VOLUME 6 | ISSUE 3 | **JULY 2015**

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

SUBSCRIPTIONS

For immediate access to the current and preceding issues in Volume 6 as well as previous volumes:

***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

INTRODUCTION

PAPERS

STUDY PROFILE

COMMENT AND DEBATE

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

SLLS Society for Longitudinal and Life Course Studies

SLLS International Conference

'Life Courses in Cross-National Comparison: Similarities and Differences'

Dublin, Ireland

18 - 21 October 2015

(Pre-conference workshop on 18.10.15, with main conference starting on 19.10.15)

The conference theme will focus on life courses in cross-national comparison this year, but will cover other areas of longitudinal and life course study: physical, psychological, social developmental and ageing processes and functioning within and across life course stages from infancy to old age; methods and findings of cohort studies; other sources of longitudinal data such as panel studies and record linkage; international comparisons; household, and income dynamics; intergenerational transfers and returns to learning; gene-environment interactions; 'mixed', and comparative methods; innovative methodology in design, measurement, data management, analysis and research practice (quantitative and qualitative).

Keynote speakers:
Prof. Elizabeth Cooksey, Ohio State University, USA
'Intergenerational Linkages, Interdisciplinary Collaborations and Interesting Findings: Lessons Learned from 50 Years of the National Longitudinal Surveys';

Prof. Richard Layte, Economic and Social Research Institute (ESRI), Ireland
'National Crises and Personal Troubles: The Great Recession in Ireland and Family Processes';

Prof. Rainer Silbereisen, Center for Applied Developmental Science (CADS), Germany
'Dealing with Perceived Uncertainties in Work and Family and their Psychosocial Consequences: Comparative Studies in Germany and Poland in Times of Social Change'.

The fee for conference registration will be discounted for SLLS members so please consider joining if you have not already done so! Please visit the membership page of the website for full details of how to join online: http://www.slls.org.uk/#!services/ch6q

More conference information will be posted on the SLLS website as it becomes available. Registration will open at the end of July 2015.

www.slls.org.uk

Editorial

John Bynner

The move from three to four issues per annum – January, April, July October - happened relatively seamlessly at the beginning of the year and is now fully established. This (July) issue is a good example of the new shape the journal will take.

The issue contains research papers addressing a range of life course topics as pursued in different countries together with a new sub-study profile and a Debate section. The latter is the first of what will be a series of debates about issues arising in the theory and practice of life course research. The final component is a conference abstracts supplement.

The papers reflect different facets of family functioning at different life course stages displaying both the range of work that is being carried out and the variety of long-term longitudinal surveys in different countries deployed to do it.

The effects of the post-2008 recession on economic vulnerability start the proceedings, drawing on the first two waves in the two cohorts – starting ages nine months and age nine - in the *Growing Up in Ireland* (GUI) longitudinal study. The differential effects of the recession on the development between the two cohorts is clearly shown in terms of vulnerability profiles with single parent families particularly subject to economic stress in the older (child) cohort. In the next paper the landscape switches to Australia where eight waves of the *Household Income and Labour Dynamics in Australia (HILDA)* survey is used to challenge the hypothesis that the amount of time spent on housework relates to selection into and out of relationships.

The following paper shifts the focus to the relationship between work-family conflict and mental disorders using age 45 data from the UK *1958 Birth Cohort Study.* Using data collected at later ages from the same source the next paper tackles the long-term adult health effects past age 50 of parental divorce. Methodology for trajectory identification follows, exemplified by analysis across eighteen waves of *British Household Panel Survey (BHPS)* data to identify trajectories of 'functional disability' in the elderly between ages 65 and 74.

The Study Profile reports the supplementation of the standard family and finance information collected in a sub-study of the 2013 US *Panel Study of Income Dynamics (PSID)* survey to serve as a resource to support research on family transfers of funds from parents to children. The new Debate section follows in which Leon Feinstein and others debate methodological criticism of his influential research based on UK *1970 Birth Cohort Study* data on 'Cognitive development and socioeconomic grades', which identified relative decline of cognitive development through the early years and childhood of children from working class family backgrounds compared with their middle class counterparts.

Finally a supplement to the issue comprises some 75 abstracts from papers and workshops presented at the 2015 *Understanding Society* scientific conference held by the UK Household Longitudinal Survey team based at the University of Essex.

The good news about expanded production is complemented by the increase in the number of submissions and people attending this year's SLLS conference in Dublin 18-21 October. SLLS members benefit from a substantial discount on the conference fee. And membership is a major part of the life-blood on which the journal's existence depends.

SLLS membership is linked to the other major component of journal support – libraries. We have launched a major push through the summer to boost the number of libraries across the world subscribing to the journal from the present 40. Paying the LLCS library subscription gives library users immediate access to all issues of the journal up to the most recent edition. If your library has yet to sign up please do all you can to persuade the ('Serials') librarian to do so, or let us know by emailing Journal Manager Cara Randall at crandall@slls.ork.uk and we will write to them ourselves. Your journal needs you!

Join our mailing lists…

Cohort Network Group

SLLS is proud to host a forum for people working in and on longitudinal studies. It aims to build on links made under the EUCCONET (*European Child Cohort Network*) whose funding for co-ordination and communication between child cohorts ended in 2013. That venture brought together researchers across the behavioural, developmental, and health and statistical sciences, and the professional data, survey and communications managers who are also an important part of the interdisciplinary teams who create and run these studies.

Key objectives of the network are the maintenance and continuation of existing studies and the facilitation of the development of new ones at local or national level, even if the aspiration for a pan-European cohort seems unrealistic.

For full details and to join the CN mailing list visit http://www.slls.org.uk/#!cohort-network/c21hq

Interdisciplinary Health Research Group

Large-scale social surveys increasingly collect biomedical data, but at present an inter-disciplinary forum concerned with making best use of these combined social and biological data, is lacking.

A preparatory meeting was held at the SLLS Annual Conference 2014, to assess whether SLLS could fill this gap. Twenty conference delegates from the social and biological sciences attended the preparatory meeting and agreed to propose to the SLLS Executive Committee that a SLLS sub-group on *Interdisciplinary Health Research* be established. The Executive Committee agreed the group with the following remit:

- To enable informed use of biomarkers by social scientists
- To enable informed use of social data by biologists
- To bring together SLLS researchers from a variety of disciplines who work on or have an interest in health and health-related issues

For full details and to join the IHR mailing list visit http://www.slls.org.uk/#!health-research/c1njv

Policy Group

Life course study and longitudinal research are potentially of central importance to the policy process. The burgeoning of major longitudinal studies throughout the world and the allocation of large-scale government funding to building longitudinal resources reflect this growing interest. In this respect, SLLS is well placed to identify the expertise and research resources needed to underpin the relevant evidence base in different policy domains. For this reason the SLLS Executive Committee decided to create a database registering members' expertise, relevant experience and policy interest areas. It acts as a source of partners for collaboration on international longitudinal research projects directed at policy issues; helps the Executive Committee respond to policy debates; and broaden the scope of our international journal, LLCS, in policy research directions.

For full details and to join the PG mailing list visit http://www.slls.org.uk/#!policy-group/c99m

www.slls.org.uk

Longitudinal and Life Course Studies 2015 Volume 6 Issue 3 Pp 230 – 244 *ISSN 1757-9597*

Family economic vulnerability & the Great Recession: an analysis of the first two waves of the Growing Up in Ireland study

Dorothy Watson Economic and Social Research Institute, Dublin and Trinity College, Dublin
Christopher T. Whelan University College Dublin
christopher.whelan@ucd.ie
Bertrand Maître Economic and Social Research Institute, Dublin and Trinity College, Dublin
James Williams Economic and Social Research Institute, Dublin and Trinity College, Dublin

(Received November 2014 Revised January 2015) http://dx.doi.org/10.14301/llcs.v6i3.331

Abstract

In this paper we make use of the first and second waves of the 2008 and 1998 cohorts of the Growing Up in Ireland study, to develop a multidimensional and dynamic approach to understanding the impact on families and children in Ireland of the Great Recession. Economic vulnerability is operationalised as involving a distinctive risk profile in relation to relative income, household joblessness and economic stress. We find that the recession was associated with a significant increase in levels of economic vulnerability and changing risk profiles involving a more prominent role for economic stress for both the 2008 and 1998 cohorts. The factors affecting vulnerability outcomes were broadly similar for both cohorts. Persistent economic vulnerability was significantly associated with lone parenthood, particularly for those with more than one child, lower levels of primary care giver (PCG) education and, to a lesser extent, younger age of PCG at child's birth, number of children and a parent leaving or dying. Similar factors were associated with transient vulnerability in the first wave but the magnitude of the effects was significantly weaker particularly in relation to lone parenthood and level of education of the PCG. For entry into vulnerability the impact of these factors was again substantially weaker than for persistent and transient vulnerability indicating a significantly greater degree of socioeconomic heterogeneity among the group that became vulnerable during the recession. The findings raise policy and political problems that go beyond those associated with catering for groups that have tended to be characterized by high dependence on social welfare.

Keywords: Economic vulnerability, life course, social class, economic class

Introduction

Ireland has seen quite remarkable macroeconomic fluctuations over the past two decades, with the fastest economic growth rate among the countries of the Organisation for Economic Co-operation and Development (OECD) during the so-called 'Celtic Tiger' boom being followed by a recession which had a more negative impact on national output in Ireland than in any other OECD country. The decade of exceptionally rapid growth from the mid-1990s saw the numbers employed expand dramatically and unemployment reduced to 4%, but was characterized by an unsustainable credit-fuelled expansion in the construction sector and an unprecedented property price boom. From 2008 onwards, the country faced the bursting of the property bubble, a collapse of property-related tax revenues, falling asset values, a major banking crisis and a ballooning fiscal deficit. This toxic combination meant that by late 2010, despite substantial increases in taxation and expenditure cuts, the Irish government had to avail itself of a 'bail-out' from the Troika (Whelan, 2010). The impact of the recession involved a decline of 13% in GDP between 2008 and 2011 and a rise in unemployment from 4% to 14% between 2007 and 2011 (Callan, Nolan, Keane, Savage & Walsh, 2013). In the context of a sharp decline in real household incomes and associated poverty thresholds, at-risk of poverty rates based on 60% of median income remained relatively stable over time. However, sharp increases were observed in levels of material deprivation and economic stress. Throughout the period 2004-2011 at-risk of poverty rates based on 60% of median income and material deprivation rates were significantly higher for children than for the remainder of the population. Both rates were higher than for most of the EU15 countries (Watson, Maître & Whelan, 2012). [1]

The research reported in this paper draws on data from the *Growing Up in Ireland* (GUI) Survey that were collected between 2007 and 2011. Data for two cohorts of children are used, one born in 1998 (the 1998 cohort) and one born in 2008 (the 2008 cohort). The availability of data for two waves for each cohort allows us to compare the pre and post-recession situations of families with infant children (age nine months and three years) and children in middle childhood (age nine and 13 years). Our analysis takes advantage of the longitudinal nature of the GUI survey and adopts a multidimensional perspective. It has been well-established, for instance, that the groups identified on the basis of low income alone differ from those found to experience multiple deprivation in terms of factors such as employment status, stage of the life course, housing tenure and urban-rural location (Nolan & Whelan, 1996). Thus, in line with emerging trends on the literature, our approach is both multidimensional and dynamic (Nolan & Whelan, 2007, 2010). A great deal of debate has taken place regarding the range of dimensions which should be incorporated in measures of childhood poverty or deprivation (Tomlinson, Walker & Williams, 2009). A crucial factor here is the position one adopts regarding the relative importance of description versus identifying causal processes. Here, rather than attempting to capture the full range of deprivations experienced by children, we focus on a multidimensional but restricted conception of economic vulnerability based on low income, household joblessness and subjective economic stress. We then proceed to explore the socio-economic factors which contribute to changing patterns of exposure to such vulnerability and their consequences. This broader measure of economic vulnerability offers a number of advantages as an indicator of longer term command over resources (Whelan & Maître, 2005).

Our research is intended to provide a contribution to the literature on the socioeconomic distribution of childhood poverty and its consequences. A good deal of the evidence relating to the extent and consequences of persistent poverty in childhood is drawn from the United States where welfare structures and child supports are distinctive (Duncan, Magnuson, Kalil & Ziol-Guest, 2012). However, as Mood and Jonsson (2014) note, increasing interest in child poverty and deprivation is related to United Nations' (UN) demands that, having ratified the Child Convention, countries monitor trends in the living conditions of children. This, as they note, has led to the development of a range of relevant welfare indices with material living conditions featuring as a central indicator of children's well-being (Bradshaw, Hoelscher & Richardson, 2006; Bradshaw & Richardson, 2009; UNICEF, 2007).

The Growing Up in Ireland (GUI) data

The GUI survey is a national longitudinal study of children. It tracks the development and wellbeing of two nationally representative cohorts of children: the 1998 cohort and the 2008 cohort. The samples were strict probability samples. The 1998 cohort of children was selected following clustering at the level of the school, while the 2008 cohort was a random sample selected from the Child Benefit records with the assistance of the Department of Social Protection. Interviews were conducted via Computer-Assisted Personal Interview (CAPI) with the primary care-giver (PCG, usually the mother), the resident secondary care-giver (SCG, usually the father), the teachers of the 1998 cohort of children at wave one and with the 1998 cohort children themselves. In the present analysis we rely on data provided by the PCG.

In this paper, data from the first two waves of both cohorts are used, when the children in the 2008 Cohort were nine months and subsequently three years old and those in the 1998 Cohort were nine and subsequently 13 years of age. The samples in the study were reweighted or statistically adjusted to ensure that they were nationally representative of the age groups in question, both cross-sectionally and longitudinally.[2] The present analysis includes the 9,793 families who responded in both waves of the 2008 cohort and the 7,423 families who responded in both waves for the 1998 cohort. The large sample sizes, the probability samples and the calibration to ensure representativeness mean that the results can be generalised to the population of children in both cohorts.

The timing of the GUI surveys in relation to the onset of 'The Great Recession' is important. The first wave of the 1998 cohort was conducted with the families of the nine-year olds between August 2007 and June 2008, slightly before the major shocks of the recession later that year. The second wave, when the children were aged 13, took place between August 2011 and March 2012. This corresponded to the deepest point of the recession, before any growth in employment was evident. The first wave of the 2008 cohort, when the children were aged nine months, occurred between September 2008 and March 2009 right at the start of the recession when unemployment was rising most sharply. The second wave, when the children were three years old, was from December 2010 to July 2011. At this stage, unemployment was still increasing and GNP was still falling but at a much slower rate.

Given the timing of the fieldwork, we would expect that the families of the 2008 cohort would already be beginning to experience the effects in terms of unemployment or concerns about employment loss in the first wave. For this reason, we might expect that the impact of the recession would be seen most clearly in the 1998 cohort since the first wave interviewing was substantially completed before the very steep rise in unemployment in the fourth quarter of 2008.

Identifying economically vulnerable groups

As knowledge of the limitations of relying solely on income to measure poverty and social exclusion has become more widespread, attention has been increasingly focused on multi-dimensional approaches (Grusky & Weeden, 2007, Nolan & Whelan, 2007, 2010). In addition to being concerned with multidimensionality, advocates of the social exclusion perspective have sought to distinguish it from the conventional income approach through its emphasis on dynamics – the manner in which processes unfold over time. Such concerns have led to the emergence, from a number of sources, of a focus on what has been termed 'vulnerability'. This involves a shift of focus from current deprivation to insecurity and exposure to risk and shock. The International Monetary Fund (IMF) (2003), the UN (2003) and the World Bank (2000) have developed a range of approaches to measuring vulnerability at the macro level. The World Bank (2000) sees vulnerability as reflecting the risk of experiencing an episode of poverty over time but also a heightened probability of being exposed to a range of disadvantages. Hanappi, Bernardi and Spini (2015), in a systematic review of social science literature using the concept of vulnerability, argue that 'vulnerability is a major concept for interdisciplinary research and potential theory development'. The authors suggest that vulnerability may remain latent until individuals or groups are challenged by critical events or the depletion of limited resources.[3]

Our objective in this paper is to focus at a micro level in order to identify families who are vulnerable to economic exclusion in the sense of being distinctive in their risk of falling below a critical resource level, living in a household characterised by a high level of joblessness and experiencing subjective economic stress. In other research, using the Survey of Income and Living (SILC) data, economic vulnerability has been measured using indicators of income poverty status, material deprivation and economic stress (Whelan & Maître, 2005, 2010, 2014). Our choice of indicators in the present study is influenced both by substantive considerations and the quality of the data available in the GUI surveys. Details of the three components of economic vulnerability are shown in figure 1. We use a similar indicator of economic stress but a slightly different measure of income level and we substitute household joblessness for material deprivation. A comparison of material deprivation levels for comparable families in GUI and SILC showed that levels of deprivation in the former were substantially below those reported in the latter. This was despite the fact that no such differences were observed in relation to the distribution of income and key socioeconomic factors that have been shown to be associated with levels of deprivation. One possibility is that social desirability considerations may have led parents to be reluctant to report material deprivation in the context of a survey focused on children. In any event, the unrealistically low levels of such deprivation in the GUI survey in comparison with comparable groups in the SILC data, led us to exclude this dimension, which has been employed in previous analyses of economic vulnerability (Whelan & Maître, 2010), and instead focus on household joblessness.

Income level refers to the quartile of equivalised household income. Equivalised income refers to total household income from all sources and all household members, net of taxes and social contributions, and adjusted for household size and composition by dividing by an equivalence scale.[4]

Income is measured by a single item answered by the PCG in GUI and is, therefore, not measured with as much precision as in SILC which is specifically designed for the purpose of collecting detailed income information.[5] Consequently, in what follows we focus on income quartiles rather than seeking to estimate numbers below conventional income poverty lines.

Economic stress is measured by a single item which has been used extensively in Irish surveys to capture 'difficulties in making ends meet'. Household joblessness is defined using the European Commission concept of 'very low work intensity', as described in figure 1.

Figure 1: The components of economic vulnerability

Variable	Description
Income level	The income quartile of the household calculated separately for each cohort in each time period. One quarter of each cohort in wave one and in wave two is found in each quartile.
Economic stress	Whether the family has 'great difficulty' or 'difficulty' in making ends meet.
Household joblessness ('very low work intensity')	The working-age adults in the household are currently in employment for less than one fifth of the available hours. Working-age adults are aged 18 to 59, excluding full-time students under age 25. The percentage of available time worked is calculated as a percentage of 35 hours, which is regarded as full-time for this purpose. This percentage is capped at 100. Note: hours worked are available for the primary and secondary care-givers only. For other adults of working age, we only know whether or not they are in employment. In calculating work intensity, we assume the work of any other adults is full-time.

The approach we adopt in analysing economic vulnerability involves an analysis of manifest indicators in order to identify underlying or latent vulnerability. We achieve this objective by the application of latent class analysis, which can be used as a tool to gain deeper understanding of the observed relationships between categorical indicators. It can be thought of as a log-linear model where there are more dimensions in the estimated table than in the observed table. Such models generate tables of expected frequencies that can be compared to the observed frequencies using goodness of fit tests. The basic idea underlying such analysis is long established and very simple (Lazarsfeld, 1950). The associations between a set of categorical variables, regarded as indicators of an unobserved typology, are accounted for by membership of a small number of latent classes. As Moisio (2004) notes, implicit in the notion of multidimensional measurement of social exclusion is the assumption that there is no one 'true' indicator of the underlying concept. Instead, we have a sample of indicators that tap different aspects of a complex phenomenon. Latent class analysis assumes that each individual is a member of one and only one of N latent classes and that, conditional on latent class membership, the manifest variables are mutually independent of each other. Conditional independence is a version of the familiar idea that the correlation between two variables may be a result of their common dependence on a third variable (McCutheon & Mills, 1998). In estimating latent class models the logic is identical but the explanatory variable is unobserved and must be identified statistically. The axiom of local independence can be seen as the defining characteristic of latent class analysis. It assumes causality running from the latent variable to the manifest indicators.

The three measured characteristics of the families shown in figure 1 yield a 16-cell table (4x2 x2) relating to multidimensional profiles of economic exclusion. Such a profile can be established for each of the four groups we consider:

- The 2008 cohort at nine months
- The 2008 cohort at three years
- The 1998 cohort at nine years
- The 1998 cohort at 13 years

Taking into account both the range of indicators and the number of groups produces a 64 cell table. Our objective is to develop a parsimonious latent class model of the underlying processes producing an allocation of individuals to the cells of this table that generates a set of expected values that provide a satisfactory fit.

Since our objective is to identify an overall economically vulnerable class that can be contrasted with the remainder of the population, we develop models with two latent classes[6]. For each model we report the likelihood ratio chi-square test (G^2) and the percentage of cases misclassified. The findings relating to a number of models are set out in table 1. The first model, which we use as a benchmark for the performance of the remaining models, allows for an association between the manifest indicators of economic vulnerability and cohort and but assumes no association between the vulnerability indicators. Not surprisingly, this model provides a poor fit to the data and misclassifies 16.6% of the cases.

A fully homogeneous latent class model reduces the G^2 for the conditional independence model by 76.8% but misclassifies 6.7% of cases. The model that allows the size of the vulnerable class to vary by cohort and time reduces the G^2 by 80.3% and misclassifies 5.9% of cases. Finally the fully heterogeneous model, which allows both the size and the profile of the vulnerable class to vary across combinations of cohorts and observation periods, reduces the conditional independence G^2 by 99.4% and misclassifies 2.8% of cases.

Table 1: Latent class model fit statistics for GUI data

	G^2	Degrees of freedom	Reduction in independence G^2	% of cases misclassified
Model*s*				
1. Conditional independence	7,647.5	45		16.6
2. Fully homogenous	1,775.3	49	76.8	6.7
3. Heterogeneous on size by cohort & time	1,508.0	46	80.3	5.9
4. Fully heterogeneous	480.3	16	99.4	2.8

Source: GUI Longitudinal '98 Cohort and '08 Cohort datasets, analysis by authors.

A satisfactory fit requires that we take into account differences in the size of the vulnerable class and vulnerability profiles relating to both cohort and period and the manner in which they interact. Having done so, the fully heterogeneous model provides a reasonably satisfactory account of the observed patterns of multidimensional economic vulnerability across cohort and time.

Levels and profiles of economic vulnerability by age group and cohort

In table 2 we show the size of the economically vulnerable class by cohort and time period and the profile of the vulnerable and non-vulnerable groups in terms of the three indicators. The sizes of the vulnerable classes are based on assigning individuals to the class in which they have the highest probability of being located.[7] The table also shows the percentage of cases that would be misclassified under the model assumptions of zero association between the manifest indicators within latent classes. The figure ranges from 2% to 4.5%, indicating a generally acceptable model fit.

Focusing first on latent class size we find that for the 2008 cohort the observed level of economic vulnerability was 19% in wave one and rose to 25% in wave two. For the older 1998 cohort the level was substantially lower in the first period of observation at 15%. However, by the second period it rose sharply to 25%. The difference in fieldwork timing, noted earlier, is important in accounting, at least in part, for the higher level of economic vulnerability among the 2008 cohort in the first wave and the sharper increases between waves for the 1998 cohort.

The conditional probabilities show, for each combination of period and cohort, the contrast between vulnerable and non-vulnerable latent classes in terms of their risk of being found in the most disadvantaged categories of each of the manifest indicators. Focusing on the younger 2008 cohort, we find that while 63% of the vulnerable class are found in the bottom quartile, this figure falls to 10% for the non-vulnerable class. The respective figures for the third and fourth deciles combined are 3% (2% + 1%) and 68% (34% + 34%). In relation to economic stress, 36% of the vulnerable class were above the relevant threshold compared to 4% of the non-vulnerable. Finally, 48% of the vulnerable class fulfilled the joblessness criterion compared to 1% of the non-vulnerable class. By the second time period when the children were three years old, a relatively similar situation existed in relation to the numbers in the bottom quartile, with respective figures of 62% and 9%, and in the third and fourth quartiles combined, with figures of 4% (4% + 0%) for the vulnerable class and 70% (34% + 36%) for the non-vulnerable class. However, economic stress levels rose among both vulnerable and non-vulnerable classes producing respective levels of 47% and 10%, so that the relative disadvantage of the vulnerable class narrowed while the absolute difference showed a modest increase. Finally, the level of joblessness showed little change for the non-vulnerable class but rose to 58% for the vulnerable class. So as the size of the vulnerable class increased, striking differences in risk profiles persisted but the impact of the recession was uneven across the component indicators.

Turning our attention to the older 1998 cohort, we find that in wave one the number found in the bottom income quartile is higher than for the younger 2008 cohort at 15% but once again is substantially higher for the vulnerable group at 69%. As in the 2008 cohort, stress levels were substantially higher for the vulnerable than the non-vulnerable group with respective figures of 31% and 2%. The contrast in terms of joblessness was even sharper than in the case of the younger cohort with a figure of 58% for the vulnerable class and 1% for the non-vulnerable class. As the size of the vulnerable class rose significantly, differentiation in terms of income quartiles became less sharp. Thus the number in the vulnerable class in the bottom quartile declined from 69% to 49% while the number in the second quartile rose from 26% to 42% and in the top half rose from 4% (4% + 0%) to 8% (7% + 1%). In relation to economic stress, we observe increases that are somewhat larger than in the case of the 2008 cohort resulting in levels of 55% and 10% for the vulnerable and non-vulnerable classes, respectively. In contrast, as with the income quartiles, differentiation in relation to household joblessness narrowed slightly with respective figures of 50% and 1% cent for the vulnerable and non-vulnerable clusters.

Table 2: Latent class size and profiles by cohort

	9 months old		3 years old		9 years old		13 years old	
% Vulnerable	18.8		25.4		14.7		24.9	
% of cases misclassified	2.45		4.51		2.04		2.21	
Economically vulnerable	Not vulner-able	Vulner-able	Not vulner-able	Vulner-able	Not vulner-able	Vulner-able	Not vulner-able	Vulner-able
	Conditional probabilities		Conditional probabilities		Conditional probabilities		Conditional probabilities	
Income quartile								
1	10%	63%	9%	62%	15%	69%	14%	49%
2	22%	34%	21%	34%	25%	26%	18%	42%
3	34%	2%	34%	4%	30%	4%	33%	7%
4	34%	1%	36%	0%	31%	0%	35%	1%
Economic stress	4%	36%	10%	47%	2%	31%	10%	55%
Joblessness	1%	48%	0%	58%	1%	58%	1%	50%
Number of cases	9679		9679		7408		7408	

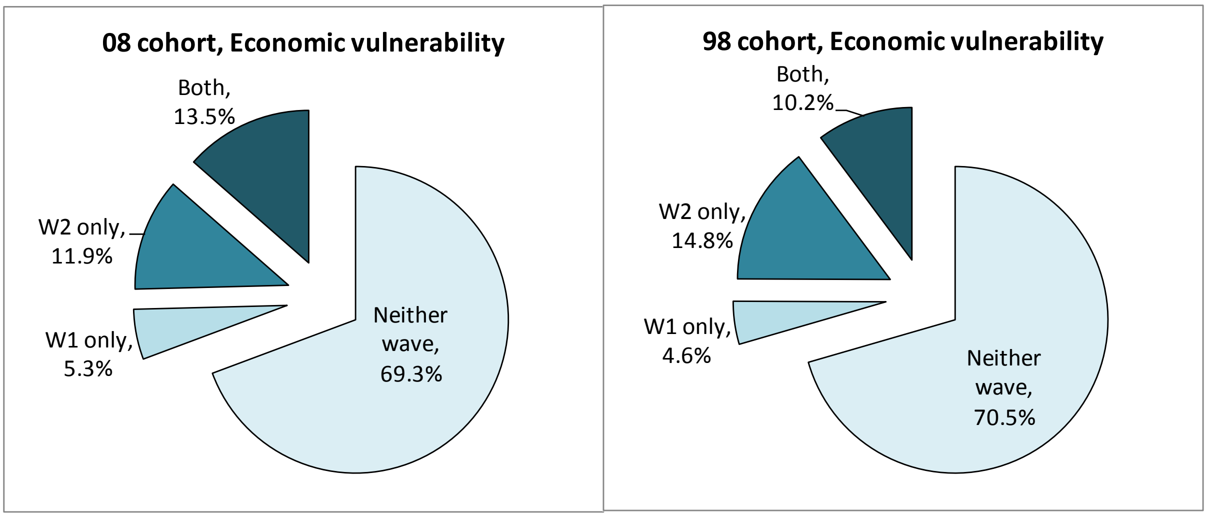
Source: GUI Longitudinal 1998 Cohort and 2008 Cohort datasets, analysis by authors.

Thus the size of the vulnerable group increased for both cohorts but more so for the older 1998 one. Members of the vulnerable group remained sharply differentiated from the non-vulnerable but trends varied across the component items. Economic stress levels increased for both vulnerable and non-vulnerable groups for both cohorts. In contrast joblessness remained minimal for the non-vulnerable in both cohorts, increased for the vulnerable 2008 group but declined for the vulnerable cluster in the 1998 cohort. This latter finding is consistent with the fact that for the 13-year old group, vulnerability was significantly more widely distributed across the income distribution as the impact of the Great Recession led to a more pervasive distribution of economic stress. This finding is consistent with analysis of the impact of the Great Recession on adult economic vulnerability levels and profiles employing EU-SILC data (Whelan & Maître, 2014).

Economic vulnerability patterns by family characteristics

At this point we examine the extent to which there was change or stability over time in the economic vulnerability status of families. As noted above, individuals are allocated to the vulnerable or non-vulnerable class on the basis of the modal assignment rule with each observation in a cell being assigned to the cluster with the largest probability. Employing this approach in figure 2, we show the proportion of children in the two cohorts who are found in each of the four economic vulnerability classes: economically vulnerable in neither period; vulnerable in wave one only; vulnerable in wave two only and vulnerable in both waves.

Figure 2: Economic vulnerability of the two cohorts



Source: GUI Longitudinal 1998 Cohort and 2008 Cohort datasets, analysis by authors. The relative risk of being vulnerable in wave 2 only or in both waves is significantly higher for the 1998 cohort.

The proportion of children who were economically vulnerable in neither period is very similar for the younger and older cohorts at about 70%. The proportion vulnerable in wave two is also very similar at about 25%. Among the younger cohort, 14% were economically vulnerable in both periods compared to 10% of the older cohort.

Inevitably, given the recession, most of the change in vulnerability status involved movements into economic vulnerability. About 12% of the younger cohort and 15% of the older cohort experienced this transition. Roughly 5% of children in both cohorts experienced movement in the opposite direction.

