social class of the two parents, where relevant, at MCS1/9months) and the pattern of results remained similar.

The association between the potential selection factors and socio-emotional wellbeing are shown in Table 1. The level of missing data for each variable is listed in Table 1.

Potential confounders

We explored a number of potential confounders which have been shown to be associated with family structure and/or child wellbeing. All were captured at 9 months: infant temperament (Carey Infant Temperament Scale), family size, birthweight (z-scores, sex and gestational age adjusted), gestational age (weeks), breast-feeding duration (months), and maternal psychological distress (using a modified Malaise inventory).

Cohort member age at MCS4, birthweight, gestational age, breastfeeding duration, and maternal psychological distress are presented as categorical measures in the descriptive statistics (Table 1), though were entered into regression models as continuous variables. Variables which were associated with the family trajectories variable and SDQ, and that significantly altered the association between them, were retained in adjusted models. These were maternal psychological distress, infant temperament and breastfeeding duration.

The association between the potential confounders and socio-emotional wellbeing are shown in Table 1. At age 7 years 6.7% (659) of children had borderline, and 7.2% (776) abnormal, SDQ scores. The prevalence of borderline-abnormal SDQ scores significantly increased from 8.1% in those never reporting poverty to 27.8% of those reporting it at all four sweeps. All of the confounders (except family size and child's age) and early-life selection factors were associated with borderline-abnormal scores. The level of missing data for each variable is listed in Table 1.

Sex was not considered to be a potential confounder (since it is not related to family structure), but instead was examined as a potential effect modifier. We found no significant interaction between family structure and sex, and so results are presented for both sexes combined.

Figure 1: Measures of family structure used in the saturated and life course models

	Life course measures of family structure	***Notation, where X=family structure; 1=MCS1, 2=MCS2, 3=MCS3, 4=MCS4***
Saturated	***"Natural couple", "lone parent", "reconstituted family"****: at MCS1, 2, 3 and 4= 81 possible permutations (55 were observed in the MCS).*	$E(Y) = \alpha + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \Theta_{12}X_1X_2 + \Theta_{23}X_2X_3 + \Theta_{13}X_1X_3 + \Theta_{14}X_1X_4 + \Theta_{24}X_2X_4 + \Theta_{34}X_3X_4 + \Theta_{123}X_1X_2X_3 + \Theta_{234}X_2X_3X_4 + \Theta_{134}X_1X_3X_4 + \Theta_{124}X_1X_2X_4 + \Theta_{1234}X_1X_2X_3X_4$
Critical periods	***MCS1:*** *Family structure at 9 months: natural couple, reconstituted couple, or lone parent*	$E(Y) = \alpha + \beta 1X1$ *Constraints:* $\beta_2X_2 = \beta_3X_3 = \beta_4X_4 = 0$; $\Theta_{12} = \Theta_{23} = \Theta_{13} = \Theta_{14} = \Theta_{24} = \Theta_{34} = \Theta_{123} = \Theta_{234} = \Theta_{134} = \Theta_{124} = \Theta_{1234} = 0$
	MCS2: *Family structure at 3 years: natural couple, reconstituted couple, or lone parent*	$E(Y) = \alpha + \beta 1X_2$ *Constraints:* $\beta_1X_1 = \beta_3X_3 = \beta_4X_4 = 0$; $\Theta_{12} = \Theta_{23} = \Theta_{13} = \Theta_{14} = \Theta_{24} = \Theta_{34} = \Theta_{123} = \Theta_{234} = \Theta_{134} = \Theta_{124} = \Theta_{1234} = 0$
	MCS3: *Family structure at 5 years: natural couple, reconstituted couple, or lone parent*	$E(Y) = \alpha + \beta 1X_3$ *Constraints:* $\beta_1X_1 = \beta_2X_2 = \beta_4X_4 = 0$; $\Theta_{12} = \Theta_{23} = \Theta_{13} = \Theta_{14} = \Theta_{24} = \Theta_{34} = \Theta_{123} = \Theta_{234} = \Theta_{134} = \Theta_{124} = \Theta_{1234} = 0$
	MCS4: *Family structure at 7 years: natural couple, reconstituted couple, or lone parent*	$E(Y) = \alpha + \beta 1X_4$ *Constraints:* $\beta_1X_1 = \beta_2X_2 = \beta_3X_3 = 0$; $\Theta_{12} = \Theta_{23} = \Theta_{13} = \Theta_{14} = \Theta_{24} = \Theta_{34} = \Theta_{123} = \Theta_{234} = \Theta_{134} = \Theta_{124} = \Theta_{1234} = 0$
Sensitive periods	***MCS1:*** *Family structure at MCS1, 2, 3 and 4: natural couple, reconstituted couple, or lone parent; mutually adjusted.*	$E(Y) = \alpha + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4$ *Constraints:* $\beta_1X_1 > \beta_2X_2, \beta_3X_3, \beta_4X_4$; $\Theta_{12} = \Theta_{23} = \Theta_{13} = \Theta_{14} = \Theta_{24} = \Theta_{34} = \Theta_{123} = \Theta_{234} = \Theta_{134} = \Theta_{124} = \Theta_{1234} = 0$
	MCS2: *Family structure at MCS1, 2, 3 and 4: natural couple, reconstituted couple, or lone parent; mutually adjusted.*	$E(Y) = \alpha + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4$ *Constraints:* $\beta_2X_2 > \beta_1X_1, \beta_3X_3, \beta_4X_4$; $\Theta_{12} = \Theta_{23} = \Theta_{13} = \Theta_{14} = \Theta_{24} = \Theta_{34} = \Theta_{123} = \Theta_{234} = \Theta_{134} = \Theta_{124} = \Theta_{1234} = 0$
	MCS3: *Family structure at MCS1, 2, 3 and 4: natural couple, reconstituted couple, or lone parent; mutually adjusted.*	$E(Y) = \alpha + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4$ *Constraints:* $\beta_3X_3 > \beta_1X_1, \beta_2X_2, \beta_3X_3$; $\Theta_{12} = \Theta_{23} = \Theta_{13} = \Theta_{14} = \Theta_{24} = \Theta_{34} = \Theta_{123} = \Theta_{234} = \Theta_{134} = \Theta_{124} = \Theta_{1234} = 0$

(Figure 1 cont'd)

	MCS4: *Family structure at MCS1, 2, 3 and 4: natural couple, reconstituted couple, or lone parent; mutually adjusted.*	$E(Y) = \alpha + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4$ Constraints: $\beta_4X_4 > \beta_1X_1, \beta_2X_2, \beta_3X_3$; $\Theta_{12} = \Theta_{23} = \Theta_{13} = \Theta_{14} = \Theta_{24} = \Theta_{34} = \Theta_{123} = \Theta_{234} = \Theta_{134} = \Theta_{124} = \Theta_{1234} = 0$
Cumulative exposures	***Number of sweeps spent in a lone parent or reconstituted family*** *(vs. natural couple): 5 categories ranging from 0-4.* ***Number of sweeps spent in lone parent family*** *(vs. natural couple or reconstituted family): 5 categories ranging from 0-4.*	$\alpha + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4$ Constraints: $\Theta_{12} = \Theta_{23} = \Theta_{23} = \Theta_{13} = \Theta_{14} = \Theta_{24} = \Theta_{34} = \Theta_{123} = \Theta_{234} = \Theta_{134} = \Theta_{124} = \Theta_{1234} = 0$
Family trajectories	***7 groups representing stability or change:*** *Natural couple throughout; Lone parent throughout; Natural to lone; Lone to natural; Lone to reconstituted; Natural to reconstituted; Other (2+ transitions [except for natural-lone-reconstituted], or started off as reconstituted family).*	*N/A*

Table 1. Prevalence of family structure, poverty and covariates, total weighted % (N) and weighted % (N) with borderline-abnormal difficulties scores*

	% (N) Total	% (N) borderline abnormal
Poverty score		
Never in poverty	55.1 (6008)	8.1 (459)
Poverty at 1 sweep	14.1 (1546)	14.4 (194)
Poverty at 2 sweeps	9.0 (984)	17.0 (161)
Poverty at 3 sweeps	8.9 (1005)	25.1 (237)
Poverty at 4 sweeps	13.0 (1444)	27.8 (370)
P value		*p*<0.001
Confounders		
Family size (MCS1, 9 months)		
1 child	42.0 (4618)	14.2 (611)
2-3 children	51.3 (5644)	13.3 (700)
4 plus children	6.8 (799)	16.0 (124)
P value		*p*=0.226
Child's sex		
Male	50.9 (5591)	16.5 (869)
Female	49.1 (5470)	11.1 (566)
P value		*p* <0.001
Child's age (MCS4, 7 years)		
6.5 years	1.9 (215)	13.8 (25)
7 years	52.2 (5792)	13.8 (747)
7.5 years	44.8 (4931)	14.0 (648)
8 years	1.0 (123)	12.9 (15)
P value		*p* = 0.977
Birthweight		
Low <2.5kg	6.0 (644)	20.13 (127)
Normal 2.5-4.5kg	92.2 (10110)	13.51 (1268)
High >4.5kg	1.8 (202)	9.21 (20)
P value		*p* =<0.001
Gestational age		
Preterm (23-31 weeks)	0.7 (76)	25.4 (22)
Moderate preterm (32-33 weeks)	0.8 (90)	13.3 (13)
Late preterm (34-36 weeks)	4.2 (444)	17.7 (68)
Early term (37-38 weeks)	14.1 (1539)	16.1 (225)
Late term (39-41 weeks)	69.0 (7579)	13.1 (934)
Post-term (42-43 weeks)	11.3 (1201)	13.0 (146)
P value		*p* =0.003
Breast-feeding duration		
Never	31.2 (3324)	18.5 (570)
1 week or less	11.8 (1458)	15.2 (211)
1-6 weeks	13.1 (1413)	17.6 (212)
6 weeks-4 months	16.3 (1812)	11.5 (203)
4 months and more	27.6 (3050)	7.6 (237)
P value		*p* =<0.001
Maternal psychological distress (MCS1, 9 months)		
Normal	86.7 (9323)	11.7 (1008)
Psychological distress	13.3 (1458)	26.8 (371)
P value		*p* =<0.001
Baby's temperament (MCS1; 9 months)		
Normal	18.5 (1972)	8.8 (168)

(Table 1 cont'd)

Difficult on 1-2 components^	68.7 (7285)	14.2 (929)
Difficult on all 3 components^	12.9 (1395)	18.2 (263)
P value		*p* =<0.001
Selection factors		
Carer age (at MCS birth)		
14-19	7.3 (695)	26.6 (179)
20-24	16.5 (1760)	21.8 (353)
25-29	28.3 (3099)	13.5 (396)
30-34	30.7 (3530)	9.6 (327)
35-39	15.0 (1714)	9.3 (160)
40-44	2.1 (261)	7.3 (19)
P value		*p* <0.001
Ethnicity (child)		
White	88.25 (9614)	13.5 (1194)
Non-white	11.8 (1421)	16.4 (237)
P value		*p* =0.036
Mother's parents separated in childhood		
No	75.0 (8446)	12.2 (981)
Yes	25.0 (2499)	19.0 (422)
P value		*p* <0.001
Maternal social class (MCS1, 9 months)		
Managerial & professional	31.1 (3530)	6.2 (222)
Intermediate	18.9 (2021)	11.0 (202)
Small employers & own accounts	4.3 (428)	12.0 (44)
Lower supervisory & technical	5.5 (615)	15.3 (87)
Semi routine & routine	34.0 (3636)	19.9 (666)
Never worked & L/T unemployed	6.6 (771)	27.1 (195)
P value		*p* <0.001
Maternal education (MCS1, 9 months)		
Higher degree	3.5 (437)	7.9 (34)
Degree	14.4 (1713)	4.4 (78)
Diploma	9.3 (1072)	9.6 (100)
A levels	9.7 (1149)	8.4 (98)
GCSE D-G	35.8 (3788)	13.3 (484)
GCSE A*-C	11.0 (1112)	21.1 (223)
Other	2.0 (234)	15.6 (39)
None	14.3 (1540)	26.7 (374)
P value		*p* <0.001

*Notes. ^positivity, receptiveness and regularity. Missing data: income poverty 74, family size 0, sex 0, child's age 0, birthweight 12, gestation 131, breastfeeding 4, maternal psychological distress 280, infant temperament 409, maternal age at birth of cohort child 2, ethnicity 26, mother's parents divorced/separated 150, social class 60, maternal education 16. *Ns are unweighted.*

Statistical Analysis

Analyses were conducted in Stata/SE 12.1 (Stata Corporation, TX), using 'svy' commands to account for the sampling design and attrition. Descriptive statistics (frequencies and weighted percentages) were estimated for all variables (Table 1). The associations between socio-emotional wellbeing and family structure and the covariates were assessed by estimating prevalence ratios (PR) and 95% confidence intervals (CIs) for borderline-abnormal SDQ scores, using Poisson regression models (and robust standard errors) (Barros & Hirakata, 2003; Lee, Chang, & Chia, 2009). Means

were estimated using linear regression, allowing for adjustment of other covariates. Analyses were carried out in the following stages:

A (1): Association between family structure and socio-emotional wellbeing (Table 2): we estimated a number of models for predicting borderline-abnormal SDQ, each containing one of the life course measures of family structure outlined earlier.

A (2): Comparing life course measures of family structure (Table 2): the life course measures of family structure were compared to a model containing the saturated measure of family structure (Table S1) using likelihood ratio (LR) tests, and Akaike information criterion (AIC)[i]. A life course measure was taken to be as parsimonious as the saturated measure when the p-value from the LR-test exceeded 0.20[ii]. Any family structure measures fitting this criterion were then examined further in stages C and D.

C: Examining family structure and poverty (Table 3): we estimated the mean number of sweeps spent in poverty, according to parsimonious life course family structure measures identified in B.

D: Unadjusted and adjusted relationships between family structure and SDQ (Table 4): we estimated unadjusted prevalence ratios (uPRs) for borderline-abnormal SDQ scores, according to the family structure measures identified in B, then adjusting for confounders (aPR[1]), selection factors (aPR[2]), and cumulative poverty (aPR[3]).

Working sample

11,538 (61% of the responding sample included in the analysis) singleton children took part in the four relevant MCS sweeps. Sixteen children were omitted from the analysis because the main respondent was not the natural mother at MCS1 (when the information on early-life factors such as breast-feeding and birthweight was collected), and a further 60 were excluded because the main respondent at MCS2-4 was not a parent (natural or otherwise). Sensitivity analyses were carried out to ensure that this did not bias the results (data not shown).

Children were further excluded from all analyses if they had missing data on family structure (52) or SDQ (349), reducing the sample size to 11,061. Further exclusions were made for analyses in C and D above, so that PRs could be directly compared before and after adjustment: income (74), or any of the selection measures or potential confounders (630; missing data listed for each variable under Table 1). This reduced the working sample to 10,357.

Results

A. Family structure and SDQ

Table 2 shows PRs for borderline-abnormal SDQ scores (at age 7 years) for each of the family structure measures (Column C), and degrees of freedom (D), p-values (E) and AIC (F) for the comparisons made with the saturated measure of family structure (which consisted of all permutations across the four sweeps [Table S1]) using LR tests. There appeared to be no *critical periods*. The PRs in lone and reconstituted families at MCS4 appeared to be higher than in earlier sweeps, however the *sensitive periods measure* did not quite reach the arbitrary level for parsimony. While there was no indication of a *cumulative effect of lone parenthood* (with all couples as the baseline), the *cumulative measure capturing the number of sweeps spent in either a lone parent or reconstituted family* provided a model fit that was as good as the saturated family structure measure. Finally, the *family trajectories* also proved to fit the data well.

In the following sections we focus on the two most parsimonious measures of family structure (family trajectories and the number of sweeps spent in a lone parent or reconstituted family).

Table 2. Percentage (N) and prevalence ratios (PRs) for borderline-abnormal (b-a) SDQ scores, according to the life course measures of family structure. N=11,061.

Column:	A	B	C	D	E	F
Family structure measure	**% (N) overall**	**%(N) b-a scores**	**PR (95% CI)**	**Degrees of freedom**	**P value**	**AIC**
Saturated measure				**54**		8555
Critical periods (unadjusted)						
MCS1 (9 months)						
Natural couple	86.9 (9,615)	11.4 (1091)	1	*3*	*P*<0.001	8611
Reconstituted family	0.2 (26)	34.6 (9)	3.05 (1.58, 2.88)			
Lone parent	12.8 (1,420)	23.6 (335)	2.08 (1.84, 2.35)			
MCS2 (3 years)						
Natural couple	86.5 (9,232)	10.8 (995)	1	*3*	*P*<0.001	8562
Reconstituted family	2.0 (220)	27.7 (61)	2.57 (1.99, 3.33)			
Lone parent	14.6 (1609)	23.6 (379)	2.19 (1.94, 2.46)			
MCS3 (5 years)						
Natural couple	79.2 (8,760)	10.5 (916)	1	*3*	*P*<0.001	8551
Reconstituted family	4.2 (459)	27.9 (128)	2.67 (2.22, 3.21)			
Lone parent	16.7 (1842)	21.2 (391)	2.03 (1.80, 2.29)			
MCS4 (7 years)						
Natural couple	75.5 (8351)	10.1 (843)	1	*3*	*P*=0.001	8540
Reconstituted family	5.9 (656)	24.9 (163)	2.46 (2.08, 2.91)			
Lone parent	18.6 (2054)	20.9 (429)	2.07 (1.84, 2.32)			
Sensitive periods (mutually adjusted)						
MCS1 (9 months)				*9*	*P*=0.177	8519
Natural couple	86.9 (9,615)	11.4 (1091)	1			
Reconstituted family	0.2 (26)	34.6 (9)	1.55 (0.79, 3.07)			
Lone parent	12.8 (1,420)	23.6 (335)	1.27 (1.08, 1.50)			
MCS2 (3 years)						
Natural couple	86.5 (9,232)	10.8 (995)	1			
Reconstituted family	2.0 (220)	27.7 (61)	1.23 (0.87, 1.76)			
Lone parent	14.6 (1609)	23.6 (379)	1.27 (1.02, 1.58)			
MCS3 (5 years)						
Natural couple	79.2 (8,760)	10.5 (916)	1			
Reconstituted family	4.2 (459)	27.9 (128)	1.23 (0.86, 1.76)			
Lone parent	16.7 (1842)	21.2 (391)	0.97 (0.75, 1.26)			
MCS4 (7 years)						
Natural couple	75.5 (8351)	10.1 (843)	1			
Reconstituted family	5.9 (656)	24.9 (163)	1.67 (1.24, 2.25)			
Lone parent	18.6 (2054)	20.9 (429)	1.64 (1.33, 2.02)^			
Cumulative lone parent/reconstituted family (baseline natural couples)						
0 sweeps	71.9 (7947)	9.6 (762)	1	5	***P*=0.63**	**8502**
1 sweeps	6.5 (718)	18.4 (132)	1.92 (1.59, 2.31)			
2 sweeps	5.9 (650)	19.2 (125)	2.01 (1.66, 2.42)			
3 sweeps	6.5 (716)	21.1 (151)	2.20 (1.85, 2.62)			
4 sweeps	9.3 (1030)	25.7 (265)	2.68 (2.33, 3.09)			
Cumulative lone parent (baseline all couples)						
0 sweeps	73.4 (8188)	10.0 (808)	1	5	*P*<0.01	8542
1 sweeps	8.7 (965)	20.2 (195)	2.03 (1.74, 2.37)			
2 sweeps	6.2 (682)	19.7 (134)	1.97 (1.64, 2.37)			
3 sweeps	5.3 (588)	20.6 (121)	2.07 (1.71, 2.50)			
4 sweeps	6.4 (708)	25.0 (177)	2.51 (2.13, 2.96)			

(Table 2 cont'd)

Family trajectories						
Natural couple throughout	71.9 (7947)	9.6 (762)	1	7	***P*=0.56**	**8506**
Lone parent throughout	6.4 (708)	25.0 (177)	2.61 (2.21, 3.07)			
Natural couple to lone	10.2 (1130)	18.1 (204)	1.88 (1.61, 2.20)			
Lone to natural couple	2.7 (300)	21.0 (63)	2.19 (1.69, 2.83)			
Lone to reconstituted	2.2 (240)	25.4 (61)	2.65 (2.04, 3.44)			
Natural to reconstituted	3.4 (371)	24.3 (90)	2.53 (2.03, 3.15)			
Other	3.3 (365)	21.5 (78)	2.23 (1.77, 2.81)			

Notes. Percentages and PRs are not weighted (in order to carry out likelihood ratio tests). ^ Significantly higher than at MCS3, but not MCS2 or 1.

B. Family structure and poverty

Mean number of sweeps spent in poverty increased incrementally with number of sweeps spent in a lone parent or reconstituted family, from 0.58 sweeps in poverty for those not living in a lone parent/reconstituted family at any sweep, to 2.83 for those living in one at all four sweeps (Table 3). Differences between groups were reduced slightly after adjustment for confounders and selection factors.

In the family trajectories measure, children living with a lone parent throughout experienced the longest periods of poverty (mean of 2.99 sweeps); unsurprisingly the mean value was similar to those identified in the cumulative measure as living with a lone parent or a reconstituted family for four sweeps (2.83 sweeps, as discussed above). However, the trajectories measure also indicated that children living in natural couple families which became reconstituted, or natural couple families which became lone, spent relatively short periods in poverty (with respective means of 1.48 and 1.73 sweeps). In contrast, lone parent families which transitioned to a reconstituted family or returned to being a natural couple, experienced relatively high rates of poverty (2.40 and 2.36 respectively). Again, differences between groups were reduced after adjustment for confounders and selection factors.