In table 3 we provide a description of patterns of vulnerability for the combined 2008 and 1998 cohorts by a range of family characteristics that previous research suggests are likely to be associated with vulnerability (e.g. Fahey, Keilthy & Polek, 2012; Watson, Maître & Whelan, 2012). These include household type, educational level and age at birth of the PCG, cohabitation and changes in household status relating to parents and number of children. We focus first on the likelihood of being vulnerable in both waves. The highest expected risk is for lone parents with two children (49%), followed by lone parents with one child (36%). It then declines sharply to 9% for a couple with 3+ children and 5% for families with 2 children and 4% for those with one child. The next most powerful influence was the educational level of the PCG. For lower second level or below, the rate was 27%. It then fell to 9% for upper second or lower third level and finally to 2% for upper third level. Persistent vulnerability was highest among those PCGs who were youngest when the child was born. For those aged less than 25 the rate was 28% and for those aged between 25-29 it fell to 15%, For no other group was the risk above 11%. Where a new parent was present in the new wave, 37% experienced persistent vulnerability, compared to 19% where a parent had died or left and 11% where there was no change. Neither cohabitation nor more children in wave two had any effect.

For those experiencing economic vulnerability in wave one only, a similar but much less sharp pattern of social differentiation emerges. The fact that such transient poverty is likely to be affected by a wide variety of relatively specific factors while at the same time a significant degree of social structuring remains evident, is consistent with the likelihood that families who experience economic vulnerability in highly favourable economic circumstances but then manage to exit this state are likely to experience recurring spells of vulnerability. In relation to family type, the highest probability of being vulnerable only in wave one was observed for lone parents with two children with a rate of 15%. This fell to 12% for their counterparts with one child and to between 3–4% for the remaining family types. Focusing on age of PCG, the highest rate of between 7–9% was observed for the under 30 groups before falling to 3–5% for older groups. For educational level the rate declined from 8% for the least educated group to 3% for the most educated. Where a new care giver joined by wave 2 the rate was highest at 16% but a parent dying or leaving had no effect.

Table 3: Observed economic vulnerability pattern by family characteristics

		W1 only	W2 only	Both	Neither
Age of PCG at birth of child	Under 25	9%	20%	28%	43%
	25-29	7%	14%	15%	64%
	(30-44 ref.)	4%	11%	8%	78%
	35-39	3%	11%	7%	78%
	40+	5%	15%	11%	69%
Household type W1	Lone parent, 1 child	12%	13%	36%	39%
	Lone parent, 2+ children	15%	12%	49%	25%
	Couple, one child	4%	10%	4%	82%
	Couple 2 children	3%	13%	5%	79%
	Couple, 3+ children	3%	16%	9%	72%
Cohabiting W1	No	5%	12%	12%	71%
	Yes	6%	18%	12%	64%
Change in parent(s)	No change, same parent(s)	5%	12%	11%	72%
	New parent	16%	11%	37%	36%
	One parent died/left	6%	32%	19%	43%
More children in Wave 2	No	5%	13%	12%	69%
	Yes	3%	12%	12%	73%
PCG Education, W1	Low 2nd or less	8%	19%	27%	46%
	Upper 2nd to lower 3rd	5%	14%	9%	73%
	Upper 3rd	3%	6%	2%	89%

Source: GUI Longitudinal 1998 Cohort (N=7,423) and 2008 Cohort (N=9,793) datasets, analysis by authors.

For those experiencing vulnerability in wave two only, a rather different pattern of socioeconomic differentiation is observed. Families with the youngest head of households were again at the highest level of risk with a rate of 20%. However, on this occasion the figure for the 40+ group at 15% is marginally higher than for the 25—29 group. For the remaining groups the figure falls to 11%. Once again risk is clearly related to education level and ranges from 19% for the lowest level of qualification to 6% for the highest. Those families where a parent left or died were sharply differentiated from other families with a risk level of 32% compared to between 11 and 12%. The loss of a parent is clearly a traumatic event which is associated with persistent vulnerability and entry into vulnerability. The arrival of a new parent has little effect on entry into vulnerability, but is strongly associated with both persistent and transient vulnerability. This may well reflect longer term processes of family instability. In contrast to the earlier findings, the pattern of differentiation by family type is extremely modest with no clear contrast between lone parent families. The highest rate of 16% is observed for couples with three or more children and the lowest of 10% for couples with one child. Cohabitation and having more children in wave two had little effect.

The probability of being vulnerable in neither wave obviously follows directly from the outcomes discussed to date and it is highest for couple with one to two children, for families with younger and better educated PCGs and where no change in parents is observed.

Economic vulnerability patterns and relative risk ratios

In extending our analysis of the factors associated with patterns of vulnerability dynamics we conducted a multinomial (polytomous) logit regression on the pooled data for both cohorts. The dependent variable is the four-category indicator of economic vulnerability dynamics: vulnerable in neither period (the reference category); transient vulnerability (vulnerable in wave one only); becoming economically vulnerable (vulnerable in wave two only) and persistent vulnerability (vulnerable in both periods). The analysis uses weighted data with standard errors adjusted for clustering and weights.[8] The results relating to relative risk ratios (RRRs) are set out in table 4 below. Goodness-of-fit statistics suitable for survey data estimators (i.e. when the *svy* prefix is used to ensure correct standard errors) are not available for multinomial regression models (StataCorp, 2013b). In the present context, since our focus is on model parameter estimates rather than comparing alternative models, the overall goodness of fit of the model is less critical. The adjusted Wald test is reported instead (Heeringa, West & Berglund, 2010). This is a test of the null hypothesis that all parameters are not significantly different from zero. The test has a p value less than .001, indicating that the null hypothesis should be rejected.

The reference category for the dependent variable is being economically vulnerable in neither wave. The impact of socioeconomic factors is relatively uniform between the 2008 and 1998 cohorts. However, a number of significant variations emerge and these are taken into account by specific interaction terms. Focusing first on the relative risk of persistent vulnerability, by far the most significant influence is family type and in particular lone parenthood. Taking a couple with one child as the benchmark we find that the RRR for lone parents with two or more children reaches 74.8. For their counterparts with only one child the value falls to 22.4. There is then a sharp drop to 4.5 for a couple with three children and to 1.92 for those with two children. The next most important influence is educational level of the PCG. With third level as the reference category, the RRR for lower second level or below is 11.8 while for upper second and lower third level it is 2.9. Age of the PCG at birth of child has a significant but less substantial impact. With the 35–39 group as the benchmark, the RRR for the under 25 group was 3.6. (However, this value relates only to the 2008 cohort and falls to 2 for the 1998 cohort). It then fell to 1.8 for the 25–29 group and to 1.6 for the 40+ group. No significant difference was observed for the 30–34 age group compared to the 35–39 age group. With the no change in parent(s) category as the reference group, the RRR for those where a parent died or left was 4.9. However, while a gross positive effect for the arrival of a new parent was observed, controlling for other factors leads to a net RRR of 0.6. In most cases, this pattern involved a formerly lone parent marrying, so that there was a shift from a one-parent to a two-parent family. The net effect of cohabitation was to raise the RRR by 1.7 and of more children to increase it by 1.5.

Transient vulnerability is the pattern where the family was economically vulnerable in wave one but became non-vulnerable in wave two. Given the recession, this pattern was relatively unusual. Turning to a consideration of factors associated with this pattern, we find that taking a couple with one child as the benchmark, the RRR rises to 8.5 for lone parents with two or more children (However, for the 1998 cohort this increases to 20.1). For lone parents with one child the RRR is 5.2. There is no significant difference between couples with different numbers of children.

Age of PCG effects were similar to those for persistent poverty with the RRR ranging from 2.3 for the under 25 group to 1.9 for those aged 25–29 and no difference between families where the PCG was older at the birth of the child. Families where a parent subsequently left or died had an RRR of 2.6. In the 2008 cohort the arrival of a parent was associated with an RRR of 1.6 but for the 1998 cohort this fell to 0.7 with this group enjoying a relative advantage rather than a disadvantage. Cohabitation raises the RRR by 1.3 but families which subsequently had more children had an RRR of 0.7.

Table 4: Relative risk ratios for economic vulnerability dynamics for the 2008 and 1998 cohort (Reference category is vulnerable in neither period)

		2008 Transient economic vulnerability (W1 only)	2008 Becoming economically vulnerable (W2 only)	2008 Persistent economic vulnerability (Both waves)
Cohort (Ref: 2008 cohort)	1998 cohort	1.000	1.435	0.380
Age of PCG at birth of child (35-39 ref.)	Under 25	2.292	3.644	3.606
	25-29	1.948	1.777	1.779
	30-34	1.000	1.000	1.000
	40+	1.000	1.553	1.576
Household Type W1 (Ref: Couple, 1 child)	Lone parent, 1 child	5.233	2.413	22.390
	Lone parent, 2+ children	8.460	4.266	74.830
	Couple, two children	1.000	1.544	1.916
	Couple, 3+ children	1.000	2.778	4.483
Cohabiting W1?	Yes	1.319	1.253	1.695
Change in parent(s) (Ref: No change)	Yes	1.619	1.000	0.567
	One parent died/left	2.601	5.055	4.935
More children in Wave 2?	Yes	0.693	1.262	1.504
PCG Education, W1 (Ref: Upper 3rd)	Lower 2nd or less	5.215	4.498	11.800
	Upper 2nd to lower 3rd	2.053	2.577	2.946
		1998 Cohort (where different)		
Age of PCG at birth of child (Ref: 35-39)	Under 25		1.665	2.005
	25-29		1.088	
	30-34			
	40+			
Household Type W1 (Ref: Couple 1 child)	Lone parent, 1 child			
	Lone parent, 2+ children	20.059		
	Couple, two children			
	Couple, 3+ children		1.878	
Cohabiting W1?	Yes			
Change in parent(s) Ref: No change, same parent(s))	Yes	0.661		
	One parent died/left		3.084	
More children in Wave 2?	Yes			
PCG Education, W1 (Ref: Upper 3rd)	Lower 2nd or less	2.435		
	Upper 2nd to lower 3rd	0.834	1.866	

Source: GUI Longitudinal 1998 Cohort (N=7,423) and 2008 Cohort (N=9,793) datasets. Note: Where cell is blank for child cohort, relative risk ratio is the same as for 2008 cohort. Where Relative risk ratio is not statistically different from the reference category for the 2008 cohort at p<=0.1, 1.000 is used in the table. Adjusted Wald test for all parameters: F (55, 10,298) = 39.89; p value <.001.

A significant contrast between the 2008 and 1998 cohorts emerges in relation to the strength of the association between the educational level of the PCG and vulnerability in the first wave. For the 2008 cohort, lower secondary qualifications are associated with an RRR of 5.2 while for upper secondary to lower third level a value of 2.1 is observed compared to the reference category of PCGs with third level qualifications. For the PCGs of the 1998 cohort, where lower levels of qualifications were more the norm, the contrast by level of education is smaller and the respective values are 2.4 and 0.8.

Becoming economically vulnerable (being vulnerable in wave two only) involves moving into vulnerability during the recession. Consistent with our earlier discussion of the gross effects, family type has a significantly weaker association with becoming economically vulnerable than with transient or persistent vulnerability. Taking a couple with one child as the reference category, the RRR for lone parents with two children is 4.3 while for those with one child it falls to 2.4. The RRR for a couple with three or more children is 2.8 while for those with two children it is 1.5. Thus lone parenthood plays a significantly weaker role for this outcome. For the 1998 cohort the only different effect is for a couple with three or more children where it falls to 1.9 compared to those with one child. Overall, family type is a much weaker factor in accounting for becoming economically vulnerable than for the other outcomes (persistent and transient vulnerability). In addition, the age of the PCG at the time the child was born has a stronger effect for the 2008 cohort than for the 1998 cohort. For the former the RRR for the 25 or under group is 3.6 and for the 25–29 category is 1.8 compared to the reference category of age 35–39. For the 1998 cohorts the respective figures are 1.7 and 1.1. For both groups the RRR for the 40+ group is 1.6. For the 2008 cohort the impact of education is similar to that for transient vulnerability with RRRs of 4.5 and 2.6 but for the 1998 cohort the latter figure relating to higher second and lower third level falls to 1.9. Cohabitation and having more children raise the RRR modestly to 1.3. In the 2008 cohort the departure or death of a parent increased the RRR of becoming economically vulnerable by 5.1 but for the 1998 cohort the RRR was lower at 3.1.

Controlling for all other variables in the model, the 1998 cohort was less likely to experience persistent vulnerability with an RRR of 0.4 but was more likely to be vulnerable in wave two only with an RRR of 1.4.

Overall, our analysis indicates that the factors associated with risk of persistent vulnerability are somewhat different to those relating to transient vulnerability or becoming economically vulnerable.

Conclusions

In this paper we have sought to develop a multidimensional and dynamic approach to understanding the impact on families and children in Ireland of an unprecedented set of economic changes associated with the Great Recession.

The first and second waves of the Growing Up in Ireland survey spanned the period from the end of Ireland's economic boom through its entry into the Great Recession. Taking advantage of this timing we developed an approach to capturing multidimensional latent economic vulnerability of families, understood as involving a distinctive risk profile in relation to relative income, household joblessness and economic stress.

The recession was associated with a significant increase in levels of economic vulnerability and changing risk profiles for both the 2008 and 1998 cohorts. While the 2008 cohort displayed higher levels of persistent and transient vulnerability in the first wave, the transition into vulnerability in wave two was more evident for the 1998 cohort. This is consistent with the fact that the timing of the surveys meant that the first wave for the 1998 cohort took place before the start of the recession while the 2008 cohort was first approached after the recession had already begun. Economic stress levels increased for both vulnerable and non-vulnerable groups for both cohorts. The sharpest change in the vulnerability profile occurred for the 1998 cohort where the level of joblessness among the vulnerable decreased and they were significantly more widely distributed across the income distribution.

The factors affecting vulnerability outcomes were broadly similar for the 2008 and 1998 cohorts. Persistent economic vulnerability was significantly associated with lone parenthood, particularly for those with more than one child, lower levels of PCG education and, to lesser extent, younger age of PCG at child's birth, number of children and a parent leaving or dying. The vulnerability of lone parent and cohabiting families is consistent with other research in Ireland on the 1998 GUI cohort (Fahey et al., 2012; Hannan & Halpin, 2014). Results reported by Fahey et al. (2012) indicate that poverty and low levels of education are important in accounting for the lower wellbeing of children in one-parent families. Hannan and Halpin (2014), similarly, point to the significance of pre-existing differences, including socioeconomic differences between family types, in accounting for the disadvantage in health and self concept faced by children in lone parent or cohabiting families (Hannan & Halpin, 2014).

The impact of educational qualifications points to the importance of continuing to emphasise education and skills acquisition, particularly for those at risk of early school leaving. In the more immediate term, the needs of lone parents outside the labour market need to be addressed. Because lone parent families have only one care giver, the challenge of balancing employment and child care is likely to be more acute. It is well established that interventions to improve the labour market skills of the unemployed bring benefits in terms of employment opportunities and future wages. What is less well understood is the mix of training, job search support and child care support that is needed to enhance the labour market prospects of lone parents. Further research is needed on the optimal mix of services and support needed to enhance the labour market prospects of lone parents as well as improving outcomes for their children.

Similar factors were associated with transient vulnerability but the magnitude of the effects was significantly weaker, particularly in relation to lone parenthood and level of education of the PCG. For entry into vulnerability the impact of these factors was again substantially weaker than for persistent and transient vulnerability indicating a significantly greater degree of socioeconomic heterogeneity among the group that became vulnerable during the recession. As a result, in the post-recession period the economically vulnerable became considerably more heterogeneous in terms of family type and educational level of the PCG. This is consistent with other research on financial stress among the adult population in Ireland which showed significant evidence of 'middle class squeeze', particularly at the middle stages of the life course (Maître, Russell and Whelan 2014; Whelan & Maitre, 2014). The findings confirm the policy and political challenges presented by the scale of the Great Recession in Ireland which go well beyond catering for groups that traditionally have been characterised by a high dependence on social welfare. The substantially increased scale of economic vulnerability and the changing vulnerability profile in which subjective economic stress plays an increased role is consistent with evidence of the dramatic increase in household indebtedness in Ireland (Russell et al., 2013). Dealing with the political pressures arising from the broad range of households who have been affected by unprecedented economic circumstances presents formidable challenges to a welfare state which has traditionally emphasised targeted provision of limited means-tested benefits for only the most economically vulnerable groups.

Acknowledgements

We would like to thank Richard Layte and two anonymous readers for constructive comments on an earlier version of this paper. Growing Up in Ireland data have been funded by the Government of Ireland through the Department of Children and Youth Affairs; and have been collected under the Statistics Act, 1993, of the Central Statistics Office. The project has been designed and implemented by the joint Economic and Social Research Institute – Trinity College Dublin Growing Up in Ireland Study Team.

References

Bradshaw,J., Hoelscher,P. & Richardson, D. (2006). *Comparing Child Well-Being in OECD Counties: Concepts and Methods.* Florence: UNICEF Innocenti Research Centre. Innocento Working Paper IWP-2006-03

Bradshaw, J. & Main, G. (2010). *PSE Measures Review Paper: Children's Deprivation Items* (Poverty and Social Exclusion in the UK: The 2011 Survey, Working Paper Series; No. 7). [s.l.] Poverty and Social Exclusion.

Bradshaw, J. & Richardson, D. (2009). An Index of Child Well-Being in Europe, *Child Indicators Research*, *2*,3, 319-51 http://dx.doi.org/10.1007/s12187-009-9037-7

Callan, T., Nolan, B., Keane, C., Savage, M & Walsh, J. R. (2013). Crisis, response and distributional impact: the case of Ireland. *IZA Journal of European Labour Studies*, 3,9, 1-17

Duncan, G.J., Magnuson, K., Kalil, A. & Ziol-Guest, K. (2012). The Importance of Early Childhood Poverty, *Social Indicators Research*, *108*, 87–98. http://dx.doi.org/10.1007/s11205-011-9867-9

Fahey, T., Keilthy, P. & Polek, E. (2012). *Family Relationships and Family Well-Being: A Study of the Families of Nine Year Olds in Ireland*. Dublin: Family Support Agency.

Grusky, D.B. & Weeden, K.A. (2007). Measuring Poverty: The Case for a Sociological Approach, in Kakawani, N. & Silber, J. (eds) *The Many Dimensions of Poverty.* Basingstoke, UK: Palgrave Macmillan.

Hannan, C., & Halpin, B. (2014). The Influence of Family Structure on Child Outcomes: Evidence for Ireland. *Economic and Social Review*, *45*, 1-24.

Heeringa, S. G., West, B. T. & Berglund, P. A. (2010). *Applied Survey Data Analysis,* 273-275. Boca Raton, FL: Chapman & Hall/CRC.

Hanappi, D., Bernardi, L. & Spini, D. (2015). Vulnerability as a heuristic concept for interdisciplinary research: assessing the thematic and methodological structure of empirical lifecourse studies, *Longitudinal and Life Course Studies*, *6*, 1, 59-97. http://dx.doi.org/10.14301/llcs.v6i1.302

IMF. (2003). *Vulnerability Indicators: A Factsheet.* Washington, DC: International Monetary Fund. http://www.imf.org/external/np/exr/facts/vul.htm

Lazarsfeld, P. F., (1950). The Logical and Mathematical Foundation of Latent Structure Analysis, in S. Stouffer, L. Guttman, E. Suchman, P. Lazarsfeld, S. Star, S. & J. Clausen. (Eds.), *Measurement and Prediction*, Princeton, NJ: Princeton University Press.

Maître, B., Russell, H. & Whelan, C. T. (2014) *Trends in Economic Stress and the Great Recession in Ireland: An Analysis of the CSO Survey on Income and Living Conditions (SILC)*. Dublin: Department of Social Protection.

McCutheoen, A. & Mills, C. (1998). Categorical Data Analysis; Log-Linear and Latent Class Models, in E. Scarborough & E. Tannenbauum (Eds.), *Research Strategies in the Social Sciences*, Oxford: Oxford University Press. http://dx.doi.org/10.1093/0198292376.003.0005

Moisio, P. (2004). A latent class application to the multidimensional measurement of poverty, *Quantity and Quality-International Journal of Methodology 38* (6), 703-17. http://dx.doi.org/10.1007/s11135-004-5940-7

Mood, C. & Jonsson, J. (2014). *Poverty and Welfare among children and their families 1968-2010,* Institute for Future Studies, Research Report 2014/2, Stockholm.

Murray, A., McCrory, C., Thornton, M., Williams, J., Quail, A., Swords, L., Doyle, E. & Harris, E. (2010). *Growing Up in Ireland, National Longitudinal Study of Children: Design, Instrumentation and Procedures for the Child Cohort*. Dublin: Department of Health and Children.

Nolan, B. & Whelan, C. T. (1996). *Resources, Deprivation and Poverty*, Oxford: Oxford University Press.

Nolan, B. & Whelan, C T. (2007). On the multidimensionality of poverty and social exclusion, In Micklewright, J., Jenkins, S. (Eds.), *Poverty and Inequality: New Directions*. Oxford University Press, Oxford.

Nolan, B. & Whelan, C T. (2010). Using Non-Monetary Deprivation Indicators to Analyse Poverty and Social Exclusion in Rich Counties: Lessons from Europe? *Journal of Policy Analysis and Management*, *29*,2, 305-323. http://dx.doi.org/10.1002/pam.20493

Russell, H., Whelan, C, T. & Maître, B (2013). *Economic Vulnerability and Severity of Debt Problems: An Analysis of the Irish EU-SILC 2008*

StataCorp (2013a). Stata: Release 13. *Statistical Software*. College Station, TX: StataCorp LP.

StataCorp (2013b). *Stata Survey Data Reference Manual, Release 13*. College Station, TX: StataCorp LP.

Thornton, M., Williams, J., McCrory, C., Murray, A. & Quail, A. (2013). *Growing Up in Ireland, National Longitudinal Study of Children: Design, Instrumentation and Procedures for the Infant Cohort at Wave 1 (9 months)*. Dublin: Department of Health and Children.

Tomlinson, Walker, R., & Williams, G. (2009), Measuring Poverty in Britain as a Multidimensional Concept, 1991 to 2003, *Journal of Social Policy, 37*, 4: 597-620

UNICEF (2007), *Child Poverty in Perspective: An overview of child well-being in rich countries. A comprehensive assessment of the lives and well-being of children and adolescents in the economically advanced nations*, Florence: UNICEF Innocenti Research Centre, Innocenti Report Card 7

UNICEF (2014), *Children of the Recession: the impact of the economic crisis on the the well-being of children,* UNICEF Innocenti Research Centre, Innocenti Report Card 12

UNITED NATIONS (2003). *Report on the World Social Situation: Social Vulnerability: Sources and Challenges*. New York: United Nations Department of Economic and Social Affairs

Watson, D., Maître, B. & Whelan, C.T. (2012), *Understanding Childhood Deprivation in Ireland*. Dublin: Department of Social Protection

Whelan, C.T & Maître, B. (2005) Vulnerability and Multiple Deprivation: Perspectives on Economic Exclusion in Europe: A Latent Class Analysis, *European Societies 7*(3),423–50

Whelan, C. T. & Maître, B (2010), Welfare Regime and Social Class Variation in Poverty and Economic Vulnerability, *Journal of European Social Policy, 20*, 4, 3, 316-332

Whelan, C. T**.** & Maître, B. (2014), The Great Recession and the Changing Distribution of Economic Vulnerability by Social Class: The Irish Case, *Journal of European Social Policy*, *24*, 5, 470-485. http://dx.doi.org/10.1177/0958928714545444

Whelan, K. (2010) Policy Lessons for Ireland's Latest Depression, *Economic and Social Review, 41* (2), 225–54.

World Bank. (2000). *World Development Report 2000-01*: *Attacking Poverty*, New York: Oxford University Press.

Endnotes

[1] The recent UNICEF (2014) report on child poverty calculated changes in poverty rates using a poverty line fixed at 60% of median income in 2008 and adjusting the 2012 line for inflation. Given the scale of the recession in Ireland, it is not surprising that it fares particularly badly on such 'peak to trough' indicators, which fail to take into account the substantial increase in household incomes during the boom, in contrast with the outcome employing conventional income poverty measures.

[2] For a discussion of design and reweighting procedures for both cohorts, see Murray et al, 2010 and Thornton et al, 2013.

[3] For a recent discussion of the use of vulnerability as a heuristic concept see Hanappi et al., (2014).

[4] The scale assigns a weight of 1 to the first adult in a household, a weight of 0.66 to each additional adult and of 0.33 to children. The equivalised household income is thus calculated as the total household net income divided by the number of equivalent adults in the household.

[5] Nevertheless, the median equivalised household incomes as measured in the GUI are very similar to the figures obtained in the SILC surveys for the comparable years and for families of children in the same age groups.

[6] Extending the analysis to allow for a larger number of latent classes would require increasing the number of categories in our indicators in order to have sufficient degrees of freedom to conduct the analysis and would lead to a significantly less parsimonious account of economic differentiation.

[7] The class sizes based on modal allocation are lower than the estimated sizes of the vulnerable classes, at 27% for the nine-month olds, 30% for the three-year olds, 19% for the nine-year olds and 30% for the 13-year olds. However, the pattern across cohorts and over time is very similar.

[8] This was accomplished using the '*svy'* routine in Stata (StataCorp, 2013b). The data are in 'wide' format, so there is one case per child. The 1998 cohort is clustered at school level. Clustering has implications for the standard errors and significance tests but not for the sizes of coefficients (Heeringa et al., 2010). There is no clustering in the 2008 cohort but adjusted standard errors are still required to take account of the weights.

Longitudinal and Life Course Studies 2015 Volume 6 Issue 3 Pp 245 – 263 ISSN 1757-9597

Time on housework and selection into and out of relationships in Australia: a multiprocess, multilevel approach

Michele Haynes Institute for Social Science Research, The University of Queensland
m.haynes@uq.edu.au
Janeen Baxter Institute for Social Science Research, The University of Queensland
Belinda Hewitt School of Social Science, The University of Queensland
Mark Western Institute for Social Science Research, The University of Queensland

(Received May 2014 Revised February 2015) http://dx.doi.org/10.14301/llcs.v6i3.303

Abstract

In this paper we investigate the impact of relationship transitions on domestic labour time using longitudinal data from eight waves of the Household, Income and Labour Dynamics in Australia (HILDA) survey. Although there is a growing body of literature on this topic, previous research has failed to adequately address selection issues relating to transitions in marital status and time on housework. A simultaneous-equations model is used to jointly examine the relationships between time on housework and marital status transitions to allow for correlation between unobserved partner and person characteristics that impact on each process. Our results show that women who transitioned from being single into marriage spend more time on housework than women who transitioned from single to cohabiting. Additionally, we find that women who spend more time on housework when single also spend more time on housework after cohabitation or marriage. But there is no evidence of selection of these women into marriage rather than cohabitation. We also found no evidence to support the hypothesis that women who do varying amounts of housework are more likely to select out of relationships. Overall we conclude that the unobserved factors influencing time spent on housework are not related to the unobserved factors influencing relationship transitions.

Keywords: multiprocess models; multilevel models; domestic labour; marital transitions; self-selection

Introduction

Previous research has shown that life course events, such as transitions into and out of relationships, are associated with large variations in women's time on domestic labour (Gupta, 1999; Baxter, Hewitt & Haynes 2008). What has not been examined, however, is whether these processes are interrelated, with factors that lead to variations in housework hours also influencing the likelihood of making certain relationship transitions. Research has shown that women who cohabit spend less time on domestic labour than women who are married (Shelton & John, 1993; South & Spitze, 1994; Baxter, 2005). Similarly, we know that women who separate from a union experience a decline in the amount of time spent on housework (Gupta 1999; Baxter, Hewitt & Haynes, 2008). What is unclear from previous research is whether these variations in housework time are due to relationship transitions, or other unobserved characteristics that are driving both relationship transitions and time spent on housework.

Less time on housework by cohabiting women may be due to the experience of cohabitation or the values, attitudes or preferences that influence both the amount of time women spend on housework and their decisions about union type. Similarly women who separate from a union may find that there is less housework to be done after separation. Alternatively the characteristics that lead women to spend less time on housework when partnered may also increase the likelihood of separation.

Disentangling cause from selection drives much social science research, for example the male marriage wage premium (Ginther & Zavodny, 2001; Dougherty, 2006), the relationship between health and unemployment (Lundin, Lundberg, Hallsten, Otosson & Hemmingsson, 2010) and the intergenerational reproduction of socioeconomic status and inequality (McLanahan & Perchesky, 2008). On the present topic, we know that women's time on housework is linked to union formation and dissolution, but we do not understand the underlying processes. The reasons for differences in housework time for women in different relationship states have been postulated rather than explicitly investigated.

The aims of this paper are twofold. First, we identify the joint influence of observed and unobserved factors on relationship transitions and housework time. Second, we examine if there is a selection effect from single into a marital or cohabiting union for women who spend more time on domestic work, or a selection effect from cohabiting or married to single for women who spend less time on housework. We use eight waves of panel data from the Household, Income, and Labour Dynamics in Australia (HILDA) survey to investigate the interrelationships between women's time on housework and relationship transitions.

We go beyond the single process model typical of most social science research to jointly model the processes of union formation and dissolution and housework hours using multiprocess, multilevel models (for examples of the application of these models to other life events see Steele, Kallis, Goldstein & Joshi, 2005; Steele, 2011). In these models observed factors are incorporated as independent variables in the equation for each process simultaneously, while unobserved person characteristics that impact on each process are permitted to co-vary. This emerging approach enables us to examine the extent to which relationship transitions and housework hours are linked and jointly determined by both observed and unobserved factors.

Time on Housework and Relationship Transitions

Women in cohabiting or marital partnerships spend more time on domestic labour than unpartnered women (Gupta, 1999; Baxter, Hewitt & Haynes, 2008; Baxter, Haynes & Hewitt, 2010). One possible explanation for this difference, according to a gender display approach, is that women in partnerships create and affirm their gender identity as partnered women by spending more time on certain female-defined housework tasks compared to single women. Female-defined housework tasks require considerable investments of time and energy on a regular basis, such as cooking, cleaning and laundry, compared to male-defined housework tasks such as lawn mowing or home repairs, which may be seasonal or irregular. Researchers have argued that married women are more likely to display gender by spending time on housework than women in other relationship states (Berk, 1985; Bittman, England, Sayer, Folbre and Matheson, 2003). The gender display approach has been recently subject to critique and re-evaluation (Sullivan, 2011), but nevertheless has been an extremely influential approach for housework research in many countries, including the United States, Europe and Australia, explaining why gender divisions of labour in households have proved so intractable (Treas & Drobnič 2010).

Other well-known explanations for women's time on housework have focused on household specialisation or bargaining often using measures of relative earnings to assess how couples make decisions about who should spend most time on unpaid housework tasks and who should specialise in paid work (Brines, 1994). Since women typically earn less than men they also typically devote a greater proportion of their time to unpaid work than men, either because they have less bargaining power or because of the perceived household utility of gender specialisation. Gupta's recent work (2006, 2007) has taken debates about earnings and housework time in a new direction by arguing that women's time on housework is related to their absolute rather than their relative earnings. This has led to a rethinking of the links between earnings and housework time. One possibility is that women use their earnings to purchase household help, thereby reducing their time on housework. Alternatively, higher earning women may have different orientations toward housework than women with lower earnings, with higher earning women less interested in doing housework, or with different preferences and beliefs about appropriate standards of tidiness and cleanliness. While Gupta's research has been supported by studies in the United States and some parts of Europe (Gupta, Evertsson, Grunow, Nermo & Sayer, 2010), recent work in an Australian context has found stronger support for relative rather than absolute earnings suggesting that institutional context may play a part in household bargaining over money and time (Baxter & Hewitt 2013).

Most previous research, with the exception of Gupta (2007) and Gupta and Ash (2009) has focused on understanding why women spend more time on housework than men rather than why partnered women spend more time on housework than single or separated women. Focusing specifically on within gender variation, shifts the focus to explaining how life course events may be related to housework patterns, and in turn lead to gender differences. It also enables consideration of how women respond to changing social contexts, in this case changes in their relationship status. As several studies have shown, time on housework is not static across the life course (Gupta 1999; Gershuny, Bittman & Brice 2005; Baxter, Hewitt & Haynes 2008). Women's and men's time on housework varies in relation to changing employment status, parenthood and marital status. In some cases, the change in housework time may be immediate, while in others it may undergo a process of "lagged adaptation" whereby men and women gradually change their behaviour in response to changing social context (Gershuny, Bittman & Brice 2005). In most cases, the evidence is clear that women's housework time changes much more than men's.

In the case of relationship transitions, changes in housework may be influenced by changes in beliefs, habits and norms about appropriate behaviours in relationships, but also changes in the household context. The transition to a partnership may be associated with other important changes in a woman's life course that occur at the same time and are related to time spent on housework. For example, the decision to cohabit or marry may be closely linked to decisions about having a child, buying a house or reducing time spent in employment. Research has shown that fertility decisions are linked to transitions between cohabitation and marriage (Steele, Kallis, Goldstein & Joshi, 2005) and the birth of a child, particularly a first child, leads to large increases in women's time on housework (Baxter, Hewitt & Haynes, 2008). These co-occurring events, perhaps reciprocally related to decisions about partnership transitions, may also lead to women spending longer amounts of time on housework.

At the same time, the characteristics that select some women into cohabiting rather than marital relationships may also influence how much time they spend on housework. Considerable research has examined the observable characteristics that lead to women's likelihood of marriage, such as education, earnings, employment characteristics and ethnicity (Oppenheimer, 1997). But there may also be unmeasured or unobservable characteristics relating to aspirations and decisions about having a child, buying a house or reducing time spent in employment that influence women's time on housework and the decision to cohabit or marry. For example, women who wish to have a large family may also prefer the more traditional and permanent union of marriage, and have plans to buy a large house to accommodate the family for a long period of time with aspirations to keep a clean and aesthetically pleasing home. On the other hand, women who intend to develop their career and do not wish to have children in the immediate future may choose to cohabit, preferring a union without long-term commitment and less focus on the home. These same fertility intentions, employment and housing decisions may also influence the time that women spend on housework with those wanting a large family, a long-term marital relationship and a larger house, more likely to spend larger amounts of time on housework both before and after union and family formation.

Additionally, the amount of housework that a woman performs when unmarried may be influenced by domestic standards, habits or orientations. These same unmeasured standards may mean that these women will be more likely to marry than those who do not have these same domestic standards or orientations. Similarly, women who are less domestically inclined may be more likely to separate from a partnership than their counterparts. In other words, the formation and dissolution of marital partnerships, and decisions about whether to cohabit or marry, may be influenced by measured covariates such as age, education, earnings and attitudes, as well as unobserved factors such as a woman's propensity to spend time on housework, or what might be termed her level of domestic proclivity[i].

In other words, unmeasured characteristics, such as housework proclivity, may be related to women's decisions about types of partnership formation. Women with less interest in domesticity and weaker housework proclivity may be more likely to form a cohabiting relationship than a marital relationship. Alternately, women with higher levels of housework proclivity may be more likely to remain in a marriage than those with weaker housework proclivity. If marriage is viewed as a more traditional and conventional form of partnering than cohabitation, it may be more attractive as a form of partnership for those with more traditional views about gender roles and those who have higher levels of domestic proclivity. These same orientations may also discourage women from leaving a marital relationship. Where such unmeasured variables jointly determine relationships and housework time, assumptions about the exogeneity of explanatory variables in standard single process models are violated and the estimation of causal effects is compromised (Cameron & Trivedi, 2005). For example, although fixed effects panel regression models automatically control for all time invariant unobservable characteristics, they do not provide an assessment of selection bias or the interrelationships among dynamic processes.

In previous research, domestic orientations are often assumed to be captured by responses to questions about gender role attitudes. However, these questions are typically framed in terms of whether men or women should do domestic work and the importance of task sharing between partners. In other words, such questions focus more on issues of gender equality within the household than orientations toward domestic work. Moreover, despite the assumption in much of the housework literature that all household tasks are unpleasant and to be avoided at all costs (Ahlander & Barr 1995), some women may experience housework as enjoyable and rewarding. While this is perhaps less plausible for certain activities, it is possible that activities such as cooking may be regarded as relaxing and pleasant and not too dissimilar to leisure activities. They may provide an outlet for aspects of creativity and imagination that are not possible in other areas, such as employment. If women enjoy doing specific kinds of household work it is likely that they will spend more time on it than women who do not, regardless of their relationship status. It is therefore important to capture unmeasured characteristics in models explaining time on household work.

In the current paper our concern is whether women with certain unmeasured characteristics are more likely to select into, or out of, relationships than others and to spend greater or lesser amounts of time on housework. We only examine women's transitions because life course events have a much greater effect on women's housework hours than men's (Baxter, Hewitt & Haynes, 2008), and focusing on within gender variations enables a stricter, within-person, examination of the relationship between life course events and time on housework than the usual approach of comparing the determinants of women and men's housework time. Women's housework hours vary considerably in relation to marital, parental and employment status with women generally increasing their hours in response to the formation of partnerships and the arrival of children, typically in combination with a reduction in paid work hours outside the home. Men's housework hours, on the other hand, tend to remain relatively stable regardless of marital, parental or employment status.

Previous research has only considered the effects of marital status transitions on housework time without considering the reciprocal influence of housework time on the type of marital status chosen.

This one-sided approach fails to acknowledge that women's aspirations for a family and a long-term relationship may be interrelated with their proclivity to do housework. Although we may expect that women's time on housework increases by differing amounts with the transition into a cohabiting relationship or marriage, it is also possible that women who are predisposed to doing more housework are more likely to marry than to cohabit outside marriage. However, if women who do more housework typically select into marriage rather than a cohabiting relationship, this could lead to biased estimates of the effects of marital transitions on housework time. We consider time spent on housework and transitions between different types of relationship status as multiple related dynamic processes. In this approach, the same unmeasured individual characteristics are assumed to affect each process. The unobserved variables are represented by individual-specific random effects in all models for housework time and marital transitions and these random effects are allowed to co-vary to adjust for selection bias in model parameter estimates.

Relationship transitions and fertility (Steele, et al. 2005, Steele, Kallis & Joshi, 2006; Upchurch, Lillard & Panis, 2002), and housing transitions and fertility (Kulu & Steele, 2013) have previously been analysed as related multistate processes by using a system of simultaneous equations with co-varying random effects. We take a similar approach in this paper by considering time spent on housework and relationship transitions as interrelated multistate processes and analyse eight waves of data from the HILDA survey. We address three key questions:

1. What is the effect of relationship transitions on women's time on housework?
2. Is there a selection effect of women who spend more time on domestic labour into partnerships compared to women who spend less time on domestic labour, and specifically into marriage rather than cohabitation?
3. Is there a selection effect of women who do less domestic labour out of partnerships, and specifically out of marriage compared to cohabitation?

Methods

Data and sample

Our data come from the first eight waves of the Household, Income and Labour Dynamics in Australia (HILDA) survey collected between 2001 and 2008. Wave one comprised 7,682 households and 13,969 individuals (Watson & Wooden, 2002a and 2002b). Households were selected using a multi-stage sampling approach, and a 66% response rate was achieved (Watson & Wooden 2002a). Within households, data were collected from each person aged over 15 years (where available) using face-to-face interviews and self-completed questionnaires, and a 92% response rate was achieved (Watson & Wooden 2002a). Waves two to eight had, respectively, response rates of 86.8%, 90.4%, 91.6%, 94.4%, 94.9%, 94.7% and 95.2% (Watson and Wooden, 2012).

Our analytic sample included all women of child-bearing age from 18-45 years. We excluded women who were widowed before and during the survey and those with missing values for housework hours, marital status and other variables measuring life course transitions, such as fertility. It was not possible to determine if a relationship transition had occurred for some relationship statuses immediately prior to wave one and therefore the final HILDA sample retained for analysis was taken from waves two to eight and comprised 3,392 women with a total of 18,376 person years and an average of 5.4 wave observations per person.

Measures

Dependent variable

HILDA collects a range of time use measures. The main outcome measure in this study is hours spent doing housework each week. Respondents were asked questions about the hours they spend in a typical week on housework (preparing meals, washing dishes, cleaning house, washing clothes). Similar measures have been used in many other housework studies (Gupta, 1999). The distribution of this variable was skewed, and was therefore logged for inclusion in the models.

Some research has relied on measures of the proportion of housework done by the respondent, or the housework share within couples operationalised as the percentage distribution or the ratio of women's housework hours to men's housework hours (Coltrane, 2000). These measures are particularly suited to research examining issues of gender equality within couples. Since our focus is on women across various relationships states, including single, and we are concerned with explaining how unmeasured characteristics, such as domestic proclivity, is associated with relationship transitions, housework hours is more appropriate in the current analyses.

Relationship transitions

Respondents were asked their current marital status at each wave, including married, cohabiting (living together but not legally married), separated, divorced, widowed and never married. We collapse marital status at each wave to three relationship states: Married, cohabiting, and single (including never married, separated and divorced) and identify eight transitions of interest between these relationship states: married – married; married – single; cohabiting – cohabiting; cohabiting – single; cohabiting – married; single – single; single – married; and single – cohabiting.

Control variables

We include two measures for children. The first is a categorical measure for the number of dependent children (defined as 18 and under) including 1 = no children, 2 = 1 child, 3 = 2 children, and 4 = 3 or more children. The second measure indicates whether the respondent had a birth between waves with 1 = no birth, and 2 = birth. We also include a range of controls that have been found to be associated with housework and relationship transitions. These include age of respondent in years and age squared, earnings, education (1 = attained bachelor degree or higher, 2 = other), employment status (1 = not employed, 2 = employed full time, 3 = employed part time) and duration in marital status (months married, cohabiting or single). We include a measure for gender role attitude in response to the statement: 'It is much better for everyone involved if the man earns the money and the woman takes care of the home and children'. Responses ranged from 1 = Strongly Disagree to 7 = Strongly Agree. This question was asked in wave one and wave five. In our models, we carried wave one values on this variable forward for waves two to four and carried wave five values forward for waves six to eight. All measures are time varying with descriptive statistics for each shown in table 1. We do not discuss the associations between these control variables and housework time from our multivariate models presented below, as our findings are similar to those reported in many other studies and these associations are not the focus of the current paper.

Analytic approach

To jointly examine the relationships between time on domestic labour and marital status transitions we use a simultaneous equations approach with correlated error variances between equations reflecting common unobserved variables. Using this approach the formation and dissolution of relationships is analysed as a multistate process that may be influenced by measured covariates and unobserved factors that measure a woman's propensity to spend time on housework, such as her degree of domesticity. For each woman in the sample data, we observe measures for housework hours and marital status on up to eight occasions and we also observe when a marital transition representing a relationship formation or dissolution occurs. In the models for marital status transitions, we include a term for lagged housework hours to investigate whether the observed measure of time spent on housework immediately prior to a transition is associated with the type of relationship transition. Because we have repeated observations on each woman, and a transition can occur more than once for a woman, then housework hours and transitions are considered to be nested within individuals and an approach using a multilevel model specification which maintains the time-ordering of the observed transitions is appropriate. In our sample data from HILDA, 1,349 relationship transitions are observed.

In an approach similar to that taken by Steele, et al. (2005, 2006) among others, we use a multiprocess multilevel model where a system of regression equations with random coefficients is estimated simultaneously. The data analysed by Steele et al. (2005, 2006) contain complete event histories for the formation and dissolution of adult de facto and marital partnerships as well as for other outcomes of interest such as the birth of a child which allows the specification of a multilevel multistate event history model to analyse duration until an event occurs. The HILDA survey does not collect a complete event history for de facto relationships and it is not easy to collect reliable retrospective data on housework hours. However we can compute the duration of the marital status at wave one and the duration of subsequent marital status events. With eight waves of data we therefore analyse the likelihood of a marital transition from wave one using multinomial models and including marital status duration as an independent variable in the model.

Our multiprocess model includes a linear mixed model for logged housework hours and different multinomial logit models with random intercepts for transitions into and out of partnerships (Pettitt, Tran, Haynes & Hay, 2006). The model component for logged housework hours (model 1) includes an indicator variable for each marital status and the indicator variables are interacted with each covariate in the model. Random coefficients are specified for the indicator variables. Three additional models are specified separately for transitions into a partnership (model 2: from single state to cohabiting or married), for transitions out of a cohabiting partnership (model 3: from cohabiting to married or single) and for transitions from marriage to single (model 4). We analyse the likelihood of a transition occurring at any point in time and hence models 2 and 3 are multinomial logit models with random intercepts and model 4 is a binary logit model with a random intercept. For models 2 to 4 the reference outcome is no transition. The random effects from models 1 to 4 are correlated and are drawn from a multivariate normal distribution with mean vector **zero** and variance-covariance matrix $\mathbf{\Sigma}$. The model specification and estimation procedure is described below.

Model 1: A linear mixed model for housework hours with random coefficients on indicator variables for single, cohabiting and married statuses.

$$\ln(y_{ti}) = b_0^k + b_1^k Age_{ti} + b_2^k Age^2{}_{ti} + b_3^k Earnings_{ti} + b_4^k Attitude_{ti} + b_5^k Birth_{ti} + b_6^k One\_child_{ti} + b_7^k Two\_children_{ti} + b_8^k More\_children_{ti} + b_9^k FTemp_{ti} + b_{10}^k PTemp_{ti} + b_{11}^k Duration_{ti} + b_{12}^k Trans1_{ti} + b_{13}^k Trans2_{ti} + \alpha_{1ki}$$

The variable Y with response y_{ti} is used to denote housework hours for woman *i* at wave number $t = 2,\ldots,8$. All covariates are interacted with the relationship state indicator variables I_k. The fixed regression constants and coefficients are denoted b_l^k where the superscript *k* denotes relationship status: s = single, c = cohabiting, m = married and $l = 0, 1,\ldots, 13$ is a variable specific number corresponding to each of the explanatory variables in the model. The superscript *k* for each of the variable coefficients indicates that the variable has been interacted with the indicator variable I_k for the corresponding relationship status producing an estimated regression coefficient for each relationship status. For the single status (k=1) the variable Trans1 represents the transition to cohabiting, and Trans2 represents the transition to marriage. For the cohabiting status (k=2) the variable Trans1 represents the transition to single and Trans2 represents the transition to marriage. For the married status (k=3) the variable Trans1 represents the transition to single and the coefficient $b_{13}^3 = 0$. The term α_{1ki} represents the individual-specific random intercept term associated with relationship status *k* in model 1 with mean zero and variance component σ_{1k}^2.

Model 2: A multinomial logit model for transitions out of the single state with no transition as the reference category and a random intercept for each transition.

$$\ln\left(\frac{\Pr(Z_{1ti}=p)}{\Pr(Z_{1ti}=0)}\right)=$$

$$g_0^p + g_1^p Age_{pti} + g_2^p Age_{pti}^2 + g_3^p Lag\_earnings_{pti} + g_4^p Degree_{pti} + g_5^p One\_child_{pti} +$$

$$g_6^p Two\_children_{pti} + g_7^p More\_children_{pti} + g_8^p Birth_{pti} + g_9^p Duration_{pti} +$$

$$g_{10}^p Lag\_HWHours_{pti} + \alpha_{2pi}$$

The variable Z_1 denotes the relationship status *p = 0, 1, 2* into which a transition is being made, where 0 = no transition, 1 = cohabiting, 2 = married. The fixed regression constants and coefficients are denoted g_l^p where p = 1, 2 and *l =0, 1,…, 10* is a variable specific number corresponding to each of the explanatory variables in the model. The term α_{2pi}represents the individual-specifc random intercept term associated with model 2 and the transition to relationship status p with mean zero and variance component σ_{2p}^2 Models 3 and 4 are specified similarly to model 2. Model 3 is a multinomial logit model for transitions out of cohabitation and model 4 is a binary logit model for transition out of marriage to separation with no transition as the reference category.

Model Estimation

For the multinomial logit models we allow the random effects across the two transition states for each of these models to co-vary. Non-zero correlations among random effects across the models may occur if the unobserved characteristics that influence a woman to do more housework in any of the relationship states also influence the decision to form or dissolve a partnership. Also, if a woman experiences several transitions across the eight waves it is possible that the propensity to undergo one type of transition may also influence the likelihood of her undergoing another transition. Therefore, all eight random effects from models 1 to 4 are specified to arise from a multivariate normal distribution with mean vector **zero** and variance-covariance matrix $\boldsymbol{\Sigma}$. That is, using the notation above for models 1 to 4, we define the simultaneous nature of the estimation process through the matrix of correlated random effects specified as

$$\boldsymbol{\alpha} = (\boldsymbol{\alpha}_{11}, \boldsymbol{\alpha}_{12}, \boldsymbol{\alpha}_{13}, \boldsymbol{\alpha}_{21}, \boldsymbol{\alpha}_{22}, \boldsymbol{\alpha}_{31}, \boldsymbol{\alpha}_{32}, \boldsymbol{\alpha}_{41}) \sim N(\mathbf{0}, \boldsymbol{\Sigma}).$$

The system of equations specified in models 1 to 4 form the multiprocess multilevel model. The parameters in each of the equations are estimated simultaneously using Markov chain Monte Carlo (MCMC) simulation (Gelman, et al., 2005) which is implemented using the freely available WinBUGS software (Spiegelhalter, et al., 1998). Non-informative normal prior distributions were specified for each of the regression parameters. A Wishart prior distribution (dimension eight) was specified for inverse $\boldsymbol{\Sigma}$. Similar methods have been used to estimate multinomial logit models with random effects for estimating the probability of employment for immigrants to Australia with time since arrival (Pettitt, et al., 2006) and the probability of employment for Australian women (Haynes, Western, Yu & Spallek, 2008).

Results

The results from the estimation of the multiprocess model defined by models 1 to 4 are shown in tables 2 to 5. All results are means of posterior distributions obtained from 46,000 MCMC simulations following a burn-in length of 4,000 simulations. Table 2 shows the estimated regression coefficients for logged housework hours from model 1. Table 3 shows the estimated regression effects for the log odds of partnership formation and table 4 shows the estimated regression effects for the log odds of partnership dissolution. Table 5 is a summary table that shows the estimated variance-covariance matrix for the eight random effects from models 1 to

4. We discuss the results for each of our research questions in turn.

What is the effect of a relationship status transition on women's time on housework?

Model 1 addresses this question in two ways. First, the positive correlations ($\rho > 0.60$) among the random effects for time spent on housework in each relationship status (shown in the top left 3 rows and columns of table 5) suggest that women with a propensity to high (low) levels of housework or domesticity when single also have a propensity to spend more (less) time on housework relative to other women, when in a partnership. Second, the regression coefficients for estimating the effect of a relationship transition on housework time (table 2) show that the formation of a relationship is associated with a significant increase in housework hours and that the effect of a transition into marriage (b=0.333, SE=0.064) is twice as high as the effect of a transition into cohabitation (b=0.169, SE=0.033). Thus, women who enter a marital union spend more hours on housework than women who enter a cohabiting union. Table 2 also shows that separation from a cohabiting relationship is associated with a significant reduction in housework hours (b= -0.104, SE=0.047).

Together, these findings suggest that women's housework hours increase with a transition into a relationship, and also if a woman spends more than average time on housework hours when she is single then she will also tend to spend more than average time on housework when she is in a cohabiting or marital relationship. This result suggests that some marital transitions do influence the change in time spent on housework but that the total amount of housework undertaken following a transition is to an extent influenced by the propensity of a woman to spend more or less than average time on housework before the transition takes place.

Is there a selection effect of women who spend more time on domestic labour into marital rather than cohabiting partnerships?

A selection effect of this type can be assessed by examining the effects of observed housework time on relationship transitions as well as the correlations among unobserved variables, measured by the random effects in model 1 and the random effects in models 2 to 4. If the random effects for time spent on housework are positively correlated with the random effects for a relationship transition then this result signals the presence of a selection effect on the relationship status, with regards to housework time. Tables 3 and 4 show that there is no significant association between housework time prior to a relationship transition and the type of relationship transition experienced. Further, table 5 shows that the correlations between the random effects from model 1 and the random effects from models 2-4 are not significantly different from zero. Hence, there is no supporting evidence that women who spend more time on domestic labour are more likely to select into marriage rather than cohabitation, from either the observable or unobservable variables related to housework time.

Is there a selection effect of women who do less domestic labour out of partnerships?

The correlations of random effects presented in table 5 show that there are no significant associations among the random effects for time spent on housework and transitions out of partnerships. Therefore there is no evidence of a selection effect of this type in our sample. Thus women who spend less time on housework in a relationship are *not* more likely to separate than women who spend more time on housework.

However, significant correlations occur among the random effects for the transition from single into cohabitation and the transition from cohabitation to marriage, the transition from cohabitation to separation and the transition from marriage to separation. While it is not surprising that women who form a cohabiting relationship will either separate or go on to marry, it is noteworthy that the unobserved characteristics of women who cohabit are positively correlated with the unobserved characteristics of women who separate from marriage ($\rho = 0.612$). This suggests that women with unobserved characteristics that are more likely to encourage them to enter a cohabiting relationship before marriage, are also more likely to encourage them to separate from marriage compared to those who marry directly. This supports previous research that has found a link between premarital cohabitation and marital separation (Dush, Cohan & Amato, 2003).

Discussion

Many life course events are interrelated. Indeed one hallmark of a life course approach is the recognition that trajectories and transitions in different life domains are interconnected rather than distinct (Han & Moen, 1999). Despite this, most previous research, including longitudinal research on the effects of marital transitions on housework time, treats this association as a single-state process that is independent of events occurring in other life domains. This can lead to selection and unobserved heterogeneity biases that call into question research findings.

We have investigated whether observed and unobserved characteristics that influence women's decisions about whether to cohabit, marry or separate, also influence their time spent on housework. This approach allows us to more accurately assess the mechanisms underpinning change in women's housework time when they transition into and out of relationships. To our knowledge this is the first paper to use multilevel, multiprocess models to examine these associations.

Overall our results suggest that movement into a relationship increases women's time on domestic labour. We find that women who transition from being single into marriage spend about twice as much time on housework as women who transition from single to cohabiting. There was no significant change in housework hours for women who transitioned from cohabiting to married. This is consistent with previous research that finds that married women do more housework than cohabiting and single women (Baxter, 2005; Shelton & John, 1993; South & Spitze, 1994). A key explanation given for this gap in housework hours between women who are cohabiting compared to those who are married is that the types of women who choose different relationship arrangements differ in their values, norms and domestic orientations (Baxter, Haynes, & Hewitt, 2010). Thus women with different preferences select into either marriage or cohabitation.

In this paper we have directly investigated this assumption examining whether the increase in housework hours is due to the characteristics of the kinds of women who partner or the result of circumstances arising after partnership. This enables us to address questions of selection and effect that have not been adequately addressed in previous studies. Moreover we were able to investigate whether unobserved characteristics relating to women's time on housework determine what kind of partnership, cohabiting or married, they enter, and whether these unobserved characteristics are related to women's decisions to leave a partnership.

Our results show that women who spend more time on housework when single, also spend more time on housework after cohabitation or marriage. But there is no evidence of selection of these women into partnerships. Women who do more housework when single are not more likely to form a partnership, and neither are they more likely to choose to marry than cohabit, compared to women who do less housework when single. We also found no evidence to support the hypothesis that women who do varying amounts of housework are more likely to separate from cohabitation or marriage. This suggests that some women are more likely to do housework irrespective of the type of relationship they are in, or transitions into and out of relationships.

Overall we conclude that the unobserved factors influencing time spent on housework are not related to the unobserved factors influencing relationship transitions. This is an important finding because it suggests that the increase in housework time experienced by women when they partner is the result of the transition into a partnership, rather than reflecting characteristics of women who partner. Eliminating selection into partnership as an explanation allows researchers to focus on the transition to partnership itself as the cause of increased housework time, which potentially yields new research questions and possibilities. Indeed if we think of the transition into a partnership as being analogous to a social mobility trajectory from an origin state to a destination state, over 45 years ago, Duncan (1966) argued that in analysing the effects of such movements, one needed to distinguish the main effects of origin and destination on an outcome from the effect of mobility per se. Duncan's work stimulated the mobility effects literature into attitudes and behaviours (Sobel, 1981) which continue today (e.g., Tolsma, de Graaf & Quillian, 2009). In this context, our research suggests a number of potential hypotheses that could be examined in the future. If the mobility process is important, rather than the destination state (partnership) for instance, different trajectories into partnership should lead to different outcomes on housework. On the other hand, if it is partnership per se that is important, the trajectory should be unimportant as long as the destination state is the same. Similarly, if there is a mobility effect on housework time, and the mobility mechanism can be directly measured (for example normative understandings of housework and gender roles that reflect origin and destination characteristics), the change in housework hours associated with a transition should not be equal to the additive effects of the mechanism in origin and destination.

In contrast to our findings for the effects of selection/unobserved heterogeneity on relationships and housework, we do find that the unobserved characteristics influencing women to cohabit prior to marriage also influence their likelihood of separation from marriage. Previous research has shown that couples are more likely to separate if they have cohabited prior to marriage compared to those who have married directly (Dush, Cohan & Amato, 2003). Our results support this finding by showing that women who select into cohabitation prior to marrying are also more likely to select out of marriage. Again, the implication of this result is that some characteristics of women (e.g. values, attitudes, preferences) are associated with cohabitation and separation and these characteristics should be incorporated into future theorising and empirical research.

More broadly, we have also endeavoured to show how multiprocess multilevel models for jointly co-occurring social events can be useful for life course analyses, and social science more generally. A great deal of basic and applied empirical work in the social sciences is motivated by the need to address questions of cause and effect (Morgan and Winship, 2007) and some have even argued that science is defined by the ability to formally specify and examine proper causal statements in theory, hypothetical populations and with real data (Heckman, 2005). With non-experimental data, unobserved heterogeneity, selection and other related issues linked to sampling variability, compromise our ability to identify causal effects, even if we can properly specify causal statements in theory and for hypothetical populations. Multiprocess multilevel models provide yet another approach to addressing some of these issues.

Acknowledgements

This paper uses unit record data from the Household, Income and Labour Dynamics in Australia (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

Ahlander, Nancy Rollins & Barr, Kathleen Slaugh. (1995). Beyond Drudgery, Power, and Equity: Toward an Expanded Discourse on the Moral Dimensions of Housework in Families. *Journal of Marriage and the Family, 57* (1), 54-68. http://dx.doi.org/10.2307/353816

Baxter, J. (2005). To Marry or Not to Marry: Marital Status and the Household Division of Labor. *Journal of Family Issues, 26* (3), 300-321.

http://dx.doi.org/10.1177/0192513X04270473

Baxter, J., Haynes, M. & Hewitt, B. (2010). Pathways into Marriage. Life course Patterns and the Domestic Division of Labor. *Journal of Family Issues, 31*, 1481-1506. http://dx.doi.org/10.1177/0192513X10365817

Baxter, J. & Hewitt, B. (2013) Economic independence or bargaining power? The relationship between women's earnings and housework time. *Feminist Economics, 19* (1), 29-53.

Baxter, J., Hewitt, B. & Haynes, M. (2008). Life Course Transitions and Housework: Marriage, Parenthood and Time on Housework. *Journal of Marriage and the Family, 70,* 259-272. http://dx.doi.org/10.1111/j.1741-3737.2008.00479.x

Berk, S. Fenstermaker (1985). *The Gender Factory*. New York: Plenum. http://dx.doi.org/10.1007/978-1-4613-2393-8

Bittman, M., England, P., Sayer, L., Folbre, N. and Matheson, G. (2003). When Does Gender Trump Money? Bargaining and Time in Household Work. *American Journal of Sociology, 109*, 186-214. http://dx.doi.org/10.1086/378341

Brines, J. (1994). Economic Dependency, Gender, and the Division of Labor at Home*. American Journal of Sociology, 100* (3), 652-688. http://dx.doi.org/10.1086/230577

Cameron, A., & Trivedi, P. (2005). *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511811241

Coltrane, Scott. (2000). Research on Household Labor: Modeling and Measuring the Social Embeddedness of Routine Family Work. *Journal of Marriage and the Family 62* (4): 1208 – 1233. http://dx.doi.org/10.1111/j.1741-3737.2000.01208.x

Dougherty, C. (2006). The Marriage Earnings Premium as a Distributed Fixed Effect**.** *Journal of Human Resources 2006, 41*, 433-443.

Duncan, Otis D. (1966). Methodological issues in the analysis of social mobility. Pp. 51-97 in Neil J. Smelser and Seymour Martin Lipset (eds.), *Social Structure and Mobility in Economic Development.* Chicago: Aldine

Dush, C., Cohan, C., & Amato, P. (2003). The Relationship between Cohabitation and Marital Quality and Stability: Change across Cohorts? *Journal of Marriage and Family, 65*, 539-549. http://dx.doi.org/10.1111/j.1741-3737.2003.00539.x

Gelman, A., Carlin, J., Stern, H. and Rubin, D. (2005). *Bayesian Data Analysis. Third Edition.* Chapman and Hall.