Table 3. Mean number of sweeps in poverty according to family structure, before and after adjustment for baseline characteristics (confounders and early-life selection factors), N=10,357

	Mean (CI) sweeps in poverty	**Adj. baseline characteristics***
No. sweeps in lone parent/reconstituted family		
0	0.58 (0.52, 0.63)	1.32 (1.02, 1.61)
1	1.58 (1.44, 1.72)	1.91 (1.60, 2.22)
2	1.74 (1.58, 1.89)	2.08 (1.77, 2.39)
3	2.19 (2.03, 2.35)	2.38 (2.09, 2.68)
4	2.83 (2.73, 2.92)	2.83 (2.55, 3.14)
Family trajectories		
Natural couple at all 4 sweeps	0.58 (0.52, 0.63)	1.33 (1.03, 1.62)
Lone at all 4 sweeps	2.99 (2.88, 3.10)	3.00 (2.71, 3.30)
Natural to lone	1.73 (1.60, 1.86)	2.14 (1.83, 2.44)
Lone to natural couple	2.36 (2.16, 2.56)	2.34 (2.01, 2.67)
Lone to reconstituted	2.40 (2.20, 2.60)	2.43 (2.12, 2.73)
Natural to reconstituted	1.48 (1.28, 1.67)	1.80 (1.50, 2.11)
Other	2.47 (2.29, 2.65)	2.54 (2.23, 2.86)

*Notes. *maternal psychological distress, breastfeeding duration in weeks, baby's temperament, maternal age at birth of MCS child, education (baseline: higher degree) and social class (baseline: managerial and professional), and whether mother's parents separated when she was < 18 years (baseline: no separation).*

C. Family structure and socio-emotional wellbeing, before and after adjustment for poverty

Table 4 presents PRs for borderline-abnormal SDQ scores according to the two life course family structure measures (number of sweeps spent in a lone parent/reconstituted family, and trajectories of family type), before and after adjustment for covariates. There was no interaction between child's sex and time spent in a lone parent/ reconstituted family (p=0.12) or family trajectories (p= 0.36)

The prevalence of borderline-abnormal SDQ increased with duration spent in a lone parent or reconstituted family, from a PR of 1.86 (1.52, 2.26) for one sweep, to 2.87 (2.46, 3.56) for all four sweeps. PRs were attenuated after adjustment for confounders, early-life selection factors and poverty (ranging from 1.27 [1.03, 1.57] to 1.41 [1.18, 1.68]).

All family structure trajectories were associated with an elevated prevalence of borderline-abnormal socio-emotional wellbeing, when compared to children who were living with a natural couple throughout. Children who were living in a family which transitioned from a natural couple to a lone parent family had the lowest PR (1.80 [1.54, 2.10]); prevalence ratios ranged from 2.5 to 2.8 for all other groups. The elevated prevalence ratios seen in all of the family trajectory groups were reduced considerably, but remained significant, after adjustment for selection factors (aPR[2]). Further attenuation occurred after adjustment for cumulative poverty (aPR[3]).

Patterns remained the same when analyses were repeated using the total SDQ score as a continuous outcome (data not shown).

Table 4. Cumulative family structure (model 1) and family trajectories (model 2). Weighted prevalence ratios (PR) and 95% confidence intervals (CIs) for borderline–abnormal SDQ scores, unadjusted and adjusted for confounders, selection factors and poverty N=10,357

	Unadjusted PR	aPR[1]	aPR[2]	aPR[3]
Model 1: Cumulative family structure (number of sweeps spent in a reconstituted/lone parent family				
No sweeps	1	1	1	1
One	1.86 (1.52, 2.26)	1.64 (1.34, 2.01)	1.35 (1.11, 1.65)	1.27 (1.03, 1.57)
Two	2.05 (1.64, 2,57)	1.69 (1.34, 2.13)	1.42 (1.13, 1.79)	1.33 (1.05, 1.69)
Three	2.33 (1.93, 2.80)	1.82 (1.52, 2.19)	1.43 (1.18, 1.71)	1.29 (1.07, 1.55)
Four	2.87 (2.46, 3.56)	2.18 (1.86, 2.54)	1.71 (1.36, 1.87)	1.41 (1.18, 1.68)
Model 1: Family trajectories				
Natural couple at all sweeps	1	1	1	1
Lone at all sweeps	2.77 (2.34, 3.29)	2.13 (1.81, 2.51)	1.57 (1.33, 1.87)	1.36 (1.13, 1.64)
Natural to lone	1.80 (1.54, 2.10)	1.53 (1.30, 1.81)	1.29 (1.09, 1.53)	1.20 (1.01, 1.43)
Lone to natural couple	2.62 (2.01, 3.40)	2.04 (1.55, 2.68)	1.50 (1.15, 1.70)	1.36 (1.03, 1.79)
Lone to reconstituted	2.66 (1.99, 3.56)	2.05 (1.53, 2.76)	1.51 (1.13, 2.01)	1.37 (1.02, 1.83)
Natural to reconstituted	2.53 (1.98, 3.22)	2.07 (1.63, 2.65)	1.71 (1.35, 2.17)	1.62 (1.27, 2.06)
Other	2.60 (2.02, 2.60)	1.95 (1.53, 2.50)	1.46 (1.13, 1.88)	1.31 (1.01, 1.69)

Notes. aPR[1]:PR for borderline-abnormal SDQ by family structure, adjusting for maternal psychological distress at MCS1, breast-feeding duration in weeks and baby's temperament. aPR[2]: adjusting for [1] and maternal age at birth of MCS child, education (baseline: higher degree) and social class (baseline: managerial and professional), and whether mother's parents separated when she was < 18 years (baseline: no separation). aPR[3]: Adjusting for [2] and cumulative poverty.

Discussion

Summary of findings

Using a structured life course approach we assessed how family structure over the first seven years might influence socio-emotional wellbeing at age 7 years. We found no evidence of critical periods for exposure to certain family structure types between birth and 7 years, although there was some indication that age 7 might be a sensitive period for exposure to lone parenthood. A cumulative measure capturing the number of sweeps spent in a lone parent family did not predict socio-emotional wellbeing as well as the saturated model, although number of sweeps in either a lone parent *or* reconstituted family did. The family trajectories measure also provided a good model fit. Thirty percent of children lived in these higher risk groups (i.e. families which were not headed by a natural couple throughout the period under study).

Through considering two measures of family structure we have observed the differences in prevalence of borderline-abnormal SDQ scores between family structure trajectories and the increase in scores associated with time spent in lone or reconstituted families. After adjustment for confounders, but before adjustment for selection or poverty, children living in a lone parent or reconstituted family for one sweep were 60% more likely to display borderline-abnormal scores than those living with both natural parents throughout. This rose to a greater than twofold risk for those living in a lone parent or reconstituted family at all four sweeps. When looking at trajectories, children who were living with both natural parents in infancy but had moved to a lone parent family by age 7 years were 53% more likely to experience borderline-abnormal scores than those living with a natural lone parent throughout. The remaining trajectory groups were all around twice as likely to experience borderline-abnormal behaviour. A reduction in PRs after adjustment for selection factors (such as the mother's social class, or whether her parents had separated when she was a child) reflects the potential importance of factors from the mother's own childhood through to the birth of the cohort child. Intervention and support in early life and across generations is therefore likely to be important for family and child wellbeing.

Comparison with other findings

An earlier analysis of the Millennium Cohort found that children who were not living with both natural parents during the preschool years were more likely to experience externalising (but not internalising) behavioural problems. As in the present analysis, children who were living in a stable lone parent family and those whose mothers re-partnered were particularly at risk, although this was to some extent confounded by maternal characteristics and poverty (Kiernan & Mensah, 2010). Two studies, examining a number of different life course processes through which family structure might influence behaviour, found that selection and the number of transitions (or 'instability') were predictive of child behaviour (Fomby & Cherlin, 2007; Magnuson & Berger, 2009). Magnuson & Berger also found that cumulative exposure to lone parent or reconstituted families and type of transition was important; children who had moved into a lone parent family by age 12 years had poorer behaviour scores than children living with both parents, while those who transitioned into a reconstituted family did not (Magnuson & Berger, 2009). Findings from the present analysis also indicate that the type of transition matters for child socio-emotional wellbeing; for example, children experiencing a transition from a natural couple to a reconstituted family had higher PRs than those who experienced the transition from a natural couple to a lone parent family. Children who experienced two or more transitions also had higher rates of borderline-abnormal SDQ scores. There was some indication that the prevalence of borderline-abnormal SDQ scores associated with lone parenthood was greater when the child were slightly older (age 7). This may be because the exposure was coterminous with the outcome, or could indicate increasing sensitivity to family changes with age, as reported in earlier work (Rogers & Pryor, 1998).

Children living in lone parent or reconstituted families are more likely to experience poverty (Department for Work and Pensions, 2012), and changes in family structure are often accompanied by changes in household income (Bradshaw & Homes, 2008; Tomlinson & Walker, 2012;) Bradshaw & Holmes, 2010; Panico, Bartley, Kelly, McMunn, & Sacker, 2010). Cross-sectional research has demonstrated that living in poverty influences children's health and wellbeing (Marmot, 2012; Ridge, 2011), and that accounting for the higher

rates of poverty in some family types reduces the association between family structure and child wellbeing (Pearce et al., 2013; Spencer, 2005). However, poverty can be fluid in nature (Graham & Power, 2004), and experiences of those who are exposed to persistent, rather than transient, poverty are likely very different (Bradshaw 2011a). Time spent in poverty has been associated with the risk of poor socio-emotional wellbeing in childhood (Kiernan & Mensah, 2009; McLeod & Shanahan, 1996; Najman et al., 2010), and our findings support this. Although one study found that cumulative poverty throughout early childhood attenuated the cross-sectional association between family structure and socio-emotional wellbeing at age three years (Kiernan & Mensah, 2009), and a second found that cross-sectional poverty mediated the association between a longitudinal measure of family structure and socio-emotional health by primary school age (Kiernan & Mensah, 2010), to our knowledge no study has examined these relationships using longitudinal measures of both family structure and poverty in the UK. In this paper we have identified subtleties in relationships not apparent from earlier analyses using cross-sectional data. For example, of the children living in a lone parent family at age 7 years, those living in a natural couple family in infancy had a substantially lower prevalence of poor socio-emotional wellbeing than those who had lived in a lone parent family throughout. However after adjustment for poverty, differences in prevalences of borderline-abnormal socio-emotional behaviour between these two groups was reduced.

Strengths and limitations

A strength of this study is that it has examined a number ways in which family structure might influence child socio-emotional wellbeing longitudinally, using a structured life course approach (Mishra et al., 2009). Findings indicate that change (or stability) in family structure is associated with child socio-emotional wellbeing, and that duration of exposure to certain family types might also be important. However, there was a degree of overlap between the two measures of family structure and it may be that these commonalities drive both associations. The family trajectories measure largely consisted of groupings specified *a priori*; in other studies, the commonly occurring trajectories may differ, which could lead to results that differ from those reported here. We assessed life course models by comparing them to a saturated model which contained all possible permutations of family structure over time. Many of these permutations were rare (see Table S1), despite the large MCS sample. It is therefore possible that the predictive power of the saturated model has been underestimated.

Longitudinal measures were derived from family structure status at four time-points and will not fully capture the experiences of all families. For example, it was not possible to account for short-term changes to family structure that may have occurred between sweeps (e.g. periods of temporary separation), or periods of lone parenthood that were likely to have been experienced in *Natural couple to reconstituted* families. Due to small numbers, it was necessary to aggregate the less common family structure types in the trajectories measure. In doing so, we may have overlooked small but informative trajectory groups. Finally, children of cohabiting couples tend to have lower levels of wellbeing than those living with married parents (Goodman & Greaves, 2010; Panico et al., 2010). However, there is evidence to suggest that this is due to "differential selection into marriage compared with cohabitation" governed by socio-economic background (Goodman & Greaves, 2010), and it was not an aim of our study to examine differences between cohabiting and married couples.

Cumulative poverty was operationalised as a count of sweeps at which income poverty was recorded; however, periods covered by the surveys ranged from nine months to two years. Short-term fluctuations in poverty status may have occurred between sweeps and these would have been overlooked. We classified poverty as <60% median national income (because it is a comparable measure across sweeps and commonly used in other studies), but it does not capture all aspects of disadvantage (Graham & Power, 2004). Changes in poverty status between sweeps may reflect very small changes in income from just above and below the threshold, which might not be expected to impact on socio-emotional wellbeing. If so, the association between poverty and socio-emotional wellbeing may have been underestimated. It is possible that poverty occurring before age 9 months may have preceded changes in family structure during pregnancy or early infancy. Earlier measures of income poverty were not available, although we were able to adjust for maternal education and social class, which are relatively stable and therefore likely to reflect prior

socio-economic circumstances. Equally, it is possible that family structure preceded some covariates (such as maternal psychological distress), and that child behavioural problems preceded relationship breakdown. However it is hard to disentangle these causal pathways, since changes in family structure tend to be a process rather than an event (Ferri & Smith, 2003). Behavioural problems in a child may put pressure on parental relations, but may also be indicative of family unrest before the relationship breakdown itself occurs.

We assessed socio-emotional wellbeing using the SDQ, which was reported by the main respondent (usually the mother). Although is a validated and reliable measure for monitoring socio-emotional wellbeing at the population level (Goodman, 1997, 2001), it is possible that parents may be more or less inclined to rate their children poorly depending on their family and economic background and own psychological state. Similarly, the borderline-abnormal cut-offs have been validated at a population level, but may also be subject to bias. When we repeated analyses using the continuous SDQ score the pattern of results remained, implying that our findings are not merely the consequence of the cut-offs used.

The MCS is a contemporary and nationally representative cohort, and therefore our findings are generalisable to the UK population. Attrition is a problem common to all longitudinal analyses. Response weights were used to account for attrition by the fourth survey (at age 7). However, of the 13,681 children included in MCS4, only 11,538 (75%) had data for all 4 sweeps. Children were more likely to have taken part in all four sweeps if at age 7 years they were living in a natural couple family (88%) compared to those living in reconstituted (77%) or lone parent families (76%). This was also more likely if they had 'normal' (85%) rather than borderline-abnormal SDQ scores (80%). Thus, our results may be subject to bias despite the use of response weights to account for attrition, possibly leading to associations being underestimated. The range of relevant information collected in the MCS allowed us to adjust for a number of early-life selection factors and confounders. However it remains possible that the observed associations between the family trajectories and socio-emotional wellbeing are due to residual confounding. We focussed on maternal early life characteristics in our analyses, because in the majority of cases where family breakdown occurred, children remained with the mother. However, we acknowledge the importance of fathers, and therefore repeated our analyses adjusting for a measure of family social class, based on the highest social class of either parent, and results were unaltered.

Finally, we employed a structured life course approach as a method for considering a number of different life course processes systematically. However, we acknowledge that other methods exist for investigating longitudinal associations in survey data, including structural equation modelling (SEM).

Conclusion

In this paper we found that children who did not live in families which were continuously headed by two natural parents were at greater risk of poorer socio-emotional wellbeing. A cumulative measure of family structure indicated an increase in prevalence of borderline-abnormal socio-emotional behaviour with the number of sweeps spent in a lone parent or reconstituted family. Findings using a family trajectories measure highlighted that the level of risk also varied depending on the type of trajectory experienced; for example, children who were living with a natural couple in infancy but were living with a lone parent by age seven years, had a lower risk than those who were living in a natural couple in infancy and a reconstituted family by age seven years. In many cases, early life characteristics and time spent in income poverty were contributing to the patterns observed, suggesting the importance of support for families early in their child's life and across generations. The UK Coalition Government continues to monitor and strive towards poverty targets established under the Child Poverty Act (Department for Work and Pensions & Department for Education, 2011). However, spending cuts, such as those to the childcare element of the Working Tax Credit and Local Housing Allowances, are likely to hit low income households with children and non-working lone parent families the hardest. Finally, children living in some family types remained at risk of poor socio-emotional wellbeing even after early life characteristics and poverty were taken into account. Future research should examine alternative mechanisms through which family structure may influence socio-emotional wellbeing.

Ethics

Research ethics approval was not required for this study as it was a secondary data analysis of the UK Millennium Cohort Study. The first sweep of the Millennium Cohort Study received ethics approval from the National Health Service Ethical Authority in February 2001 (MREC/01/6/19), and the second and third received approval from the London Multi-Centre Research Ethics Committee in September 2004 (MREC/03/2/022) and December 2005 (05/MRE02/46).

Acknowledgements

We would like to thank all the Millennium Cohort families for their participation, and the director of the Millennium Cohort Study and colleagues in the management team at the Centre for Longitudinal Studies, Institute of Education, University of London. We would like to thank the MCS user group within the Population, Policy and Practice Programme: Carol Dezateux, Lucy Griffiths, Tim Cole, Helen Bedford, Phillippa Cumberland, Jamie Fagg, Francesco Sera, Mario Cortina-Borja, Emeline Rougeaux, and Marco Geraci; and members of the Policy Research Unit in the Health of Children, Young People and Families: Terence Stephenson, Ruth Gilbert, Russell Viner, Miranda Wolpert, Amanda Edwards, Steve Morris, Helen Roberts, and Becky Fauth. Thanks also to the parents who are part of National Children's Bureau Families Research Advisory Group (FRAG) for commenting on preliminary findings, to Lidia Panico, Mark Petticrew and Margaret Whitehead for their advice early on in the project, and finally to the reviewers who commented on the paper.

Funding

This work was supported by the Policy Research Unit in the Health of Children, Young People and Families, which is funded by the Department of Health Policy Research Programme. This is an independent report commissioned and funded by the Department of Health. The views expressed are not necessarily those of the Department. The Population, Policy and Practice Programme was formed in 2014, incorporating the activities of the Centre for Paediatric Epidemiology and Biostatistics (CPEB). The CPEB was supported in part by the Medical Research Council in its capacity as the MRC Centre of Epidemiology for Child Health (award G0400546). Research at the UCL Institute of Child Health and Great Ormond Street Hospital for Children receives a proportion of the funding from the Department of Health's National Institute for Health Research Biomedical Research Centres funding scheme. The Millennium Cohort Study is funded by grants to former and current Directors of the study from the Economic and Social Research Council (Professor Health Joshi, Professor Lucinda Platt and Professor Emla Fitzsimons) and a consortium of government funders. The study sponsors played no part in the design, data analysis and interpretation of this study, the writing of the manuscript, or the decision to submit the paper for publication and the authors' work was independent of their funders.

References

Adams, N., Carr, J., Collins, J., Johnson, G., & Matejic, P. (2012). Households below average income: an analysis of the income distribution 1994/95 - 2010/11. London: Department for Work and Pensions.

Amato, P., & Keith, B. (1991). Parental divorce and the well-being of children: a meta-analysis. *Psychological Bulletin, 110* (1), 26-46. http://dx.doi.org/10.1037/0033-2909.110.1.26

Barros, A.J., & Hirakata, V. N. (2003). Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. *BMC Medical Research Methodology, 3*(21). http://dx.doi.org/10.1186/1471-2288-3-21

Birnie, K., Martin, R.M., Gallacher, J., Bayer, A., Gunnell, D., Ebrahim, S., Ben-Shlomo, Y. (2011). Socio-economic disadvantage from childhood to adulthood locomotor function in old age: a lifecourse analysis of the Boyd Orr and Caerphilly prospective studies. *Journal of Epidemiology and Community Health, 65*, 1014-1023. http://dx.doi.org/10.1136/jech.2009.103648

Bradshaw, J. (2011a). Child poverty and deprivation. In J. Bradshaw (Ed.), *The well-being of children in the UK* (Third edition, pp. 27-52). Bristol: The Policy Press. http://dx.doi.org/10.1332/policypress/9781847428370.003.0003

Bradshaw, J. (2011). Demography of childhood. In J. Bradshaw (Ed.),*The well-being of children in the UK* (Third edition, p13-25.). Bristol: The Policy Press.

Bradshaw, J., & Holmes, J. (2010). Child poverty in the first five years of life. In K. Hansen, Joshi, H. &, Dex,S. (Ed.), *Children of the 21st century: the first five years* (p13-31). Bristol: The Policy Press.

Bradshaw, J., & Homes, J. (2008). Family poverty assessed at three years old. *Working Paper* (Vol. 2008/7). London: Centre for Longitudinal Studies, Institute of Education.

Conger, R. D., Conger, K., & Martin, M. (2010). Socioeconomic status, family processes,and individual development. *Journal of Marriage and the Family, 72*(3), 685–704. http://dx.doi.org/10.1111/j.1741-3737.2010.00725.x

Connelly, R., & Platt, L. (2014). Cohort profile: UK Millennium Cohort Study (MCS). *International Journal of Epidemiology.* http://dx.doi.org/ 10.1093/ije/dyu001

Cooper, R., Mishra, D.G., & Kuh,D. (2011). Physical activity across adulthood and physical performance in midlife: findings from a British birth cohort. *American Journal of Preventive Medicine. 41*(4), 376-384. http://dx.doi.org/10.1016/j.amepre.2011.06.035

Department for Work and Pensions, & Department for Education. (2011). A new approach to child poverty: tackling the causes of disadvantage and transforming families' lives. Norwich: The Stationery Office.

Dunn, J., Deater-Deckard, K., Pickering, K., O'Connor, T., & Golding, J. (1998). Children's adjustment and prosocial behaviour in step-, single-parent, and non-step-family settings: findings from a community study. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 39*(8), 1083-1095. http://dx.doi.org/10.1111/1469-7610.00413

Evans, T., & Kohli, H., S. (1997). Socio-economic status and the prevention of child home injuries: a survey of parents of preschool children. *Injury Prevention, 3*, 29-34. http://dx.doi.org/10.1136/ip.3.1.29

Ferri, E., & Smith, K. (2003). Partnerships and parenthood. In E. Ferri, J. Bynner & M. Wadsworth (Eds.), *Changing Britain, changing lives: three generations at the turn of the century* (pp. 105-132). London: Bedford Way Papers.

Fomby, P., & Cherlin, A. (2007). Family instability and child well-being. *American Sociological Review, 72*(2), 181-204. http://dx.doi.org/10.1177/000312240707200203

Giles, L., Davies, M., Whitrow, M., Rumbold, A., Lynch, J., Swayer, M., & Moore,V. (2011). Structured regression analysis of life course processes: an example exploring how maternal depression in early childhood affects children's subsequent internalizing behavoiur. *Annals of Epidemiology, 21*(9), 654-659. http://dx.doi.org/10.1016/j.annepidem.2011.04.002

Goodman, A., & Greaves, E. (2010). Cohabitation, marriage and child outcomes *IFS commentary* (Vol. C114). London: Institute for Fiscal Studies.