Gershuny, J., Bittman, M., Brice, J. (2005). Exit, Voice and Suffering: Do Couples Adapt to Changing Employment Patterns? *Journal of Marriage and Family, 67* (3): 656-665. http://dx.doi.org/10.1111/j.1741-3737.2005.00160.x

Ginther, D.K., and Zavodny, M. (2001). Is the Male Marriage Premium Due to Selection? The Effect of Shotgun Weddings on the Return to Marriage. *Journal of Population Economics*, *14*, 313-328. http://dx.doi.org/10.1007/s001480000058

Gupta, S. (1999). The Effects of Transitions in Marital Status on Men's Performance of Housework. *Journal of Marriage and the Family, 61*, 700-711. http://dx.doi.org/10.2307/353571

Gupta, S. (2006). Her Money her Time: Women's Earnings and their Housework Hours. *Social Science Research, 35,* 975-999. http://dx.doi.org/10.1016/j.ssresearch.2005.07.003

Gupta, S. (2007). Autonomy, Dependence, or Display? The Relationship between Married Women's Earnings and Housework. *Journal of Marriage and Family, 69,* 399-417. http://dx.doi.org/10.1111/j.1741-3737.2007.00373.x

Gupta, S. and Ash, M. (2009). Whose money, whose time? A nonparametric approach to modeling time spent on housework in the United States. *Feminist Economics, 14* (1), 93-120. http://dx.doi.org/10.1080/13545700701716664

Gupta, S., Evertsson, M., Grunow, D., Nermo, M., and Sayer, L. (2010). Economic Inequality and Housework. In Treas, J., and Drobnič. (Eds). *Dividing the Domestic. Men, Women, and Household Work in Cross-National Perspective* (105-122). Stanford: Stanford University Press. http://dx.doi.org/10.11126/stanford/9780804763578.003.0006

Hakim, C. (2000). *Work, Life-Style Choices in the Twenty First Century*: *Preference Theory*. Oxford: Oxford University Press.

Han, S-K., & Moen, P. (1999). Work and Family Over Time: A Life Course Approach. *Annals of the American Academy of Political and Social Science*. *562*, 98-110. http://dx.doi.org/10.1177/0002716299562001007

Haynes, M. Western, M., Yu, L. and Spallek, M. (2008). Analysing Nominal Data from a Panel Survey: Employment Transitions of Australian Women. 103rd Annual Meeting of the American Sociological Association, 1-4 August, 2008, Boston, USA.

Heckman, J. (2005). The Scientific Model of Causality. *Sociological Methodology*, *35*, 1-97. http://dx.doi.org/10.1111/j.0081-1750.2006.00163.x

Kulu, H. and Steele, F. (2013). Interrelationships between childbearing and housing transitions in the family life course. *Demography*, *50*, 1687-1714. http://dx.doi.org/10.1007/s13524-013-0216-2

Lundin, A., Lundberg, I., Hallsten, L., Otosson, J., & Hemmingsson, T. (2010). Unemployment and Mortality – A Longitudinal Prospective Study on Selection and Causation in 49321 Swedish Middle-Aged Men. *Journal of Epidemiology and Community Health*, *64*, 22-28. http://dx.doi.org/10.1136/jech.2008.079269

McLanahan, S., & Perchesky, C., (2008). Family Structure and the Reproduction of Inequalities. *Annual Review of Sociology*, *34*, 257-276. http://dx.doi.org/10.1146/annurev.soc.34.040507.134549

Morgan, S.,and Winship, C. (2007). *Counterfactuals and Causal Inference*: *Methods and Principles for Social Research*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511804564

Oppenheimer, V. (1997). Women's Employment and the Gain to Marriage: The Specialization and Trading Model. *Annual Review of Sociology*, *23*, 431-453. http://dx.doi.org/10.1146/annurev.soc.23.1.431

Pettitt, A., Tran, T., Haynes, M., & Hay, J. (2006). A Bayesian Hierarchical Model for Categorical Longitudinal Data from a Social Survey of Immigrants. *Journal of the Royal Statistical Society - Series A, 169* (1), 97-114. http://dx.doi.org/10.1111/j.1467-985X.2005.00389.x

Shelton, B. & John, D. (1993). Does Marital Status Make a Difference? Housework among Married and Cohabiting Men and Women. *Journal of Family Issues*, *14* (3), 401-420. http://dx.doi.org/10.1177/019251393014003004

Sobel, M. (1981). Diagonal Mobility Models: A Substantively Motivated Class of Designs for the Analysis of Mobility Effects. *American Journal of Sociology*, *46*, 893-906. http://dx.doi.org/10.2307/2095086

South, S. J. & Spitze, G. (1994). Housework in Marital and Nonmarital Households. *American Sociological Review, 59* (3), 327-347. http://dx.doi.org/10.2307/2095937

Spiegelhalter, D., Thomas, A. & Best, N. (1998). WinBUGS: Bayesian Inference using Gibbs Sampling, Manual version 1.2, Imperial College, London and Medical Research Council Biostatistics Unit, Cambridge.

Steele, F., Kallis, C., Goldstein, H. & Joshi, H. (2005). The relationship between childbearing and transitions from marriage and cohabitation in Britain. *Demography*, *42* (4), 647-673. http://dx.doi.org/10.1353/dem.2005.0038

Steele, F., Kallis, C. & Joshi, H. (2006). The formation and outcomes of cohabiting and marital partnerships in early adulthood: the role of previous partnership experience. *Journal of the Royal Statistical Society – Series A*, *169* (4), 757-779. http://dx.doi.org/10.1111/j.1467-985X.2006.00420.x

Steele, F. (2011). Multilevel Discrete-Time Event History Models with Applications to the Analysis of Recurrent Employment Transitions. *Australian and New Zealand Journal of Statistics*, *53* (1), 1-26. http://dx.doi.org/10.1111/j.1467-842X.2011.00604.x

Sullivan, O. (2011). An End to Gender Display through the Performance of Housework? A Review and Reassessment of the Quantitative Literature using Insights from the Qualitative Literature. *Journal of Family Theory and Review, 3,* 1-13. http://dx.doi.org/10.1111/j.1756-2589.2010.00074.x

Tolsma, J., de Graaf, N. & Quillian, L., (2009). Does Intergenerational Social Mobility affect Antagonistic Attitudes towards Ethnic Minorities? *British Journal of Sociology*, *60*, 257-277. http://dx.doi.org/10.1111/j.1468-4446.2009.01230.x

Treas, J., and Drobnič. (Eds) (2010). *Dividing the Domestic. Men, Women, and Household Work in Cross-National Perspective.* Stanford: Stanford University Press. http://dx.doi.org/10.11126/stanford/9780804763578.001.0001

Upchurch, D., Lillard, L. and Panis, C. (2002). Nonmarital childbearing: Influences of education, marriage and fertility. *Demography*, *39*, 311-329. http://dx.doi.org/10.1353/dem.2002.0020

Watson, N. & Wooden, M. (2002a). The Household, Income and Labour Dynamics in Australia (HILDA) Survey. HILDA Project Technical Paper Series. No, 1/02, May 2002.

Watson, N. & Wooden, M. (2002b). The Household, Income and Labour Dynamics in Australia (HILDA) Survey: Wave 1 survey methodology. Melbourne: The University of Melbourne.

Watson, N. and 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.

Endnotes

[1] This argument is similar to Hakim's (2000) view that some women may have home-centred preferences.

Table 1: Means, standard deviations, proportions of housework hours and model covariates for pooled sample of women aged 18-45 years, HILDA waves 2-8[a].

	HILDA (2 – 8)	
Model Variables	Mean[b]	SD[c]
Housework hours	17.81	13.3
Housework hours (logged)	2.60	0.8
Age	36.93	7.9
Earnings ($10,000)	2.80	2.10
Bachelor Degree or higher (1=yes)	30%	
Number of children <18		
None	38%	
1 child	12%	
2 children	27%	
3 + children	23%	
Employment status:		
Not employed	25%	
Employed Full time	39%	
Employed Part time	36%	
Relationship Status Duration (Months)		
Married	159	97
Cohabiting	34	24
Single	46	26
Transition variables		
Birth	5%	
Remain married	55.6%	
Remain cohabiting	11.1%	
Remain single	26.0%	
Single-Married	0.7%	
Single-Cohabiting	2.6%	
Married-Separated	1.2%	
Cohabiting-Separated	1.3%	
Cohabiting-Married	1.5%	
Woman-years	18,376	
Number of women	3,393	

[a] Analytic sample includes data from wave 2 as not all partnership transitions are available at wave 1
[b] Means reported for continuous measures and percentages (%) reported for categorical measures
[c] Standard deviations reported only for continuous measures.

Table 2: Estimated coefficients and standard errors from models of change in log housework hours for single, cohabiting and married women aged 18-45 years.[a]

Variable	Single State Model for Log Housework Hours		Cohabit State Model for Log Housework Hours		Married State Model for Log Housework Hours	
	Coeff	SE	Coeff	SE	Coeff	SE
Constant	2.198*	0.040	2.498*	0.065	2.749*	0.031
Age	0.014*	0.002	0.008*	0.003	0.005*	0.002
Age squared	-0.001*	0.0002	-0.0005	0.0003	-0.0003	0.0002
Earnings (log '0,000s)	0.105*	0.030	-0.024	0.037	-0.107*	0.017
Male breadwinner attitudes	0.012*	0.002	0.001	0.005	0.007*	0.002
Bachelor degree (1=yes)	0.055	0.032	-0.005	0.044	0.016	0.025
Birth of child						
No birth						
Birth	0.105	0.058	0.023	0.050	0.094*	0.024
Number of children<18						
None						
1 child	0.492*	0.044	0.335*	0.050	0.173*	0.028
2 children	0.465*	0.043	0.516*	0.050	0.300*	0.025
3+ children	0.535*	0.046	0.529*	0.061	0.363*	0.027
Employment						
Not employed						
Employed Full-time	-0.304*	0.031	-0.373*	0.044	-0.406*	0.022
Employed Part-time	-0.193*	0.028	-0.195*	0.039	-0.177*	0.018
Duration in current marital status	-0.0002	0.0003	-0.0004	0.0005	0.00009	0.0002
Trans S-C	0.169*	0.033				
Trans S-M	0.333*	0.064				
Trans C-S			-0.104*	0.047		
Trans C-M			0.033	0.045		
Trans M-S					-0.068	0.044

a. Estimated values are means of posterior distributions obtained from 46,000 MCMC simulations, following a burn-in of 4,000.
b. * indicates that the 95% credible interval for the estimated regression coefficient does not contain zero.

Table 3: Estimated coefficients and standard errors from models of log odds of partnership formation for women aged 18-45 years.[a]

Variable	Model for Transition Single-Married		Model for Transition Single-Cohabiting		Model for Transition Cohabiting-Married	
	Coeff	SE	Coeff	SE	Coeff	SE
Constant	-5.215*	0.831	-2.390*	0.169	-1.711*	0.353
Age	-0.064*	0.022	-0.068*	0.009	-0.057*	0.015
Age_squared	-0.006*	0.002	-0.004*	0.0009	-0.006*	0.002
Lag earnings (log ‘0,000s)	0.018	0.062	0.064*	0.022	0.081	0.071
Bachelor degree (1=yes)	0.703*	0.365	0.114	0.133	0.475*	0.221
Birth of child						
No birth						
First birth	0.237	0.767	1.793*	0.273	0.036	0.316
Number of children <18						
None						
1 child	-0.039	0.476	-0.232	0.199	-0.014	0.270
2 children	0.281	0.441	-0.533*	0.206	-0.827*	0.302
3+ children	0.346	0.473	-0.663*	0.238	-0.349	0.315
Duration in previous marital status	0.008*	0.005	0.002	0.002	0.006	0.005
Lag housework hours (log)	0.006	0.018	0.004	0.023	0.006	0.018

a. Estimated values are means of posterior distributions obtained from 46,000 MCMC simulations, following a burn-in of 4,000.
b. * indicates that the 95% credible interval for the estimated regression coefficient does not contain zero.

Table 4: Estimated coefficients and standard errors from models of log odds of partnership dissolution for women aged 18-45 years.[a]

Variable	Model for Transition Cohabiting-Single		Model for Transition Married-Single	
	Coeff	SE	Coeff	SE
Constant	-3.990*	0.407	-3.943*	0.258
Age	-0.010	0.014	-0.009	0.016
Age_squared	-0.002*	0.002	-0.001	0.002
Lag earnings (log '0,000s)	-0.206*	0.061	-0.047	0.076
Bachelor degree (1=yes)	-0.410	0.226	-0.581*	0.197
Birth of child				
No birth				
Birth	-0.993*	0.433	-1.157*	0.460
Number of children <18				
None				
1 child	0.503	0.255	-0.446	0.282
2 children	-0.342	0.279	-0.206	0.220
3+ children	-0.221	0.295	-0.347	0.239
Duration in previous marital status	-0.016*	0.005	-0.002*	0.001
Lag housework hours (log)	0.004	0.023	-0.006	0.030

a. Estimated values are means of posterior distributions obtained from 46,000 MCMC simulations, following a burn-in of 4,000.
b. * indicates that the 95% credible interval for the estimated regression coefficient does not contain zero.

Table 5: Estimated random-effects covariance matrix from the multi-process model (includes estimates of correlation in [ ]).

	Housework hours for single state	Housework hours for married	Housework hours for cohabiting	Likelihood of S-M transition	Likelihood of S-C transition	Likelihood of C-M transition	Likelihood of C-S transition	Likelihood of M-S transition
Housework hours for single state	0.263*							
Housework hours for married	0.135* [0.605]	0.189*						
Housework hours for cohabiting	0.172* [0.696]	0.126* [0.601]	0.232*					
Likelihood of S-M transition	0.027 [0.021]	-0.096 [-0.089]	0.096 [0.081]	6.119*				
Likelihood of S-C transition	0.069 [0.133]	-0.008 [-0.018]	0.023 [0.047]	0.654 [0.262]	1.019*			
Likelihood of C-M transition	0.018 [0.025]	-0.031 [-0.051]	0.046 [0.068]	2.110 [0.610]	0.339* [0.240]	1.954		
Likelihood of C-S transition	0.027 [0.047]	0.018 [0.037]	-0.031 [0.057]	-0.085 [-0.031]	0.479* [0.424]	-0.266 [0.170]	1.253	
Likelihood of M-S transition	0.077 [0.122]	-0.011 [0.020]	0.044 [0.074]	0.952* [0.312]	0.762* [0.612]	0.325 [0.188]	0.426 [0.309]	1.522*

Work-family conflict as a predictor of common mental disorders in the 1958 British birth cohort

Tahera Razavi Department of Psychology, School of Biological and Chemical Sciences, Queen Mary, University of London, UK

Charlotte Clark Centre for Psychiatry, Wolfson Institute of Preventive Medicine, Barts & The London School of Medicine & Dentistry, Queen Mary University of London, UK

Stephen A Stansfeld Centre for Psychiatry, Wolfson Institute of Preventive Medicine, Barts & The London School of Medicine & Dentistry, Queen Mary University of London, UK

s.a.stansfeld@qmul.ac.uk

(Received November 2014 Revised May 2015) http://dx.doi.org/10.14301/llcs.v6i3.327

Abstract

The impact of work-family conflict on common mental disorders (CMD) has been examined in cross-sectional studies. The current paper examines work-family conflict and its effect on CMD in a large nationally representative longitudinal sample. This study uses data from the 1958 British birth cohort study, a longitudinal, prospective cohort study of men and women born in a single week in 1958. At 45 years 9,297 individuals were followed up and 9,008 individuals were working. In this sample work-family conflict, sociodemographic factors and the number of hours worked were assessed at age 42 years. The Revised Clinical Interview Schedule (CIS-R) was used to assess CMD, as classified by the International Statistical Classification of Diseases, 10th Revision (ICD-10), in cohort members at age 45 years. Work-family conflict was prospectively associated with an increased risk of common mental disorders (OR=1.76 95% CI 1.36-2.20) adjusting for gender, marital status, social class and educational qualifications. However there was no significant prospective association between the number of hours worked and the prevalence of CMD in this cohort. These results suggest that work-family conflict is a risk factor for future common mental disorder and that in order to prevent common mental disorder this should be considered in job planning. There is a need for more prospective studies with more detailed measures of work-family conflict to confirm these results.

Keywords: work; work family conflict; working hours; mental health; cohort studies; depression; longitudinal studies

Introduction

The balance between paid work and the life outside work has become an important focus in Britain (Lewis, 2003; Taylor, 2002; Wheatley, 2012). This issue has arisen due to the increase in non-traditional, dual-career households and single-parent households who do not confine work and life commitments to traditional gender roles (Byron, 2005).

The complex balance between work and life requires developing equilibrium between both activities and roles outside work or leisure activities and paid employment (Guest, 2002; Lowry & Moskos, 2008). The interference between the fulfilment of each of these roles is considered to generate work-life conflict (Reynolds, 2005). This conflict is considered to have direction, signifying that work activities can impact on life activities and vice versa - in this direction it is often termed family-work conflict (Greenhaus & Parasuraman, 1999; Frone, Russell, & Cooper, 1992).

Work and family life can both influence the prevalence of mental health conditions. Work-life conflict and family-work conflict have been cross-sectionally associated with mood, anxiety and substance use disorders (Frone, 2000; Grzywacz & Bass, 2003; Seto, Morimoto, & Maruyama, 2004; Wang, Lesage, Schmitz, & Drapeau, 2006; Obidoa, Reeves, Warren, Reisine, & Cherniack, 2011; Nitzche, Jung, Pfaff, & Driller, 2012). These studies suggest that work-life imbalance can be considered to be an aspect of the multifactorial aetiology of common mental disorders (Kendler, Prescott, Myers, & Neale, 2003; Hämmig & Bauer, 2009). Common mental disorders, most frequently, depressive and anxiety disorders are the mental health conditions most likely to arise from psychosocial stressors.

Several structural factors may influence work-life balance such as gender, marital status and socioeconomic status (Barnett & Hyde, 2001; Cinamon & Rich, 2002; Duxbury, Higgins & Lee, 1994). It may be more difficult for women to attain a balance between work and life, because they often take the largest share of domestic duties and childcare (Frone, 2000; Canivet et al., 2010). However Byron (2005) found that gender alone was a poor predictor of work-family conflict. Widowed and divorced males and females have higher incidence rates of mental disorders compared to single and married males and females (Bebbington, 1987). Duxbury, Higgins and Lee (1994) found that single parents faced similar work-family conflict as those who were married. Similarly, Byron (2005) concluded that marital status was not a good predictor of work-family conflict.

Socioeconomic status is often associated with poor health (Adler & Ostrove, 1999; Hemingway, Nicholson, Stafford, Roberts, & Marmot, 1997). Lorant et al., (2003) found that individuals in the lowest socioeconomic group were 2.5 times more likely to have depressive disorder than those in the highest socioeconomic group. Clark et al., (2011) suggest that financial crises such as being in debt are associated with prevalence of common mental disorders (CMD). Social disadvantage may mean that people are unable to afford the resources that could buffer the strain involved in carrying out work and non-work roles simultaneously.

Time balance and allocation is also considered to influence the difficulties faced in attaining stability between work and life outside work. Spurgeon, Harrington & Cooper (1997) concluded that long working hours, specifically working hours greater than fifty hours per week were associated with risks to mental health. Virtanen et al., (2011) found that there was a significantly greater risk of depressive and anxiety symptoms among those working over fifty-five hours per week as compared to those working between thirty-five to forty hours per week. Part of this increased risk may be due to working irregular hours as working rotating shifts and having irregular, changing hours has been related to greater risk of common mental disorders than working regular schedules (Martens, Nijhuis, Van Boxtel, & Knottnerus, 1999).

However, although the impact of life activities on work activities and vice versa may conflict with one another they can also have a positive or complementary effect on each other (Greenhaus & Parasuraman, 1999). Equally Lewis (2003) advocates that people may enjoy working and gain a sense of satisfaction from it, and suggests that there are individual differences in the balance between work and life.

In summary, the aforementioned studies suggest that there is an association between work-life conflict and prevalence of mental ill health. However as these findings rely upon cross-sectional methodology they are insufficient as evidence to assess causal relationships (Casper, Eby, Bordeaux, Lockwood, & Lambert, 2007; Hämmig, Gutzwiller& Bauer, 2009).

Also, prior research regarding work-life conflict has often limited the research population to specific categories or professions (Hämmig et al., 2009) and many studies are not nationally representative as they are limited to subpopulations (Seto et al., 2004; Shimazu, Bakker, Demerouti & Peeters, 2010; Takeuchi & Yamazaki, 2010; Janzen & Kelly, 2012; Grant-Vallone & Ensher, 2001).

In the present study, we use a nationally representative sample, the 1958 British birth cohort of people born in England, Scotland and Wales in one week in 1958 to examine the association of work-family conflict and working hours on common mental disorders.

The aims of this study are:

1. To assess whether work-life conflict is associated with common mental disorders three years later.
2. To assess whether the number of hours worked is associated with common mental disorders three years later.

A relatively short three year follow up allows for immediate effects of work-life conflict expected on common mental disorders to be captured and is brief enough for work and family responsibilities not to have changed much in the interval.

Method

Sample

The 1958 British birth cohort originated as a study into perinatal mortality assessing over 17,000 births in England, Wales and Scotland, in a single week in March 1958. Data were attained on 17,414 (98.2%) of the birth of the total 17,733 births in the first survey (Parsons, Power, & Manor, 2001). Nine further follow-up surveys have been conducted when the cohort members were aged seven, 11, 16, 23, 33, 42, 45, 50 and 55 years (Power & Elliott, 2006) collecting data from parents, medical examinations and records, school, and from the cohort members. The South East Multi-Centre Research Ethics Committee provided ethical approval for the survey. These analyses utilise data from the 42 year and 45 year surveys.

Measures

Sociodemographic Factors

Cohort members' marital status was assessed at the 42-year survey and classified as single, married, separated, divorced or widowed. Registrar General social class was assessed at the 42-year survey, classified as social class I or II, III non-manual, III manual, IV or V (Power & Elliott, 2006). Cohort members self-reported their educational qualification level as "none", "O-levels" and "A-levels or higher".

Work-family conflict and working hours

Current main activity was defined through whether cohort members were self-employed, not working, part-time paid employees or full-time paid employees. This paper focuses on cohort members who were full-time or part-time paid employees.

Cohort members were given a questionnaire regarding psychosocial work characteristics at 42 years. Work-family conflict was assessed by a self-report question about job demands. Cohort members were asked to answer the question "Do you think the demands of your work interfere with the demands of home and family life?" with either "yes" or "no". The emotional demands of the cohort member's job were reported by the question "How mentally/emotionally demanding is work?". In response, cohort members were asked to state "a lot", "a moderate amount" or "very little". Hours worked per week were self-reported, and excluded time for meal breaks and overtime.

Common Mental Disorders

Common mental disorders were assessed during the 45-year follow-up by the revised Clinical Interview Schedule (CIS-R) (Lewis, Pelosi, Araya, & Dunn, 1992) administered by a nurse using a computer–assisted personal interview conducted in the cohort members home (Stansfeld, Clark, Rodgers, Caldwell, & Power, 2008). Disorders were assessed in the preceding week according to the International Statistical Classification of Diseases, 10th Revision (ICD-10), which allowed the classification of common mental disorders.

Diagnoses included were depressive episodes, classified as mild, moderate or severe, generalised anxiety disorder (GAD), phobias (agoraphobia and social phobia) and panic disorder (Clark, Rodgers, Caldwell, Power, & Stansfeld, 2007). It was also possible to have a co-morbid diagnosis of the aforementioned conditions. A dichotomous outcome, 'any diagnosis' was based on the presence or absence of any of these diagnoses.

Statistical Analyses

Analyses were performed using SPSS Version 18.0 (SPSS Inc., USA). In order to assess the longitudinal association of work-family conflict, working hours and socioeconomic factors on common mental disorders, binary logistic regression analyses were used.

Initially the prevalence of common mental disorders was established and co-occurrence was assessed through the use of chi-square tests between the diagnosis of common mental disorders and work-family conflict, working hours, and socioeconomic factors. Logistic regression analyses predicting the relationship of work-family conflict and hours worked per week on CMD were conducted, initially unadjusted, and then adjusted for gender, social class and marital status. All estimates were expressed as odds ratios with 95% confidence intervals in relation to the reference group.

The associations of marital status, social class and gender on work-family conflict were initially explored through chi-square tests, followed by regression analyses on work-family conflict and marital status, social class and gender in the 42 year sample. Regression analyses were also conducted on work-family conflict and work hours both with and without overtime. These were also expressed as odds ratios with 95% confidence intervals in relation to the reference group.

Results

Descriptives of the sample

Of the original 18,558 cohort members, 11,419 cohort members completed the 42-year follow up and 9,297 cohort members completed the 45-year biomedical survey, of whom 562 (6%) had a diagnosis of at least one CMD. Of the working sample of 9,008 individuals, 50% were female, 72% were married, 57% were employed full time and 35% were of manual social class.

Prevalence of sociodemographic factors, work-family conflict and working hours and unadjusted odds for common mental disorder

Table 1 shows the odds of common mental disorders (CMD) associated with sociodemographic factors. Being single was associated with a 52% increase in odds for a CMD compared to being married or remarried. Furthermore, being divorced, separated or widowed was associated with almost double the odds of a CMD. Females had a 55% increase in odds for a CMD.

Social class was also associated with increased odds for CMD; social classes IV and V had a 52% increase in odds compared with classes I and II. However, social class III non-manual and social class III manual did not have increased odds compared with classes I and II. Being educated up to O-level qualifications was associated with a 26% increase in odds for a CMD compared to having A-levels or higher: those with no qualifications had almost a three-fold increase in odds for a CMD.

Those who were employed part-time had significantly higher odds of a CMD compared with those employed full-time. Those who were unemployed had a three-fold increase in odds for a CMD compared with those employed full-time. In the working sample of 9,008 individuals, those who experienced work-family conflict had a 38% increase in odds for CMD, however mental and emotional demands of work were not significantly associated with CMD. In this sample neither the number of hours worked, not including overtime, nor the amount of overtime worked were found to be associated with CMD.

Table 1. Unadjusted odds ratio (95% confidence interval) for diagnosis of common mental disorders at 45 years by sociodemographic factors at 42 years

	N (%)	OR	95% CI	
Marital Status	9008			
Married/remarried	6481 (71.9)	1.00		
Single	1085(12.0)	1.52*	1.18	1.95
Divorced/separated/widowed	1442 (16.0)	1.82*	1.47	2.25
Gender	9297			
Male	4622 (49.7)	1.00		
Female	4675 (50.3)	1.55*	1.31	1.85
Social class at 42 year	7752			
I and II	3422 (44.1)	1.00		
III non-manual	1645 (21.2)	1.27	0.97	1.67
III manual	1529 (19.7)	0.94	0.70	1.28
IV and V	1156 (14.9)	1.52**	1.14	2.03
Educational qualifications	8087			
A-levels or higher	3843 (47.5)	1.00		
O-levels	3622 (44.8)	1.26	1.03	1.54
None	622 (7.7)	2.73*	2.06	3.63
Employment status	9008			
Full-time	5115 (56.8)	1.00		
Part-time	1525 (16.9)	1.42	1.11	1.81
Not working	1227 (13.6)	3.22***	2.61	3.98
Self employed	1141 (12.7)	0.98	0.72	1.34
Does work interfere with family life?	7777			
No	3600 (46.3)	1.00		
Yes	4177(53.7)	1.38**	1.12	1.71
How mentally/emotionally demanding is work?	7779			
Very little	935 (12.0)	1.00		
Moderate amount	2542 (32.7)	0.74	0.51	1.08
A lot	4302 (55.3)	1.20	0.86	1.66

Any children aged 0-16?	8998			
Yes	6193 (68.8)	1.00		
No	2805 (31.2)	1.38***	1.15	1.65
How long does it take to travel to work?	7304			
0-30 minutes	5134 (70.3)	1.00		
30 minutes-1 hour	1410 (19.3)	0.95	0.72	1.26
Over 1 hour	500 (6.8)	1.26	0.85	1.86
Works at home	260 (3.6)	1.04	0.59	1.84
Total hours worked per week for those that do not work overtime and for those that do work overtime	6291	0.99	0.98	1.00
Usual hours worked per week for those that do not work overtime	2608			
31-50 hours	1484 (56.9)	1.00		
Over 51 hours	180 (6.9)	0.43	0.16	1.20
Less than 31 hours	944 (36.2)	1.09	0.75	1.57
Usual hours worked per week including paid and unpaid overtime	3683	0.99	0.98	1.00
Total hours worked excluding overtime	6581	0.99	0.98	1.00
Any overtime worked, paid and unpaid?	6297			
No overtime	2864 (45.5)	1.00		
1-10 hours	2534 (40.2)	0.96	0.75	1.23
Over 11 hours	899 (14.3)	0.95	0.67	1.35

***p≤ .001
**p≤ .01
*p≤ .05

In the working sample at 42 years, work-family conflict was greater in women, single, separated, divorced and widowed people, those of manual social class, part time workers and those without children (table 2). Those who had a travel time to work of over thirty minutes did not experience significantly more work-family conflict. However, those who worked less than thirty-one hours had a two-fold increase of work-family conflict in comparison with those who worked the norm of 31-50 hours, whereas those who worked over 51 hours had a decreased odds of work-family conflict. Correspondingly those who worked overtime were significantly less likely to experience work-family conflict than those who did not work overtime

Table 2. Unadjusted odds ratio (95% confidence interval) for work-family conflict by sociodemographic factors at 42 years

	N (%)	OR	95% CI	
Marital Status	9614			
Married/remarried	6988 (72.7)	1.00		
Single	1144 (11.9)	1.32***	1.17	1.50
Divorced/separated/widowed	1482 (15.4)	1.26***	1.13	1.42
Gender	9616			
Male	4549 (47.3)	1.00		
Female	5067 (52.7)	2.01***	1.85	2.18
Social class at 42 years	9581			
I and II	4112 (42.9)	1.00		
III non-manual	2043 (21.3)	3.01***	2.70	3.36
III manual	1937 (20.2)	1.90***	1.70	2.12
IV and V	1489 (15.5)	4.03***	3.56	4.57
Employment status	9616			
Full-time	6323 (65.8)	1.00		
Part-time	1876 (19.5)	3.27***	2.92	3.65
Self employed	1417 (14.7)	0.98	0.87	1.10
Any children aged 0-16?	9606			
Yes	6499 (67.7)	1.00		
No	3107 (32.3)	1.59***	1.25	1.74
How long does it take to travel to work?	9028			
0-30 minutes	6337 (70.2)	1.00		
30 minutes-1 hour	1748 (19.4)	0.62***	0.56	0.69
Over 1 hour	611 (6.8)	0.43***	0.36	0.52
Works at home	332 (3.7)	0.85	0.68	1.06
Total hours worked per week for those that do not work overtime and for those that do work overtime	7746	0.94***	0.94	0.95
Usual hours worked per week for those that do not work overtime	3265			
31-50 hours	1837 (56.3)	1.00		
Over 51 hours	231 (7.1)	0.17***	0.12	0.24
Less than 31 hours	1197 (36.7)	2.23***	1.90	2.62

Usual hours worked per week including paid and unpaid overtime	4481	0.95***	0.94	0.95
Total hours worked excluding overtime	8120	0.95***	0.94	0.95
Any overtime worked, paid and unpaid?	7752			
No overtime	3576 (46.1)	1.00		
1-10 hours	3056 (39.4)	0.56***	0.50	0.61
Over 11 hours	1120 (14.4)	0.19***	0.16	0.22

***p≤ .001
**p≤ .01

Adjusted analyses of work-family conflict and CMD

Table 3 shows the associations between work-family conflict, working hours and CMD at 45 years adjusted for social class, gender, marital status and educational qualifications. After adjustment, the odds of work-family conflict for CMD three years later increased and the odds for mental and emotional demands of work also increased and became statistically significant. Upon adjustment, work hours and overtime were not significantly associated with CMD.

Table 3. Adjusted odds ratio (95% confidence interval) for diagnosis of common mental disorders at 45 years by sociodemographic factors at 42 years

Adjusted for social class, gender, marital status and qualifications				
	N	OR	95% CI	
Employment status	6829			
Full-time	4456 (65.3)	1.00		
Part-time	1382 (20.2)	1.17	0.86	1.59
Self employed	991 (14.5)	0.87	0.61	1.24
Does work interfere with family life?	6827			
No	3147 (46.1)	1.00		
Yes	3680 (53.9)	1.73***	1.36	2.20
How mentally/emotionally demanding is work?	6828			
Very little	804 (11.8)	1.00		
Moderate amount	2240 (32.8)	0.92	0.60	1.39
A lot	3784 (55.4)	1.72**	1.17	2.54

View on security of current job	6811			
Very secure	2419 (35.5)	1.00		
Fairly secure	3642 (53.5)	1.19	0.93	1.52
Not very secure	750 (11.0)	1.32	0.91	1.91
Any children aged 0-16?	6825			
Yes	4769 (69.9)	1.00		
No	2056 (30.1)	1.15	0.89	1.48
How long does it take to travel to work?	6423			
0-30 minutes	4514 (70.3)	1.00		
30 minutes-1 hour	1243 (19.4)	1.08	0.80	1.47
Over 1 hour	432 (6.7)	1.56*	1.02	2.39
Works at home	234 (3.6)	1.09	0.58	2.03
Total hours worked per week for those that do not work overtime and for those that do work overtime	5529	1.00	0.99	1.01
Usual hours worked per week for those that do not work overtime	2273			
31-50 hours	1281 (56.4)	1.00		
Over 51 hours	152 (6.7)	0.58	0.20	1.66
Less than 31 hours	840 (37.0)	0.87	0.54	1.39
Usual hours worked per week including paid and unpaid overtime	3256	1.00	0.98	1.01
Total hours worked excluding overtime	5785	1.00	0.99	1.01
Any overtime worked, paid and unpaid?	5534			
No overtime	2493 (45.0)	1.00		
1-10 hours	2252 (40.7)	1.13	0.87	1.49
Over 11 hours	789 (14.3)	1.27	0.86	1.88

***p≤ .001
**p≤ .01
*p≤ .05

Discussion

In these analyses of the 1958 birth cohort we found that work-family conflict was predictive of common mental disorders, whereas the amount of hours spent at work did not increase the risk of CMD three years later. The analyses are longitudinal and the associations regarding work-family conflict persist after adjustment for marital status, social class, gender and educational qualifications.

Work-family conflict was more frequent in women, those who were single, widowed, divorced or separated, those of less advantaged social class and those working part-time (<31 hours per week).

Implications

Work-family conflict and mental health

In previous cross-sectional studies (Wang et al., 2006; Frone, 2000) work-family conflict was associated with the prevalence of CMD. The current study demonstrates that work-family conflict is a significant risk indicator for CMD and a contributor to the complex aetiology of CMD (Kendler et al., 2003). The longitudinal methodology in the current paper adds to the evidence that the association of work-family conflict and the prevalence of CMD may be causal.

Work-family conflict and family structure

Despite the social transformations regarding male and female roles in society (Simon, 2002) women had almost twice the odds of work-family conflict compared with men. This was in accordance with Duxbury, Higgins and Lee (1994) who found that women had higher levels of conflict between work and family than men. Equally a potential explanation for the findings may be due to non-traditional roles for males and females. For women, family roles have been traditionally considered to be the core role and paid employment considered to be the non-traditional role. Therefore those women who work may face increased anxiety due to their opinion on their adequacy of fulfilling traditional family roles such as partner or parent (Duxbury & Higgins, 1991).

In contrast to Duxbury, Higgins and Lee (1994), those who were single, divorced, separated and widowed were more likely to experience work-family conflict than those who were married. This may be due to factors such as household duties and domestic help associated with children, as those who are married may share those responsibilities (Bedeian, Burke and Moffett, 1988). Equally it may be due to other factors such as spouse support moderating the effects of work- family conflict (Fu & Shaffer, 2001). It may be the case that residing with a partner is a more relevant signifier than marital status and thus the unexplained finding of higher work-family conflict in 'single' respondents may still reflect the presence of co-habiting partners. Social class had a prominent impact on work-family conflict as social classes III non-manual and manual, IV and V had an increased risk of CMD compared to those who were in social classes I and II. This may be due to a lack of resources to cope with both work and non-work roles and feeling financially insecure (Singh-Manoux, Adler & Marmot, 2003).

Work-family conflict and working hours

There was a significant three-fold risk of work-family conflict for those who work part-time as compared to those who work full-time. This may be because part-time work entails irregular hours, which may interrupt family routines (Wheatley, 2012). Equally it may be due to a potential difference in income between those who work full-time and those who work part-time as financial issues are often a cause for concern in relation to CMD (Clark et al., 2011). Part-time work may also be chosen if there are other potentially stressful roles such as caring which may contribute both to increased work-family conflict and increased CMD risk. By contrast, the ability to work overtime may also be possible only if you have fewer commitments outside work.

We did not find a significant association between the number of hours worked per week and CMD. Spurgeon et al., (1997) suggested that those who work over 50 hours per week would be at greater risk of CMD. In addition Virtanen et al., (2011) indicated that individuals who work over 55 hours per week were more likely to be at risk of CMD than those who work 35-40 hours. However, greater hours can be assumed to indicate greater income and this may be considered to reduce the risk of CMD as there is reduced financial pressure. Greater hours may also indicate greater choice over the length of the working day in non-manual occupations. In accordance with this, Singh-Manoux et al., (2003) also indicate that household income and feeling financially secure is associated with health as they are aspects of social status.

Individuals who work long hours may enjoy doing so and it may provide meaning to life (Taylor, 2002). Similarly Bedeian et al., (1988) advocated that work may be central to individuals' lives as it is rewarding not only financially but also socially and emotionally.

Equally Greenhaus and Powell (2006) suggest that work aspects and family aspects have additive effects on wellbeing. Barnett and Hyde (2001) propose that multiple roles are not harmful and are generally beneficial for both males and females in terms of mental health. This may be because involvement in both work and family roles may provide compensatory satisfaction from conflict in one of the other roles (Greenhaus & Powell, 2006).

As well as work-life conflict other forms of job demands were assessed, namely mental and emotional demands at work. The association between CMD and mental and emotional demands became significant after adjustment for gender, marital status, social class and qualifications. This association may be dependent on the nature of the work as those who were in social class III manual were less likely to experience work-family conflict than those who were in social class III non-manual. This may be because many social class III manual jobs are associated with low psychological strain as individuals may be self-employed and have greater control over their jobs (Stansfeld & Candy, 2006).

Strengths and limitations

This study has a number of limitations. Work-life conflict was assessed using a single item and this would be better assessed through a multi-item questionnaire (Carlson, Kacmar & Williams, 2000). There are a number of questionnaires that deal with work-life conflict in more detail – the use of one of these would have given a more reliable measure of work-family conflict. The item available to us did not fully capture the three aspects of work-life conflict: time-based, strain-based and behaviour-based conflict (Carlson et al, 2000). The generalisability of the 1958 British birth cohort to the wider population must also be considered. These analyses are of cohort members at 45 years old. This may be a late stage in life for work-family conflict, which may be more acute when for caring for younger dependent children (Shimazu et al., 2010; Seto et al., 2004). Thus our findings may not apply across time periods or to other age groups as people face different issues at different career stages (Bedeian et al., 1988).

Equally attrition needs to be taken into account. Although Atherton, Fuller, Shepherd, Strachan and Power (2008) advocate that the 1958 sample is broadly representative, they suggest there was underrepresentation of some groups: individuals who were psychologically distressed and individuals in lower social classes were prone to loss from the cohort. Consequently increased associations may be observed in the general population. Different associations with work-family conflict may also be observed for anxiety, depression and phobias, but this study did not have the statistical power to analyse associations for the different disorders.

These initial analyses could be refined by further work in a number of ways. Multiple imputation techniques could be used to address missing data. In longitudinal studies it is appropriate to take account of the time varying nature of social class and marital status and multiple waves of data would allow this. It would have been ideal to adjust our analyses for common mental disorders at 42 years (baseline) or to restrict the sample to those without common mental disorders at baseline. Unfortunately, there was no comparable measure of common mental disorder at age 42 years.

In spite of these limitations, studies of this nature allow us to begin to understand the potential causal links between mental health, work-family conflict, working hours and family structure. This study is one of few which examine temporal relationships between common mental disorders and work-life conflict.

Clinical and public health implications

These findings are pertinent as the balance between work and life is considered to be an important current focus in Britain (Taylor, 2002). Common mental disorders are the second most significant cause of sickness absence from work for longer than 21 days (Nieuwenhujisen, Verbeek, deBoer, Blonk, & van Dijk, 2006). As a result work-life policies have been implemented (Wheatley, 2012) which may benefit from further understanding of which factors increase risk of common mental disorders. Work-family conflict was found to be a significant risk factor for CMD. Although work hours did not directly impact the prevalence of CMD they had a significant impact on work-family conflict, which has not been seen in previous studies.

Future research

Further studies are required in order to assess work-family balance as a whole as family-work conflict was not assessed (Byron, 2005). Additionally further research could include more on family structure and socioeconomic status, such as total household income, in order to formulate a more complete understanding of the factors that contribute to the prevalence of common mental disorders within the family. Moreover, further analyses examining work-family conflict and common mental disorders may be studied by occupation in order to understand whether associations differ by occupation as they do by social class. Similarly, additional analyses exploring gender differences may be beneficial. Moreover, to take full advantage of the richness of multiple waves of data available in the cohort it would be useful to carry out structural equation modelling of the predictors of work-family conflict and their antecedents.

Conclusions

In conclusion, work-family conflict was a significant longitudinal predictor for common mental disorders. The number of hours worked was not associated with the prevalence of CMD, though increased work hours are associated with less work-family conflict. Furthermore, working overtime was associated with reduced risk of work-family conflict. The findings suggest that reducing the work-family conflict that individuals face may reduce the risk of CMD. Further research into factors that cause work-family conflict is required.

Acknowledgements

These analyses were carried out as part of an undergraduate dissertation; no additional funding was required. We thank the data providers: Centre for Longitudinal Studies, Institute of Education and National Birthday Trust Fund, National Children's Bureau, City University Social Statistics Research Unit (original data producers).

References

Adler, N. E., & Ostrove, J. M. (1999). Socioeconomic status and health: what we know and what we don't. *Annals of the New York Academy of Sciences*, *896*, 3-15. http://dx.doi.org/10.1111/j.1749-6632.1999.tb08101.x

Atherton, K., Fuller, E., Shepherd, P., Strachan, D. P., & Power, C. (2008). Loss and representativeness in a biomedical survey at age 45 years: 1958 British birth cohort. *Journal of Epidemiology and Community Health*, *62*, 216-223. http://dx.doi.org/10.1136/jech.2006.058966

Barnett, R. C., & Hyde, J. S. (2001). Women, men, work, and family. *American Psychologist*, *56*, 781-796. http://dx.doi.org/10.1037/0003-066X.56.10.781

Bebbington, P. (1987). Marital status and depression: a study of English national admission statistics. *Acta Psychiatrica Scandinavica*, *75*, 640-650. http://dx.doi.org/10.1111/j.1600-0447.1987.tb02849.x

Bedeian, A. G., Burke, B. G., & Moffett, R. G. (1988). Outcomes of work-family conflict among married male and female professionals. *Journal of Management*, *14*, 475-491. http://dx.doi.org/10.1177/014920638801400310

Byron, K. (2005). A meta-analytic review of work–family conflict and its antecedents. *Journal of Vocational Behavior*, *67*, 169-198. http://dx.doi.org/10.1016/j.jvb.2004.08.009

Canivet, C., Östergren, P. O., Lindeberg, S. I., Choi, B., Karasek, R., Moghaddassi, M., & Isacsson, S. O. (2010). Conflict between the work and family domains and exhaustion among vocationally active men and women. *Social Science & Medicine*, *70*, 1237-1245. http://dx.doi.org/10.1016/j.socscimed.2009.12.029

Carlson, D.S., Kacmar, K.M., Williams LJ. (2000). Construction and initial validation of a multidimensional measure of work-family conflict. *Journal of Vocational Behavior*, 56, 249-276. http://dx.doi.org/10.1006/jvbe.1999.1713

Casper, W. J., Eby, L. T., Bordeaux, C., Lockwood, A., & Lambert, D. (2007). A review of research methods in IO/OB work-family research. *Journal of Applied Psychology*, *92*, 28-43.

http://dx.doi.org/10.1037/0021-9010.92.1.28
Cinamon, R. G., & Rich, Y. (2002). Gender differences in the importance of work and family roles: Implications for work–family conflict. *Sex Roles*, *47*, 531-541. http://dx.doi.org/10.1023/A:1022021804846
Clark, C., Pike, C., McManus, S., Harris, J., Bebbington, P., Brugha, T. Jenkins, R., Meltzer, H., Weich, S., & Stansfeld, S. (2011). The contribution of work and non-work stressors to common mental disorders in the 2007 Adult Psychiatric Morbidity Survey. *Psychological Medicine*, *1*, 1-14.
Clark, C., Rodgers, B., Caldwell, T., Power, C., & Stansfeld, S. (2007). Childhood and adulthood psychological ill health as predictors of midlife affective and anxiety disorders: the 1958 British Birth Cohort. *Archives of General Psychiatry*, *64*, 668-678. http://dx.doi.org/10.1001/archpsyc.64.6.668
Duxbury, L. E., & Higgins, C. A. (1991). Gender differences in work-family conflict. *Journal of Applied Psychology*, *76*, 60-73. http://dx.doi.org/10.1037/0021-9010.76.1.60
Duxbury, L., Higgins, C., & Lee, C. (1994). Work-Family Conflict A Comparison by Gender, Family Type, and Perceived Control. *Journal of Family Issues*, *15*, 449-466. http://dx.doi.org/10.1177/019251394015003006
Frone, M. R. (2000). Work–family conflict and employee psychiatric disorders: The national comorbidity survey. *Journal of Applied Psychology*, *85*, 888-895. http://dx.doi.org/10.1037/0021-9010.85.6.888
Frone, M. R., Russell, M., & Cooper, M. L. (1992). Antecedents and outcomes of work-family conflict: testing a model of the work-family interface. *Journal of Applied Psychology*, *77*, 133-161.
Frone, M. R., Russell, M., & Cooper, M. L. (1997). Relation of work–family conflict to health outcomes: a four-year longitudinal study of employed parents. *Journal of Occupational and Organizational Psychology*, *70*, 325-335. http://dx.doi.org/10.1037/0021-9010.77.1.65
Fu, C. K., & Shaffer, M. A. (2001). The tug of work and family: Direct and indirect domain-specific determinants of work-family conflict. *Personnel Review*, *30*, 502-522. http://dx.doi.org/10.1108/EUM0000000005936
Grant-Vallone, E. J., & Ensher, E. A. (2001). An examination of work and personal life conflict, organizational support, and employee health among international expatriates. *International Journal of Intercultural Relations*, *25*, 261-278. http://dx.doi.org/10.1016/S0147-1767(01)00003-7
Greenhaus, J. H., & Parasuraman, S. (1999). *Research on work, family, and gender: Current status and future directions.* In G. N. Powell (Ed.), *Handbook of gender & work*, *391–412*. Thousand Oaks, CA: Sage. http://dx.doi.org/10.4135/9781452231365.n20
Greenhaus, J. H., & Powell, G. N. (2006). When work and family are allies: A theory of work-family enrichment. *Academy of Management Review*, *31*, 72-92. http://dx.doi.org/10.5465/AMR.2006.19379625
Grzywacz, J. G., & Bass, B. L. (2003). Work, Family, and Mental Health: Testing Different Models of Work-Family Fit. *Journal of Marriage and Family*, *65*, 248-261. http://dx.doi.org/10.1111/j.1741-3737.2003.00248.x
Guest, D. E. (2002). Perspectives on the study of work-life balance. *Social Science Information*, *41*, 255-279. http://dx.doi.org/10.1177/0539018402041002005
Hämmig, O., & Bauer, G. (2009). Work-life imbalance and mental health among male and female employees in Switzerland. *International Journal of Public Health*, *54*, 88-95. http://dx.doi.org/10.1007/s00038-009-8031-7
Hämmig, O., Gutzwiller, F., & Bauer, G. (2009). Work-life conflict and associations with work-and nonwork-related factors and with physical and mental health outcomes: a nationally representative cross-sectional study in Switzerland. *BMC Public Health*, *9*, 435. http://dx.doi.org/10.1186/1471-2458-9-435

Harrington, J. M. (2001). Health effects of shift work and extended hours of work. *Occupational and Environmental Medicine*, *58*, 68-72. http://dx.doi.org/10.1136/oem.58.1.68

Hemingway, H., Nicholson, A., Stafford, M., Roberts, R., & Marmot, M. (1997). The impact of socioeconomic status on health functioning as assessed by the SF-36 questionnaire: the Whitehall II Study. *American Journal of Public Health*, *87*, 1484-1490. http://dx.doi.org/10.2105/AJPH.87.9.1484

Janzen, B. L., & Kelly, I. W. (2012). Psychological Distress Among Employed Fathers Associations With Family Structure, Work Quality, and the Work–Family Interface. *American Journal of Men's Health*, *6*, 294-302. http://dx.doi.org/10.1177/1557988311435835

Kendler, K. S., Prescott, C. A., Myers, J., & Neale, M. C. (2003). The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Archives of General Psychiatry*, *60*, 929-937. http://dx.doi.org/10.1001/archpsyc.60.9.929

Lewis G, Pelosi AJ, Araya R, & Dunn G. (1992). Measuring psychiatric disorder in the community: a standardized assessment for use by lay interviewers. *Psychological Medicine*, *22*, 465-86. http://dx.doi.org/10.1017/S0033291700030415

Lewis, S. (2003). The integration of paid work and the rest of life. Is post-industrial work the new leisure? *Leisure Studies*, *22*, 343-345. http://dx.doi.org/10.1080/02614360310001594131

Lorant, V., Deliège, D., Eaton, W., Robert, A., Philippot, P., & Ansseau, M. (2003). Socioeconomic inequalities in depression: a meta-analysis. *American Journal of Epidemiology*, *157*, 98-112. http://dx.doi.org/10.1093/aje/kwf182

Lowry, D., & Moskos, M. (2008). Mobile phones, spillover and the 'work–life balance'. *Mobility and Technology in the Workplace*, 167-179.

Martens, M. F. J., Nijhuis, F. J. N., Van Boxtel, M. P. J., & Knottnerus, J. A. (1999). Flexible work schedules and mental and physical health. A study of a working population with non-traditional working hours. *Journal of Organizational Behavior*, *20*, 35-46. http://dx.doi.org/10.1002/(SICI)1099-1379(199901)20:1<35::AID-JOB879>3.0.CO;2-Z

Nieuwenhuijsen, K., Verbeek, J. H., de Boer, A. G., Blonk, R. W., & van Dijk, F. J. (2006). Predicting the duration of sickness absence for patients with common mental disorders in occupational health care. *Scandinavian Journal of Work, Environment & Health*, 67-74. http://dx.doi.org/10.5271/sjweh.978

Nitzsche, A., Jung, J., Pfaff, H., & Driller, E. (2012). Employees' negative and positive work–home interaction and their association with depressive symptoms. *American Journal of Industrial Medicine,* 1-8.

Obidoa, C., Reeves, D., Warren, N., Reisine, S., & Cherniack, M. (2011). Depression and work family conflict among corrections officers. *Journal of Occupational and Environmental Medicine*, *53*, 1294-1301. http://dx.doi.org/10.1097/JOM.0b013e3182307888

Parsons, T. J., Power, C., & Manor, O. (2001). Fetal and early life growth and body mass index from birth to early adulthood in 1958 British cohort: longitudinal study. *British Medical Journal*, *323*, 1331-1335. http://dx.doi.org/10.1136/bmj.323.7325.1331

Power, C., & Elliott, J. (2006). 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

Power, C., Hertzman, C., Matthews, S., & Manor, O. (1997). Social differences in health: life-cycle effects between ages 23 and 33 in the 1958 British birth cohort. *American Journal of Public Health*, *87*, 1499-1503. http://dx.doi.org/10.2105/AJPH.87.9.1499

Reynolds, J. (2005). In the Face of Conflict: Work-Life Conflict and Desired Work Hour Adjustments. *Journal of Marriage and Family*, *67*, 1313-1331.

http://dx.doi.org/10.1111/j.1741-3737.2005.00219.x

Seto, M., Morimoto, K., & Maruyama, S. (2004). Effects of work-related factors and work-family conflict on depression among Japanese working women living with young children. *Environmental Health and Preventive Medicine*, *9*, 220-227. http://dx.doi.org/10.1007/BF02898103

Shimazu, A., Bakker, A. B., Demerouti, E., & Peeters, M. C. (2010). Work-family conflict in Japan: how job and home demands affect psychological distress. *Industrial Health*, *48*, 766-774. http://dx.doi.org/10.2486/indhealth.MS1131

Simon, R. W. (2002). Revisiting the Relationships among Gender, Marital Status, and Mental Health. *American Journal of Sociology*, *107*, 1065-1096. http://dx.doi.org/10.1086/339225

Singh-Manoux, A., Adler, N. E., & Marmot, M. G. (2003). Subjective social status: its determinants and its association with measures of ill-health in the Whitehall II study. *Social Science &Medicine*, *56*, 1321-1333. http://dx.doi.org/10.1016/s0277-9536(02)00131-4

Spurgeon, A., Harrington, J. M., & Cooper, C. L. (1997). Health and safety problems associated with long working hours: a review of the current position. *Occupational and Environmental Medicine*, *54*, 367-375. http://dx.doi.org/10.1136/oem.54.6.367

Stansfeld, S. A., Clark, C., Rodgers, B., Caldwell, T., & Power, C. (2008). Childhood and adulthood socio-economic position and midlife depressive and anxiety disorders. *The British Journal of Psychiatry*, *192*, 152-153. http://dx.doi.org/10.1192/bjp.bp.107.043208

Stansfeld, S., & Candy, B. (2006). Psychosocial work environment and mental health—a meta-analytic review. *Scandinavian Journal of Work, Environment & Health*, 443-462. http://dx.doi.org/10.5271/sjweh.1050

Takeuchi, T., & Yamazaki, Y. (2010). Relationship between work–family conflict and a sense of coherence among Japanese registered nurses. *Japan Journal of Nursing Science*, *7*, 158-168. http://dx.doi.org/10.1111/j.1742-7924.2010.00154.x

Taylor, R. (2002). *The future of work-life balance*. Economic and Social Research Council.

Virtanen, M., Ferrie, J. E., Singh-Manoux, A., Shipley, M. J., Stansfeld, S. A., Marmot, M. G., & Kivimäki, M. (2011). Long working hours and symptoms of anxiety and depression: a 5-year follow-up of the Whitehall II study. *Psychological Medicine*, *41*, 2485-2494. http://dx.doi.org/10.1017/S0033291711000171

Wang, J. L., Lesage, A., Schmitz, N., & Drapeau, A. (2006). The relationship between work stress and mental disorders in men and women: findings from a population-based study. *Journal of Epidemiology and Community Health*, *62*, 42-47. http://dx.doi.org/10.1136/jech.2006.050591

Weich, S., & Lewis, G. (1998). Poverty, unemployment, and common mental disorders: population based cohort study. *British Medical Journal*, *317*, 115-119. http://dx.doi.org/10.1136/bmj.317.7151.115

Wheatley, D. (2012). Work-life balance, travel-to-work, and the dual career household. *Personnel Review, 41*, 813 – 831. http://dx.doi.org/10.1108/00483481211263764

Longitudinal and Life Course Studies 2015 Volume 6 Issue 3 Pp 279 – 302 *ISSN 1757-9597*

The effect of parental divorce on the health of adult children

Jason R. Thomas Department of Sociology and Criminology, Penn State University, US
jrt17@psu.edu
Robin S. Högnäs Department of Sociology, University of Louisville, US

(Received August 2013 Revised November 2014) http://dx.doi.org/10.14301/llcs.v6i3.267

Abstract

Decades of research have produced evidence that parental divorce is negatively associated with offspring outcomes from early childhood, through adolescence, and into the adult years. This study adds to the literature on the effects of parental divorce by examining how the timing of a parental divorce influences the total effect on adult health. Furthermore, we look at how this long-term effect of parental divorce depends on mediators such as the family's socioeconomic status, parental involvement, cognitive test scores, behavioural problems, smoking, and the offspring's own experience with divorce. The analyses use data from the National Child Development Study, which includes nine waves of data beginning at birth in 1958 and continuing through age 50. Results from a structural equation model suggest that a parental divorce experienced before age seven does influence adult health by operating primarily through family socioeconomic status and smoking in adulthood.

Keywords: parental divorce, health, life course

Introduction

There are numerous non-clinical determinants of adult health that affect exposure and resilience to health insults and toxic environments. These range from behaviours such as smoking, diet and exercise to social and economic factors. In recent years, increasing attention has focused on early conditions and how they shape the life course trajectories leading towards adult health and mortality. While some have pointed to the role of physical health in early life (Almond & Currie, 2011), others have highlighted the early developmental difficulties with cognitive and social skills (Conti & Heckman, 2010). Much progress has been made on this front, as exemplified by the growing body of research looking at the effects of parental divorce (Troxel & Matthews, 2004), an experience associated with a wide range of outcomes indicative of the child's wellbeing (Amato, 2010). However, questions still remain about the importance of when the parental divorce occurs, as well as the pathways through which parental divorce operates to influence the individual's health in adulthood.

To date, the literature on the negative effects of parental divorce on adult health has provided evidence concerning mortality (Hayward & Gorman, 2004; Martin, Friedman, Clark, & Tucker, 2005; Preston, Hill, & Drevenstedt, 1998; Schwartz et al., 1995; Larson & Halfon, 2013), cancer (Hemminki & Chen, 2006), and the number of health problems experienced by respondents (Maier & Lachman, 2000). Although Schwartz et al. (1995) find an effect of parental divorce on the risk of mortality the generalisability of this result is in question because the data come from a prospective study of children in California recruited for their high cognitive ability. It is also worth noting the evidence of a negative association between parental divorce and the emotional and mental health of the children when they reach adulthood (Chase-

Lansdale, Cherlin & Kiernan, 1995; Cherlin, Chase-Lansdale & McRae 1998).

We build on this previous work by addressing two issues that have yet to be tackled by this growing literature. First, previous work has failed to examine whether the timing of parental divorce influences the effect on the child's adult health. Many aspects of child development reflect a cumulative process where early advantage leads to greater levels of advantage at older ages (Cunha & Heckman, 2008; Heckman, 2007). Insults (or investments) experienced at a young age may be particularly influential, relative to similar exposures at older ages, if they lower the rate of developmental growth. The gap between two trajectories, one growing at a slower rate due to a developmental insult, will increase with age, and the difference will be larger if the exposure occurs earlier in life (Lansford et al., 2006). To the extent that developmental outcomes affect adult health, the timing of influential events will help determine the size of health disparities later in life. Furthermore, developmental outcomes are more vulnerable at some ages than others (Hetherington, Stanley-Hagan, & Anderson, 1989; Lansford et al., 2006; Woodward, Fergusson, & Belsky, 2000). It follows that the pathways connecting parental divorce to adult health may vary depending on when the event occurs and which developmental outcomes are the most vulnerable. This heterogeneity in the accumulation of disadvantage and in the types of vulnerability to parental divorce suggests differences in the long-term effects of parental divorce associated with the age at which the family disruption occurred. We begin to fill this gap by focusing on three different age intervals (from birth to age seven, age seven to 11, and age 11 to 16), and separately assessing the total effect of a parental divorce during each interval on the child's adult health.

Previous studies have considered variables that mediate the relationship between parental divorce and adult health, but do so by sequentially including mediators into a series of regression models. Our contribution moves beyond this approach, specifying a structural equation model that identifies the pathways through which parental divorce influences adult health. The list of mediating variables includes family socioeconomic status, parental involvement, cognitive test scores, behavioural problems, smoking, and divorce. Finally, we decompose the total effect into the contributions of each mediator to show that parental divorce primarily operates through the family's socioeconomic status and the child's smoking behaviour to generate adverse health outcomes.

Background

The crude divorce rate in Britain began to climb around 1960 and continued to increase before leveling off in the early 1980s in Scotland, and in the mid-1990s in England and Wales. The crude divorce rate then fluctuated before declining in recent years in each of the countries (National Records of Scotland http://www.gro-scotland.gov.uk/statistic; Office for National Statistics 2013). Even with the recent decline, a large fraction of children are exposed to the dissolution of parental relationships and the associated risks for negative outcomes. While parental divorce is not directly tied to adult physical health, numerous mediators create a connection.

Socioeconomic status (Banks, Marmot, Oldfield, & Smith, 2006; Elo, 2009; Meara, Richards, & Cutler, 2008), health behaviours (Cutler & Lleras-Muney, 2010; Pampel, Krueger, & Demney, 2010), and social support (Cutrona, 1996; House, Landis, & Umberson, 1988) all predict adult health outcomes and are associated with parental divorce. Individuals with greater education, wealth, and occupational status tend to be healthier than those with fewer socioeconomic resources (Elo, 2009). Research suggests that socioeconomic attainment depends, at least in part, on the early development of cognitive and non-cognitive skills (Conti & Heckman, 2010; Almond & Currie, 2005; Goodman, Joyce, & Smith, 2011). Moreover, the acquisition of cognitive and non-cognitive skills tends to be hampered by parental divorce (Forehand, Neighbors, Devine, & Armistead, 1994; Kurdek, Fine, & Sinclair, 1994; Potter, 2010; Tillman, 2007), and the associated decrease in economic resources is one of the primary explanations accounting for this result (Carlson & Corcoran, 2001; McLanahan & Sandefur, 1994; Thompson, Hanson, & McLanahan, 1994).

The dependence of early skill acquisition on parental divorce also rests upon noneconomic mechanisms. For example, the level and quality of parental involvement in a child's life, particularly with respect to schooling, also facilitates academic performance (Izzo, Weissberg, Kasprow, & Fendrich, 1999; Miedel & Reynolds, 2003), a common indicator of early skills. Parental divorce tends to reduce the amount of contact the child has with their noncustodial parents (Kelly, 2007; McLanahan & Sandefur, 1994; Tach, Mincy, & Edin, 2010).[1] Previous work also suggests that some children suffer psychologically and/or emotionally as a result of their parents' divorce and, consequently, the child's academic achievement suffers as well (Forehand, Neighbors, Devine, & Armistead, 1994; Kurdek, Fine, & Sinclair, 1994; Potter, 2010; Tillman, 2007).