Goodman, A., Rowe, R., & Gan, Y. (2010). SDQ: information for researchers and professionals about the Strengths and Difficulties questionaires 2012, from http://www.sdqinfo.com/c3.html. Last access October 2014.

Goodman, R. (1997). The Strengths and Difficulties questionnaire: a research note. *Journal of Child Psychology and Psychiatry, 38*(5), 581-586. http://dx.doi.org/10.1111/j.1469-7610.1997.tb01545.x

Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*(11), 1337-1345. http://dx.doi.org/10.1097/00004583-200111000-00015

Graham, H., & Power, C. (2004). Childhood disadvantage and adult health: a lifecourse framework. London: Health Development Agency.

Gustafsson, P., Persson, M., & Hammarstrom, A. (2011). Socio-economic disadvantage and body mass over the life course in women and men: results from the Northern Swedish Cohort. *European Journal of Public Health, 22*(3), 322-327. http://dx.doi.org/10.1093/eurpub/ckr061

Hallqvist, J., Lynch, J., Bartley, M., Lang, T., & Blane, D. (2004). Can we disentangle life course processes of accumulation, critical period and social mobility? An analysis of disadvantaged socio-economic positions and myocardial infarction in the Stockholm Heart Epidemiology Program. *Social Science and Medicine, 58 (8)*, 1555-1562. http://dx.doi.org/10.1016/S0277-9536(03)00344-7

Hansen, K. (2014). Millennium Cohort Study: a guide to the datasets (eighth edition). First, second, third, fourth and fifth surveys. London: Centre for Longitudinal Studies.

Hardy, R., Mishra, G., & Kuh, D. (2008). Body mass index trajectories and age at menopause in a British birth cohort. *Maturitas, 59 (4)*, 304-314. . http://dx.doi.org/10.1016/j.maturitas.2008.02.009

Kiernan, K., & Mensah, F. (2009). Poverty, maternal depression, family status and children's cognitive and behavioural development in early childhood: a longitudinal study. *Journal of Social Policy, 38 (4)*, 565-588. http://dx.doi.org/10.1017/S0047279409003250

Kiernan, K., & Mensah, F. (2010). *Partnership trajectories, parent and child well-being.* Bristol: The Policy Press.

Lee, J., Chang, C., H., & Chia, K., S. (2009). A practical guide for multivariate analysis of dichotomous outcomes. *Annals of the Academy of Medicine, Singapore, 38(80)*, 714-719.

Lin, L., Leung, G. M., Hui, L.L., Lam,T.H., & Schooling, C. M. (2011). Is informal child care associated with childhood obesity? Evidence from Hong Kong's "Children of 1997" birth cohort. *International Journal of Epidemiology, 40(5)*, 1238-1246. http://dx.doi.org/10.1093/ije/dyr086

Magnuson, K., & Berger, L. (2009). Family structure, states and transitions: associations with children's wellbeing during middle childhood. *Journal of Marriage and the Family, 71*(3), 575-591. http://dx.doi.org/10.1111/j.1741-3737.2009.00620.x

Marmot, M. (2012). *The health effects of poverty*. London: Child Poverty Action Group.

McLeod, J., Shanahan ,M. J. (1996). Trajectories of poverty and children's mental health. *Journal of Health and Social Behaviour, 37*(September), 207-220. http://dx.doi.org/10.2307/2137292

McMunn, A., Nazroo, J., Marmot, M., Boreham, R., & Goodman, R. (2001). Children's emotional and behavioural well-being and the family environment: findings from the Health Survey for England. *Social Science & Medicine, 53(4)*, 423-440. http://dx.doi.org/10.2307/2137292

Menard, S. (1995). *Applied Logistic Regression Analysis*. Thousand Oaks, CA: Sage.

Mishra, G., Nitsch, D., Black, S., De Stavola, B., Kuh, D., & Hardy, R. (2009). A structured approach to modelling the effects of binary exposures across the life course. *International Journal of Epidemiology, 3(2)8*, 528-537. http://dx.doi.org/10.1093/ije/dyn229

Murray, E., Mishra, G,. Kuh, D., Guralnik, J., Black, S., & Hardy, R. (2011). Life course models of socioeconomic position and cardiovascular risk factors: 1946 birth cohort. *Annals of Epidemiology, 21(8)*, 589-597. http://dx.doi.org/10.1016/j.annepidem.2011.04.005

Najman, J., Clavrino, A., McGee, T.R., Bor, W., Williams, G.M., & Hayatbakhsh, M.R. (2010). Timing and chronicity of family poverty and development of unhealthy behaviours in children: a longitudinal study. *Journal of Adolescent Health, 46(6)*, 538-544. http://dx.doi.org/10.1016/j.jadohealth.2009.12.001

Panico, L., Bartley, M., Kelly, Y., McMunn, A., & Sacker, A. (2010). Changes in family structure in early childhood in the Millennium Cohort Study. *Population Trends, 142,* 75-89. http://dx.doi.org/10.1057/pt.2010.32

Pearce, A., Lewis, H., & Law, C. (2013). The role of poverty in explaining health variations in 7-year-old children from different family structures: findings from the UK Millennium Cohort Study. [Research Support, Non-U.S. Gov't]. *Journal of Epidemiology andCommunity Health, 67*(2), 181-189. http://dx.doi.org/10.1136/jech-2012-200970

Plewis, I. (2004). Millennium Cohort Study: technical report on sampling. 3rd ed. (Vol. 3). London: Centre for Longitudinal Studies.

Ridge, T. (2011). The everyday costs of poverty in childhood: a review of qualitative research exploring the lives and experiences of low-income children in the UK. *Children and Society, 25(1)*, 73-84. http://dx.doi.org/10.1111/j.1099-0860.2010.00345.x

Rodgers, B., & Pryor, J. (1998). Divorce and separation: The outcomes for children. York: Joseph Rowntree Foundation.

Spencer, N. (2005). Does material disadvantage explain the increased risk of adverse health, educational, and behavioural outcomes among children in lone parent households in Britain? A cross sectional study. *Journal of Epidemiology and Community Health, 59*(2), 152-157. http://dx.doi.org/10.1136/jech.2004.020248

Tomlinson, M., & Walker, R. (2012). *Things get better (and sometimes worse): poverty dynamics and children*. London: Child Poverty Action Group.

Wadsworth, J., Burnell, I., Taylor, B., & Butler, N. (1985). The influence of family type on children's behaviour and development at age five years. *Journal of Child Psychology and Psychiatry, 26*(2), 245-254. http://dx.doi.org/10.1111/j.1469-7610.1985.tb02263.x

West, D., Leung, G.M., Jiang, C.Q., Elwell-Sutton, T.M., Zhang, W.S., Lam, T.H., Cheng, K.K., & Schooling, C.M. (2012). Life-course origins of social inequalities in adult immune cell markers of inflammation in a developing southern Chinese population: the Guangzhou Biobank Cohort Study. *BMC Public Health, 12*(269. http://dx.doi.org/10.1186/1471-2458-1112-1269

Wills, A., Hardy, R.J., Black, S., & Kuh, D.J. (2010). Trajectories of overweight and body mass index in adulthood and blood pressure at age 53: the 1946 British birth cohort. *Journal of Hypertension, 28*(4), 679-686. http://dx.doi.org/10.1097/HJH.0b013e328335de7b

Endnotes

[i] We opted to use the AIC rather than the Bayesian information criterion (BIC), because the AIC is less likely to favour models with fewer degrees of freedom .

[ii] To our knowledge there is no consensus over the size of p-value that should be used, and we adopted a conservative cut-off of 0.20, as used by Gustafsson et al (2011).

Realising health data linkage from a researcher's perspective: following up the 6-Day Sample of the Scottish Mental Survey 1947

Caroline E Brett Department of Psychology, University of Edinburgh
caroline.brett@ed.ac.uk
Ian J Deary Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of Edinburgh

(Received August 2013 Revised July 2014) http://dx.doi.org/10.14301/llcs.v5i3.266

Abstract

Health and wellbeing in old age are influenced by genetic, environmental and social factors throughout the life course. At present, few longitudinal studies offer information from childhood through to old age. Data linkage between multiple sources of health data enhances the value of existing longitudinal data. Regulations governing access to personal data for health research exist to protect the privacy and confidentiality of data on behalf of the individual. This paper outlines the process of obtaining permission for data linkage from a researchers' perspective, using a case study which offers an unusual opportunity to understand life course influences such as socio-economic status, childhood deprivation and measured intelligence on health and wellbeing in old age in an entire year-of-birth population. The Scottish Mental Survey 1947 (SMS1947, n = 70,805) has childhood intelligence data from individuals born in 1936 and attending schools in Scotland in June 1947. Representative sub-groups of the SMS1947 provided additional sociological information. The 6-Day Sample (n = 1,208), born on 6 days of 1936, were followed up for 16 years to age 27. Their younger siblings also took an intelligence test and were followed up for several years. Our team's planned research on the SMS1947 falls into two distinct parts. The first is a revival of the 6-Day Sample study involving tracing Sample members and inviting survivors to a follow-up study. The second part aims to carry out linkage between existing data on the SMS1947, its sub-groups, and the younger siblings, and morbidity and mortality data from central databases in Scotland and in England and Wales. We conclude by offering some recommendations for simplifying the process of obtaining permission to access linked health data, and place these into the context of the shifting landscape of data linkage in the UK and beyond.

Keywords: Scottish Mental Survey; intelligence; ageing; health; wellbeing; life course epidemiology; data linkage; health outcomes

Introduction

There is a growing acknowledgement amongst researchers in epidemiology and public health of the influence of biological, physical and social exposures across the life course on subsequent health and wellbeing (Blane, Netuveli & Stone, 2007; Foresight Mental Capacity and Wellbeing Project, 2008; Kuh, Ben-Shlomo, Lynch, Hallqvist & Power, 2003; Kuh, Cooper, Hardy, Richards & Ben-Shlomo, 2014; Lynch & Davey Smith, 2005; Richter & Blane, 2013). Life course research provides an inter-disciplinary framework for understanding how genetic as well as environmental factors such as socio-economic status, occupational hazards and childhood deprivation influence medical and social

inequalities in health and mortality. Life course epidemiology relies heavily on the use of longitudinal studies, which enable researchers to track the health and social trajectories of individuals over time.

Cohort studies are particularly valuable if they follow the same individuals from early life into adulthood (Pearson, 2011). They can reduce confounding by factors such as age or geographical location. The British birth cohorts, for example, offer a wealth of information from birth or early years through adulthood and into early old age, and have helped to address key questions on health inequalities, mental wellbeing, health behaviours, and genetic influences on health and ageing (Cooper et al., 2012; Kuh et al., 2012; Pearson, 2011; Power & Elliott, 2005; Wadsworth, Kuh, Richards & Hardy, 2006).

The oldest British birth cohort, born in 1946, reached retirement age in 2011 (Wadsworth et al., 2006). This restricts, for the moment, the usefulness of the British cohorts in answering questions about health in older age. A number of longitudinal cohort studies exist specifically to investigate ageing processes. A recent systematic review identified 67 longitudinal studies of older adults across all continents (Seematter-Bagnoud & Santos-Eggimann, 2006). Most participants were aged 50 or above at the time of recruitment, and although the length of follow-up varied from 2 to more than 30 years, the majority followed participants up for less than 10 years. Both factors limit the usefulness of these studies in investigating life course influences on ageing. In the UK, longitudinal studies of ageing include the Whitehall II study of retired civil servants (Marmot & Brunner, 2004), and the English Longitudinal Study of Ageing (Steptoe, Breeze, Banks & Nazroo, 2012). Both have produced many findings on health and wellbeing from the latter stages of working life into retirement and old age. However, they include minimal information from childhood and early adulthood, which are key periods of transition and exposure to social and environmental influences (Foresight Mental Capacity and Wellbeing Project, 2008).

Another valuable resource to life course epidemiology is population-based data collected by government or military organisations. These sometimes contain detailed medical, psychological and cognitive records on large numbers of individuals. For example, many epidemiologists have made use of Swedish conscription data, which is not only detailed but involves over 1 million young men followed for decades (Batty et al., 2007, 2009). The Vietnam Experience Study, although smaller in number, has been used to investigate determinants of health and mortality (Batty et al., 2008; Weiss, Gale, Batty & Deary, 2009). Udjus (1964) conducted innovative research on the relationships between physical growth and intelligence, education and family size using data from Norwegian conscripts (Tanner, 1966; Udjus 1964). However, military-based studies exclude women, and individuals who do not meet strict medical criteria. By contrast, data on cognitive function and educational attainment are available for much of the population through schools and examination boards, and can greatly contribute to longitudinal studies (e.g. Booth et al., in press; Calvin, Fernandes, Smith, Visscher & Deary, 2010; Deary, Strand, Smith & Fernandes, 2007).

Data linkage to enhance longitudinal research

The majority of longitudinal studies still rely on data provided by the participants themselves, either by self-report or the collection of objective health measures such as biomarkers. However, the value of these data is reduced by attrition and the inaccuracies of self-report data. Maintaining regular contact with longitudinal study participants is costly in terms of time, effort and resources. Researchers are, therefore, increasingly turning to alternative data sources. One such source is the computerised administrative health (and non-health) data which most developed countries now collect routinely. The primary purpose of these data is to enhance planning and policy decision-making by providing detailed information on the population's health and service use. This type of health data commonly comprises death certification, hospital admissions, hospital consultations and prescription records, with increasing coverage of consultation at general practitioners, disability assessment and sickness absence. Their value easily extends to medical research, providing extensive population-based information on the aetiology of disease, its course and outcome. In longitudinal cohort studies, linkage between research data and routinely-collected health records can be used to verify the quality and accuracy of self-report data such as medical history. However, administrative health data collected at point of service use is limited in scope, including

only individuals who consult health services. Furthermore, unlike some military or conscript data, health records generally contain scant information on health behaviours or function.

The research value of administrative health data can be further enhanced when data from different sources relating to the same individuals are combined. Data linkage of this kind is not a new concept (Dunn, 1946). The linkage of health data with non-health data such as social care, benefits, crime or census records, has a multitude of benefits, not only to medical research but also in terms of public health, cost efficiency and administrative accuracy (Holman et al., 2008, Scottish Government, 2010). The use of large, multiple, population-based administrative data sets conserves patient privacy by using a single identifier, thereby reducing the need for personal identifiers such as name and date of birth (Holman et al., 2008). Due to the organisational and collaborative effort required, very few such data sets exist internationally, and most by necessity tend to exist at a regional rather than country-wide level. Linked data sets currently exist in Manitoba, Canada; British Columbia, Canada; Rochester, Minnesota; Oxford, England; Scotland; Western Australia; and New South Wales, Australia (Roos, Menec & Currie, 2004; Scottish Government, 2010). One exception is Finland, which holds personal data on its population in large, linked data sets (Gissler & Haukka, 2004). In the UK, the Longitudinal Study – England and Wales (Goldring & Newman, 2010), the Northern Ireland Longitudinal Study (O'Reilly, Rosato, Catney, Johnston & Brolly, 2012) and the Scottish Longitudinal Study (Boyle et al., 2009) have successfully linked health and non-health data, including Census data, on a representative sample of these countries' populations. These linked data sets represent excellent examples of collaborative working and have enabled researchers and policy makers to address important research questions around health inequalities that would otherwise have been impossible to answer.

Rapid technological advances in recent years have made it possible to collect, store and analyse vast amounts of personal data and this has led to a worldwide debate around issues of privacy and confidentiality (Regidor, 2004). Data protection legislation was introduced in countries throughout the world, with the explicit aim of protecting the individual's right to privacy while recognising the need to collect and store personal data (e.g. European Union (EU) Data Protection Directive 95/46/EC, 1995; UK Data Protection Act, 1998). Health data collected at the point of use is subject to the common law duty of confidentiality. Patients use health services with the expectation that the information they are providing will be kept confidential, and registered clinicians are bound by the statutes and regulations of their professions. As a consequence, additional legislation governs the use of personal data relating specifically to health. In the UK, the government-commissioned Caldicott Report of 1997 highlighted six key principles and made 16 recommendations about the flow and handling of patient information within the NHS (The Caldicott Committee, 1997). Caldicott Guardians were appointed within each NHS organisation to uphold these principles. More recently, information governance frameworks incorporating legal rules, guidance and best practice were put in place to guard against the inappropriate use of patient information. In Finland, the Personal Data Act stipulates that health and social information can only be gathered by informed consent from the individual, with the exception of data collected for statistics and historical or scientific research (Gissler & Haukka, 2004).

The issue of consent

The majority of debate concerning the use of health data in research, centres around the issue of consent. Mason and Laurie (2010) argue that "it is undeniable that consent remains the primary policy device in legitimating medical research". Indeed, all medical and epidemiological research that requires the presence of individuals should be carried out in accordance with the ethical principles of research on human subjects laid out in the Declaration of Helsinki (1964) (Laurie & Sethi, 2011; Regidor, 2004). Research subjects must give voluntary consent, after having been informed of the purpose, methods and possible risks and benefits of the research study in question. Of course, in many epidemiological or longitudinal studies using health data, obtaining informed consent from each and every individual is not possible. This is particularly the case in studies with large numbers of subjects, where obtaining consent would be too costly and time-consuming, or where subjects have died, or in studies using historical data (Laurie & Sethi, 2011; Regidor, 2004). The requirement to obtain informed consent can also introduce bias, as those who do

not provide consent often systematically differ from those who do (Al-Shahi Salman et al., 2014; Kho, Duffett, Willison, Cook & Brouwers, 2009). Further issues arise in longitudinal research beginning in childhood, where consent needs to be updated once respondents reach an appropriate age. The EU Directive (1995) currently makes provision for circumstances in which obtaining informed consent may not be possible. Article 81 states that although consent must ideally be obtained for all research using personal data, exceptions could be made for research serving "exceptionally high public interests" and that "cannot possibly be carried out otherwise" (Directive 95/46/EC, 1995, Mansell, 2013).

The ethical and moral implications of the use of health data without consent have been discussed in detail by Regidor (2004). Regidor highlights the impact that differing interpretations of the EU Directive have on access to personal medical data in countries across the EU. Often, restrictions on access to data are placed by the data custodians themselves, with considerable variation within and between countries. Even where data sets are linked, researchers may be required to apply for separate approval from each data custodian, as occurs in the Western Australia Data Linkage System (Holman et al., 2008), or for further approvals where the data set is particularly large or contains particularly sensitive health information, as with the Finnish health and social welfare registers (Gissler & Haukka, 2004). In the United States, legislation varies by state, rendering it almost impossible to conduct population level health research and introducing the possibility that key demographic or social groups are under-represented in medical research (Melton, 1997; Regidor, 2004). Progress, some argue, is informed less by the benefits of such research and more by the perceived risks (Flowers & Ferguson, 2010; Laurie & Sethi, 2011; Regidor, 2004). Certainly, the administrative burden of complying with the requirements of research regulation and management, causes delay and wastage in medical research (Al-Shahi Salman et al., 2014).

Data linkage in the UK

The situation in the UK appears, at first, to be more promising. In recent years there has been an increasing acknowledgement within the government of the value of health data in life course research. The UK's current Prime Minister argued for the opening up of patient data for research (Cabinet Office and Prime Minister's Office, 2011; Walker, 2011). There is a recognised need for systems to enable smooth, safe and secure linkage between research studies and health and other records. Recent initiatives such as the Medical Research Council's funding of eHealth Centres, the Economic and Social Research Council's funding of four Administrative Data Research Centres across the UK, and infrastructure initiatives such as the ScottisH Informatics Programme (SHIP), the NIHR Research Capability Programme in England, and the Welsh Secure Anonymised Information Linkage System (SAIL), are all working towards the availability of electronic health records for research purposes within a secure and ethically-approved environment.

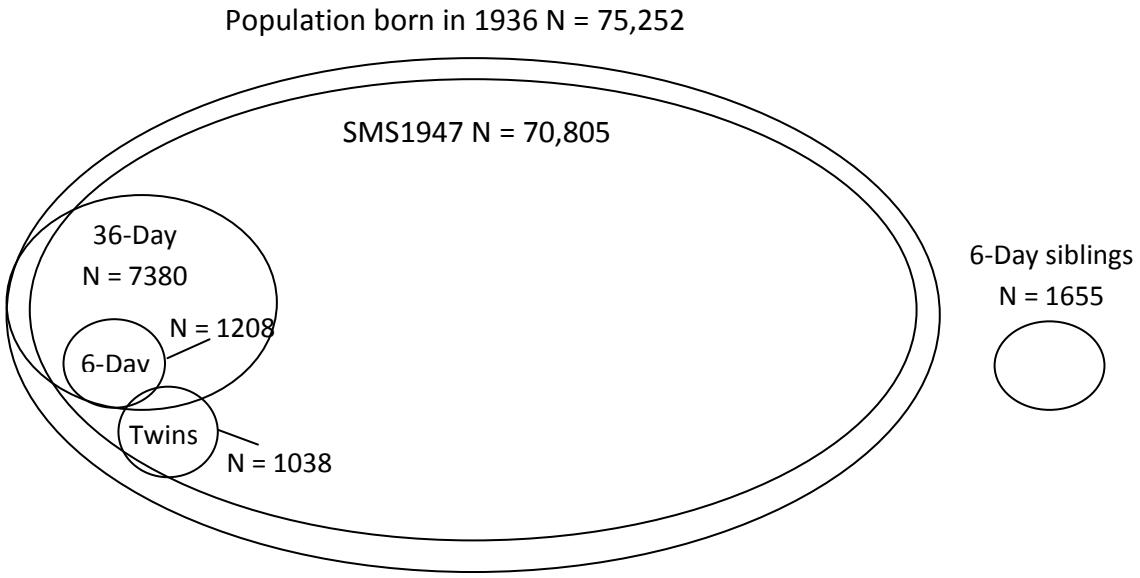
Clearly then, there is potential value in data linkage, there are databases and information structures in place to support it, and there is encouragement to do it from the scientific community and the government. However, the reality of gaining permission to carry out data linkage is often complex, with inconsistencies in the interpretation of legislation and the restrictions placed on access to different types of health data (Laurie & Sethi, 2011; Regidor, 2004). In this paper, we present a case study of the processes involved in our obtaining permission to link health data, held within the UK, with existing life course data from the Scottish Mental Survey 1947. We will then make recommendations based on our experience and discuss the shifting landscape of data linkage in the UK and what it might mean for researchers.

The Scottish Mental Survey 1947 and its proposed research uses

The Scottish Mental Survey 1947 (SMS1947) took place on 4th June 1947 (Deary, Whalley & Starr, 2009; Mental Survey Committee, 1949). It tested the intelligence of almost all children born in 1936 and attending Scottish schools. It applied a version of the Moray House Test No. 12 to 70,805 children. It was organised by the Scottish Council for Research in Education. Three subsamples were tested afterwards to collect more information. The 7,380 children born on the first three days of each month had a four-page 'Sociological Schedule' filled in; they were named the 36-Day Sample (Mental Survey Committee, 1953, 1958). The Sociological Schedule data were also collected on all 542 twin

pairs born in 1936. The 1,208 children born on the first day of the even-numbered months had, in addition to the Sociological Schedule, an individual Stanford-Binet IQ test and home visits and interviews almost every year until they were 27. They were named the 6-Day Sample (MacPherson, 1958; Maxwell, 1961, 1969). Their younger siblings had the Stanford-Binet IQ test applied as they reached age 11. Eventually, 1,554 of these siblings were tested. Figure 1 illustrates the relationships between the sub-groups of the SMS1947.

Figure 1. Venn diagram of the Scottish Mental Survey 1947 and its sub-groups



The 6-Day Sample offers a rich and detailed source of information on childhood and early adulthood in a representative sample of the Scottish population born in 1936. The existing data has already been used to answer research questions on relationships between education and cognition (Deary, Batty, Pattie & Gale, 2008; Calvin, Crang, Paterson & Deary, 2014; Paterson, Pattie & Deary, 2010, 2011).

No members of the SMS1947 had been contacted since youth, and their data had been largely unused for decades (Deary et al., 2009). The 6-Day Sample study has two ideas. First, it aims to re-contact surviving members of the 1,208-strong 6-Day Sample in their 70s, and to collect data on their lives since childhood and on their current health and cognitive capabilities. Second, it aims to link the childhood data from the whole Scottish Mental Survey 1947 (70,805-strong) anonymously to health records in the UK. We next go into more detail on each of these broad aims.

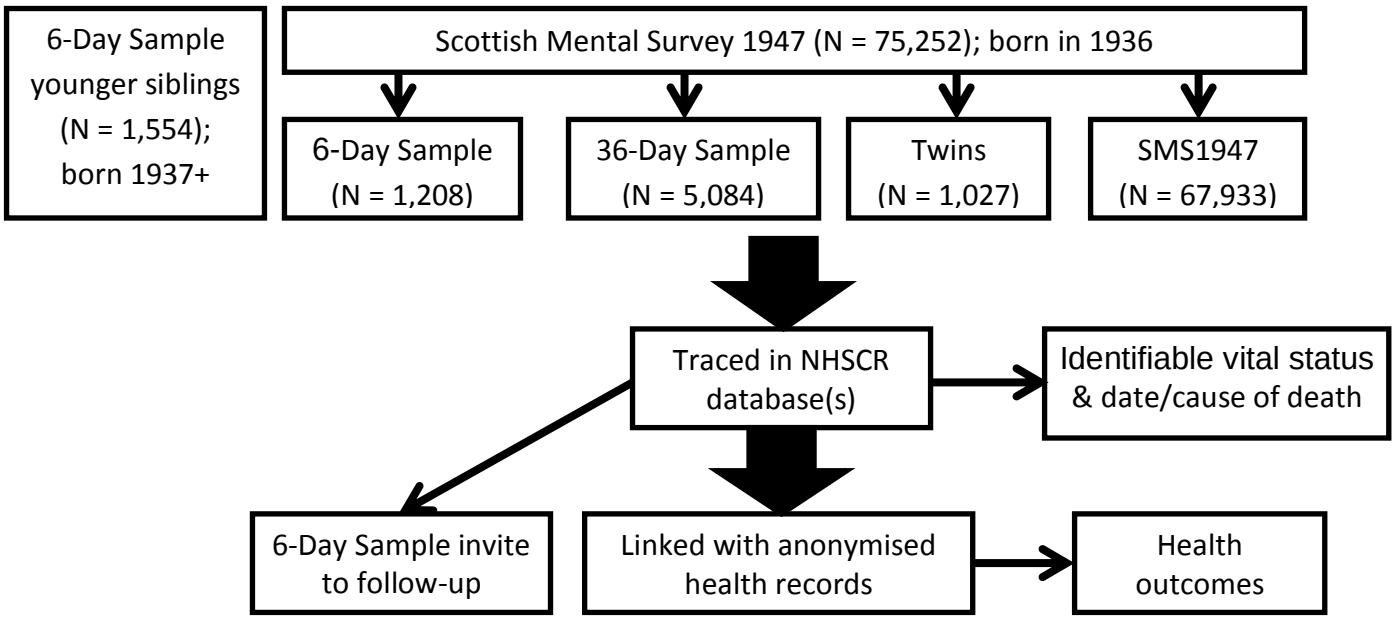
The first aspect of the study meant that the 1,208 members of the 6-Day Sample had to be traced through the NHS Central Register in Scotland, and in England and Wales. We considered that we could not contact people who had emigrated permanently from the UK. Then, identified, surviving members of the 6-Day Sample would be invited, through the Registrar General for Scotland and their equivalent in England and Wales, to a follow-up study, carried out by post and telephone. This would involve answering questions about their lives, health and wellbeing, carrying out some basic physical tests at home, and completing some simple cognitive tests over the telephone. Sub-groups of the follow-up study participants would be invited to three sub-studies: an in-depth life course interview, a validation clinic visit, and a urinary cortisol study.

The second aspect of the study involved proposed linkage to UK health records: all members of the 6-Day Sample study would be located in the Scottish Morbidity Records and equivalent databases in England and Wales. Information within these databases would be linked anonymously (to the researchers) with the information gathered during childhood and early adulthood. Given the effort that this would require, it was decided to undertake this process not only for the 6-Day Sample but for the entire Scottish Mental Survey 1947 group (70,805 individuals, which increased to 75,252 when we included those who did not sit the intelligence test in 1947), and the 6-Day Sample's younger siblings.

Childhood intelligence data were available for the majority of these individuals and the study plan represents what could be the most comprehensive epidemiological study of the associations between childhood intelligence and subsequent health outcomes ever conducted.

Figure 2 illustrates the study aims in diagrammatic form.

Figure 2. Diagram of the 6-Day Sample study aims



The permissions process: summary

A summary of the organisations we applied to is provided in Table 1 and a narrative of the process is in Appendix 1 (available in the Supplementary File). In total, 7 separate regulatory bodies or organisations were applied to for permission to receive data from four different sources: the NHS Central Register (NHSCR), Scottish Morbidity Records (SMR), Hospital Episode Statistics (HES) and the Office for National Statistics (ONS). Seven amendments or re-submissions were made as a result of changes to the study protocol or other supporting documents. The name and function of two of the bodies in England and Wales changed while our application was being processed. In total, 210 documents were sent in support of applications, amendments or re-submissions (not all of which are detailed in Table 1). From beginning to end, the process took 538 days between January 2012 and July 2013. Figure 3 shows a timeline of the permissions process.

The permissions process: ethical issues

The main ethical issues in our study were capacity and consent.

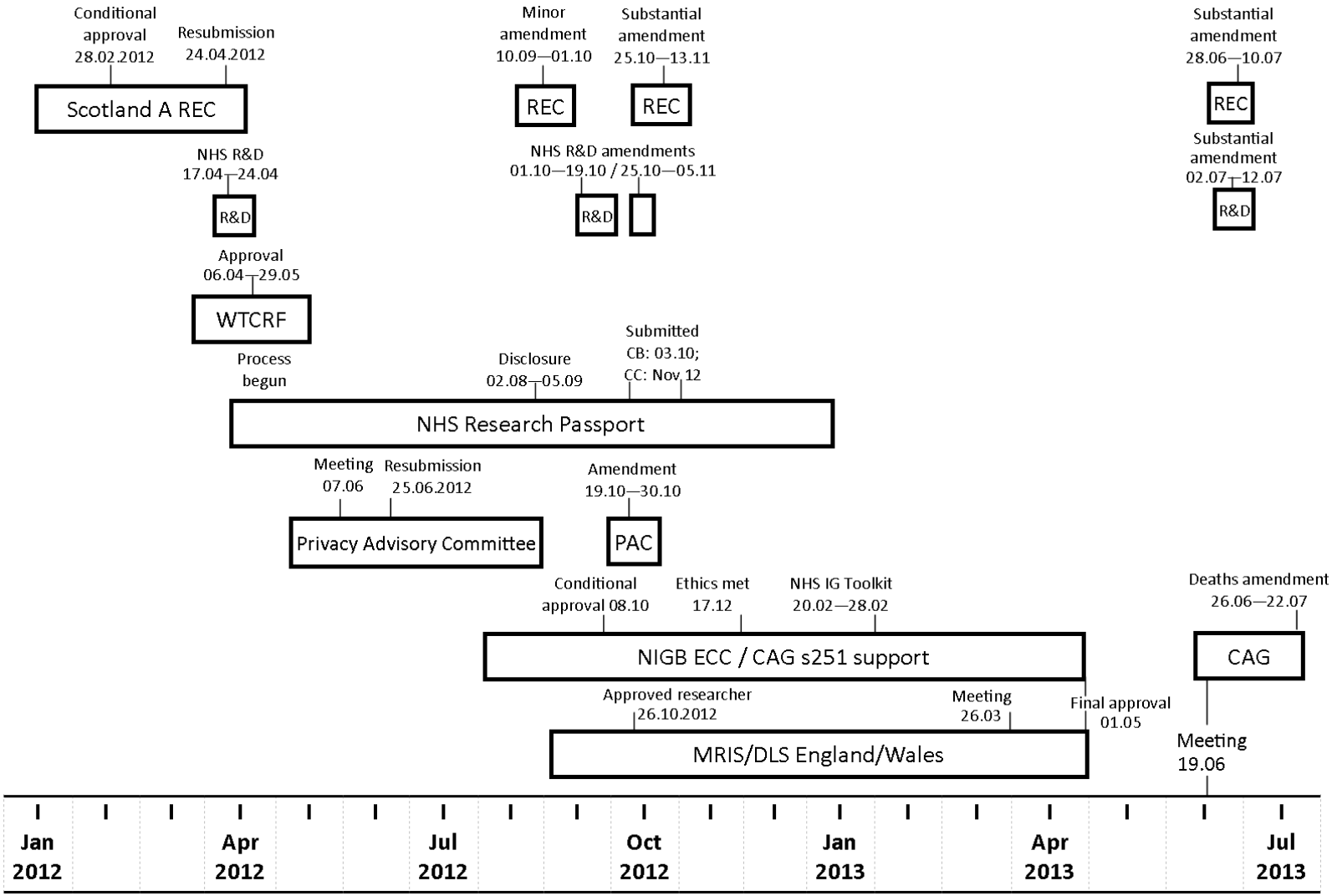
First, there was the possibility that, due to their age, some participants might be suffering from dementia or other disease that might impair their ability to give informed consent for the follow-up study. In order to maintain the representativeness of the 6-Day Sample, we wished to include these people in the study where possible, and we received ethical approval from a national Research Ethics Committee (REC) in Scotland to do so. However, participants in England/Wales were subject to different legislation and, rather than delay recruitment further by applying to an English REC, we elected to exclude the (very small) number of participants lacking mental capacity in England/Wales.

Second, the study relied on our using personal data without explicit consent from participants. For the linkage aspects, this was necessary in order to maintain the representativeness of the sample – not only would it be practically impossible to obtain consent from relatives of participants who had died, or from participants who had emigrated, but obtaining consent might have introduced systematic bias into our sample (Al-Shahi Salman et al., 2014). For the follow-up study, personal data held within NHSCR databases was required to send out the invitations – although we ourselves did not receive this information directly.

Table 1. Summary of the permissions process

Organisation	Permission sought	Supporting documents	Time to initial approval	Amendments	Number of documents for amendments	Time to final approval
NHS Research Ethics Committee Scotland A	General ethical approval	29	86 days	2	13	268 days
NHS Lothian R&D	General study approval	27	7 days (NB: REC approval first)	2	12	202 days
Wellcome Trust Clinical Research Facility	Follow-up study approval	13	53	0	N/A	53 days
NHS Research Passport	Permission to speak to participants	8	275 days	0	N/A	275 days
Privacy Advisory Committee	Data custodian for Scottish Morbidity Records & NHSCR data	30	113 days	1	2	169 days
Ethics and Confidentiality Committee of the NIGB (now the Confidentiality Advisory Group of the HRA)	Support under section 251 of the Health and Social Care Act (2006) to use health data without consent	49 (including NHS IG Toolkit)	203 days	1	0	339 days
Data Linkage Service, Health and Social Care Information Centre	Data custodian for Hospital Episode Statistics & ONS deaths and cancer data	26	230 days	1	11	230 days
Summary	--	**182**	--	**7**	**38**	**538 days**

Figure 3. Timeline of the permissions process for the 6-Day Sample study



Acronyms used: REC = Research Ethics Committee; R&D = NHS Lothian Research & Development; WTCRF = Wellcome Trust Clinical Research Facility; PAC = Privacy Advisory Committee; NIGB = National Information Governance Board; ECC = Ethics and Confidentiality Committee; MRIS = Medical Research Information Service; DLS = Data Linkage Service; CAG = Confidentiality Advisory Group.

In Scotland, permission was sought from the Privacy Advisory Committee, which provides advice, on a non-statutory footing, to the owners of Scottish health data, the NHS Information Services Division (ISD) and National Records of Scotland. In England/Wales, permission was sought from the Ethics and Confidentiality Committee of the National Information Governance Board (this has since been renamed and repositioned as the Confidentiality Advisory Group of the Health Research Authority). This body has statutory powers, under section 251 of the Health and Social Care Act 2006, to give approval for the use of medical data without consent.

In their review of current practices for the use of personal information for medical research in the UK, Laurie and Sethi (2011) summarised it as a 'consent or anonymise, with authorisation' approach. The 6-Day Sample study incorporated all three approaches – consent for the follow-up study, in which individuals provided data directly to us, anonymisation for the data linkage aspects, and authorisation for the provision of identifiable death data, for which neither consent nor anonymisation were appropriate.

The permissions process: challenges

A main challenge to us as researchers was gaining understanding of the legislation and framework for data linkage. This was not helped by the complexity of the current landscape, and the often unclear and conflicting guidelines (Laurie & Sethi, 2011). Although we embarked on a steep and successful learning curve, navigating the complexities inevitably delayed the process. One inconsistency which caused delay was the differing interpretation of the Statistics and Registration Service Act 2007 between Scotland and England/Wales. This led to some confusion over whether we were to receive identifiable or anonymised cause of death data.

Another substantial challenge was information governance. As researchers, we have a duty to ensure that the data provided by research participants is taken care of. To this end, practices were already in place to protect the data, store it securely, and prevent unauthorised access or loss. However, the majority of these practices were transmitted and maintained by members of the research team and had not been written in the form required by the external organisations. We submitted our Ethics and Confidentiality Committee (ECC) application on Friday 17th August 2012, having spent several weeks producing a written 'System Level Security Plan' as required. On Monday 20th August 2012 the NHS Information Centre put into force a rule that stated that all researchers applying for secondary use of health data needed to complete the NHS Information Governance Toolkit. The full implications of this were not clear until many months later. The Toolkit is designed for completion by NHS Trusts in order to ensure that they are complying with legislation and the common law duty of confidentiality. We were advised to complete this at an institution level – i.e. on behalf of the whole of the University of Edinburgh. However, despite being part of a large institution with the associated benefits in terms of infrastructure and expertise, the existing University-wide data protection policies applied mostly to the handling of student data and were not sufficient for the handling of the particular research data of our project. In addition, the University has a diverse research community and practices and infrastructure varied between different research groups and departments. As a result, we were obliged to complete the Toolkit on behalf of just our research study. Although we only had to make minimal alterations to our existing practices, the Toolkit required evidence of all information governance practices. With some assistance from computing officers, we created 19 new documents in order to meet the Toolkit's requirements. And we passed.

The final challenge was perhaps the most intriguing. Although we weren't the first study to conduct data linkage on individuals in Scotland and in England/Wales, it transpired that we were the first study to request data from the three countries on the same individuals across time. Therefore, considerable effort was expended in fostering collaborative relationships between the organisations providing the data, and putting systems in place to ensure smooth linkage between the two external data sets and our own existing data.

Discussion

This case study—which is fully described in the narrative account in Appendix 1— outlines the lengthy process involved in obtaining permission to link life course data from an existing research study

with nationally-held health databases. Nevertheless, all permissions were eventually granted and the linkage and follow-up is now underway. This is not the first attempt in the UK to 'revive', many years later, a study first conducted in childhood. Both the Boyd-Orr cohort (Gunnell, Frankel, Nanchahal, Braddon & Davey Smith, 1996; Martin, Gunnell, Pemberton, Frankel & Davey Smith, 2005) and the Aberdeen Children of the 1950s study (Batty et al., 2005) accomplished this successfully. The complexities of tracing and contacting research participants decades later contrast with cross-sectional studies taking a snapshot of a population sample at a specific time-point, and those longitudinal cohort studies that have successfully maintained continuous contact with participants since baseline testing. As the second-oldest birth cohort in the UK with childhood data (the oldest being its predecessor, the Scottish Mental Survey 1932 (Deary et al., 2009; Scottish Council for Research in Education, 1933)), and one of only a handful of year-of-birth cohorts in the world, the Scottish Mental Survey 1947 offers a rare opportunity to link childhood intelligence (and, to a lesser extent, deprivation) with future health outcomes across the life course and into older age. The sub-groups of the SMS1947, in particular the 6-Day Sample, represent an opportunity to investigate the influences of psycho-social circumstances, social and environmental exposures, and occupational characteristics in early adulthood on subsequent health and social mobility.

We began this process with a novel yet simple idea: to trace an entire year-of-birth population and link information collected in childhood with subsequent health outcomes, and to invite a sub-group of this population to a follow-up study of health and wellbeing in old age. The process has been long with many challenges along the way. Protecting the privacy and confidentiality of personal data held within administrative databases is paramount. Most health data is routinely collected for the purposes of monitoring a population's health and evaluating health service use and utility. Although this is done without individuals' explicit consent, the common law duty of confidentiality means that individuals have the right to expect their personal data collected during the course of any contact with health services to be protected and their confidentiality maintained. Regulating access to health data is a necessary means of achieving this. Whereas no researcher would disagree with the need for regulation, it could be argued that, at times, different interpretations and implementations of legislation governing personal health data adds unnecessary complexity to the process of obtaining approval.

Recommendations

Reflecting on our experience, there are a number of recommendations we could make that would streamline the permissions process and reduce wastage.

A data linkage information portal for researchers

Each country or administrative area could create an information centre as a single point of entry for researchers wishing to use their data for medical research. This portal should detail all the existing data sets, their ownership and provenance, access restrictions and information on how researchers can obtain permission to access the data. There is precedent for such a centre; the Finnish Information Centre for Register Research (http://rekisteritutkimusen.wordpress.com/) fulfils this purpose in Finland and linked administrative data sets such as that in Western Australia (http://www.datalinkage-wa.org/) also use this kind of entry point. However, these are the exceptions rather than the rule.

A centralised system for applying for permissions

Many countries have recognised and responded to the need for a more streamlined process of ethical review (Al-Shahi Salman et al., 2014); however, for the most part, the complexities of regulation and legislation involved in data linkage have to date prevented this from being extended to include review of applications to access health data for research purposes. An important aspect of this system would be the review of supporting documents, which frequently change in response to feedback from regulatory bodies and need to be re-reviewed by ethics committees. This process, for our study at least, was time-consuming and prone to human error – for example, simply forgetting to double check the version number and date on the 29th supporting document led to a two week delay in obtaining NHS R&D approval. A more efficient system would involve uploading supporting documents and submitting them for review centrally.

More guidance and training for researchers

Many researchers, like ourselves, are unfamiliar with the legislation and processes surrounding the use of data for which we are not directly responsible, and would benefit from outside guidance. Ideally, this would involve a single point of contact throughout the process with someone with expertise in the area. Although most researchers are already taking steps to protect their own research data, basic information governance training – including an overview of relevant legislation – should be mandatory for all researchers embarking on this process. Experienced researchers themselves might be a rich source of guidance and training. Initiatives such as Cohort and Longitudinal Studies Enhancement Resources (CLOSER: www.closerprogramme.co.uk) provide a platform for collaboration between studies, enabling them to pool resources and expertise, provide training, facilitate data linkage, encourage harmonisation across studies and increase the impact of longitudinal studies on policy and practice. The case study presented here has already provided input to the Scottish Government's data linkage strategy and plans (Brett & Deary, 2013; Deary, 2013).

A single repository of linked datasets which researchers can access in a secure environment

As with the existing linked administrative datasets in Western Australia, Finland and other countries, linking health and non-health datasets using a single identifier reduces the need for personal identifiers, reduces the likelihood of linkage errors, and protects the privacy of individuals. Accessing data in a secure environment or 'safe haven' reduces the need for such strict information governance requirements within a researcher's own environment, and enables access to those who lack the infrastructure to provide a secure environment. However, secure environments need to meet researchers' needs. Occasionally, several researchers within a team work in parallel on the same dataset and need to be able to pass results or even data between themselves. The data provided to researchers needs to be in a format they can use for analysis to avoid unnecessary and costly hours spent manipulating the data within the safe haven. Often, researchers wish to use their own written software or code in order to manipulate data, which can be prohibited or difficult to implement within a secure environment. The Secured Unified Research Environment (SURE), a remote-access computing facility funded by the Australian and New South Wales Governments as part of the Population Health Research Network, provides a blueprint for a researcher-friendly secure environment (https://www.sure.org.au). Where a secure environment is not available or practical (for example, where researchers are geographically distant from a suitable safe haven), legally binding data use agreements provide an alternative. Unfortunately, these do not always ensure that the organisation holding the data retains control over its use, which can lead to difficulties such as that currently experienced by the HSCIC in England (House of Commons Health Committee, 2014b).