In addition to socioeconomic pathways leading to adult health, parental divorce may operate through health behaviours, particularly smoking. Furstenberg and Kiernan (2001) analyse the same data used here and demonstrate that parental divorce is associated with higher odds of smoking at age 33. As previously discussed, one of the mechanisms at work relies on socioeconomic status and the elevated risk of smoking associated with lower status (Cutler & Lleras-Muney, 2010; Pampel, Krueger, & Demney, 2010). However, Furstenberg and Kiernan (2001) still find an association after conditioning on measures of the child's cognitive ability, behavioral problems, and family socioeconomic status. This result is also consistent with earlier work by Wolfinger (1998), who concludes that less social control, lower socioeconomic status, and psychological problems associated with a parental divorce do not account for the observed relationship between parental divorce and smoking. While genetic factors or self-medicating delayed onset of psychological problems may be playing a role (Troxel & Matthews, 2004), more work is needed to understand this particular pathway.

Finally, we turn to pathways involving social support, particularly as it is manifested through the child's own experience with divorce. Parents play an important role in shaping the next generation more generally via the modeling of family norms, roles, and relationships observed during childhood and beyond (Barber, 2001; Baumrind, 1986; Bengston, 1975). For example, those who experience their parents' divorce will be more likely to later divorce themselves (McLanahan & Bumpass, 1988) in part because they are socialised to have a lower commitment to the institution of marriage (Amato & DeBoer 2001). It is also the case that offspring of divorced parents are more likely themselves to have lower socioeconomic status, which is strongly associated with both the risk of divorce and poor health (Adams, Hurd, McFadden, Merrill, & Riberio, 2003; Albouy & Lequien, 2009; Amato, 2010). Moreover, divorce has been found to be directly associated with declines in health — both men and women report lower self-assessed health following divorce compared to those who remain married (Williams & Umberson, 2004). In summary, the intergenerational transmission of divorce is another potential pathway through which a parental divorce can indirectly influence the adult health of the offspring.

While we model the pathways described above, there are additional mediators that may play an important role in generating a total effect of parental divorce on the child's adult health (for an excellent discussion of mediators associated with adult mental health see Maughan & McCarthy, 1997). For example, declines in parenting quality associated with the parent's adjustment to the divorce, stressful life changes (such as family transitions and moving), and inter-parental conflict (Amato, 1993) also link parental divorce to the child's wellbeing. There are also many other sources of social support and types of health behaviours that we neglect. These omissions reflect the limitations in the information available to us in the NCDS as well as our desire to analyse a model that is not bereft of parsimony.

Before moving on to the analysis, it is important to note how the problem of selection may bias estimates of the total effect of parental divorce. Previous research has shown associations between indicators of the child's wellbeing measured both before and after the parental divorce, and that conditioning on the lagged value of the outcome significantly diminishes the negative effect of parental divorce (Amato, 2010; Cherlin et al., 1991). While it may be that pre-divorce conflict between the parents creates a stressful environment that leads to poorer outcomes for the child, selection may be contributing to the observed associations. It will be shown that we only find convincing evidence of an effect on adult health among those who experience a parental divorce by age seven. The only antecedent variables that are available are observed at the birth of the child, leaving us with no information to assess the role of selection.[3]

Research questions, data, and methods

We are interested in how parental divorce relies on mediating variables (described above) to produce a total effect on adult children's health. Our analysis begins by focusing on the timing of the parental divorce and estimating the total effect on the child's adult health for each of the following age intervals when the parental divorce occurred: birth to age seven; age seven to age 11, and age 11 to age 16. These age intervals are based on the available information rather than important stages of child development. We then assess how mediating variables account for the total effect of parental divorce. The mediators explored here include the family's economic resources, parental investment, cognitive skills, emotional and psychological problems, health behaviours, and the intergenerational transmission of divorce. We only find convincing evidence for an effect of parental divorce that occurs before age seven, thus our analysis of mediating variables is restricted to this group.

To examine the total effect of parental divorce on adult children's health, we use data from the 1958 National Child Development Study (NCDS), a prospective longitudinal study of nearly all (98%) children born in the week of March 3-9, 1958 in England, Scotland, and Wales. Follow-up waves were collected at ages seven, 11, 16, 23, 33, 42, 46, and 50. See Power and Elliott (2006) and Ferri (1993) for detailed descriptions of the NCDS. From the baseline sample (N = 17,415), we select an analytic sample using the following criteria: (1) biological children of couples married at birth (N = 16,662); (2) the cohort member did not experience a parental death or become adopted by age 16 (N = 15,767); and (3) the cohort member was not known to have died (N = 14,637).[2] Thus, the 7,511 males and 7,126 females included in our analyses constitute 84% of the original birth cohort sample. We assume the incomplete data in our analytic sample are missing at random and multiply impute the missing information.[4]

In an effort to be consistent with prior research, while also reducing the limitations of any single health measure, we examine three different health outcomes at age 50. First, we use self-rated health which we treat as an ordered categorical variable with three outcomes: 1=excellent; 2=very good/good; and 3=fair/poor. The categories of fair and poor health are collapsed to obtain a group size large enough for calculating reliable estimates. Although self-rated health is a commonly used measure, it is not without limitations. Thus, we include two additional health measures: (1) adult children's number of health problems[5], which we also treat as an ordered categorical variable with three outcomes: 1=no reported health problems; 2=one or two health problems; and 3=three or more health problems. Finally, our third measure of adult children's health at age 50 is the physical functioning scale drawn from the 36-Item Short Form Survey (Ware, Snow, Kosinski, & Gandek, 1993).[6] The skewed distribution of this variable led us to recode it as a dichotomous indicator where a value of 1 indicates one standard deviation below the gender-specific mean of the observed cohort members.

Father's social class, observed at the start of the age interval when the parental divorce takes place, is an exogenous variable included in our models to adjust for a possible spurious association between parental divorce and adult health (Alwin & Hauser, 1975). Father's social class is measured using the Registrar-General's Social Class schema and is treated as an ordinal measure where higher values indicate higher status: (1) unskilled manual, unemployed, or no male head of household; (2) partly skilled manual; (3) skilled manual; (4) skilled non-manual; (5) managerial and technical; and (6) professional. Unemployed fathers are included with the lowest value for social class (i.e., 1 – unskilled manual or unemployed).

Recall that our analysis of the pathways predicting adult health is restricted to those experiencing a parental divorce by age seven. Our mediators include three measures of family's socioeconomic status (SES) at age seven following parental divorce. First, we use the social class of the male head of the household. If there is no male head of household, as is more common after a parental divorce, or if the head is unemployed, we assign the lowest value for social class (i.e., 1=unskilled manual, unemployed, or no male head of household). Next, we include an ordered categorical measure for crowding in participant's home: up to 1 person per room=1; between 1 and 1.5 persons per room=2; and over 1.5 persons per room=3. Finally, our analysis includes a dummy indicator for the family experiencing financial difficulties, as reported by the interviewer who visited the participant's home at age seven.

Additional mediators include the following measures of parental involvement observed at age seven: does the mother (or father) read to the child (0=hardly ever; 1=occasionally; 2=every week); does the mother (or father) take the child out – e.g., for walks, outings, picnics, visits, shopping (0=hardly ever; 1=occasionally; 2=most weeks); and maternal reports of the father's role in the management of the child (0=primarily left to the mother; 1=significant role, but less than the mother; 2=equal to the mother). For the first two variables, reading to the child and taking the child out, we sum the values for the mother and father to produce a composite score. If the mother (father) did not live in the household, the question refers to the "permanent mother substitute" ("male head of household"). For families where there is either no mother or no father figure, a zero is assigned for that parent figure. With respect to our third measure of parental involvement, if there is no mother figure in the household, the father's role in managing the child is assigned a value of 3, the highest level of involvement.

The total effect of parental divorce may also be mediated by the child's cognitive and emotional development observed at age 11. Cognitive development is measured using the average of standardized test scores for reading comprehension, maths, and general tests on verbal and non-verbal skills. Emotional development is measured using the Bristol Social Adjustment Guide (BSAG), a count of the number of behaviours and attitudes indicative of social maladjustment, unsettledness, and emotional problems as identified and reported by participants' teachers (Stott, 1969). The influential work of Ghodsian (1977) has led many researchers to use two underlying factors, internalizing and externalizing problems, in their analysis. We, however, are interested in the total contribution of the social and emotional problems captured by the BSAG measure and, thus, follow the suggestion of Stott (1963) by creating an ordered, categorical measure: 1 – a score of 0-9 (stable); 2 – a score of 10-19 (unsettled); and 3 – a score of 20 or more (maladjusted). Our final two mediators consist of a dummy indicator for whether or not the participant experienced their own divorce by age 50; and a categorical variable constructed by summing the number of times the participant reported being a current smoker at ages 23, 33, 42, 46, and 50 (ranging from 0 to 6). Finally, we include a dummy indicator of parental smoking when the cohort member is 16 years old, which is used as a predictor of participants' smoking behaviour.

We begin by estimating the total effect of parental divorce on adult health using the model depicted in figure 1. This model includes direct effects of both parental divorce and father's social class on adult health at age 50, and a direct effect of father's social class on parental divorce. For each of the three health outcomes at age 50, we estimate gender-specific versions of the model in figure 1 for participants who experience a parental divorce between birth and age seven, between age seven and age 11, and between age 11 and age 16. Participants whose parents were continuously married from birth to age 16 are in the reference group in each model.

Figure 1. Path model for the total effect of a parental divorce on health at age 50

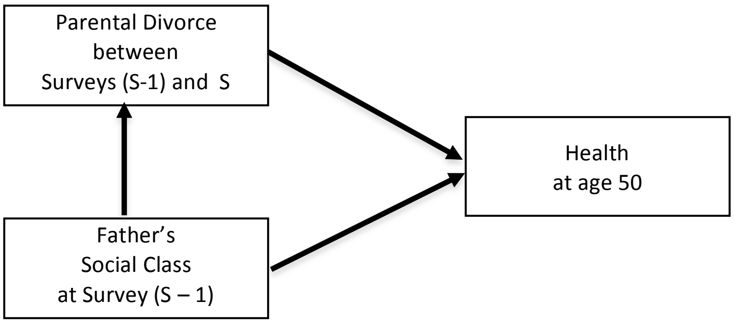
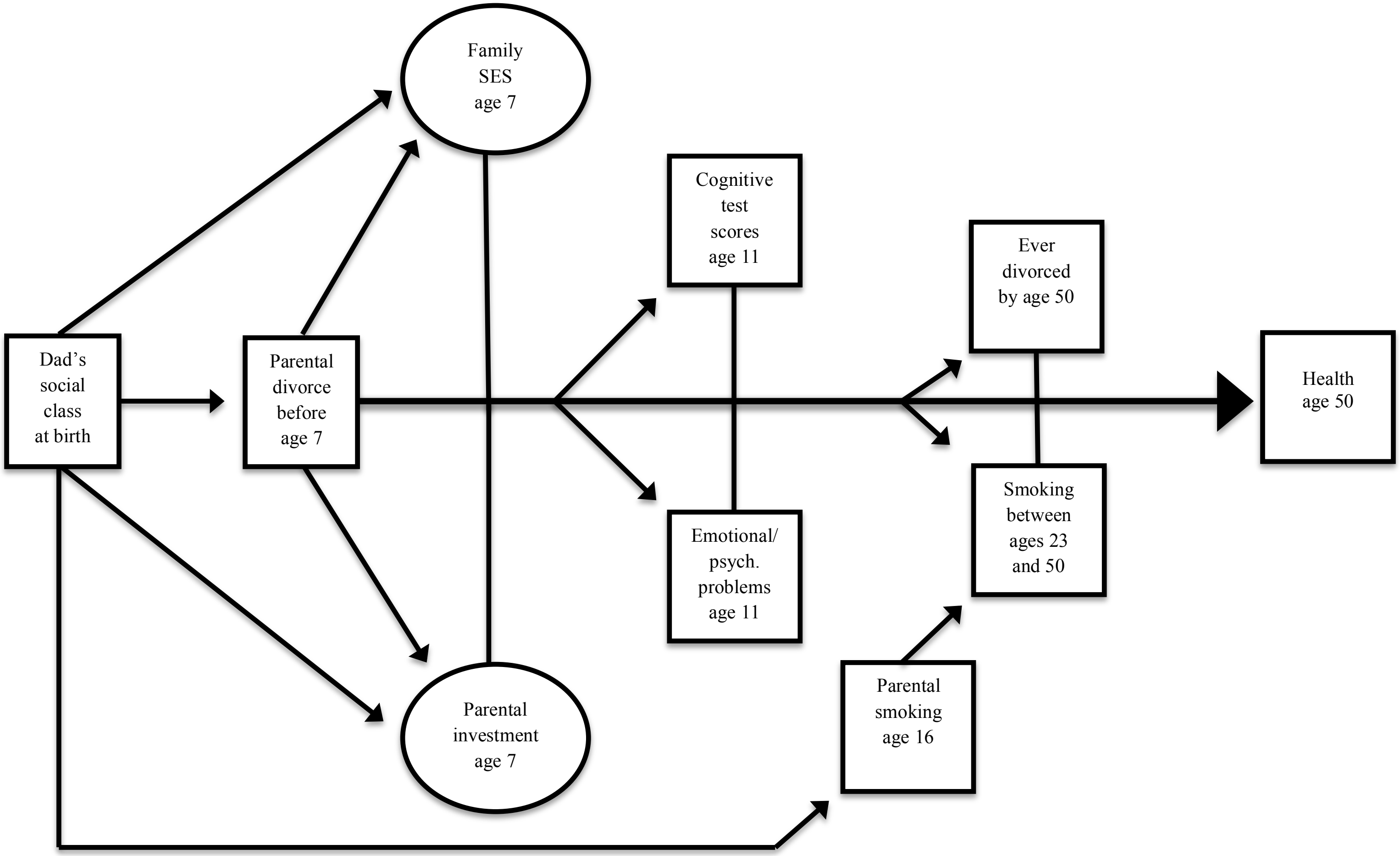


Figure 2 presents the structural equation model for parental divorce by age seven and the pathways leading to adult health. Our choice of mediating variables is motivated by previous research described earlier, but we are obviously restricted by the information available in the NCDS. The mediating variables included here are plausibly related to the consequent variables, and thus we adopt an exploratory approach to analyse the unique configuration of variable (and conditional associations) included in the model. More specifically, the model includes all of the direct effects of each variable on the subsequent variables, with the exception of father's social class at birth (an antecedent of parental divorce) and parental smoking (an exogenous variable). This specification undoubtedly will include pathways that are not statistically significant, but protects against excluding those that are relevant to the total effect of parental divorce on the child's adult health.

Figure 2. Structural model of the pathways linking parental divorce to health at age 50



Notes. Latent variables indicated with circles (see text for a description of the measurement model). Correlations between variables measured during the same wave are included in the model (arrows left out to avoid clutter). With the exception of parental smoking at age 16, the model includes direct effects of each variable on subsequent variables observed at older ages (as indicated by the connection to the bold path in the middle of the figure).

Our model includes two latent variables that represent the family's SES and the amount of parental involvement at age seven. Family SES is measured using the social class of the male head of the household, crowding in the cohort member's home, and the indicator of the family experiencing financial difficulties (all age seven). Parental involvement is measured by the following manifest variables: reading to the child; taking the child out; and the father's role in managing the child. Separate models are fitted to each of the three health outcomes at age 50, and the total effects are decomposed into the shares associated with each of the mediating variables. Parameter estimates are obtained from M*plus* version 7.11 (Muthén & Muthén, 2012) using weighted least squares with categorical outcomes modeled as latent variables with a probit specification. All models are estimated separately for men and women to allow for gender differences in the effects of a parental divorce (Amato, 2001), as well as for differences in smoking, health, and health assessment (Arber & Cooper, 1999; Macintyre, Hunt, & Sweeting; 1996; Peto et al., 2000).

Results

Descriptive statistics including the observed and imputed means and the percentage of missing data for each variable are presented in table 1. Several variables are missing for more than 40% of the cases, and when pooling information across waves to describe smoking and divorce histories in adulthood, the amount of missing information approaches 60%. We follow the advice of White, Royston, and Wood (2010) and impute 60 data sets to guard against the loss of power and to achieve an appropriate level of reproducibility for our results. By comparing the means from the imputed data to those of the observed data, we see that cohort members who experienced a parental divorce are more likely to have missing information, particularly during the youngest age interval. Similarly, less desirable outcomes (e.g., financial difficulties, low cognitive test scores, exhibiting behavioural problems) are also associated with missingness. Overall, however, the means from the imputed data are similar to those from the observed data.

Table 1. Means for observed and ten multiply imputed (MI) data sets N = 14,637

	Observed Data		MI
	Mean	% Missing	Mean
Male (0 = no, 1 = yes)	0.513	0%	0.513
Parental divorce by age 7 (0-1)	0.036	33%	0.054
Parental divorce between ages 7 & 11 (0-1)	0.022	33%	0.024
Parental divorce between ages 11 & 16 (0-1)	0.029	33%	0.034
Male head of households social class at birth (1 – 6)	3.157	0%	3.157
Male head of households social class at age 7 (1 – 6)	3.234	16%	3.243
Male head of households social class at age 11 (1 – 6)	3.241	21%	3.217
Financial difficulties for family at age 7 (0-1)	0.073	23%	0.086
Measure of crowding in the household at age 7 (1 – 3)	1.562	18%	1.572
Mother/father reads to the child at age 7 (0 – 4)	2.403	17%	2.402
Mother's report of father involvement at age 7 (0-2)	1.470	15%	1.460
Mother/father take child out at age 7 (0 – 2)	1.525	16%	1.517
Test scores at age 11 (standardized)	0.034	19%	-0.003
Social maladjustment at age 11 (1 – 3)	1.448	19%	1.478
Parents smoke when child is age 16 (0-1)	0.730	34%	0.736
Cohort member smokes in adulthood (0 – 6)	0.962	57%	1.123
Cohort member divorces by age 50 (0-1)	0.393	46%	0.468
Self-rated health at age 50 (1 – 3)	2.037	43%	2.102
Number of health problems at age 50 (1 – 3)	1.558	43%	1.558
Low physical functioning at age 50 (0-1)	0.117	48%	0.151

Notes: The numbers in parentheses indicate the range for the categorical variables. Multiple imputation estimates are averaged across 60 multiply imputed data sets. See the text for more details.

Total effect of parental divorce

Estimates of the total effect of parental divorce on adult health are presented in table 2. Moving across the columns, from left to right, we see gender-specific results for those who experienced a parental divorce by age seven, between ages seven and 11, and between ages 11 and 16. Results from our models predicting self-rated health at age 50 are presented in the first four rows. For each age interval, higher levels of social class are associated with lower chances of reporting poor health for both males and females. Conversely, parental divorce is positively associated with worse health, but the estimates are only statistically significant for males and females during the youngest age interval. This positive effect on worse health also extends to females between the ages of 11 and 16 at the $p < 0.10$ level of statistical significance. The indices of model fit suggest a close replication of the observed covariance matrix, which is the case for all of the models in table 2. Another finding that holds across all of the models in the table is that father's social class has a negative effect on parental divorce (results not shown). Turning to the middle panel in table 2, we see that parental divorce is positively associated with the number of health problems reported at age 50, but the estimates are not statistically significant. There are, however, statistically significant and positive effects of parental divorce experienced by age seven on low physical functioning (see the bottom panel of table 2).[7]

Table 2. Total effects of parental divorce, conditional on father's social class, estimated from gender-specific structural equation models of three health outcomes at age 50: self-rated health (SRH), # of health problems (NHP), and a measure of low physical functioning (PHF) – higher values for the dependent variable indicate worse health

	Parental Divorce between ages 0 and 7		Parental Divorce between ages 7 and 11		Parental Divorce between ages 11 and 16	
SRH	**Males**	**Females**	**Males**	**Females**	**Males**	**Females**
Father's Social Class	-0.208***	-0.241***	-0.202***	-0.251***	-0.207***	-0.263***
Parental divorce	0.139**	0.103**	0.056	0.043	0.025	0.088*
CFI/TLI	1.00/1.00	1.00/1.00	1.00/1.00	1.00/1.00	1.00/1.00	1.00/1.00
RMSEA	0.00	0.00	0.00	0.00	0.00	0.00
NHP	**Males**	**Females**	**Males**	**Females**	**Males**	**Females**
Father's Social Class	-0.050*	-0.104***	-0.044	-0.112***	-0.031	-0.089***
Parental divorce	0.076	0.076	0.058	0.045	0.003	0.052
CFI/TLI	1.00/1.00	1.00/1.00	1.00/1.00	1.00/1.00	1.00/1.00	1.00/1.00
RMSEA	0.00	0.00	0.00	0.00	0.00	0.00
PHF	**Males**	**Females**	**Males**	**Females**	**Males**	**Females**
Father's Social Class	-0.238***	-0.227***	-0.272***	-0.277***	-0.256***	-0.260***
Parental divorce	0.147**	0.150**	0.103	0.093	0.002	0.058
CFI/TLI	1.00/1.00	1.00/1.00	1.00/1.00	1.00/1.00	1.00/1.00	1.00/1.00
RMSEA	0.00	0.00	0.00	0.00	0.00	0.00
Smallest N	7,024	6,611	6,756	6,458	6,860	6,515

*Notes: Children of continuously married parents up to age 16 serve as the reference group for the parental divorce variable. P-values (two-tailed): * p < .10; ** p < .05; *** p < .01. Direct effects of father's social class on parental divorce not shown (all estimates are negative and statistically significant). Estimates are averaged across 60 multiply imputed data sets with standard errors corrected for variance within and between the data sets. CFI is the comparative fit index, TLI is the Tucker-Lewis index, and RMSEA is the root mean squared error of approximation. Results were obtained using the Mplus statistical software (version 7.11).*

Mediators of the effect of parental divorce on the child's adult health

We now turn to the structural model in figure 2 and the relative importance of different pathways stemming from a parental divorce that occurs by age seven. Estimates from the latent variable model for self-rated health are presented in tables 3 and 4 for males and females, respectively. Table 5 contains the corresponding estimates from the measurement model for both males and females. Most of the results are similar across the models for the three different health outcomes, thus we only present the results for self-rated health. Later, we present a decomposition of the total effect of parental divorce on each of the three health outcomes at age 50.

Table 3. Estimates from the latent variable model for the effects of a parental divorce by age seven on self-rated health at age 50 for males

				Outcomes			
	Family SES age 7	Parental Involvement age 7	Cognition age 11	Behavioral problems age 11	Smoking Adulthood	Divorce by age 50	Self-rated health age 50
Father's SC at birth	0.674***	0.112***					
	(0.026)	(0.018)					
Parental divorce by age 7	-0.352***	-0.481***	0.002	0.147**	0.240***	0.156*	0.034
	(0.021)	(0.053)	(0.034)	(0.055)	(0.063)	(0.080)	(0.069)
Family SES age 7			0.337***	-0.206***	-0.007	-0.091***	-0.069**
			(0.017)	(0.021)	(0.026)	(0.026)	(0.023)
Parental involvement age 7			0.032	0.017	0.066	0.034	-0.008
			(0.028)	(0.043)	(0.051)	(0.052)	(0.046)
Cognition age 11					-0.113***	-0.086**	-0.111***
					(0.027)	(0.032)	(0.026)
Behavioral problems age 11					0.121***	0.050	0.078**
					(0.026)	(0.033)	(0.027)
Smoking in adulthood							0.229***
							(0.024)
Parental Smoking					0.215***		
					(0.033)		
Divorce by age 50							0.006
							(0.026)

*Notes: Direct effects of father's social class on parental divorce not shown (all estimates are negative and statistically significant). Standard errors in parentheses. P-values: * p < .05; ** p < .01. Estimates are averaged across 60 multiply imputed data sets with standard errors corrected for variance within and between the data sets. The smallest sample size across the 60 imputed data sets is N = 7,024. Root mean square error of approximation = 0.058. Comparative fit index = 0.953. Tucker-Lewis index = 0.906. Results were obtained using the Mplus statistical software (version 7.11).*

Before discussing the effects of parental divorce among males, shown in table 3, we note that father's social class has a negative and statistically significant effect on parental divorce in all of our models (estimates not shown to save space). The results suggest that, net of father's social class at birth, parental divorce is associated with lower levels of both family SES and parental involvement (measured after the divorce). Conditional on family SES and parental involvement, a parental divorce is not associated with lower cognitive test scores at age 11, but it is predictive of more behavioural problems at that age. Parental divorce also has positive and statistically significant direct effects on smoking and divorce in adulthood. Finally, we see that the mediating variables account for practically all of the total effect. Family SES at age seven, cognition and behavioural problems at age 11, and smoking in adulthood all exert statistically significant, direct effects on self-rated health and thus serve as important mediators in extending the influence of parental divorce to health at age 50.

Table 4. Estimates from the latent variable model for the effects of a parental divorce by age 7 on self-rated health at age 50 for females

	Family SES age 7	Parental Involvement age 7	Cognition age 11	Outcomes Behavioral problems age 11	Smoking Adulthood	Divorce by age 50	Self-rated health age 50
Father's SC at birth	0.654*** (0.022)	0.062*** (0.014)					
Parental divorce by age 7	-0.407*** (0.046)	-0.372*** (0.038)	-0.017 (0.040)	0.157** (0.069)	0.208*** (0.067)	0.161** (0.081)	0.004 (0.067)
Family SES age 7			0.359*** (0.017)	-0.203*** (0.023)	-0.114*** (0.027)	-0.072** (0.028)	-0.097*** (0.023)
Parental involvement age 7			-0.013 (0.047)	0.012 (0.079)	0.129 (0.082)	0.095 (0.085)	-0.003 (0.077)
Cognition age 11					-0.062** (0.028)	-0.077** (0.035)	-0.123*** (0.029)
Behavioral problems age 11					0.149*** (0.029)	0.086** (0.033)	0.078** (0.027)
Smoking in adulthood							0.201*** (0.023)
Parental Smoking					0.183*** (0.033)		
Divorce by age 50							0.049* (0.027)

*Notes: Direct effects of father's social class on parental divorce not shown (all estimates are negative and statistically significant). Standard errors in parentheses. P-values: * p < .05; ** p < .01. Estimates are averaged across 60 multiply imputed data sets with standard errors corrected for variance within and between the data sets. The smallest sample size across the 60 imputed data sets is N = 6,611. Root mean square error of approximation = 0.060. Comparative fit index = 0.946. Tucker-Lewis index = 0.893. Results were obtained using the Mplus statistical software (version 7.11).*

The results for females (see table 4) are generally similar to those of males. Notable differences include a more important role of family SES in discouraging women to smoke, relative to men. Similarly, behavioural problems may be more influential on the risk of divorce for females compared to males, and divorce for women appears to have a stronger effect on self-rated health than that of men. With respect to the measurement model (see table 5), our results are very similar for females and males. We also note that, among both females and males, the model provides a reasonable fit to the data (see model fit statistics at the bottom of table 3 and 4). The values of the Tucker-Lewis Index (0.893 for women and 0.906 for men) reflect the penalty for model complexity, which is consistent with several of the estimated coefficients being statistically insignificant or marginally significant. Given the exploratory nature of our model, however, we proceed with the implications of this model.

Table 5. Estimates from the measurement model for the effects of a parental divorce by age 7 on self-rated health at age 50 for males and females

	Latent Variables			
	Family SES at age 7		Parental involvement at age 7	
	Males	**Females**	**Males**	**Females**
Father's social class measured at age 7	1.00	1.00		
	--	--		
Family has financial difficulty	-1.025***	-0.991***		
	(0.099)	(0.089)		
Crowding in the household measured at age 7	-0.540***	-0.541***		
	(0.027)	(0.027)		
Parents read to child measured at age 7			1.00	1.00
			--	--
Parents take child out with them most weeks measured at age 7			2.204***	2.783***
			(0.358)	(0.437)
Mother's report of father involvement measured at age 7			0.620***	0.897***
			(0.074)	(0.109)

*Standard errors in parentheses. P-values: * p < .05; ** p < .01. Estimates are averaged across 10 multiply imputed data sets with standard errors corrected for variance within and between the data sets. The smallest sample size across the 60 multiply imputed data sets is N = 7,024 and 6,611 for males and females (respectively). For males, the root mean square error of approximation = 0.058. Comparative fit index = 0.946. Tucker-Lewis index = 0.893. For females, the root mean square error of approximation = 0.060. Comparative fit index = 0.953. Tucker-Lewis index = 0.906. Results were obtained using the Mplus statistical software (version 7.11).*

Decompositions of the total effect of parental divorce by age seven on the three different health outcomes at age 50 are presented in tables 6 and 7 for males and females, respectively. The top row of results in each table consists of the total effect of parental divorce, the indirect effects associated with each of the mediators, and, finally, the direct effect of parental divorce on self-rated health at age 50. The next row contains the proportion of the total effect that is accounted for by the corresponding mediating variable in that column. Finally, in the third row of the tables, we try to illustrate the substantive importance of the effect sizes by showing the change in the predicted probability of fair/poor health that is associated with each component. These changes are calculated by adding the effect size to the mean of the latent variable, and reporting the associated change in the predicted probability. This calculation is a very crude approximation and should only serve as a rough guide in interpreting the results

Table 6. Total, direct, and indirect effects of parental divorce by age 7 on three health outcomes at age 50 for males: self-rated health (SRH), # of health problems (NHP), and physical functioning (PHF) – higher values for the dependent variable indicate worse health

Dependent variable	Total effect of parental divorce	Indirect effects of parental divorce by age 7						Direct effect of parental divorce
		Parental involvement	Family SES	Cognitive test scores	Emotional problems	Smoking in adulthood	Divorce by age 50	
Self-rated health	0.151**	-0.003	0.049***	0.000	0.015**	0.055***	0.001	0.034
% of total effect	100%	-2%	32%	0%	10%	36%	1%	23%
Expected change in predicted probability of fair/poor health	0.043	-0.001	0.013	0.000	0.004	0.015	0.000	0.010
# of health problems	0.072	-0.010	0.011	0.001	0.012*	0.006	-0.001	0.055
% of total effect	100%	-14%	15%	1%	17%	8%	-1%	76%
Expected change in predicted probability of 3+ health problems	0.022	-0.003	0.003	0.000	0.003	0.002	0.000	0.017
Physical functioning	0.166**	-0.018	0.074***	-0.001	0.019**	0.031**	0.001	0.059
% of total effect	100%	-11%	45%	-1%	11%	19%	1%	35%
Expected change in the predicted probability of low physical functioning[a]	0.036	-0.003	0.015	0.000	0.004	0.006	0.000	0.008

*Notes: (a) Physical functioning is defined by being one standard deviation below the gender-specific mean of the observed cohort members. P-values: * p < .05; ** p < .01. Estimates are averaged across 60 multiply imputed data sets with standard errors corrected for variance within and between the data sets. The smallest sample size across the 60 multiply imputed data sets is N = 7,024. Root mean squared approximation = 0.061 (SRH); 0.059 (NHP); and 0.061 (PHF). Comparative fit index = 0.949 (SRH); 0.948 (NHP); and 0.949 (PHF). Tucker-Lewis index = 0.898 (SRH); 0.901 (NHP); and 0.898 (PHF). Results were obtained using the MPlus statistical software (version 7.11).*

The findings suggest that, among males, smoking in adulthood and family SES each account for over 30% of the total effect of parental divorce, corresponding to a one or two percentage point change in the probability of reporting fair/poor health. We find a similar pattern among females, but the decline in family SES associated with a parental divorce accounts for over half of the total effect of parental divorce and a two percentage point increase in reporting fair/poor health, while smoking in adulthood contributes an additional 30%, or roughly a one percentage point increase in reporting fair/poor health. Before moving on, we wish to comment briefly on the role of cognition. While it does not play an important role on its own, cognition is involved with some of the most powerful explanatory pathways running through family SES. In other words, pathways that include both family SES and cognition account for 10% (males) and 15% (females) of the total effect of parental divorce.

Table 7. Total, direct, and indirect effects of parental divorce by age 7 on three health outcomes at age 50 for females: self-rated health (SRH), # of health problems (NHP), and physical functioning (PHF) – higher values for the dependent variable indicate worse health.

Dependent variable	Total effect of parental divorce	Indirect effects of parental divorce by age 7: Parental involvement	Family SES	Cognitive test scores	Emotional problems	Smoking in adulthood	Divorce by age 50	Direct effect of parental divorce
Self-rated health	0.142***	-0.012	0.080***	0.002	0.018**	0.042**	0.008	0.004
% of total effect	100%	-8%	56%	1%	13%	30%	6%	3%
Expected change in predicted probability of fair/poor health	0.041	-0.003	0.022	0.001	0.005	0.012	0.002	0.002
# of health problems	0.081	-0.010	0.035	0.000	0.012*	0.012**	0.007	0.024
% of total effect	100%	-12%	43%	0%	15%	15%	9%	30%
Expected change in predicted probability of 3+ health problems	0.026	-0.003	0.011	0.000	0.004	0.004	0.002	0.008
Physical functioning	0.190**	-0.030	0.091***	0.004	0.017*	0.022**	0.003	0.083
% of total effect	100%	-16%	48%	2%	9%	12%	2%	44%
Expected change in the predicted probability of low physical functioning[a]	0.046	-0.006	0.021	0.001	0.004	0.005	0.001	0.021

*Notes: (a) Physical functioning is defined by being one standard deviation below the gender-specific mean of the observed cohort members. P-values: * p < .05; ** p < .01. Estimates are averaged across 60 multiply imputed data sets with standard errors corrected for variance within and between the data sets. The smallest sample size across the 60 multiply imputed data sets is N = 6,611. Root mean squared approximation = 0.060 (SRH); 0.060 (NHP); and 0.060 (PHF). Comparative fit index = 0.946 (SRH); 0.945 (NHP); and 0.945 (PHF). Tucker-Lewis index = 0.893 (SRH); 0.891 (NHP); and 0.892 (PHF). Results were obtained using the MPlus statistical software (version 7.11).*

Turning briefly now to the results for the number of health problems reported at age 50, found in the fourth row of results in tables 6 and 7, we are reminded that the total effect of parental divorce is not statistically significant. It may be worth noting that behavioural problems may provide some link between characteristics early in life and subsequent health problems in adulthood, but the results provide, at best, only very weak support for this claim. Finally, the bottom three rows of results in tables 6 and 7 correspond to the effects of parental divorce on low physical functioning at age 50. The results suggest that, for both males and females, nearly half of the total effect depends on the decline in family SES that occurs after a parental divorce. Our crude approximation suggests that this change is associated with 1.5 percentage point increase in the probability of males being one standard deviation below the mean score for males on the physical functioning scale. The corresponding increase for females is roughly two percentage points. Again, smoking and behavioural problems also account for a significant portion of the total effect, but the contributions are each less than 20% of the total effect.

Discussion

The growing literature on the early origins of adult health is uncovering a wide array of conditions and experiences that can slant the life course trajectory towards less desirable outcomes later in life. We add to this literature by providing evidence of a total effect of parental divorce on adult health, and by identifying the most important mediators that transmit the total effect. Among individuals in the 1958 NCDS, it is primarily the cohort members who experience a parental divorce before age seven that experience a long-lasting effect on their adult health. The evidence suggests that a decline in family SES experienced after a parental divorce is the most important change in the early environment that perpetuates negative outcomes. Subsequent declines in the accumulation of cognitive skills help complete the link to poorer health at age 50. This finding is consistent with the idea that SES allows parents to make investments via economic, social, and cultural capital, which have important returns for cognitive ability and, ultimately, adult health (Conti & Heckman, 2010; Farkas, 2003). Pathways stemming from a parental divorce and directly running through smoking in adulthood generally make the second largest contribution to the total effect. Parental divorce also operates through behavioural problems to influence adult health, although the relative size of this contribution is smaller. These final two pathways are connected, suggesting that smoking may be a behavior adopted by children to deal with psychological and emotional stresses associated with a parental divorce (although, see Wolfinger, 1998).

We conducted separate analyses for women and men, and found that the story was very similar. One notable exception is that the relative importance of smoking is much larger among males, relative to females, in our models of self-rated health. This difference is primarily due to the finding that family SES exerts a strong, negative effect on the smoking behaviour of female cohort members in adulthood. Among men, however, there is no direct link, thus family SES and smoking do not work in tandem to produce worse health at age 50, which is the case for females.

Despite the meager effect size supported by our analysis, further study of the role of parental divorce on adult health is warranted because our analysis may understate the importance of this early experience. A major limitation of our study is that the measures of parental divorce lack precise information on timing and the number of family transitions experienced. A reasonable hypothesis is that the negative effects of a parental divorce are exacerbated if the event occurs during a critical period of child development, most likely at a relatively early age. The windows of time we are looking through (i.e., before age seven, and between ages seven and 16) are potentially too broad for us to identify the most vulnerable or influential periods. Furthermore, experiencing multiple parental divorces or changes in family structure may also strengthen the total effect on the child's adult health (Fomby & Cherlin, 2007). That said, we were unable to assess the problem of selection biases (Amato, 2010; Cherlin et al., 1991), and thus our estimates of the total effect of parental divorce may be overstated.

As discussed earlier, our analysis also suffers from the lack of information on potentially important mediators, such as parenting quality, inter-parental conflict, and the number of stressful life events associated with a parental divorce (Amato, 1993). Furthermore, our measurement of parental involvement could be improved by including additional types of investments made by parents, such as helping with schoolwork or hiring tutors. Neglecting some mediators and measuring others with insufficient information biases the assessment of the relative contribution of the mediators. Future research should move in this direction and extend our analysis of how parental divorce influences child health in adulthood. Such efforts will be useful for developing strategies to intervene and offset the negative consequences of early life stressors.

Finally, we return to our finding that the long-term effect of parental divorce on the child's adult health pertains only to those in the youngest age group. This result points to the conceptualization of child development as a cumulative process where early experiences with the disadvantages associated with parental divorce bend life course trajectories toward less desirable outcomes. More research is need to understand how parental divorce alters the developmental trajectories of children (e.g., Lansford et al., 2006), particularly with respect to long-term effects, such as adult health. A natural extension is to view adult health as a dynamic process, where both the intercept and slope of health trajectories are functions of early life experiences, such as parental divorce. We believe this to be a useful way forward for strengthening our understanding of the long-term effects of parental divorce.

Acknowledgements

This research was supported by the National Institutes of Health, Population Research Institute Center Grant, R24HD041025, as well as the center grant to the Center for Demography and Ecology (5R24HD04783). Financial support was also provided by the National Institute of Aging (5P30AG017266). The second author's time was also supported by an ARRA supplement to NICHD grant #R01HD57894.

References

Adams, P., Hurd, M.D., McFadden, D.L., Merrill, A., & Riberio, T. (2003). Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status. *Journal of Econometrics, 112*, 3-56. http://dx.doi.org/10.1016/S0304-4076(02)00145-8

Albouy, V., & Lequien, L. (2009). Does compulsory education lower mortality? *Journal of Health Economics, 28*, 155-168. http://dx.doi.org/10.1016/j.jhealeco.2008.09.003

Almond, D., & Currie, J. (2011). Human capital development before age five. *Handbook of Labor Economics, 4,* 1315-1486.

Alwin, D.F. & Hauser, R.M. (1975). The decomposition of effects in path analysis. *American Sociological Review, 40,* 37-47. http://dx.doi.org/10.2307/2094445

Amato, P.R. (1993). Children's adjustment to divorce: theories, hypotheses, and empirical support. *Journal of Marriage and the Family, 55,* 23-38. http://dx.doi.org/10.2307/352954

Amato, P.R. (2001). Children of divorce in the 1990s: an update of the Amato and Keith (1991) meta-analysis. *Journal of Family Psychology, 15,* 355-370. http://dx.doi.org/10.1037/0893-3200.15.3.355

Amato, P.R. (2010). Research on divorce: continuing trends and new developments. *Journal of Marriage and Family, 72*, 650-666. http://dx.doi.org/10.1111/j.1741-3737.2010.00723.x

Amato, P.R., & DeBoer. D.D. (2001). The transmission of marital instability across generations: relationship skills or commitment to marriage? *Journal of Marriage and Family, 63*, 1038-1051.

http://dx.doi.org/10.1111/j.1741-3737.2001.01038.x
Arber, S., & Cooper, H. (1999). Gender differences in health in later life: the new paradox? *Social Science & Medicine, 48,* 61-76. http://dx.doi.org/10.1016/S0277-9536(98)00289-5

Banks, J., Marmot, M., Oldfield, Z., & Smith, J.P. (2006). Disease and disadvantage in the United States and in England. *Journal of the American Medical Association, 295*, 2037-2045. http://dx.doi.org/10.1001/jama.295.17.2037

Barber, J.S. (2001). The intergenerational transmission of age at first birth among married and unmarried men and women. *Social Science Research, 30*, 219-247. http://dx.doi.org/10.1006/ssre.2000.0697

Baumrind, D. (1986). Authoritarian versus authoritative parental control. *Adolescence, 3*, 255-272.

Bengtson, V.L. (1975). Generation and family effects in value socialization. *American Sociological Review, 40*, 358-371. http://dx.doi.org/10.2307/2094463

Carlson, M.J. & Corcoran, M.E. (2001). Family structure and children's behavioral and cognitive outcomes. *Journal of Marriage and Family, 63,* 779-792. http://dx.doi.org/10.1111/j.1741-3737.2001.00779.x

Chase-Lansdale, P. L., Cherlin, A.J., & Kiernan, K.E. (1995). The long-term effects of parental divorce on the mental health of young adults: a developmental perspective. *Child Development, 66*, 1614-1634. http://dx.doi.org/10.2307/1131900

Cherlin, A.J., Chase-Lansdale, P.L., & McRae, C. (1998). Effects of divorce on mental health throughout the life course. *American Sociological Review, 63*, 239-249. http://dx.doi.org/10.2307/2657325

Cherlin, A., Furstenberg Jr., F., Chas-Linsdale, P.L., Kiernan, K.E., Robins, P.K., Morrison, D.R., Teitler, J.O. (1991). Longitudinal studies of effects of divorce on children in Great Britain and the United States. *Science*, *252, 1386-1389*. http://dx.doi.org/10.1126/science.2047851

Conti, G., & Heckman, J.J. (2010). Understanding the early origins of the education-health gradient: a framework that can also be applied to analyze gene-environment interactions. *Perspectives on Psychological Science*, *5*, 585-605. http://dx.doi.org/10.1177/1745691610383502

Cunha, F. & Heckman, J.J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources, 43*, 738-782. http://dx.doi.org/10.1353/jhr.2008.0019

Cutrona, C.E. (1996). *Social Support in Couples: Marriage as a Resource in Times of Stress*. Thousand Oaks, CA: Sage Publications. http://dx.doi.org/10.4135/9781483327563

Cutler, D.M. & Lleras-Muney, A. (2010). Understanding differences in health behaviors by education. *Journal of Health Economics, 29*, 1-18. http://dx.doi.org/10.1016/j.jhealeco.2009.10.003

Elo, I.T. (2009). Social class differentials in health and mortality: patterns and explanations in comparative perspective. *Annual Review of Sociology, 35*, 553-572. http://dx.doi.org/10.1146/annurev-soc-070308-115929

Farkas, G. (2003). Cognitive skills and noncognitive traits and behaviors in stratification processes. *Annual Review of Sociology, 29,* 541-562. http://dx.doi.org/10.1146/annurev.soc.29.010202.100023

Ferri, E. (1993). *Life at 33: The Fifth Follow-up of the National Child Development Study*. London: National Children's Bureau, City University Economic and Social Research Council.

Fomby, P., & Cherlin, A.J. (2007) Family instability and child well-being. *American Sociological Review*, *72*, 181-204.

http://dx.doi.org/10.1177/000312240707200203

Forehand, R., Neighbors, B., Devine, D., & Armistead, L. (1994). Interparental conflict and parental divorce: the individual, relative, and interactive effects. *Family Relations, 43*, 387-393. http://dx.doi.org/10.2307/585369

Furstenberg, F.F. & Kiernan, K.E. (2001). Delayed parental divorce: how much do children benefit? *Journal of Marriage and Family, 63,* 446-457. http://dx.doi.org/10.1111/j.1741-3737.2001.00446.x

Gohdsian, M. (1977). Children's behaviour and the BSAG: Some theoretical and statistical considerations. *British Journal of Clinical Psychology, 16*, 23-28. http://dx.doi.org/10.1111/j.2044-8260.1977.tb00999.x

Goodman, Alissa, Robert Joyce, and James P. Smith. 2011. The Long Shadow Cast by Childhood Physical and Mental Problems on Adult Life. *Proceedings of the National Academy of Sciences* 108(15):6032-6037. http://dx.doi.org/10.1073/pnas.1016970108

Hayward, M.D., & Gorman, B.K. (2004). The long arm of childhood: the Influence of early-life social conditions on men's mortality. *Demography, 41*, 87-107. http://dx.doi.org/10.1353/dem.2004.0005

Heckman, J.J. (2007). The economics, technology, and neuroscience of human capability formation. *Proceedings of the National Academy of Sciences, 104,* 13250-13255. http://dx.doi.org/10.1073/pnas.0701362104

Hemminki, K., & Chen, B. (2006). Lifestyle and cancer: effect of parental divorce. *European Journal of Cancer Prevention, 15,* 524-530. http://dx.doi.org/10.1097/01.cej.0000220633.93104.64

Hetherington, E.M., Stanley-Hagan, M., & Anderson, E.R. 1989. Marital transitions: a child's perspective. *American Psychologist, 44,* 303-312. http://dx.doi.org/10.1037/0003-066X.44.2.303

House, J.S., Landis, K.R., & Umberson, D. (1988). Social Relationships and Health. *Science, 241*, 540-545. http://dx.doi.org/10.1126/science.3399889

Izzo, C.V., Weissberg, R.P., Kasprow, W.J., & Fendrich, M. (1999). A longitudinal assessment of teacher perceptions of parent involvement in children's educational and school performance. *American Journal of Community Psychology, 27,* 817-839. http://dx.doi.org/10.1023/A:1022262625984

Kelly, J.B. (2007). Children's living arrangements following separation and divorce: insights from empirical and clinical research. *Family Process, 46*, 35-52. http://dx.doi.org/10.1111/j.1545-5300.2006.00190.x

Kurdek, L.A., Fine, M.A., & Sinclair, R.J. (1994). The relation between parenting transitions and adjustment in young adolescents. *Journal of Early Adolescence, 14*, 412-432. http://dx.doi.org/10.1177/0272431694014004001

Lansford, J.E., Malone, P.S., Castellino, D.R., Dodge, K.A., Pettit, G.S., & Bates, J.E. (2006). Trajectories of internalizing, externalizing, and grades for children who have and have not experienced their parents' divorce or separation. *Journal of Family Psychology, 20,* 292-301. http://dx.doi.org/10.1037/0893-3200.20.2.292

Larson, K., & Halfon, N. (2013). Parental divorce and adult longevity. *International Journal of Public Health, 58,* 89-97. http://dx.doi.org/10.1007/s00038-012-0373-x

Macintyre, S., Hunt, K., & Sweeting, H. (1996). Gender differences in health: are things really as simple as they seem? *Social Science & Medicine, 42,* 617-624. http://dx.doi.org/10.1016/0277-9536(95)00335-5

Maier, E.H., & Lachman, M.E. (2000). Consequences of early parental loss and separation for health and well-being in midlife. *International Journal of Behavioral Development, 24,* 183-189. http://dx.doi.org/10.1080/016502500383304

Martin, L.R., Friedman, H.S., Clark, K.M., & Tucker, J.S. (2005) Longevity following the experience of parental divorce. *Social Science Medicine, 61,* 2177-2189. http://dx.doi.org/10.1016/j.socscimed.2005.04.027

Maughan, B., & McCarthy, G. (1997). Childhood adversities and psychosocial disorders. *British Medical Bulletin, 53,* 156-169. http://dx.doi.org/10.1093/oxfordjournals.bmb.a011597

McLanahan, S., & Bumpass. L. (1998). Intergenerational consequences of family disruption. *American Journal of Sociology, 94*, 130-152. http://dx.doi.org/10.1086/228954

McLanahan, S., & Sandefur, G. (1994). *Growing Up with a Single Parent: What Hurts, What Helps*. Cambridge, MA: Harvard University Press.

Meara, E.R., Richards, S., & Cutler, D.M. (2008). The gap gets bigger: changes in mortality and life expectancy, by education, 1981-2000. *Health Affairs, 27*, 350-360. http://dx.doi.org/10.1377/hlthaff.27.2.350

Miedel, W.T., & Reynolds, A.J. (2003). Parent involvement in early intervention for disadvantaged children: does it matter? *Journal of School Psychology, 37,* 379-402. http://dx.doi.org/10.1016/S0022-4405(99)00023-0

Muthén, L.K., & Muthén, B.O. (2012). Mplus user's guide (7th ed). Muthén & Muthén, Los Angeles.

Office for National Statistics. (2013). *Vital Statistics: Population and Health Reference Tables, Summer 2013 update*. Accessed 27th August 2013 from http://www.ons.gov.uk/ons/rel/vsob1/vital-statistics--population-and-health-reference-tables/summer-2013-update/rtd-annual-table.xls

Pampel, F.C., Krueger, P.M., & Demney, P.T. (2010). Social disparities in health behaviors. *Annual Review of Sociology, 36*, 349-370. http://dx.doi.org/10.1146/annurev.soc.012809.102529

Peto, R., Darby, S., Deo, H, Silcocks, P., Whitley, E., & Doll, R. (2000). Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. *British Medical Journal, 321,* 323-329. http://dx.doi.org/10.1136/bmj.321.7257.323

Potter, D. (2010). Psychosocial well-being and the relationship between divorce and children's academic achievement. *Journal of Marriage and Family, 72*, 933-946. http://dx.doi.org/10.1111/j.1741-3737.2010.00740.x

Power, C., & Elliott, J. (2006). 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

Preston, S.H., Hill, M.E., & Drevenstedt, G.L. (1998). Childhood conditions that predict survival to advanced ages among African-Americans. *Social Science Medicine, 47*, 1231-1246. http://dx.doi.org/10.1016/S0277-9536(98)00180-4

Royston, P., White, I.R. (2011). Multiple imputation by chained equations (MICE): Implementation in Stata. *Journal of Statistical Software, 45,* 1-20.

Schwartz, J.E., Friedman, H.S., Tucker, J.S., Tomlinson-Keasy, C., Wingard, D.L., & Criqui, M.H. (1995). Sociodemographic and psychosocial factors in childhood as predictors of adult mortality. *American Journal of Public Health, 85*, 1237-1245. http://dx.doi.org/10.2105/AJPH.85.9.1237

Stott, D.H. (1963). The social adjustment of children: Manual of the Bristol social adjustment guides (2nd ed). London: University of London Press.

Stott, D.H. (1969). The social adjustment of children: Manual of the Bristol social Adjustment guides (3rd ed). London: University of London Press.

Tach, L., Mincy, R., & Edin, K. (2010). Parenting as a 'Packaged Deal': relationships, fertility, and nonresident father involvement among unmarried parents. *Demography, 47*, 181-204. http://dx.doi.org/10.1353/dem.0.0096

Thompson, E., Hanson, T.L., & McLanahan, S.S. (1994). Family structure and child well-being: economic resources vs. parental behaviors. *Social Forces, 73*, 221-242. http://dx.doi.org/10.1093/sf/73.1.221

Tillman, K.H. (2007). Family structure pathways and academic disadvantage among adolescents in stepfamilies. *Sociological Inquiry, 77*, 383-424. http://dx.doi.org/10.1111/j.1475-682X.2007.00198.x

Troxel, W.M., & Matthews, K.A. (2004). What are the costs of marital conflict and dissolution to children's physical health? *Clinical Child and Family Psychological Review, 7,* 7-29. http://dx.doi.org/10.1023/B:CCFP.0000020191.73542.b0

Ware Jr., J.E., Snow, K.K., Kosinski, M., & Gandek, B. (1993). SF-36 Health Survey Manual and Interpretation Guide. Boston, MA: New England Medical Center, The Health Institute.

White, I.R., Royston, P., Wood, A.M. (2010). Multiple imputation using chained equations: issues and guidance for practice. *Statistics in Medicine, 30,* 377-399. http://dx.doi.org/10.1002/sim.4067

Williams, K., & Umberson, D. (2004). Marital status, marital transitions, and health: a gendered life course perspective. *Journal of Health and Social Behavior*, *45*, 81-98. http://dx.doi.org/10.1177/002214650404500106

Wolfinger, N.H. (1998). The effects of parental divorce on adult tobacco and alcohol consumption. *Journal of Health and Social Behavior, 39,* 254-269. http://dx.doi.org/10.2307/2676316

Woodward, L., Fergusson, D.M., & Belsky, J. (2000). Timing of parental separation and attachment to parents in adolescence: results of a prospective study from birth to age 16. *Journal of Marriage and Family, 62*, 162-174 http://dx.doi.org/10.1111/j.1741-3737.2000.00162.x

Endnotes

[1] It should be noted, however, that step-parent investments in non-biological children potentially offset negative outcomes associated with fewer investments from noncustodial, biological parents (Amato, 1993).

[2] There are too few deaths to estimate the effect of parental divorce on mortality with much precision. An alternative approach is to group the deceased with those reporting the worst health outcomes at age 50. We do not adopt this strategy because of the concentration of observed deaths at younger ages and the concomitant lack of information on the intermediate variables leading up to health at age 50. Thus, our results are only generalisable to those who survive to age 50. Unobserved (and thus unknown) deaths are, however, included in our analysis via multiple imputation, which raises some question about our results; although, we have conducted the same analyses using only participants observed at age 50 and reached similar conclusions.

[3] We did explore the use of low birth weight, but the inclusion of this variable did not change the results and was not associated with parental divorce.

[4] We implement multiple imputation by chained equations using the *ice* command in Stata version 12.1 (Royston & White, 2011). All of the variables used in the analysis are included in the imputation equations, along with many other measures related to socioeconomic status, health, divorce, and the likelihood of missing data (e.g., the percent of missing data for each case, and the number of times the cohort member has moved). Additional details are available from the authors upon request.

[5] The 16 health problems are: asthma or wheezy bronchitis; seasonal or perennial allergic rhinitis; (sugar) diabetes; convulsions, fits, or epileptic seizures; recurrent backache, prolapsed disc, or sciatica; cancer or leukaemia; problems with hearing; problems with eyesight; high blood pressure; migraines; eczema or other skin problems; chronic fatigue syndrome; problems with stomach, bowels, or gall bladder; problems with bladder or kidneys; and cough or bringing up phlegm.

[6] The physical functioning scale is based on ten items, inquiring if the cohort member's health has limited them in the following activities: (1) vigorous activities such as running, lifting heavy objects, or participating in strenuous sports; (2) moderate activities such as moving a table, pushing a vacuum cleaner, bowling, or playing golf; (3) lifting or carrying groceries; (4) climbing several flights of stairs; (5) climbing one flight of stairs; (6) bending, kneeling, or stooping; (7) walking more than one mile; (8) walking half a mile; (9) walking 100 yards; and (10) bathing or dressing yourself. Responses to each item are coded with values 0 -- "no, not limited at all"; 50 – "yes, limited a little", and 100 – "yes, limited a lot", then the average taken across the items is calculated to provide the value for the physical functioning scale.

[7] A related question concerns the need to focus on health at age 50, as opposed to an earlier measure of health. In other words, does all of the total effect of parental divorce on health at age 50 operate through health at a younger age? We examine this hypothesis by estimating models that include self-rated health at age 33 as a mediator between parental divorce and our three health measures at age 50. We find that among males parental divorce has a statistically significant direct effect on self-rated health age 50, net of self-rated health at age 33. Similar results hold for the direct effect of parental divorce on low physical functioning at age 50, net of self-rated health at age 33, and for males and females.

Longitudinal and Life Course Studies 2015 Volume 6 Issue 3 Pp 303 – 318 **ISSN 1757-9597**

Trajectories of functional disability for the elderly in Britain

Robert French Centre for Multilevel Modelling, University of Bristol
R.French@bris.ac.uk
Fiona Steele Department of Statistics, London School of Economics and Political Science

(Received August 2014 Revised June 2015) http://dx.doi.org/10.14301/llcs.v6i3.317

Abstract

This study uses an innovative approach to characterise trajectories of functional disability over the final stages of the life course. We use data from the British Household Panel Survey (BHPS), an annual household survey of all adults in a representative sample of British households from 1991-2008. The analysis focuses on the sub-sample of elderly household members who were aged from 65 to 74 in any of the 18 waves of data, with a final sample of 3,671 individuals contributing a total of 13,982 person years. As in previous research, we estimate latent growth curves, but extend the standard model to incorporate a measurement model for the latent outcome variable 'functional disability'. We identify accelerating trajectories of functional disability for a representative sample of elderly individuals separately by gender. We show that socio-occupational classification is associated with the level of initial functional disability and to a lesser extent the increase in functional disability with age. The contribution of this paper is to explore the use of a measurement model to exploit the variation between items in discriminatory power for identifying an individual's functional disability. Further we are able to explicitly test for temporal measurement invariance in functional disability i.e. to what extent the items consistently measure the latent variable as people age.

Keywords: Ageing, activities of daily living, health trajectories, Britain, British Household Panel Survey (BHPS), structural equation model (SEM), growth model, measurement model, temporal measurement invariance

Introduction

The ageing population in the UK is the result of a reduced birth rate and delayed mortality. Delayed mortality may result in a change in the prevalence of morbidity, either increasing (Verbrugge, 1984) or decreasing (Fries, 1980), which has implications for health care costs. For a well-informed policy response to this ageing population one needs well-defined measures of health for the elderly, and to establish how these measures progress with age, and how the level and nature of change with age differs between individuals.

The ageing process is typically represented by a trajectory of declining health, defined by increasing disability (Grundy & Glaser, 2000), or diminishing quality of life (Zaninotto, Falaschetti, & Sacker, 2009), self-rated health (Sacker, Worts, & McDonough, 2011), physical performance (Payette et al., 2011), or ability to carry out everyday activities (Haas, 2008). In this study we are concerned with a functional definition of health - how far health limits an individual's ability to enjoy a normal life - rather than a medical definition or diagnosis, since it allows comparability between individuals across a variety of different health conditions (Burchardt, 2000). This is typically measured using questions regarding individuals' ability to undertake everyday tasks over several domains. The first core set of such questions is the activities of daily living (ADL) (Katz, Ford, Moskowitz, Jackson, & Jaffe, 1963), though the ADL term is now used generically to describe a wide variety of question sets that attempt to capture the same construct. Extensions to ADL include the instrumental activities of daily living (iADL) (Lawton & Brody, 1969) which includes higher level tasks and SF-36 (Ware & Sherbourne, 1992) which captures social functioning. Because ADL are measured over different domains it is useful to combine these items into a single metric of functional disability for analysing changes in health.

We argue that the methods used to combine the ADL scores for models of change in functional disability have two important limitations. Firstly, these studies use simple aggregations of individual items, such as the sum of ADL scores, to create the single metric for analysing change, typically assigning equal or arbitrarily-chosen differential weights to each activity. This approach ignores variability between items in their relative difficulty and in their ability to discriminate between individuals with different levels of physical functioning. Secondly, previous studies have assumed that the difficulty and discriminatory power of items is the same at each age, widely referred to as *temporal/longitudinal measurement invariance*. Departures from this assumption due to changes in the relationship between the observed items and the underlying construct with age observed in cross-sectional studies (LaPlante, 2010) will confound attempts to identify growth patterns.

We propose a measurement model to more fully capture the underlying structure of functional disability. The measurement model includes parameters representing the difficulty and discriminatory power of each item. Making explicit the relationship between the items and the latent construct (physical functioning) in this way also allows exploration and testing of temporal measurement invariance. Provided that temporal measurement invariance holds, the measurement model can then be combined with a growth model for latent physical functioning. We allow our models of change in functional disability to differ by gender and socioeconomic status. We estimate separate models for each gender because we expect functional disability trajectories for women to show worse health for biological, psychological and sociological factors (Nathanson, 1975). We allow trajectories to vary by socioeconomic status as we expect a social gradient through accrued exposure to risk factors, both in terms of direct effects from certain types of employment, but also from the indirect risk factors and mediating factors associated with class (Nilsson, Avlund, & Lund, 2010).

Review of approaches to modelling trajectories of physical functioning

Trajectories of functional disability can be estimated using either a multilevel model (MLM) or structural equation model (SEM). In their simplest form, these models are equivalent (Curran, 2003; Steele, 2008). Both allow for individual-specific trajectories with normally distributed latent variables representing individual departures from the intercept and slope of an overall growth curve, and both can be extended to allow for nonlinear growth. These latent variables are usually referred to as random effects in MLM and factors in SEM.

In a MLM for growth the repeated health measurements are viewed as a two-level hierarchical structure with occasions nested within individuals and age is treated as a time-varying explanatory variable (Goldstein & Woodhouse, 2001). The advantage of the MLM approach is that it is very flexible, with possible extensions to the basic growth curve including allowance for additional levels of clustering and between-individual variation in the timing of measurements at a given occasion. Individuals not present at all measurement points can be included under a 'missing at random' assumption (Little & Rubin, 2002), however those with missing items within a wave require multiple imputation of missing values in order for that wave to be included.

In a SEM for growth, the measures at each occasion are treated as the observed indicators of the unobserved latent growth factors, i.e. latent variables for the individual-specific intercepts and slopes. The advantage of the SEM approach is the ability to include additional latent variables, for example to allow for measurement error in outcomes or covariates. It is also straightforward using SEM software to incorporate individuals with incomplete data within waves.