The management of public perception of the use of personal data for medical research

Public perception of the use of personal data, either with or without consent, is influenced by the media. Decisions on the use of personal data for medical research, especially without consent, rely on a delicate balance between the protection of privacy and confidentiality, and research carried out for the 'public good' (Gissler & Haukka, 2010; Laurie & Sethi, 2011; Regidor, 2004,). All too often, this is inadequately portrayed in the media and focus lands on the negative aspects of data used in this way. In the UK, the Prime Minister David Cameron's enthusiasm for opening up NHS data has taken the form of the care.data project, which aims to link all NHS data together and greatly enhance medical research and permit access to the private sector. However, some argue that its implementation to date has left a lot to be desired, with the media focussing on the commercial exploitation of health data (Goldacre, 2014a, 2014b). The result is that the project has been 'put on hold' (House of Commons Health Committee, 2014a, 2014b) and all access to HES data is currently under review (Whitfield, 2014). In contrast, the Western Australian Data Linkage System (WADLS) has, almost from its inception, encouraged public involvement in research based on linked data through active consultation, public surveys, consumer representation on management committees and the publication of regular bulletins in the public domain outlining the WADLS and its research findings (Brook, Rosman & Holman, 2008; Holman et al., 2008).

Data linkage and data sharing have important roles to play in life course epidemiology. Linked data have the power to identify trends in population health and understand the aetiology of disease, and to inform better decisions about care and improve quality and safety standards within health and social care. Data linkage can unlock the potential of routinely-collected

health and non-health data dating back up to 50+ years. The increasing use of data linkage necessitates important decisions about ownership, access and confidentiality. Stringent controls need to be in place to ensure that access to health records is granted according to public interest and confidentiality is maintained throughout. However, as evidenced by the present case study, this process is often complex due to inconsistencies in legislation and interpretation of legislation across countries and data custodians. The process of obtaining permission is lengthy and often difficult to navigate, which inevitably leads to delays and wastage (Al-Shahi Salman et al., 2014). In the 6-Day Sample study, wastage occurred where a consideration of the costs and effort required to obtain additional permissions (for example, to link to health records held in Wales and Northern Ireland) was considered disproportionate to the potential gains. Nevertheless, it is important for us to state that all individuals within the regulatory bodies were helpful and expressed interest in the study and its aims.

Aside from these legislative and administrative complexities, research using routinely collected health records is not without its problems. There are two main sources of bias arising from the use of health records. First, the quality of health records needs to be considered. Aside from the inherent bias arising from the inclusion of only individuals who consult health services, health records rely on accurate completion by clinicians and hospital coders. Studies have shown cause of death certification to be problematic in terms of omissions and errors (Middleton et al., 2011). Indeed, the process is currently undergoing reform in England and Wales (Office for National Statistics, 2012). Second, linkage errors occur, particularly where probabilistic matching techniques are employed (Harron, Wade, Gilbert, Muller-Pebody & Goldstein,et al., 2014). Separating linkage from analysis, as generally occurs in research using linkage to health records, means that researchers are unable to evaluate the impact such errors might have on their results (Harron, Wade, Muller-Pebody, Goldstein & Gilbert, 2012; Harron et al., 2014).

The landscape of data linkage is constantly changing. Concerns have been raised over the revision of the EU Data Protection Directive, which may result in even more stringent controls on the use of personal data without consent, rendering studies such as the 6-Day Sample less likely to take place (Mansell, 2013). The situation in the UK is promising. Politicians and decision-makers are recognising the added value of data linkage (Cabinet Office and Prime Minister's Office, 2011; Walker, 2011) and encouraging funding to be channelled in this direction. Four Farr Institutes of Health Informatics Research are in the process of being set up across the UK. In Scotland, the Government commissioned a Data Sharing and Linkage Service in collaboration with data custodians and researchers, and its proposed design – as well as the proposed implementation of the Farr Institutes – meets many of the recommendations detailed above (Data Sharing and Linkage Service, 2013).

Given the great encouragement to undertake data linkage, and the initiatives that are taking place throughout the world, we hope this case study will be valuable for those researchers who, like our team's investigators, mostly come from the social and medical sciences and who wish to conduct life course research on human health and wellbeing. We concede that the process might be more straightforward for experts in data linkage. We offer this report as a guide and warning to researchers in the field, and as a stimulus to the greater simplification and harmonisation of organisations, processes and legislation.

Acknowledgements

Caroline Brett and the 6-Day Sample study are supported by a grant from the cross-council Lifelong Health and Wellbeing Initiative (G1001401, Lifelong health and wellbeing of the 'Scotland in Miniature' cohort: the 6-Day Sample of the Scottish Mental Survey 1947). Ethical approval for the study was given by the Scotland A-Research Ethics Committee; ref: 12-SS-0024. The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1). Funding from the Biotechnology and Biological Sciences Research Council (BBSRC) and Medical Research Council (MRC) is gratefully acknowledged.

References

Al-Shahi Salman, R., Beller, E., Kagan, J., Hemminki, E., Phillips, R. S., Savulesco, J., MacLeod, M., Wisely, J., & Chalmers, I. (2014). Increasing value and reducing waste in biomedical research regulation and management. Lancet, 383, 176-185.http://dx.doi.org/10.1016/S0140-6736(13)62297-7

Batty, G. D., Morton, S. M. B., Campbell, D., Clark, H., Davey Smith, G., Hall, M., Macintyre, S., & Leon, D. A. (2005). The Aberdeen Children of the 1950s cohort study: background, methods and follow-up information on a new resource for the study of life course and intergenerational influences on health. Paediatric and Perinatal Epidemiology, 18, 221-239. http://dx.doi.org/10.1111/j.1365-3016.2004.00552.x

Batty, G. D., Wennerstad, K. M., Davey Smith, G., Gunnell, D., Deary, I. J., Tynelius, P., & Rasmussen, F. (2007). IQ in early adulthood and later cancer risk: Cohort study of one million Swedish men. Annals of Oncology, 18, 21-28. http://dx.doi.org/10.1093/annonc/mdl473

Batty, G. D., Shipley, M. J., Mortensen, L. H., Boyle, S. H., Barefoot, J., Grønbæk, M., Gale, C. R., & Deary, I. J. (2008). IQ in late adolescence/early adulthood, risk factors in middle age and later all-cause mortality in men: the Vietnam Experience Study. Journal of Epidemiology and Community Health, 62, 522-531. http://dx.doi.org/10.1136/jech.2007.064881

Batty, G. D., Wennerstad, K. M., Davey Smith, G., Gunnell, D., Deary, I. J., Tynelius, P., & Rasmussen, F. (2009). IQ in early adulthood and mortality by middle age. Epidemiology, 20, 100-109. http://dx.doi.org/10.1097/EDE.0b013e31818ba076

Blane, D., Netuveli, G., & Stone, J. (2007). The development of life course epidemiology. Review Epidemiology Sante Publique, 55, 31-38. http://dx.doi.org/10.1016/j.respe.2006.12.004

Booth, J. N., Tomporowski, P. D., Boyle, J. M. E., Ness, A. R., Joinson, C., Leary, S. D., & Reilly, J. J. (in press). Obesity impairs academic attainment in adolescence: findings from ALSPAC, a UK cohort. International Journal of Obesity.

Boyle, P., Feijten, P., Feng, F., Hattersley, L., Huang, Z., Nolan, J. & Raab, G. (2009). Cohort Profile: The Scottish Longitudinal Study (SLS), International Journal of Epidemiology 38, 385-392. http://dx.doi.org/10.1093/ije/dyn087

Brett, C. E., & Deary, I. J. (2013, October). Navigating record linkage in Scotland and England and Wales: reviving the 6-Day Sample study. Paper presented at CPD session, Public Health Directorate, Scottish Government. Edinburgh, UK.

Brook, E. L., Rosman, D. L. & Holman, C. D. J. (2008). Public good through data linkage: measuring research outputs from the Western Australia Data Linkage System. Australian and New Zealand Journal of Public Health, 32, 19-23. http://dx.doi.org/10.1111/j.1753-6405.2008.00160.x

Cabinet Office and Prime Minister's Office. (2011). PM speech on life sciences and opening up the NHS. Accessed on 18th July 2013 from https://www.gov.uk/government/speeches/pm-speech-on-life-sciences-and-opening-up-the-nhs.

Calvin, C. M., Crang, J. A., Paterson, L., & Deary, I. J. (2014). Childhood evacuation during World War II and subsequent cognitive ability: The Scottish Mental Survey 1947. Longitudinal and Life Course Studies, 5, 227-244. http://dx.doi.org/10.14301/llcs.v5i2.255

Calvin, C. M., Fernandes, C., Smith, P., Visscher, P. M., & Deary, I. J. (2010). Sex, intelligence and educational achievement in a national cohort of over 175,000 11-year-old schoolchildren in England. Intelligence, 38, 424-432. http://dx.doi.org/10.1016/j.intell.2010.04.005

Cooper, R., Frank, J., Leyland, A., Hardy, R., Lawlor, D. A., Wareham, N. J., Dezateux, C., & Inskip, H. (2012). Using cohort studies in lifecourse epidemiology. Public Health, 126, 190-192. http://dx.doi.org/10.1016/j.puhe.2011.12.002

Data Protection Act. (1998). Chapter 29.

Data Sharing and Linkage Service. (2013). Analysis of responses to the Technical Consultation on the design of the Data Sharing and Linkage Service. Edinburgh: Scottish Government.

Deary, I. J. (2013, October). Reusing historical data: the Scottish Mental Surveys of 1932 and 1947. Scottish Government Data Linkage Framework: Data Linkage Exemplars, Public Acceptability, Research Results and Pathfinder Projects (Keynote Lecture). Edinburgh, UK.

Deary, I. J., Batty, G. D., Pattie, A. & Gale, C. R. (2008). More intelligent, more dependable children live longer. Psychological Science, 19, 874-880. http://dx.doi.org/10.1111/j.1467-9280.2008.02171.x

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. Intelligence, 35, 451-456. http://dx.doi.org/10.1016/j.intell.2006.09.003

Deary, I. J., Whalley, L. J., & Starr, J. M. (2009). A lifetime of intelligence: follow-up studies of the Scottish Mental Surveys of 1932 and 1947. Washington, DC: American Psychological Association.

Declaration of Helsinki (1964). Recommendations guiding medical doctors in biomedical research involving human subjects. Adopted by the 18th World Medical Assembly, Helsinki, Finland. The New England Journal of Medicine, 271, 473.

Directive 95/46/EC. (1995). Director 95/46/EC of the European Parliament and of the Council of 24 processing of personal data and on the free movement of such data. Official Journal of the European Communities No L, 38, 31-55.
Dunn, H. L. (1946). Record linkage. American Journal of Public Health, 36, 1412-1416. http://dx.doi.org/10.2105/AJPH.36.12.1412
Flowers, J., & Ferguson, B. (2010). The future of health intelligence: Challenges and opportunities. Public Health, 124, 274-277. http://dx.doi.org/10.1016/j.puhe.2010.03.006
Foresight Mental Capacity and Wellbeing Project. (2008). Final project report. London: The Government Office for Science.
Gissler, M., & Haukka, J. (2004). Finnish health and social welfare registers in epidemiological research. Norsk Epidemiologi, 14, 113-120.
Goldacre, B. (2014a). The NHS plan to share our medical data can save lives – but must be done right. The Guardian, 21st February 2014.
Goldacre, B. (2014b). Care.data is in chaos. It breaks my heart. The Guardian, 28th February 2014.
Goldring, S., & Newman, J. (2010). The ONS Longitudinal Study – a prestigious past and a bright future. Population Trends, 139, 4-10. http://dx.doi.org/10.1057/pt.2010.2
Gunnell, D. J., Frankel, S., Nanchahal, K., Braddon, F. E. M., Davey Smith, G. (1996). Lifecourse exposure and later disease: a follow-up study based on a survey of family diet and health in pre-war Britain (1937-39). Public Health, 110, 85-94. http://dx.doi.org/10.1016/S0033-3506(96)80052-7
Harron, K., Wade, A., Muller-Pebody, B., Goldstein, H., & Gilbert, R. (2012). Opening the black box of record linkage. Journal of Epidemiology and Community Health, 66, 1198. http://dx.doi.org/10.1136/jech-2012-201376
Harron, K., Wade, A., Gilbert, R., Muller-Pebody, B., & Goldstein, H. (2014). Evaluating bias due to data linkage error in electronic healthcare records. BMC Medical Research Methodology, 14:36. http://dx.doi.org/10.1186/1471-2288-14-36
Holman, C. D. J., Bass, A. J., Rosman, D. L., Smith, M. B., Semmens, J. B., Glasson, E. J., Brook, E. L., Trutwein, B., Rouse, I. L., Watson, C. R., de Klerk, N. H., & Stanley, F. J. (2008). A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system. Australian Health Review, 32, 766-777. http://dx.doi.org/10.1071/AH080766
House of Commons Health Committee. (2014a). Oral evidence: Care.data database, HC 1105 Tuesday 25 February 2014.
House of Commons Health Committee. (2014b). Oral evidence: Handling of NHS Patient Data 1105 Tuesday 8 April 2014.
Kho, M. E., Duffett, M., Willison, D. J., Cook, D. J., & Brouwers, M. C. (2009). Written informed consent and selection bias in observational studies using medical records: Systematic review. British Medical Journal, 338, b866. http://dx.doi.org/10.1136/bmj.b866
Kuh, D., Ben-Shlomo, Y., Lynch, J., Hallqvist, J., & Power, C. (2003). Life course epidemiology. Journal of Epidemiology and Community Health, 57, 778-783. http://dx.doi.org/10.1136/jech.57.10.778
Kuh, D., Cooper, R., Richards, M., Gale, C., von Zglinicki, T., & Guralnik, J. (2012). A life course approach to healthy ageing: The HALCyon programme. Public Health, 126, 193-195. http://dx.doi.org/10.1016/j.puhe.2012.01.025
Kuh, D., Cooper, R., Hardy, R., Richards, M., & Ben-Shlomo, Y. (2014). A life course approach to healthy ageing. Oxford University Press.
Laurie, G., & Sethi, N. (2011). Information governance of use of health-related data in medical research in Scotland: Current practices and future scenarios. SHIP Core Programme 2 Working Paper No 1. Edinburgh: SHIP.
Lynch, J. & Davey Smith, G. (2005). A life course approach to chronic disease epidemiology. Annual Review of Public Health, 26, 1-35. http://dx.doi.org/10.1146/annurev.publhealth.26.021304.144505
MacPherson, J. S. (1958). Eleven-year-olds grow up: A seven-year follow-up of a representative sample from the 1947 Scottish Mental Survey. London: University of London Press.
Marmot, M., & Brunner, E. (2004). Cohort profile: The Whitehall II study. International Journal of Epidemiology, 34, 251-256. http://dx.doi.org/10.1093/ije/dyh372
Mansell, P. (2013). EU data-protection amendments could undermine health research. PharmaTimes online, October 24 2013. Accessed 7th March 2014 from http://www.pharmatimes.com/article/13-10-24/EU_data-protection_amendments_could_undermine_health_research.aspx.
Martin, R. M., Gunnell, D., Pemberton, J., Frankel, S. & Davey Smith, G. (2005). Cohort profile: The Boyd Orr cohort – an historical cohort study based on the 65 year follow-up of the Carnegie Survey of Diet and Health (1937-39). International Journal of Epidemiology, 34, 742-749. http://dx.doi.org/10.1093/ije/dyi124
Mason, J. K., & Laurie, G. T. (2010). Law and medical ethics, 8th edition. Oxford: Oxford University Press.
Maxwell, James. (1961). The level and trend of national intelligence: The contribution of the Scottish Mental Surveys. London: University of London Press.

Maxwell, James. (1969). Sixteen years on: A follow-up of the 1947 Scottish Survey. London: University of London Press.
Melton, L. S. (1997). The threat to medical-records research. The New England Journal of Medicine, 337, 1466-1470. http://dx.doi.org/10.1056/NEJM199711133372012
Mental Survey Committee. (1949). The trend of Scottish intelligence: A comparison of the 1947 & 1932 surveys of the intelligence of eleven-year-old pupils. London: University of London Press.
Mental Survey Committee. (1953). Social implications of the 1947 Scottish Mental Survey. London: University of London Press.
Mental Survey Committee. (1958). Educational and other aspects of the 1947 Scottish Mental Survey. London: University of London Press.
Middleton, D., Anderson, R., Billingsly, T., Bande Mangaliso Virgil, N., Wimberly, Y., & Lee, R. (2011). Death certification: issues and interventions. Open Journal of Preventive Medicine, 1, 167-170. http://dx.doi.org/10.4236/ojpm.2011.13022
Office for National Statistics. (2012). Death certification reform: A case study on the potential impact on mortality statistics, England and Wales. Retrieved from https://www.gov.uk/government/publications/death-certification-reforms-update-newsletter-issue-2
O'Reilly, D. Rosato, M., Catney, G., Johnston, F. and Brolly, M. (2012). Cohort description: The Northern Ireland Longitudinal Study (NILS), International Journal of Epidemiology, 41, 634-641. http://dx.doi.org/10.1093/ije/dyq271
Paterson, L., Pattie, A., & Deary, I. (2010). Post-school education and social class destinations in Scotland in the 1950s. Longitudinal and Life Course Studies, 1(4), 371-393.
Paterson, L., Pattie, A., & Deary, I. J. (2011). Social class, gender and secondary education in Scotland in the 1950s. Oxford Review of Education, 37(3), 383-401. http://dx.doi.org/10.1080/03054985.2011.564847
Pearson, H. (2011). Study of a lifetime. Nature, 471, 20-24. http://dx.doi.org/10.1038/471020a
Power, C., & Elliott, J. (2005). Cohort profile: 1958 British birth cohort (National Child Development Study). International Journal of Epidemiology, 35, 34-41. http://dx.doi.org/10.1093/ije/dyi183
Regidor, E. (2004). The use of personal data from medical records and biological materials: ethical perspectives and the basis for legal restrictions in health research. Social Science & Medicine, 59, 1975-1984. http://dx.doi.org/10.1016/j.socscimed.2004.02.032
Richter, M., & Blane, D. (2013). The life course: Challenges and opportunities for public health research. International Journal of Public Health, 58, 1-2. http://dx.doi.org/10.1007/s00038-012-0436-z
Roos, L. L., Menec, V., & Currie, R.J. (2004). Policy analysis in an information-rich environment. Social Science & Medicine, 58, 2231–41. http://dx.doi.org/10.1016/j.socscimed.2003.08.008
Seematter-Bagnoud, L., & Santos-Eggimann, B. (2006). Population-based cohorts of the 50s and over: A summary of worldwide previous and ongoing studies for research on health in ageing. European Journal of Ageing, 3, 41-59. http://dx.doi.org/10.1007/s10433-006-0022-4
Scottish Council for Research in Education. (1933). The intelligence of Scottish children: A national survey of an age-group. London: University of London Press.
Scottish Government. (2010). Linking social care, housing & health data. Data linkage literature review 2010. Edinburgh: Scottish Government.
Steptoe, A., Breeze, E., Banks, J., & Nazroo, J. (2012). Cohort profile: The English Longitudinal Study of Ageing. International Journal of Epidemiology, 42, 1640-1648. http://dx.doi.org/10.1093/ije/dys168
Tanner, J. M. (1966). Galtonian Eugenics and the study of growth: The relation of body size, intelligence test score, and social circumstances in children and adults. The Eugenics Review, 58, 122-135.
The Caldicott Committee (1997). Report on the review of patient-identifiable information. London: Department of Health.
Udjus, L. G. (1964). Anthropometric changes in Norwegian men in the twentieth century. Oslo: Universitetsforlaget.
Wadsworth, M., Kuh, D., Richards, M., & Hardy, R. (2006). Cohort profile: The 1946 National Birth Cohort (MRC National Survey of Health and Development). International Journal of Epidemiology, 35, 49-54. http://dx.doi.org/10.1093/ije/dyi201
Walker, D. (2011). Administrative data linkage: the key to improving public policy? Accessed 18th July 2013 from http://www.alliance4usefulevidence.org/administrative-data-linkage-the-key-to-improving-public-policy/.
Weiss, A., Gale, C. R., Batty, G. D., & Deary, I. J. (2009). Emotionally stable, intelligent men live longer: The Vietnam Experience Study Cohort. Psychosomatic Medicine, 71, 385-394. http://dx.doi.org/10.1097/PSY.0b013e318198de78
Whitfield, L. (2014). HSCIC reviews HES access. Ehealth Insider, 17th April 2014. Retrieved from: http://www.ehi.co.uk/news/ehi/9364/hscic-reviews-hes-access

How wrong were we? Dependent interviewing, self-reports and measurement error in occupational mobility in panel surveys

Francisco Perales, Institute for Social Science Research, The University of Queensland
f.perales@uq.edu.au

(Received February 2014 Revised September 2014) http://dx.doi.org/10.14301/llcs.v5i3.295

Abstract

Occupation is a central concept in sociology and economics, and individual change in occupation is of major importance to literatures on wage determination, human capital, careers and social mobility. The collection of occupational data in surveys, particularly panel surveys, is challenging due to measurement error, and observed rates of occupational mobility are argued to be overestimated. We use a methodological discontinuity in the collection of occupational data from independent interviewing (respondents are asked to describe their occupation each year) to dependent interviewing (respondents are shown their previous response and only asked to describe their occupation if this has changed) and information on self-reported occupational changes in two panel surveys to estimate the degree of error in occupational mobility in panel data. We also test whether observed patterns differ by the level of aggregation of occupational classifications and examine the external validity of different measures of occupational mobility through their predicted impacts on selected labour market outcomes. Results indicate that occupational mobility is dramatically lower under dependent than independent interviewing (particularly for highly disaggregated occupational classifications) and that there is an evident mismatch between respondents' self-reports of occupational switches and mobility measures inferred from changes in occupational codes. The impacts of occupational changes on earnings and job satisfaction are more consistent with theoretical predictions under dependent than independent interviewing and when occupational mobility is inferred from respondents' self-reports. These findings have important implications for survey design, question the validity of existing studies on occupational change and call for further research.

Keywords: Occupation, occupational mobility, measurement error, dependent interviewing, self-reports, wages, job satisfaction, panel data

Introduction

The study of occupation has a long tradition in Sociology and Economics, and occupation serves as a core concept in research on wage determination, human capital acquisition, careers, social stratification and social mobility, and other fields such as social and friendship networks or cultural consumption (Blau & Duncan, 1967; Chan & Goldthorpe, 2007; De Beyer & Knight, 1989; Kambourov & Manovskii, 2008; Perales, 2013; Polachek, 1981; Sorensen, 1974). Occupations, understood as identifiable lines of economic activity requiring specific skills, knowledge and duties (Lee, Carswell & Allen, 2000), are however not easy to delineate or capture empirically. This is evidenced by the large number of existing occupational classifications, the diversity of structural principles used in creating these and the evident heterogeneity in the number of hierarchical levels and occupational titles they encompass. Besides, in post-modern societies, occupations are fluid

entities which emerge, evolve, disappear and reappear with changes in the structure of work (Abbott, 1989; Sobek, 1996). As a consequence, classifications often become obsolete and are superseded by new versions or completely new instruments.

In social research there is a fundamental problem in how occupational data are to be collected in surveys. Typically, survey participants are prompted to describe the line of work that they do, and their *verbatim* answers are later coded by a person (the coder) or a computer program as a given occupational unit of a previously determined occupational classification. This process is, however, subject to measurement error from a number of sources.[i] These include incomplete descriptions reported by respondents, inaccurate records made by interviewers and, more acutely, incorrect allocation of such reports into occupational units by coders (Laurie & Moon, 2003; Lynn & Sala, 2006; Mellow & Sider, 1983; Moscarini & Thomsson, 2007; Watson & Summerfield, 2009). For the latter, a common method of quality control is to calculate the degree to which different coders agree on the proper classification of respondents' descriptions of their occupation, also known as the inter-coder reliability (Groves, 1989). Research suggests that agreement rates for occupational data between two separate coders are far from optimal, and as low as 50% in some social surveys (Elias, 1997; Campanelli, Thomson, Moon & Staples, 1997; Laurie & Moon, 2003).[ii]

A large share of the academic interest in occupations has an inherent longitudinal dimension.[iii] We are interested in what prompts people to switch the type of work they do (Evans, 1999; Harper, 1995), in how individuals move across successions of occupations over their life courses and develop a 'career' (Budoki & Dex, 2010; Jacobs, 1999) and on the impact occupational changes have on labour market outcomes such as earnings and job satisfaction (Kambourov & Manovskii, 2008; Longhi & Brynin, 2010; Wilson & Green, 1990). Obtaining robust answers to these questions requires the use of longitudinal data, with panel datasets being considered the 'gold standard'. Unfortunately, the complexities inherent in the measurement of occupation outlined before are exacerbated in the context of panel data. Typically, the method used to collect information on occupation in panel surveys is the same as that used in cross-sectional surveys: each year respondents are asked about the line of work they are engaged in, with their answers coded into a certain occupational unit. Since the process of occupational coding is error prone *each year*, this implies not only that there is a high probability of coding an individual into the wrong occupational unit in any given survey wave, but also that the probability of coding an individual into a wrong occupational unit in *at least* one of two subsequent years is even higher (Hill, 1994; Sullivan, 2009). For example, a person working continuously as a "Court and Legal Clerk" (detailed code 5992 of the current Australian occupational classification) might have been coded as such in one year and as a "Secretary" (code 5212) in the next, due to standard coding error. Furthermore, even at an acceptable rate of coding reliability of 80%, coder variance means that in a panel dataset, one would expect to observe spurious mobility episodes for a shocking 20% of cases in a sample of occupational 'stayers' who give the exact same occupational description in two consecutive survey waves (Campanelli et al., 1997; Lynn and Sala, 2006). As a result, when researchers use changes in occupational codes across panel survey waves as evidence of individuals changing occupations, the likelihood of identifying spurious change can be alarmingly high. Despite this being potentially catastrophic for longitudinal research using occupational data, there is surprisingly little literature and even less empirical evidence on the issue - with the exceptions of Isaoglu (2010), Lynn & Sala (2006) and Moscarini & Thomsson (2007). A reason for this is that there are usually no means to tell whether the observed occupational switches are spurious and a product of survey error, or genuine changes in the nature of the work that individuals perform over time. This has elicited a debate over whether the conspicuously large rates of occupational mobility reported in past research are indeed genuine (Isaoglu, 2010; Longhi & Brynin, 2010). Additionally, findings stemming from research studies which take a non-critical approach to the quality of occupational mobility data in panel studies might be biased.

In this paper, we add to the limited body of existing knowledge and attempt to shed light over these issues by undertaking separate analyses of two different panel surveys: the British Household Panel Survey (BHPS) and the Household, Income and Labour Dynamics in Australia (HILDA) Survey.

First, we estimate the degree of error in yearly rates of occupational mobility inferred from changes in occupational codes across adjacent survey waves using data from the BHPS. To do so, we exploit a change in the methodology used to collect information on occupation in this study from independent interviewing (respondents are asked to describe their occupation each year) to dependent interviewing (respondents are shown their previous response and only asked to describe their occupation if this has changed). Second, we capitalize on the availability of a self-reported measure of occupational mobility in the HILDA Survey to compare the prevalence of year-on-year occupational mobility as denoted by changes in occupational codes across adjacent survey waves and respondents' self-reports. Additionally, we test whether any observed patterns differ by the level of aggregation of occupational data and examine the external validity of different measures of occupational mobility through their predicted impacts on selected labour market outcomes.

Key findings indicate that occupational mobility rates inferred from occupational codes are implausibly high when occupation-related panel data are collected via independent interviewing, do not match respondents' self-reports and fall substantially with dependent interviewing. The performance of occupational mobility measures improves when dependent interviewing is used, and self-reported occupational mobility measures perform better than measures based on changes in occupational codes under independent interviewing. These results have important implications for the collection of longitudinal data on individuals' occupation of work, and cast doubts on the validity of existing evidence on occupational mobility from studies that take for granted the quality of occupational change data in panel surveys.

The remaining sections in this paper will sequentially address themes in occupational mobility research; useful features of occupation data in the BHPS (i.e. introduction of dependent interviewing) and the HILDA Survey (i.e. availability of a self-reported measure of occupational change); the nature and structure of occupational classifications in these datasets; new empirical evidence on measurement error in occupational mobility in panel studies; and a discussion of the associated findings.

Background

Occupational mobility in the academic literature

The concept of 'occupation' is broader than that of 'job'. Occupation denotes a line or type of work, without reference to a specific workplace, firm or employer. Hence, occupation and job changes cannot be equated. Employees can change jobs without changing occupations - for instance, a primary school teacher who moves from school A to school B - but the opposite is improbable. In sociology, research on occupational mobility is, to a large extent, embedded in the literatures on status attainment and occupational (or class) careers. Occupations drive the structure of economic and social welfare and movement across occupations is seen as a vehicle for upward social mobility (Blau & Duncan, 1967; Goldthorpe, 1987). Additionally, occupational change might be a channel for career adjustment as a response to poor early career decisions or evolving preferences (Longhi & Brynin, 2010). In economics, mobility has more often been studied at the level of jobs than of occupations. The discussion revolves around the notion of 'utility' using a rational choice theoretical framework (Borjas, 1981). In very simplified terms, changing job and/or occupation is a risky and potentially costly process. Workers only incur such changes if there is an inherent motivation, that is, if the expected utility from doing so exceeds the utility of staying in the current position, minus the costs associated with the change (Booth & Francesconi, 1999). Such costs include direct economic costs, opportunity costs and psychological costs (Groot & Veberne, 1997). In empirical research designs, evidence of labour market mobility being associated with increases in social standing and utility gains is gathered using proxies such as wage gains and growth in job satisfaction (see e.g. Longhi & Brynin, 2010; Parrado, Caner & Wolff, 2007; Wilson & Green, 1990).[iv]

Empirically, there are different ways through which occupational mobility can be identified in survey data. Most commonly, longitudinal data are used and a comparison of the codes attributed to respondents in different survey waves undertaken to distinguish mobility from stability. While early sociological interest in status attainment emphasized divergences between occupation at labour market entry and current or last occupation (Featherman, Lancaster Jones & Hauser, 1975), interest has progressively shifted to the analysis of

more complex successions of several adjacent occupational spells (Budoki & Dex, 2010; Jacobs, 1999). American and British evidence suggests that occupational mobility is pervasive. Parrado et al. (2007) and Kambourov and Manovskii (2008) report that in the US, between 10% and 20% of workers change occupations each *year*, with rates varying with the degree of aggregation used to measure occupation. Moscarini and Thomsson (2007) report even higher rates of up to 35% per *month* using highly disaggregated occupational data. Raw (uncorrected) estimates of 11-45% have been reported for Germany (Isaoglu, 2010; Longhi & Brynin, 2010) and of 20-53% for Britain (Longhi & Brynin, 2010; Lynn & Sala, 2006). For the reasons outlined before, the degree of error in occupational mobility reported in previous research is likely to be high, with existing research suggesting that implementation of dependent interviewing practices might reduce misclassification of occupational movers (Lynn & Sala, 2006; Moscarini & Thomsson, 2007) and proposing potential corrections when information on job changes is also available (Isaoglu, 2010; Longhi & Brynin, 2010).

Dependent interviewing and occupational mobility: the case of the British Household Panel Survey

The British Household Panel Survey (BHPS) is an annual household panel survey which began in 1991 and is representative of individuals living in Britain that year (Taylor, Brice, Buck & Prentice-Lane, 2010). This is one of the largest and best-established household panel surveys in the world and is part of the Cross National Equivalence File (CNEF). Up to wave 15 of the survey (2005) occupational data were collected by independent interviewing. Regardless of their responses in previous years, survey respondents were asked to answer the following question: *"What was your (main) job last week? Please tell me the exact job title and describe fully the sort of work you do"*. In its 16th wave (2006) the study experienced a change in the way in which data on employment characteristics and household finances, including data on occupation, were collected. The new methodology consisted of (proactive) dependent interviewing. In broad terms, the routing process involved respondents being shown their report of their occupation in their last interview and asked whether the same description applied to their current occupation. If respondents answered 'yes', the interview continued as normal. If respondents answered 'no', they were then prompted to answer the independent interviewing question reproduced above. More detailed information on the motivation, aims and considerations on the introduction of dependent interviewing on the BHPS as well as on the actual procedure can be found in Jäckle, Laurie and Uhrig (2007).

This seemingly small change in the way in which occupation-related information is collected should minimize the risk that respondents' answers are coded into the wrong occupational group and reduce the emergence of spurious occupational changes. This methodological discontinuity is used to explore to what degree previous estimates of occupational change in that study were artificial, and the product of coding error.

Self-reported occupational mobility: the case of the Household, Income and Labour Dynamics in Australia Survey

The Household, Income and Labour Dynamics in Australia (HILDA) Survey is also a major annual panel dataset (Watson & Wooden, 2012). This survey was designed using the BHPS as a model, thus resembling it in its structure, and is also part of the CNEF. The HILDA Survey tracks a representative sample of the Australian population since 2001. Unlike the BHPS, information on occupation is collected in the 'traditional' manner, i.e. via independent interviewing, throughout the entire life of the survey. The actual questionnaire item is placed after a battery of questions on the current main job and reads: *"What kind of work do you do in this job? That is, what is your occupation called and what are the main tasks and duties you undertake in this job? Please describe fully"*.

The HILDA Survey data are complementary to the BHPS data and enable us to gain further insights into occupational mobility. Specifically, the HILDA Survey includes a question which asks respondents: *"Has your occupation changed since [date of last interview]? Note that a promotion or a change in employer does not necessarily mean a change in occupation"*. This will allow us to estimate the degree of misclassification in occupational mobility when this is inferred from a comparison of respondents' occupational codes across survey releases.

Occupational data in the British Household Panel Survey and the Household, Income and Labour Dynamics in Australia Survey

The BHPS analysis covers the period 1991-2008.[v] The information on occupation of employment in the BHPS is coded into several classifications. In all survey years, occupational data are available using the 1990 Standard Occupational Classification (SOC90). SOC90 clusters jobs into occupations in terms of their content and the educational qualifications, training and work experience required to carry out the associated tasks and duties. It has a hierarchical structure divided into four levels, each denoting a different degree of occupational aggregation. It features 9 major occupational groups, 22 sub-major groups, 77 minor occupational groups and 371 unit groups. A revision of this classification, SOC2000, is available in the data since wave 10 (2001). SOC2000 is also arranged hierarchically, but the number of occupations at each aggregation level differs slightly from that in SOC90. SOC2000 features 9 major groups, 27 sub-major groups, 81 minor groups and 353 unit groups. The change in the survey practices used to collect occupational data in the BHPS from dependent to independent interviewing took in place in 2006. Thus, we can test the impact this had on occupational mobility inferred from wave-on-wave changes in occupational codes from both the SOC90 and SOC2000 classifications. The use of both schemes will enable us to gather more robust evidence that any effects found are genuine.

The analysis of the HILDA Survey covers the period 2001-2012. In the HILDA Survey occupational data are collected using the 2006 Australian and New Zealand Standard Classification of Occupations (ANZSCO06) in all survey waves. In ANZSCO06, occupations are organized on the basis of their similarities in terms of both skill level and skill specialization into four levels of occupational aggregation. These include 8 major groups, 43 sub-major groups, 97 minor groups and 358 unit groups. Note that information on respondents' detailed occupation (i.e. the 3- and 4-digit levels) in the HILDA Survey is only available in its 'unconfidentialised' version.

Throughout the paper we compare results at all available levels of aggregation for the occupational classifications used. This is important because, as the specificity of occupational units increases, the probability of measurement error should also increase (Elias, 1997; Laurie & Moon, 2003; Lynn & Sala, 2006; Sullivan, 2009).

Research hypotheses

Based on the above discussion, we formulate several simple, testable hypotheses:

Hypothesis 1 (BHPS) – The introduction of dependent interviewing will reduce the observed rates of occupational mobility in the British labour market captured in the BHPS data.

Hypothesis 2 (BHPS) – The fall in observed occupational mobility after the introduction of dependent interviewing in the BHPS data will be more apparent when using highly disaggregated versions of occupational classifications.

Hypothesis 3 (HILDA) – Rates of occupational change inferred from respondents' self-reports will be lower than those inferred from changes in occupational codes across survey waves.

Hypothesis 4 (BHPS) – The impacts of occupational changes on labour market outcomes (namely, wages and job satisfaction) will be more consistent with theoretical predictions after the introduction of dependent interviewing.

Hypothesis 5 (HILDA) – The impacts of occupational mobility on wages and job satisfaction will be more consistent with theory when occupational mobility is inferred from respondents' self-reports rather than from changes in occupational codes**.**

Results

Empirical evidence

We now present the results of our empirical enquiries and test the above hypotheses. First, we will present trends in occupational mobility rates in Britain and Australia under different interview conditions (dependent and independent interviewing) and using different specifications of occupational change (different occupational codes vs. self-reports). Second, we will examine for which measures the predicted impacts of occupational mobility on labour market outcomes most closely resemble the expectations from sociological and economic theory.

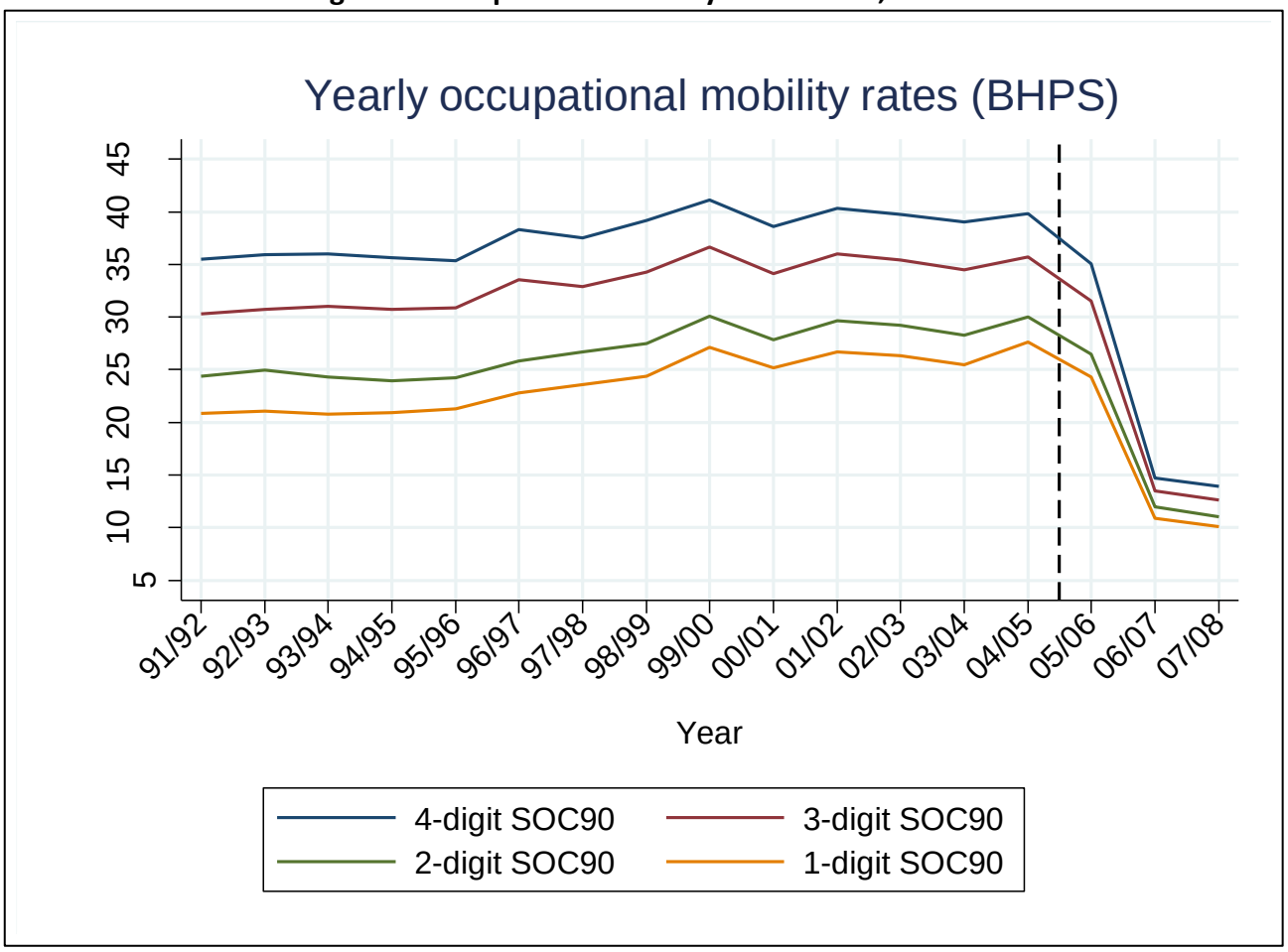
All subsequent analyses include only employees who were interviewed and in paid employment at times t and t-1, and exclude individuals who move from self-employment, unemployment or inactivity into employment (or *vice versa*), and those for which there is missing data in either of the adjacent survey waves.

Trends in occupational mobility in Britain and Australia

Figure 1 shows trends in occupational mobility in Britain between 1991 and 2008 using BHPS data. Each data point gives the percentage of workers who changed occupations *between a given survey wave and the previous*, inferred by such individuals having a different occupational code. Each coloured line shows the results for a different level of aggregation of the SOC90 occupational classification. The discontinuous vertical line denotes the time point when the methodology employed to collect occupation-related information in the BHPS changed from independent to dependent interviewing.

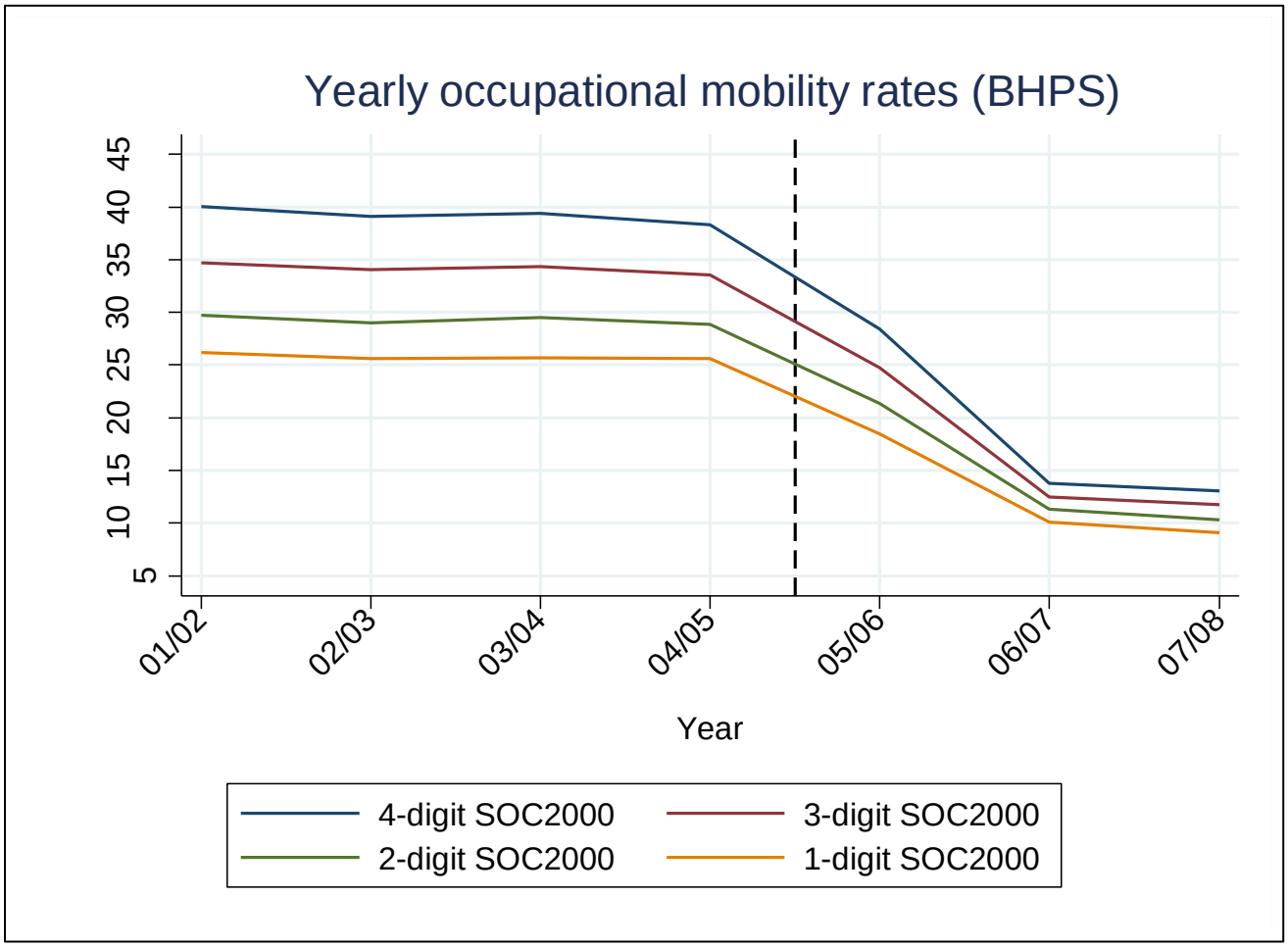
Many interesting findings emerge from this simple trend graph. First, in the period between 1992 and 2005, occupational change seems conspicuously high. For instance, between 1999 and 2000 over 40% of all workers appear to have changed detailed occupational units. Second, in the same time period there is evidence of a small increase (of around 5 percentage points) in mobility rates. Third, the more disaggregated the classification used, the higher the rates of occupational mobility observed. In the 1992-2005 period, about 35-40% of workers seem to change detailed occupational units each year, but 'only' 20-25% change major occupational groups. As lines of work become more disaggregated they become more similar to each other. Thus, one would expect mobility rates within major occupational groups and across detailed occupational units to be high. However, the extremely large size of the divergences suggests that measurement error exists. Fourth and most significantly, the introduction of dependent interviewing radically reduces the occupational mobility rates observed in the BHPS data. For example, rates at the 4-digit level of occupational aggregation, which were about 40% in the 2004-2005 period, fall to about 15% in the 2007-2008 period.

Figure 1. Occupational mobility in the BHPS, SOC90



It is possible that the break in the time series associated with the introduction of dependent interviewing reported above is confined to the occupational classification we used (SOC90). This poses the question: does the pattern also emerge for other classifications? Results in Figure 2 suggest that the trend is also apparent when using SOC2000. For those years in which the SOC2000 classification is available (2002-2008), the pattern is very similar to that in Figure 1: overall occupational mobility drops dramatically after the introduction of dependent interviewing.

Figure 2. Occupational mobility in the BHPS, SOC2000



Unexpectedly, the reduction in occupational mobility occurring after the introduction of dependent interviewing in Figures 1 and 2 seems to have taken place in an odd, step-wise fashion: rates reduce only slightly in 2006, drop more substantially in 2007, and stabilize in 2008. While this pattern is counter-intuitive, there are three additive and inter-related reasons why the introduction of dependent interviewing practices in wave 16 (2006) of the BHPS did not result in an immediate fall in occupational mobility rates. First, 1,152 (or 18%) of the in-scope respondents did not have valid occupational descriptions to be fed forward from wave 15 for dependent interviewing in wave 16 (despite having seemingly valid occupational codes in wave 15). The occupation information for these respondents in wave 16 was subsequently collected via independent interviewing (Jäckle et al. 2007), for which we have shown that longitudinal consistency is poor. This increased the observed rate of occupational mobility for wave 16. In contrast, in waves 17 (2007) and 18 (2008) just 115 (or 2%) and 62 (or 1%) of the in-scope respondents had no valid occupational descriptions to be fed forward, respectively. Second, for no apparent reason, a non-negligible number of wave 16 respondents were assigned different occupational codes, despite having confirmed that their occupation was the same as in wave 15 via dependent interviewing. For

example, this affected the detailed SOC2000 codes of 1,359 wave 16 respondents. The number of entries affected by this inconsistency reduced markedly in waves 17 (175 respondents) and 18 (156 respondents). Third, substantive 'cleaning' work was undertaken to streamline wave 15 verbatim occupational descriptions to be used during dependent interviewing in wave 16, and this might have prompted some respondents to erroneously infer and report an occupational change. As documented by Jäckle et al. (2007, p. 21): *"responses […] underwent an editing process to ensure that the descriptions were readable, to correct spelling errors and where necessary to shorten the description. For example, if at the last interview a respondent had reported their job title plus a lengthy job description of the tasks they actually did in their job, this was edited to include the job title and the key feature of their job"*. Reportedly, this was *"a non-trivial task"* that *"at subsequent waves should be significantly reduced, as only those responses where new information has been entered will need to be checked"* (Jäckle et al. 2007, p. 22). Therefore, the unexpected trend in occupational mobility rates following the introduction of dependent interviewing in the BHPS is simply the product of errors associated with the complex transition from independent interviewing**.**

Suspecting that spurious occupational changes are commonplace in panel data, a handful of previous studies have attempted to partially correct for this by considering job mobility information in conjunction with changes in occupational codes to identify occupational mobility (see e.g. Isaoglu, 2010; Longhi & Brynin, 2010). In these studies, changes in occupational codes across waves are only taken as evidence of occupational mobility if respondents incur also a *job* change between such waves. In the BHPS, job changes can be inferred from a separate question reading: *"What was the date you started working in your present position? If you have been promoted or changed grades, please give me the date of that change. Otherwise please give me the date when you started doing the job you are doing now for your present employer"*. Note however that answers to this question are also likely to be 'noisy', due to respondents being unsure about the exact meaning of terms such as 'position', 'grade' or 'employer', recall biases and other survey errors (Brown & Light, 1992).

In Figure 3 (SOC90) and Figure 4 (SOC2000) we present trends in occupational mobility in Britain using the definition of occupational change that incorporates information on job changes. Three aspects stand out. First, within the independent interviewing period, rates of occupational mobility more than halve relative to those in Figures 1 and 2. Second, the downward trend after the implementation of dependent interviewing in the BHPS, despite being less marked, is still highly visible. Third, within the dependent interviewing period, rates of occupational mobility are comparable to those in Figures 1 and 2. This suggests that under independent interviewing conditions, using job mobility information might partially correct for spurious occupational mobility due to measurement error. While it could be argued that a sharp fall in occupational mobility could also be due to rapid and marked exogenous labour market changes, the BHPS data was collected before the great economic recession affecting Britain from 2008.

Figure 3. Occupational mobility conditional on job change in the BHPS, SOC90

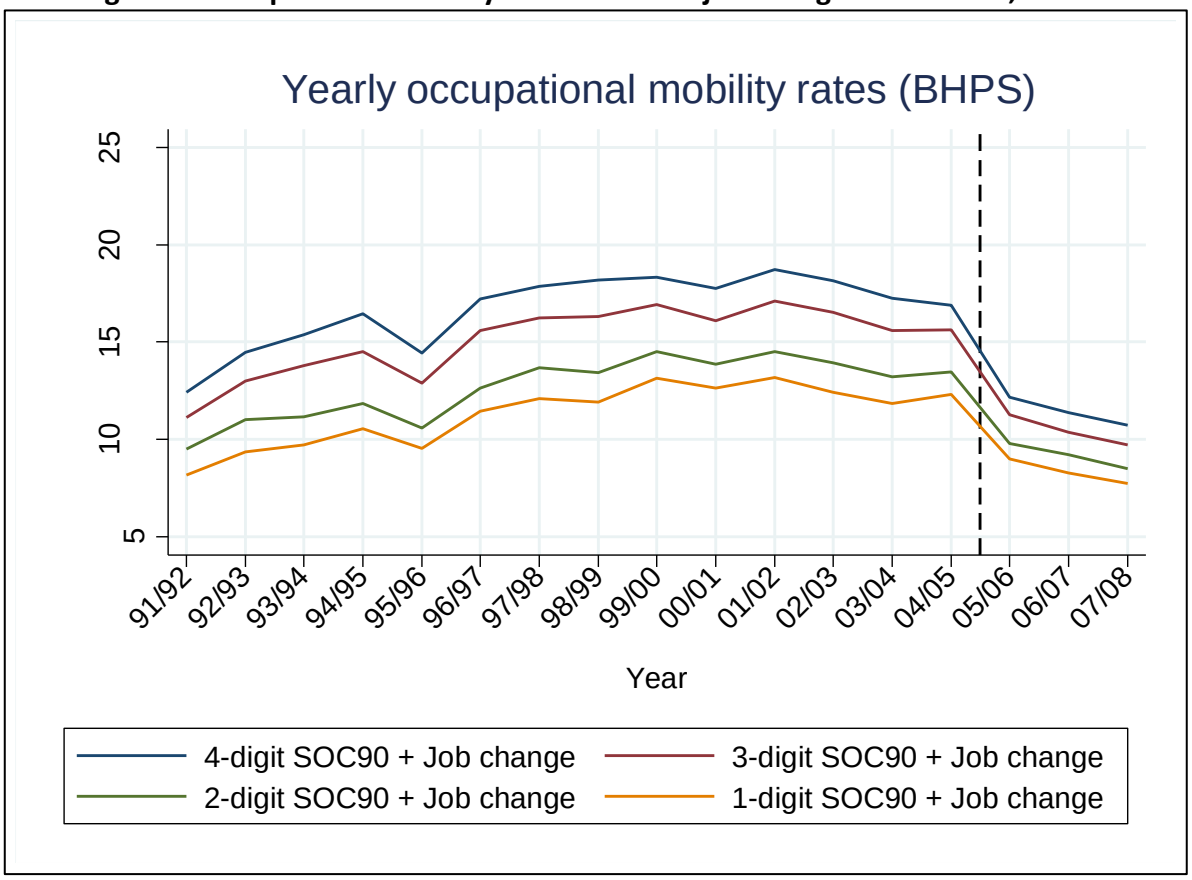
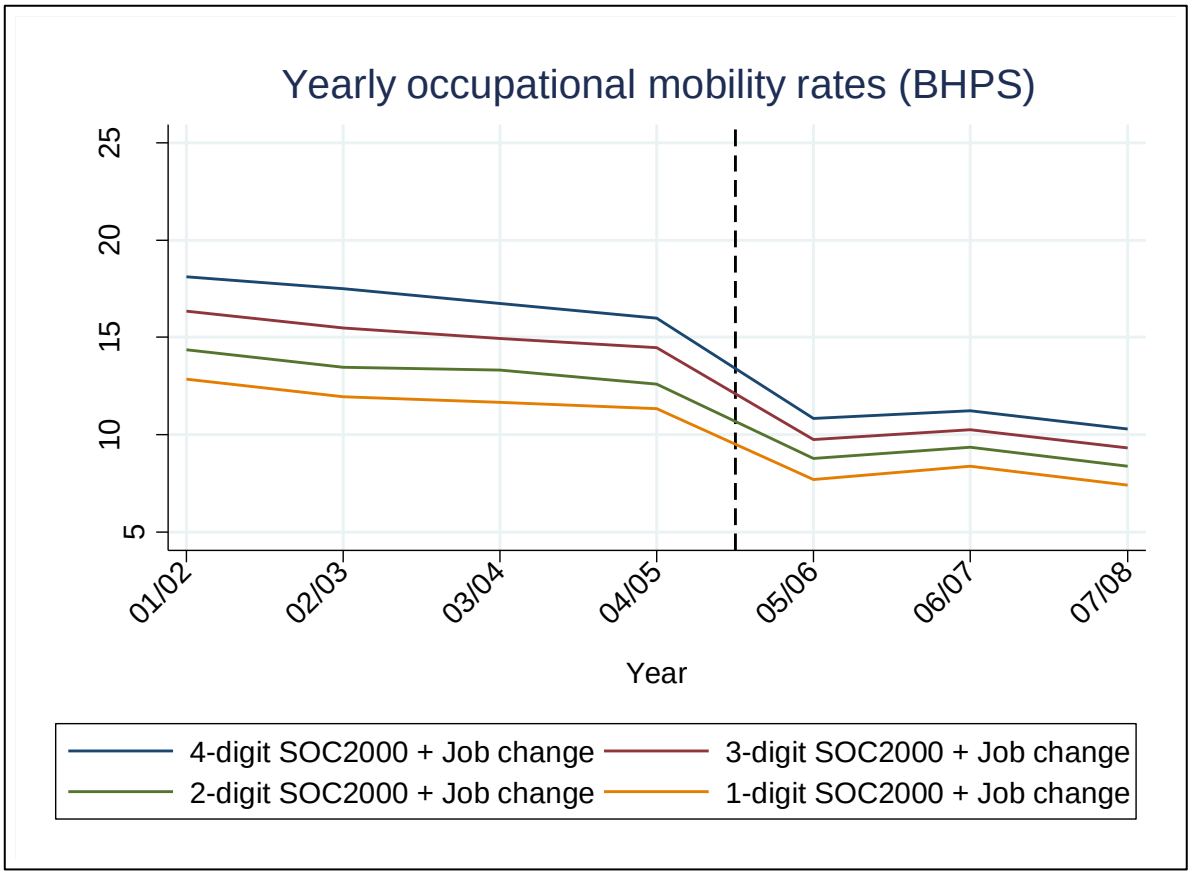


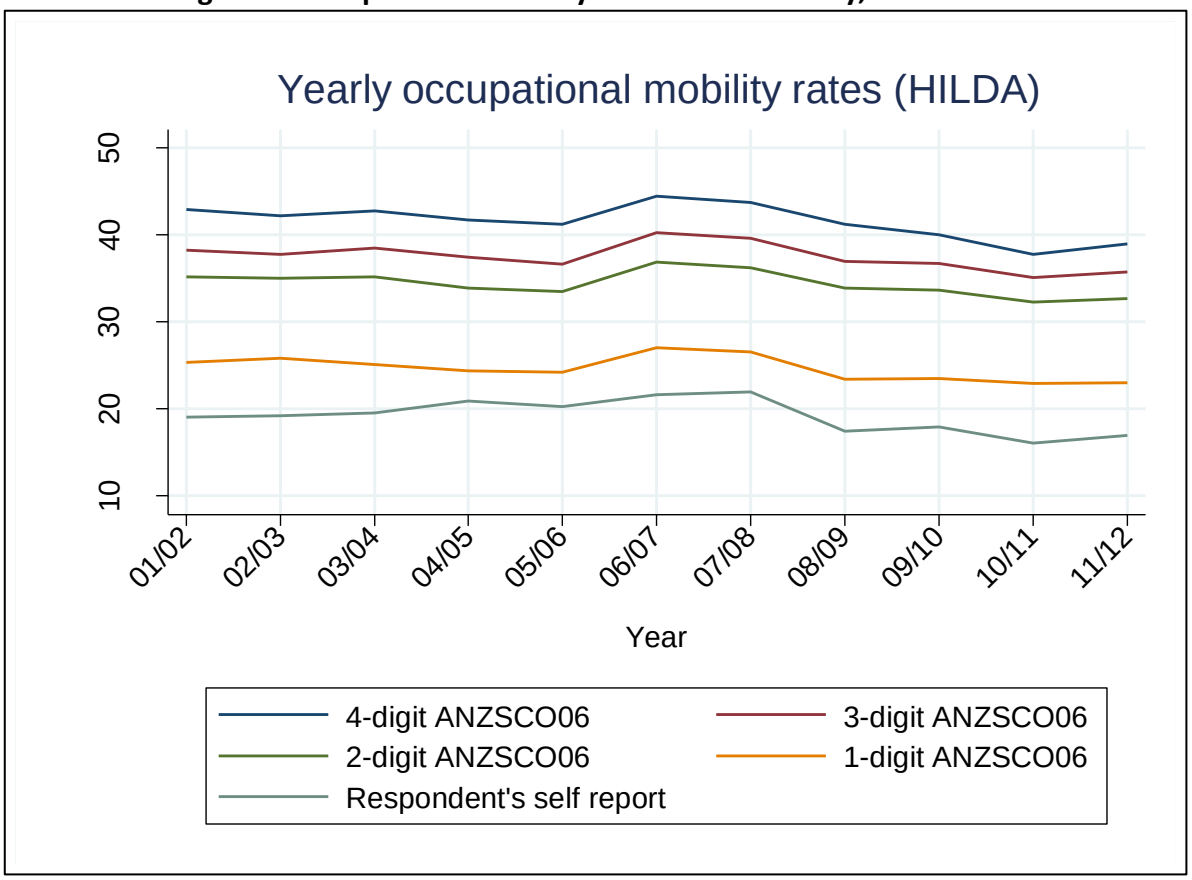
Figure 4. Occupational mobility conditional on job change in the BHPS, SOC2000



We now turn to discuss occupational mobility trends in Australia using the HILDA Survey data. Figure 5 shows trends for the period 2001-2012. As for the BHPS, we infer mobility using changes in occupational codes across adjacent survey instalments at each level of aggregation of the ANZSCO06 occupational classification. Additionally, we show trends based on the variable included in the survey that considers respondents' self-reports. Before we do this, it is worth noting how much the assessments of occupational mobility vary across measures. The measure of occupational change relying on respondents having different occupation codes and the measure based on respondents' self-reports disagree in 25-30% of cases and pairwise correlations between these are just 0.31-0.34.

The lines in the graph indicate that there is also a suspiciously high rate of occupational mobility in the Australian labour market – similar to that for Britain – which also increases with occupational disaggregation. Around 40% of workers appear to switch detailed occupational units each year, whereas around 25% seemingly move across major occupational groups. This is unsurprising, given that the HILDA Survey uses independent interviewing to collect occupational data throughout the life of the panel. A more interesting, novel piece of information emerges when examining the rates of occupational change calculated using respondents' self-reports. These are substantially lower than those calculated using occupational code mismatches across adjacent survey years and indicate that around 20% of all workers change occupations each year. The discrepancies with the 4-digit level of aggregation – arguably the level at which respondents judgments operate (Moscarini & Thomsson, 2007) – are very high (20-25 percentage points). This strongly suggests that 'true' levels of occupational change are indeed lower than those typically inferred from occupational codes in panel studies. It is also worth noting that, unlike what was observed for the BHPS, there are no major shifts in the trend in the Australian panel data. To the extent that the labour markets in these two countries and the two panel surveys are comparable, this constitutes supporting evidence that it is the survey practice of dependent interviewing that is behind the shifts observed in Figures 1 to 4 for Britain.

Figure 5. Occupational mobility in the HILDA Survey, ANZSCO06



The impact of occupational mobility on labour market outcomes

Testing the external validity of survey measures requires assessing their relative performance in adhering to theoretical predictions. As discussed before, the sociological and economic literatures expect occupational mobility to be a vehicle for upward social mobility and, consequently, be accompanied by rises in wages and 'utility'. The latter notion is empirically approximated using information on job satisfaction (Frey & Stutzer, 2002). In this part of the analysis, we test how the different measures of occupational change perform in predicting wages and job satisfaction in both Britain and Australia. Logically, the most valid measures will show the strongest partial correlations with these labour market outcomes.

We begin by looking at the BHPS using wage and job satisfaction change-score models of the following form:

$$\ln(W)_{it} - \ln(W)_{it-1} = a + C_{it-1,t}\beta + X_{it}\gamma + e_{it}$$

$$JS_{it} - JS_{it-1} = a + C_{it-1,t}\beta + X_{it}\gamma + e_{it}$$

…where i and t subscripts stand for individual and time, a is an intercept, ln(W) is the natural logarithm of hourly wages, JS denotes overall job satisfaction (on a scale from 1-7), C is a given measure of occupational change, X is a vector of control variables, e is the usual stochastic regression error term, and β and γ are regression coefficients or vectors of coefficients to be estimated. Hourly wages are calculated using information on the number of usual weekly work hours and the usual gross monthly pay in the current main job, and adjusted for inflation using Consumer Price Indexes. The X vector comprises control variables used in recent research on occupational mobility (Longhi & Brynin, 2010): age group, gender, highest educational qualification, partnership status, number of children, part-time work, firm size, employment contract and private sector employment.

For simplicity, we estimate these equations using standard ordinary least squares regression. More complex models are possible using these data. For example, one could model job satisfaction as an ordered outcome, account for selection into employment and allow for unobserved heterogeneity. However, because the focus of this article is elsewhere and for simplicity's sake, we choose to estimate relatively simple and generalizable specifications. The models are fitted once for year 2004 (two years before the introduction of dependent interviewing) and once for year 2008 (after the introduction of dependent interviewing). We choose time periods that are two rather than one year away from the introduction of dependent interviewing because of the odd behaviour of occupational mobility rates in the year immediately following its implementation in 2006 observed in Figures 1 and 2. Results are presented in Table 1.

In the wage equation for 2004 (column 1) occupational mobility is captured by changes in occupational codes across adjacent survey years under *independent* interviewing conditions. Here, a change in occupation has no statistically significant impact on wage changes all else being equal, irrespective of the occupational classification (SOC90 or SOC2000) and level of occupational aggregation (1 to 4 digits) used. In the wage equation for year 2008 (column 2) occupational mobility is captured by changes in occupational codes across adjacent survey years under *dependent* interviewing. In these models, a change of occupation is associated with an increase of 2-3% in hourly wages, which is in most cases statistically different from zero. Therefore, using changes in occupational codes across adjacent survey waves as evidence of occupational mobility yields results which are more consistent with theory under dependent than independent interviewing. This hints that the former is a more desirable way to collect occupational data in panel studies.

We now turn our attention to the job satisfaction equations in columns 3 and 4. Occupational mobility is associated with increased job satisfaction under both interview practices, but more strongly so under dependent than independent interviewing. In the BHPS, job satisfaction is measured on a scale from 1 to 7. The magnitude of the impacts is of 0.12-0.19 units in the equation for year 2004 under independent interviewing, and a much higher 0.49-0.58 units in the equation for year 2008 under dependent interviewing. Again, this is evidence in support of the idea that, if occupational change is inferred by changes in occupational codes, dependent interviewing is a more desirable survey feature than independent interviewing.[vi]

Table 1. Impacts of occupational change on wages and job satisfaction in the BHPS

Occupational mobility measured as a…	Y=Change in…			
	Log hourly wages		Job satisfaction	
	2004	2008	2004	2008
4-digit SOC90 change	−0.004	0.033**	0.190***	0.550***
3-digit SOC90 change	−0.001	0.037**	0.181***	0.551***
2-digit SOC90 change	−0.003	0.021	0.156***	0.536***
1-digit SOC90 change	−0.005	0.032*	0.125**	0.491***
4-digit SOC2000 change	−0.001	0.026*	0.177***	0.570***
3-digit SOC2000 change	0.001	0.025*	0.175***	0.567***
2-digit SOC2000 change	0.006	0.024(*)	0.188***	0.582***
1-digit SOC2000 change	0.008	0.023(*)	0.183***	0.544***
n	*5,192*	*4,847*	*5,233*	*4,910*

*Notes. OLS models. 1991-2008. Control variables include age group, gender, highest educational qualification, partnership status, number of children, part-time work, firm size, employment contract, and private sector employment. Significance levels: (*) 0.1, * 0.05, ** 0.01, *** 0.001.*

Similar regression models to those estimated above are fitted using all 12 available waves of the HILDA Survey data. This time, we compare the performance in predicting change in wages and job satisfaction of measures of occupational mobility based on individuals having different occupational codes across adjacent survey years and respondents' self-reports. Results are presented in Table 2.

Column 1 contains the results of wage equations. When using occupational mobility measures in which change is inferred from occupational codes, the predicted impacts of occupational moves on wage changes are very small (0.4%-0.6%) and sometimes not (or only marginally) statistically significant. In contrast, the measure of mobility constructed from self-reports performs much better: it predicts a growth of 1.8% in hourly wages following occupational switches and the associated coefficient is statistically significant at the 99.9% level.

Table 2. Impacts of occupational change on wages and job satisfaction in the HILDA Survey

Occupational mobility measured as a…	Y=Change in…	
	Log hourly wages	Job satisfaction
4-digit ANZSCO06 change	0.006*	0.233***
3-digit ANZSCO06 change	0.006*	0.255***
2-digit ANZSCO06 change	0.006(*)	0.264***
1-digit ANZSCO06 change	0.004	0.282***
Self-reported occupation change	0.018***	0.553***
n (observations)	*60,874*	*63,907*
n (individuals	*12,832*	*13,153*

*Notes. OLS models with standard errors adjusted for the clustering of observations within individuals. 2001-2012. Control variables include age group, gender, highest educational qualification, partnership status, number of children, part-time work, firm size, employment contract, and private sector employment. Significance levels: (*) 0.1, * 0.05, ** 0.01, *** 0.001.*

Column 2 contains the results of equivalent job satisfaction models. Note that overall job satisfaction has a different range in the HILDA Survey of 0-10, and so the magnitude of the reported effects is not comparable to that of the effects reported for the BHPS. As for the BHPS, all measures of occupational mobility are associated with growth in job satisfaction. However, all else being equal, self-reported occupational changes are more strongly associated with job satisfaction (they increase it by 0.55 units) than are occupational mobility measures based on changes in occupational codes (0.23-0.28 units).

These results provide clear evidence that, under independent interviewing, self-reported measures of occupational change deliver results which are more consistent with theoretical expectations than measures based on occupational codes. This suggests that the former might be a better construct to capture occupational mobility in panel surveys, and studies other than the HILDA Survey would benefit from its inclusion.

Discussion and conclusion

Longitudinal data on occupation of work is of key importance to research in the sociology of work and labour economics, but the quality of information on occupational mobility in panel surveys is suspected to be poor due to measurement error. In this paper we have examined the consistency of data on occupational change in panel studies via separate analyses of the BHPS and the HILDA Survey.

Consistent with our first and main research hypothesis, analysis of BHPS data evidences that the switch from independent to dependent interviewing in the collection of occupation data in this survey substantially reduced the implausibly high, previously observed mobility rates. This suggests that, when occupation-related data are collected via independent interviewing, measurement error is high. Also as expected, the fall in mobility rates was more visible for more disaggregated versions of occupational classifications than for more aggregated versions, implying that the error is more pervasive in detailed schemas. The latter is unfortunate, as highly disaggregated occupational units convey the most valuable information to social and behavioural scientists. Using supplementary information on job mobility to separate spurious from 'true' occupational switches is a useful though imperfect tool to correct for these issues under independent interviewing conditions.

Analysis of the HILDA Survey further indicated that respondents' self-reports of occupational switches often conflict with mobility inferred from changes in occupational codes across survey waves and depict more modest yearly rates of occupational mobility. This is consistent with our third hypothesis and constitutes supporting evidence for the argument that, under independent interviewing conditions, inferring occupational mobility from changes in occupational codes across survey waves is problematic and does not necessarily capture 'true' change. Adding to this view and in line with hypotheses 4 and 5, the impacts of occupational changes on earnings and job satisfaction are more consistent with theoretical predictions after the introduction of dependent interviewing in the BHPS and when occupational mobility is inferred from respondents' self-reports in the HILDA Survey.

Substantively, our preferred estimates suggest that the 'true' extent of year-on-year occupational mobility is 10-15% in Britain and 15-20% in Australia, and that occupational switches have non-negligible effects on the wages and overall job satisfaction of workers in both countries.

Based on these results, we conclude that measurement error in occupational mobility in panel surveys can be very high, to a level that is intolerable and has the potential to severely distort research findings. This has important implications. Our results are applicable to other panel datasets which employ the 'traditional' method of independent interviewing to collect occupation-related data, including the HILDA Survey, the German Socio-Economic Panel (see Isaoglu, 2010) and other studies affiliated to the CNEF, and should be taken into account in the design and re-design of panel surveys that collect data on occupation of work. In particular, the implementation of dependent interviewing practices and the inclusion of survey questions capturing perceived occupational mobility are desirable courses of action. The patterns reported here also highlight the need to revise findings from existing research studies that take the quality of occupational mobility data in panel surveys for granted. Due to the statistical noise produced by this type of measurement error, 'true' relationships might have gone unnoticed, whereas correlation of the errors

with individual- and occupation-level traits might have given rise to spurious associations (Sullivan, 2009). Additionally, any survey measures derived from occupational mobility/stability data, such as tenure in an occupation as a proxy for occupation-specific human capital, will be affected (Kambourov & Manovskii, 2009; Sullivan, 2009).

There are, however, *caveats* to this research which must be kept in mind when interpreting its findings and which might hopefully inspire further inquiry. First, in our BHPS analyses we do not observe rates of occupational mobility under independent and dependent interviewing synchronically, that is, for the same survey year. This implies that, although this is highly unlikely, we cannot explicitly rule out that the reported breaks in the trends in occupational mobility after the adoption of dependent interviewing practices are partially the product of real changes in the prevalence of this phenomenon, or of some other mechanism such as non-random panel attrition.

Second, implementation of dependent interviewing in the BHPS might not only be associated with a decrease in misclassification in occupational reports, but also with an increase in other survey errors, such as acquiescence bias (Hill, 1994; Jäckle, 2009; Lugtig & Lensvelt-Mulders, 2013; Mathiowetz & McGonagle, 2000). Given the routing in the relevant BHPS module, if undecided or unknowledgeable respondents do have a tendency to agree with the statement presented to them, then dependent interviewing would produce downwards-biased estimates of change. For instance, respondents might report occupational stability even if some new element, such as managerial duties, has come into their jobs and effectively places then into a different occupation. Alternatively, agreeing might be an act of 'satisficing' to avoid having to provide a new occupational description (a relatively lengthy process). However, Jäckle et al. (2007) found no evidence of this sort of satisficing behaviour in evaluating the introduction of dependent interviewing in the BHPS, and acquiescence bias is usually most problematic in long batteries of similarly structured questions and in telephone interviews, neither of which is the case here.

Third, self-reported measures of occupational mobility are not perfect either: respondents might lack knowledge on what exactly constitutes a change in occupation, as opposed to both stability and other types of mobility (e.g. job changes and employer changes). Given how difficult it is to delimit detailed occupations, these measures may in fact understate 'true' mobility levels if respondents tend to report stability when unsure. Such measures may also be polluted by measurement error due to poor recalls of the timing of switches relative to the previous interview date, with the resulting errors likely to be more prevalent when the two events take place within a short time period.[vii]

It would be difficult to test the degree to which these mechanisms affect our findings using existing data. However, the robustness of our results could be thoroughly examined via survey experimentation. Devices such as the UK Household Longitudinal Study: *Understanding Society* Innovation Panel (Buck & McFall, 2012) could be used for such an endeavour. In addition, the refinement of models that allow for misclassification in occupational reports is urgently needed to improve the analysis of existing longitudinal datasets in which occupation-related information has been collected in non-optimal ways (Sullivan, 2009). It is highly likely that the type of measurement error in panel datasets considered in this paper is confined to occupation-related information, and hence exploring its incidence in other work-related information, such as data on industry of work (Lynn & Sala, 2006), would prove a fruitful avenue for further research. Additionally, future studies could advance the findings reported here by examining whether longitudinal inconsistencies in occupational reports are random or, instead, affect certain types of individuals and occupations more than others.

Acknowledgements

The author would like to thank Nicole Watson, Noah Uhrig and Wojtek Tomaszewski for helpful comments and suggestions on an earlier version of this paper, Simonetta Longhi for sharing her expertise on job changing in BHPS data, and three anonymous LLCS referees for comments and suggestions that added substantive value to this work. The BHPS data used in this paper were made available through the ESRC Data Archive. The BHPS data were originally collected by the ESRC Research Centre on Micro-social Change at the University of Essex (now incorporated within the Institute for Social and Economic Research). Neither the original collectors of the data nor the Archive bear any responsibility for the analyses or interpretations presented here. This paper also uses unit record data from the HILDA Survey. The HILDA Project was initiated and is funded by the Australian Government Department of Social Services (DSS) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this paper, however, are those of the author and should not be attributed to either DSS or the Melbourne Institute.

References

Abbott, A. (1989). The new occupational structure: What are the questions? *Work and Occupations*, *16* (3), 273-291. http://dx.doi.org/10.1177/0730888489016003002

Blau, P. M., & Duncan, O. D. (1967). *The American occupational structure*. New York: Wiley.

Booth, A. L., & Francesconi, M. (1999). Job mobility in 1990s Britain: Does gender matter? ISER Working Papers. Colchester: University of Essex.

Borjas, G. (1981). Job mobility and earnings over the life cycle. *Industrial & Labor Relations Review, 34* (3), 377–388. http://dx.doi.org/10.2307/2522783

Brown, J., & Light, A. (1992). Interpreting panel data on job tenure. *Journal of Labor Economics, 10* (3), 219-257. http://dx.doi.org/10.1086/298286

Buck, N., & McFall, S. (2012). Understanding Society: Design overview. *Longitudinal and Life Course Studies,* 3 (1), 5-17.

Bukodi, E., & Dex, S. (2010). Bad start: Is there a way up? Gender differences in the effect of initial occupation on early career mobility in Britain. *European Sociological Review*, *26* (4), 431-446. http://dx.doi.org/10.1093/esr/jcp030

Campanelli, P., Thomson K., Moon, N., & Staples, T. (1997). The quality of occupational coding in the UK. In L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.) *Survey measurement and process quality* (pp. 437-457). New York: Wiley.

Chan, T. W., & Goldthorpe, J. H. (2007). Social stratification and cultural consumption: Music in England. *European Sociological Review*, *23* (1), 1-29. http://dx.doi.org/10.1093/esr/jcl016

De Beyer, J., & Knight, J. B. (1989). The role of occupation in the determination of wages. *Oxford Economic Papers*, 41 (3), 595-618.

Elias P. (1997). Occupation classification (ISCO-88): Concepts, methods, reliability, validity and cross-national comparability. *OECD Labour Market and Social Policy Occasional Papers 20*, OECD Publishing. http://dx.doi.org/10.1787/304441717388

European Foundation for the Improvement of Living and Working Conditions. (1997). *Voluntary and forced job mobility in Europe*. Accessed 21st of February 2014 from: http://www.eurofound.europa.eu/pubdocs/2007/12/en/1/ef0712en.pdf.

Evans, P. (1999). Occupational downgrading and upgrading in Britain. *Economica*, *66* (261), 79-96. http://dx.doi.org/10.1111/1468-0335.00157

Featherman, D. L., Lancaster Jones, F., & Hauser, R. M. (1975). Assumptions of mobility research in the United States: The case of occupational status. *Social Science Research*, *4* (4), 329-360. http://dx.doi.org/10.1016/0049-089X(75)90002-2

Frey, B. S., & Stutzer, A. (2002). What can economists learn from happiness research? *Journal of Economic Literature*, *40* (2), 402-425. http://dx.doi.org/10.1257/jel.40.2.402

Groot, W., & Verberne, M. (1997). Aging, job mobility, and compensation. *Oxford Economic Papers*, *49* (3), 380-403. http://dx.doi.org/10.1093/oxfordjournals.oep.a028615

Goldthorpe, J. H. (1987). *Social mobility and class structure in modern Britain* (2nd edition). Oxford: Clarendon.

Groves, R. M. (1989). *Survey errors and survey costs*. Wiley, New York. http://dx.doi.org/10.1002/0471725277

Harper, B. (1995). Male occupational mobility in Britain. *Oxford Bulletin of Economics and Statistics, 57* (3), 349-369. http://dx.doi.org/10.1111/j.1468-0084.1995.mp57003005.x

Hill, D. H. (1994). The relative empirical validity of independent and dependent data in a panel survey. *Journal of Official Statistics*, *10 (*4), 359-380.

Isaoglu, A. (2010). Occupational affiliation data and measurement errors in the German Socio-Economic Panel. *SOEP Papers on Multidisciplinary Panel Data Research 318*. Berlin: Deutsches Institut für Wirtschaftsforschung.
Jäckle, A., Laurie, H., & Uhrig, S. C. N. (2007). The introduction of dependent interviewing on the British Household Panel Survey. *ISER Working Papers, 2007-07*. Colchester: University of Essex.
Jäckle, A. (2009). Dependent interviewing: A framework and application to current research. In P. Lynn (Ed.) *Methodology of longitudinal surveys* (pp. 93-112). Chichester: John Wiley. http://dx.doi.org/10.1002/9780470743874.ch6
Jacobs, S. (1999). Trends in women's career patterns and in gender occupational mobility in Britain. *Gender, Work and Organization*, *6* (1), 32-46. http://dx.doi.org/10.1111/1468-0432.00067
Kambourov, G., & Manovskii, I. (2008). Rising occupational and industry mobility in the United States: 1968-97. *International Economic Review*, *49* (1), 41-79. http://dx.doi.org/10.1111/j.1468-2354.2008.00473.x
Kambourov, G., & Manovskii, I. (2009). Occupational specificity of human capital. *International Economic Review, 50* (1), 63-115. http://dx.doi.org/10.1111/j.1468-2354.2008.00524.x
Kish, L. (1965). *Survey sampling*. New York: John Wiley & Sons.
Laurie, H., & Moon, N. (2003). The reliability of coding occupational descriptions: Measurement issues in a CAPI panel survey. *ISER Working Papers, 2003-13*. Colchester: University of Essex.
Lee, K., Carswell, J., & Allen, N. (2000). A meta-analytic review of occupational commitment: Relations with person and work-related variables. *Journal of Applied Psychology*, *85* (5), 799-811. http://dx.doi.org/10.1037/0021-9010.85.5.799
Longhi, S., & Brynin, M. (2010). Occupational change in Britain and Germany. *Labour Economics, 17* (4), 655-666. http://dx.doi.org/10.1016/j.labeco.2010.02.001
Lugtig, P., & Lensvelt-Mulders, G. J. L. M. (2013). Evaluating the effect of dependent interviewing on the quality of measures of change. *Field Methods* (forthcoming).
Lynn, P., & Sala, E. (2006). Measuring change in employment characteristics: The effects of dependent interviewing. *International Journal of Public Opinion Research*, *18* (4), 500-509. http://dx.doi.org/10.1093/ijpor/edl013
Mathiowetz, N. A., & McGonagle, K. A. (2000). An assessment of the current state of dependent interviewing in household surveys. *Journal of Official Statistics, 16* (4), 401-418.
Mellow, W., & Sider, H. (1983). Accuracy of response in labor market surveys: Evidence and implications. *Journal of Labor Economics*, *1* (4), 331-344. http://dx.doi.org/10.1086/298016
Moscarini, G., & Thomsson, K. (2007). Occupational and job mobility in the US. *Scandinavian Journal of Economics*, *109* (4), 807-836. http://dx.doi.org/10.1111/j.1467-9442.2007.00510.x
Parrado, E., Caner, A., & Wolff, E. N. (2007). Occupational and industrial mobility in the United States. *Labour Economics*, *14* (3), 435-455. http://dx.doi.org/10.1016/j.labeco.2006.01.005
Perales, F. (2013). Occupational feminization, specialized human capital and wages: Evidence from the British labour market. *Work, Employment and Society*, *27* (4), 600-620. http://dx.doi.org/10.1177/0950017012460305
Polachek, S. W. (1981). Occupational self-selection: A human capital approach to sex differences in occupational structure. *Reviews of Economics and Statistics*, *63* (1), 60-69. http://dx.doi.org/10.2307/1924218
Sobek, M. (1996). Work, status, and income: Men in the American occupational structure since the late nineteenth century. *Social Science History*, *20* (2), 169-207. http://dx.doi.org/10.2307/1171236
Sorensen, A. B. (1974). A model for occupational careers. *American Journal of Sociology*, *80* (1), 44-57. http://dx.doi.org/10.1086/225761
Sullivan, P. (2009). Estimation of an occupational choice model when occupations are misclassified. *Journal of Human Resources*, *44* (2), 495-535. http://dx.doi.org/10.1353/jhr.2009.0030
Taylor, M. F., Brice, J., Buck, N., & Prentice-Lane, E. (2010). *British Household Panel Survey user manual volume A: Introduction, technical report and appendices*. Colchester: University of Essex.
Watson, N., & Summerfield, M. (2009). Quality of the occupation and industry coding in the HILDA Survey. *HILDA Discussion Paper Series 3/09*. Melbourne: Institute of Applied Economic and Social Research, University of Melbourne.
Watson, N., & Wooden, M. (2012). The HILDA Survey: A case study in the design and development of a successful household panel study. *Longitudinal and Life Course Studies*, *3* (3), 369-381.
Wilson, R. M., & Green, C. (1990). Occupation, occupational change and movement within the income distribution. *Eastern Economic Journal*, *16* (3), 209-220.

Endnotes

[i] From a *total survey error* perspective, four sources of error affect survey data: sampling, coverage, non-response, and measurement error (Kish, 1965; Groves, 1989). Here, we are concerned about the last of these components: measurement error.

[ii] Inter-coder reliability rates vary substantially with the complexity of the occupational classification used, especially with the number of occupational units (Campanelli, Thomson, Moon & Staples, 1997). As a result, there is not a widely accepted threshold of what constitutes an acceptable inter-coder reliability rate. Nevertheless, agreement rates over 75-80% are generally considered acceptable (Elias, 1997).

[iii] A large share of the sociological literature is devoted to examining *inter*-generational occupational mobility (i.e. the transmission of occupational attainment from parents to their offspring). Here, the focus is however on *intra*-generational (i.e. within-generation) occupational mobility.

[iv] The picture becomes substantially more complex when the reasons for the move are considered, most importantly whether separations are voluntary (i.e. worker-initiated) or forced (i.e. employer-initiated). While the former sort of moves should be associated with substantial wage gains, the latter need not. We cannot and do not incorporate this distinction into our analyses. However, existing research suggests that voluntary moves are three to four times more prevalent than forced moves (European Foundation for the Improvement of Living and Working Conditions, 1997; Harper, 1995). We would therefore expect the net effect of mobility on wages to be positive.

[v] After 2008, BHPS respondents were incorporated into a new panel survey (the UK Household Longitudinal Study: *Understanding Society*). However, use of the BHPS subsample of *Understanding Society* for our purposes is problematic. First and foremost, there was a two-year gap between 2008 and 2010 when BHPS respondents were not interviewed. This would make it difficult to infer year-on-year occupational changes. Second, *Understanding Society* data are only available for up to Wave 3, and thus its usage would add little to our analyses. Third, the global financial crisis reached the British labour market in 2008 and may have exerted a profound exogenous impact in long-term mobility trends.

[vi] We also estimated models using the occupational mobility measures that incorporate information on job changes. Results, not shown but available upon request, suggest that under independent interviewing these perform better than mobility measures based solely on changes in occupational codes in predicting change in job satisfaction, but not wage changes.

[vii] Additionally, it is possible that reverse causality is at play when respondents answer the question on self-reported occupational mobility in the HILDA Survey: individuals might be more inclined to remember a change or even catalogue a change as such if this was accompanied by a boost in either wages or job satisfaction.

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.

DOIs. Do NOT include DOIs in the boibliography – they are added during editing as resolvable URLs.

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 APA 6th edition 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.

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.

CONTENTS

EDITORIAL

PAPERS

LLCS 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