Several studies have fitted a latent growth curve model to trajectories of functional disability. Li (2005) estimates a two-level random effects model of 'ADL disability' using the Michigan's Medicaid Waiver Program of individuals aged 65+ measured every few months from 1999 to 2003. They find evidence of an accelerating trajectory of ADL disability for the whole sample. Park et al. (2008) use a similar model of 'functional status' for the

University of Alabama at Birmingham Ageing Study which surveys individuals aged 65+ every 6 months from 1999 to 2004. They also find increasing and accelerating functional disability. Mendes de Leon et al. (2002) estimate a MLM of 'ADL disability' using the Women's Health and Ageing Study which follows women aged 65+ for 24 consecutive weekly assessments in 1992. They find a linear increase in functional disability over this much shorter time frame. Haas (2008) estimates a latent growth curve of 'functional limitations' using the Health and Retirement Study which follows individuals aged 61-71 at baseline annually from 1992-2002, and find an increasing and accelerating trajectory for functional limitations.

These papers use unconditional models as a baseline and to identify functional form, and conditional models to quantify how these trajectories differ by individual characteristics. Stuck et al. (1999) review the individual risk factors for ADL decline. In this study we focus on two of the most common: gender and SES. Although females live longer than males, women generally have higher reported illness (Nathanson, 1975). There are many reasons for these differences, for example biological factors such as genes and hormones make males more susceptible to diseases that result in death, e.g. heart disease, while women are more likely to suffer from conditions which impact on reported health but not death, e.g. arthritis (Case & Paxson, 2005). Moreover, there are gender differences in acquired risks: for example men are more likely to smoke and drink while females are more likely to be overweight and face stress (Verbrugge, 1989). The SES gradient in health arises from direct risk factors associated with occupation, e.g. physical hazards and psychosocial stressors at work, but also from risk behaviours associated with class, e.g. smoking and heavy alcoholic drinking (Feng et al., 2013). Over the life course we expect the SES effect to increase as exposure lengthens (Sacker, Clarke, Wiggins, & Bartley, 2005), but once an individual retires the SES effects accrued during the working life may diminish as exposure to certain risk factors associated with work cease (House, Kessler, & Herzog, 1990).

Measurement of physical functioning in longitudinal studies

All of the methods of studying longitudinal change in functional disability discussed above use a single health outcome variable created using answers to a series of questions. The simplest approach to creating a single measure from these multiple questions is to sum ADL scores on each question. For example, Li (2005) uses questions on eight activities, with responses coded between zero (no limitation) and four (maximum limitation). These scores are summed across the eight items to generate the functional disability outcome measure. Using the total ADL score is problematic since each component is given equal weight, thus ignoring variation in the discriminatory power of the different items. Others studies use ad hoc methods to assign different weights to the items. For example, Holstein et al. (2006) measure levels of difficulty for 12 ADL items, and use these to create four categories of functional disability: (i) individuals who can manage all items without difficulty, (ii) individuals who can manage every activity but some with difficulty, (iii) those who need help in at least one category, and (iv) those who need help with two or more activities. Such an approach compounds the problem of equal weighting of different items by then using arbitrary thresholds for categorisation; it also ignores much of the information contained in the responses. We propose to use a measurement model to generate a single metric for functional limitations, which has the advantage that it allows each of the activities to have its own relationship with the latent outcome variable, rather than imposing equal or arbitrary weights.

Methods

Data and measures

Data for the study are from the British Household Panel Survey (BHPS), an annual household survey of all adults in a representative sample of British households from 1991-2008 (Institute for Social and Economic Research, 2010). Elderly household members (aged 65 or over) were asked additional questions on their ability to carry out activities of daily living and these formed the sample for analysis. We have 1,712 males and 1,959 females, contributing a total of 13,982 person years. Individuals not present for all waves of the survey were still included in the analysis, as were cases with missing data on some of the ADL measures for a particular year. Wave non-response and missing data on ADL items are handled using maximum likelihood methods under a missing at random assumption (Little & Rubin, 2002) in Mplus (Muthén & Muthén, 1998-2012).

Our observed indicators of physical functioning are the ADL items: 'get in and out of bed', 'cut toenails', 'get up and down stairs or steps', 'bath, shower or wash all over', 'get around the house', and 'walk down the road'. The score for each ADL item was constructed from responses to two questions: whether the individual is able to carry out an ADL (Q1 coded unaided, aided or not at all) and, for those who answered "unaided", the level of difficulty in performing the ADL (Q2 coded very easy, fairly easy, fairly difficult or very difficult). Thus for each ADL we can construct a six-point score, ranging from zero for those with the least disability who could carry out the ADL unaided (Q1) and very easily (Q2) to five for those with the most disability who could not carry out the ADL at all (Q1 only).

Although measurements of ADL were available for all individuals aged 65 and above, we only include individuals at ages 65 to 74. For example, with 18 years of data, an individual aged 65 at wave one may have ADL measurements until they were aged 83 (at wave 18), but we discard data for ages 75 and above. Likewise individuals that entered the survey aged 75 or above were not included in our analysis. By restricting analysis to smaller, more homogenous age-groups we are more likely to satisfy the measurement invariance assumption - discussed in the following section - that the measurement model is consistent as individuals age. For the same reason we also estimate separate models for male and females, thus avoiding the assumption that the measurement model has the same form for both genders.

We allow the level and rate of change of the trajectories to differ by SES. The measure of SES used is the National Statistics Socio-Economic Classification (NS-SEC) (Office for National Statistics, 2010) which categorises each individual's final occupation into eight classes. The 'never worked and long term unemployed' category is excluded because this group is likely to have health issues and hence trajectories that are rather different from the majority of the population. We would expect a social gradient in functional disability using NS-SEC due to accumulation of exposure to risk factors over the working life.

Longitudinal structural equation model (SEM) of physical functioning

In this paper we use a type of SEM known as a multiple indicator growth model (Chan, 1998; Hancock, Kuo, & Lawrence, 2001; Wu, Liu, Gadermann, & Zumbo, 2010). The model consists of two simultaneously estimated components: a measurement model relating responses on the six observed ADL items to a latent variable representing physical functioning, and a growth model for change in the latent variable with age. Separate SEMs were fitted for men and women.

Measurement model

Let y_{rti} denote the response on item r at age t for individual i. A general longitudinal measurement model can be written

$$y_{rti} = \alpha_{rt} + \lambda_{rt} f_{ti} + \epsilon_{rti} \qquad (1)$$

where f_{ti} is the latent functional disability at age t for individual i, α_{rt} are intercepts, λ_{rt} are coefficients or factor loadings, and ϵ_{rti} are residuals. The age-specific factors f_{ti} and residuals ϵ_{rti} are each assumed to follow multivariate normal distributions. We allow for autocorrelation in both functional disability and individual items across ages. We assume that the covariance between items at a given age t is explained by the common factor f_{ti}, so that $\text{cov}(\epsilon_{rti}, \epsilon_{sti}) = 0$ for $r \neq s$. To fix the location and scale of f_{ti} we impose the identification constraints $\alpha_{1t} = 0$ and $\lambda_{1t} = 1$.

The model of equation (1), which we refer to as model 1, allows for changes in the underlying structure of functional disability with age through the inclusion of age–specific intercepts and loadings. However, under this model individual trajectories in f_{ti} are difficult to interpret because changes in the true level of physical functioning with age are confounded with changes in its measurement. Before estimating growth trajectories for f_{ti} we therefore test for temporal measurement invariance by considering two increasingly restricted forms of equation (1). In model 2, factor loadings for the same item are constrained to be equal across ages ($\lambda_{rt} = \lambda_r$). This model assumes *metric invariance* which can be tested by comparison with the base model 1. We then consider model 3 with the additional restriction that the intercepts for the same item are fixed across ages ($\alpha_{rt} = \alpha_r$). A comparison of model 3 and model 2 tests for *scalar invariance*. The combination of metric and scalar invariance in model 3 is sometimes referred to as *strong* *invariance*, which is widely considered as an essential prerequisite for examining temporal change in f_{ti}.

Testing for temporal measurement invariance

We test the overall fit of the measurement models using chi-squared (χ^2) tests, comparing each model with the saturated model which has unconstrained means and covariance matrix. Although the χ^2 test is widely used, there are several limitations relevant to our study: (i) the χ^2 test statistic is dependent on sample size and sensitive to the size of the correlations between the observed items, with large samples and correlations leading to higher values of χ^2, (ii) in a multi-group model (or repeated observation of the same group over time) the χ^2 test is sensitive to even minor deviations between the groups' sample covariance matrices, and (iii) the test is based on the assumption that the observed variables have a multivariate normal distribution, with departures from normality leading to higher values of χ^2 (Kline, 2005; Vandenberg & Lance, 2000). These problems with the χ^2 test have led to the development of numerous fit indices which are usually considered alongside the χ^2 test, many of which are based on the χ^2 with adjustments for sample size and model complexity.

For each of these alternative tests of model fit, Vandenberg and Lance (2000) specify the traditional values required to infer good model fit alongside the more stringent thresholds proposed by Hu and Bentler (1999). We consider both of these thresholds in our analysis. The first of the alternative tests is the Tucker-Lewis index (TLI) (Tucker & Lewis, 1973) which is less susceptible to sample size and favours parsimonious models. Values of the TLI range between 0 and 1 with higher values indicating better fit, and a traditional threshold of 0.9 or above and a more stringent threshold of 0.95 or above for a good model fit. The second alternative test of fit is the root mean square error of approximation (RMSEA) (Steiger, 1990) which does not require a null model and also adjusts for model complexity. The RMSEA also ranges from 0 to 1, but with values close to zero indicating a better fit. The traditional threshold value for an acceptable model fit is 0.08 or less, and a more stringent threshold of 0.06 or less. The third alternative test is the standardised root mean square residual (SRMR) (Bentler, 1995) which is sensitive to model specifications among the factor covariances. The SRMR again ranges from 0 to 1, with lower values indicating better model fit, the traditional threshold for good model fit is 0.10 or less, and a more stringent threshold of 0.08 or less.

In addition to the χ^2 test and alternative tests for assessing *absolute* model fit described above, Vandenberg and Lance (2000) suggest two ways for evaluating *relative* model fit, in our case the change in model fit arising from adding the temporal measurement invariance constraints of models 2 and 3. The first test is based on the change in the chi-squared ($\Delta\chi^2$), where a non-significant difference between models indicates that the additional temporal measurement invariance constraint does not lead to a deterioration in model fit. The second approach is to examine the change in the comparative fit index (ΔCFI). Cheung and Rensvold (1999) provide guidelines on model fit suggesting that a ΔCFI value closer to zero than -0.01 indicates that the more restrictive model is an adequate fit (i.e. the invariance hypothesis should not be rejected), a ΔCFI of between -0.01 and -0.02 indicates researchers should be suspicious about the invariance assumption, and ΔCFI of less than -0.02 suggests that the invariance constraint should be rejected.

Latent growth models with SES effects

The measurement model shown in equation (1) specifies the relationship between an individual's latent functional disability f_{ti} at age t and their responses on the observed ADL items. Age is centred at the baseline age of 65. The second part of the SEM (commonly referred to as the 'structural' model) is a growth model for change in this latent variable with age. We consider a nonlinear growth model in which f_{ti} changes as a quadratic function of age and additionally depends on dummy variables for SES x_{mi} $(m = 2, 3, \ldots, 7)$, taking the first category as the reference. Growth models with a cubic polynomial in age were also considered, but the addition of the cubic term did not lead to a significant improvement in model fit for any of the four samples. We therefore present results for quadratic models. The growth model can be expressed as

$$f_{ti} = \beta_{0i} + \beta_{1i}t + \beta_{2i}t^2 + \sum_{m=2}^{7} \gamma_{0m}x_{mi} + \sum_{m=2}^{7} \gamma_{1m}x_{mi}\,t + \sum_{m=2}^{7} \gamma_{2m}x_{mi}t^2 + e_{ti} \qquad (2)$$

The intercept and coefficients of the quadratic function in age, $\beta_{ki} = \beta_k + u_{ki}$ $(k = 0,1,2)$, are composed of a fixed part β_k common to all individuals and an individual-specific random effect u_{ki}, where the random effects (u_{0i}, u_{1i}, u_{2i}) are assumed to follow a trivariate normal distribution. The e_{ti} are independent normally distributed time-varying residuals. The main effects of SES, the coefficients γ_{0m} of x_{mi}, allow baseline functional disability (at age 65, $t = 0$) to depend on SES, while the coefficients of the interactions between SES and t and t^2 (γ_{1m} and γ_{2m}) allow the rate of change in functioning with age to vary with SES.

This SEM (Model 4) which combines the measurement model of equation (1) and growth model of equation (2) is the main model of interest. We also estimate a second SEM (Model 5) which constrains the factor loadings to be equal for all items. This is akin to modelling the growth of a functional disability measure which is simply the sum of the scores on each of the items. Thus contrasting Model 4 with Model 5 allows us to see the effect of failure to allow for differences in the discriminatory power of the ADL items when modelling functional disability trajectories.

Results

Measurement models and evidence for temporal measurement invariance

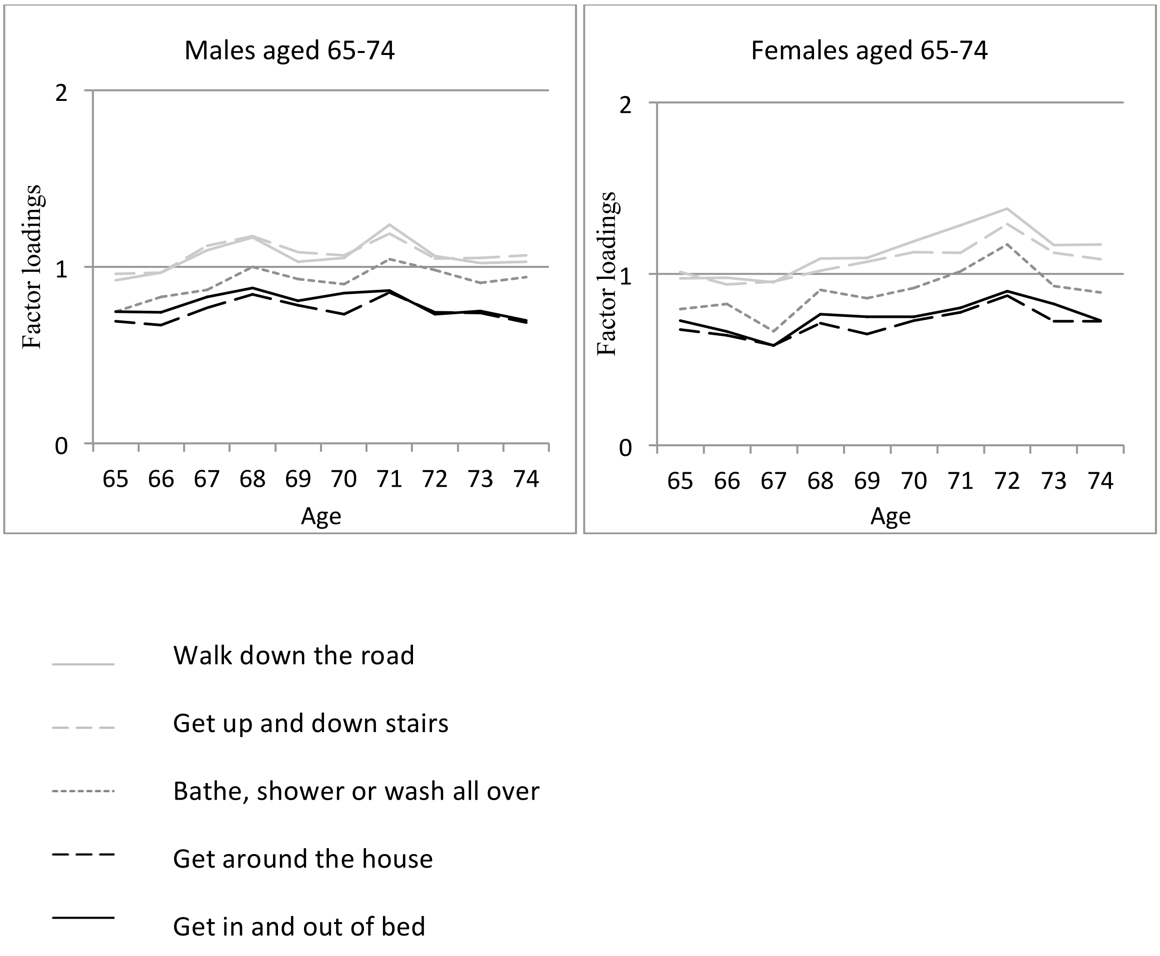
To test for temporal measurement invariance in our data we estimate three versions of the measurement model with increasingly rigorous constraints. Model 1 is a simple measurement model with no measurement invariance constraints i.e. factor loadings and item intercepts are allowed to vary with age. Absolute model fit statistics for Model 1 are presented in the first panel of Table 1. For both gender groups the χ^2 test indicates significant differences between Model 1 and the baseline saturated model (with parameters for the means, variances and covariances for the 6 ADL items measured at 10 time points). The TLI gave weak evidence of good model fit with values below the more stringent threshold for both gender sub-samples, with females just above the less stringent while males were below even this threshold. The RMSEA provided the strongest evidence of good model fit, with values for the well below the more stringent threshold for both samples. The SRMSR also provides evidence of good model fit, with values below the more stringent threshold.

Table 1: Tests for temporal measurement invariance

	Males	Females
Absolute fit of Model 1:		
Chi-square test statistic, 1395 df (χ^2)	4830	4317
TLI	0.892	0.910
RMSEA	0.038	0.033
SRMSR	0.078	0.070
Change in model fit between Model 1 and Model 2:		
Chi-square test statistic, 45 df ($\Delta\chi^2$)	116	132
Change in CFI (ΔCFI)	-0.002	-0.002
Change in model fit between Model 2 and Model 3:		
Chi-square test statistic, 45 df ($\Delta\chi^2$)	161	169
Change in CFI (ΔCFI)	-0.002	-0.003
n	*1712*	*1959*

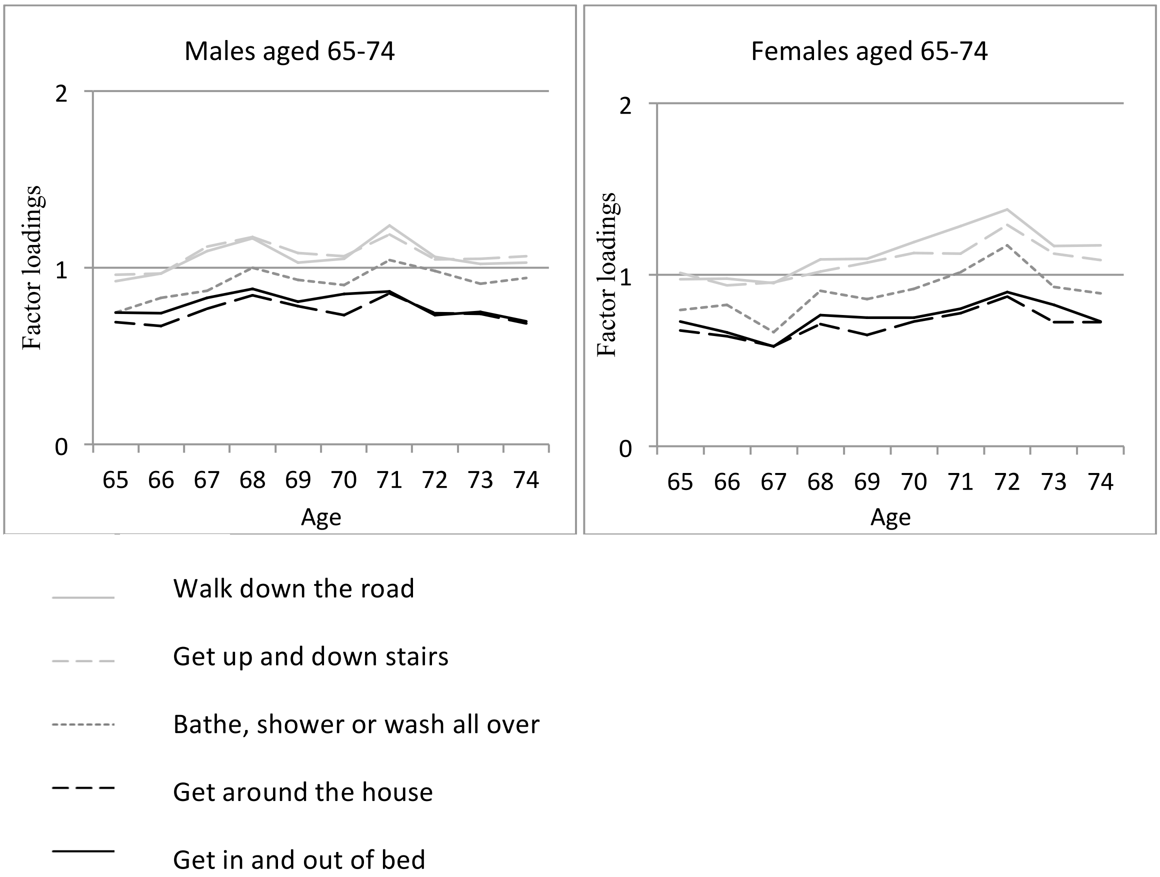
Model 2 is a restricted version of model 1 with the factor loadings for each item constrained to be equal for all ages. Figure 1 shows how the estimated factor loadings of model 1 are broadly similar over ages though with a slight upwards trend (which is consistent with all activities becoming more difficult as individuals get older), so it seems reasonable that constraining these to be equal over time may be a sensible assumption. We formally test whether this is assumption holds by comparing the change in model fit between models 1 and 2, in other words whether the differences in the factor loadings of the measurement model by age shown in figure 1 are sufficiently large to make a significant change to model fit. The tests of change in model fit between model 1 and model 2 are shown in the second panel of table 1. The $\Delta\chi^2$ between models 1 and 2 suggests that imposing time invariant factor loadings leads to a significantly worse model fit. However we see only a small $\Delta\mathrm{CFI}$, far below the threshold for metric invariance. Overall we conclude that there is some evidence of metric invariance.

Figure 1: Trajectories of factor loadings (λ_r), when allowed to vary by age (Model 1)



Model 3 is a more restricted version of model 2 in which the intercepts for each item are constrained to be the same for all ages. The intercepts were allowed to vary by age in models 1 and 2. Figure 2 shows the estimates of the item intercepts by age for model 2. These show a downward trend in the item intercepts over time. We formally test for scalar invariance - that is whether constraining the item intercepts to be equal over ages is a reasonable assumption - by examining the change in model fit statistics between model 2 and model 3 (see the third panel of table 1). As seen for the contrast between models 1 and 2 the $\Delta\chi^2$ indicates a significantly worse model fit, but a very small $\Delta\mathrm{CFI}$ supports the assumption of scalar invariance.

Figure 2: Item intercepts α_r, when allowed to vary by age (Model 2)



Model 4 is our main model of interest, a full SEM combining a measurement model of the same specifications as model 3 (assuming metric and scalar invariance) with a growth model. We interpret the measurement model parameters below and in the following section we contrast the growth model parameters of model 4 with the alternative SEM which approximates a growth curve model fitted to an unweighted sum of scores on the ADL items (model 5). Model 5 is similar to the growth models fitted in most previous research, but with functional disability as a latent variable rather than a sum score.

An important consideration when evaluating differences in parameter estimates across sub-samples (male vs. female) or model specifications (model 4 vs. model 5) is that these may be due in part to differences in the variance of the physical functioning factor. Suppose, for example, that we wish to compare the factor loading for a particular ADL item for two groups. Even if the underlying relationship between the ADL response and the factor is the same for each group, the estimated factor loading will be of smaller magnitude in the group with the largest factor variance. Standardised factor loadings and growth model coefficients can be computed to take account of such scaling effects (see Supplementary Data for details). We present unstandardised factor loadings and item intercepts for the measurement model component of the SEM in table 2, and unstandardised model estimates for all growth model parameters of models 4 and 5 in table 3. Between gender comparisons can be made as the factor variance is fairly similar across genders. However, because the factor variance changes according to whether or not the factor loadings in the measurement model are permitted to vary across ADL items, we present a separate set of standardised estimates for the overall SES effects of Models 4 and 5 in Table 4.

The measurement model component of the SEM

The factor loadings λ_r and item intercepts α_r of the measurement part of the growth model are shown in table 2. The factor loadings are interpreted as the expected change in the observed ADL item for a one-unit change in the factor, and represent the discriminatory power of the items in terms of the functional disability latent variable. The factor loading for the first item: 'cutting toenails' is constrained to one to fix the scale of the latent variable. 'Walking down the road' and 'managing steps' have the largest factor loadings indicating they are best at discriminating between individuals with different levels of functioning. 'Getting around the house' has the lowest factor loadings, followed by 'getting in and out of bed' then 'bathe, shower or wash all over' i.e. these are the least discriminatory items for changes in functional disability.

Table 2: Factor loadings λ_r and item intercepts α_r for the measurement part of the SEM (Model 4). Standard errors are given in brackets.

	Males	Females
Factor loadings (λ_r):		
Cut toenails	1	1
Walk down the road	1.013 (0.023)	1.058 (0.024)
Get up and down stairs or steps	1.034 (0.023)	1.026 (0.024)
Bath, shower or wash all over	0.871 (0.019)	0.839 (0.020)
Get in and out of bed	0.755 (0.017)	0.703 (0.017)
Get around the house	0.716 (0.016)	0.664 (0.016)
Item intercepts (α_r):		
Cut toenails	0	0
Walk down the road	-0.519 (0.039)	-0.841 (0.050)
Get up and down stairs or steps	-0.425 (0.039)	-0.611 (0.049)
Bath, shower or wash all over	-0.489 (0.033)	-0.692 (0.040)
Get in and out of bed	-0.444 (0.028)	-0.594 (0.033)
Get around the house	-0.442 (0.027)	-0.596 (0.031)

The item intercepts represent the difficulty of the items. We constrain the first item 'cutting toenails' to zero, and this is the least difficult item because the estimated intercepts for the other items are all negative. For both genders 'walking down the road' is the most difficult item, followed by 'bathing, showering and washing'; the other categories ('managing stairs or steps', 'getting in and out of bed' and 'getting around the house') have roughly equal values for each gender. The intercepts are larger in magnitude for females, which is consistent with the literature on poorer female health.

The growth model component of the SEM

The parameter estimates for the growth model component of the full SEM (model 4) are shown in the left half of table 3. The coefficients of the SES dummy variables are interpreted as contrasts with the reference group 'routine occupations' at the baseline age in the sample. Functional disability at baseline (β_0) is greater for females. The intercept variances, $\mathrm{var}(u_{0i})$, are interpreted as the between-individual variance in the level of physical functioning at age 65 $(t = 0)$ for each gender. We see a slightly larger baseline variance for females.

Table 3: Growth model parameters and model fit statistics for Models 4 and 5, SEMs with unequal and equal factor loadings across ADL items.

	Model 4: Growth model from SEM with unequal factor loadings for ADL items				Model 5: Growth model from SEM with equal factor loadings for ADL items			
	Male		Female		Male		Female	
Parameter estimates								
Intercept growth factor mean (β_0)	1.098***	(0.067)	1.443***	(0.065)	1.102***	(0.058)	1.478***	(0.055)
Slope growth factor mean (β_1)	0.060***	(0.023)	0.053**	(0.021)	0.048**	(0.019)	0.034*	(0.018)
Quadratic growth factor mean (β_2)	-0.001	(0.002)	0.002	(0.002)	-0.001	(0.002)	0.002	(0.002)
Effects of NS-SEC on intercept (γ_{0m}):								
Routine occupations (reference)								
Semi-routine occupations	0.028	(0.096)	-0.023	(0.082)	0.027	(0.079)	-0.017	(0.065)
Lower supervisory and technical occupations	-0.047	(0.097)	0.205*	(0.120)	-0.046	(0.079)	0.157*	(0.095)
Small employers and own account workers	-0.041	(0.091)	-0.127	(0.125)	-0.031	(0.074)	-0.093	(0.098)
Intermediate occupations	-0.386***	(0.136)	-0.168*	(0.086)	-0.305***	(0.111)	-0.121*	(0.068)
Lower managerial and professional occupations	-0.185**	(0.088)	-0.160*	(0.084)	-0.133*	(0.071)	-0.125*	(0.066)
Higher managerial and professional occupations	-0.247**	(0.103)	-0.536**	(0.209)	-0.188**	(0.084)	-0.380**	(0.163)
Effects of NS-SEC on coefficient of t (γ_{1m}):								
Routine occupations (reference)								
Semi-routine occupations	-0.030	(0.035)	-0.002	(0.030)	-0.027	(0.029)	-0.001	(0.025)
Lower supervisory and technical occupations	0.009	(0.036)	0.003	(0.043)	0.012	(0.030)	0.002	(0.035)
Small employers and own account workers	-0.060*	(0.034)	-0.031	(0.046)	-0.052*	(0.029)	-0.023	(0.038)
Intermediate occupations	-0.034	(0.048)	-0.047	(0.031)	-0.024	(0.041)	-0.036	(0.026)
Lower managerial and professional occupations	-0.063**	(0.032)	-0.021	(0.031)	-0.051*	(0.027)	-0.011	(0.025)
Higher managerial and professional occupations	-0.107***	(0.038)	-0.031	(0.075)	-0.087***	(0.032)	-0.026	(0.062)
Effects of NS-SEC on coefficient of t^2 (γ_{2m}):								
Routine occupations (reference)								
Semi-routine occupations	0.001	(0.004)	-0.001	(0.003)	0.002	(0.003)	-0.001	(0.003)
Lower supervisory and technical occupations	-0.002	(0.004)	-0.006	(0.004)	-0.002	(0.003)	-0.005	(0.004)
Small employers and own account workers	0.005	(0.004)	0.001	(0.005)	0.005*	(0.003)	0.001	(0.004)
Intermediate occupations	0.005	(0.005)	0.004	(0.003)	0.004	(0.004)	0.003	(0.003)
Lower managerial and professional occupations	0.005*	(0.003)	-0.001	(0.003)	0.004	(0.003)	-0.001	(0.003)
Higher managerial and professional occupations	0.009**	(0.004)	0.005	(0.008)	0.007*	(0.004)	0.004	(0.007)
Intercept growth factor variance, $\mathrm{var}(u_{0i})$	0.586***	(0.043)	0.667***	(0.047)	0.358***	(0.023)	0.367***	(0.022)
Slope growth factor variance, $\mathrm{var}(u_{1i})$	0.006***	(0.001)	0.009***	(0.001)	0.004***	(0.001)	0.005***	(0.001)
Covariance between factor mean and slope, $\mathrm{cov}(u_{0i}, u_{1i})$	-0.010**	(0.005)	-0.007	(0.005)	-0.007**	(0.003)	-0.006**	(0.003)
Residual variance for the factor, $\mathrm{var}(e_i)$	0.217***	(0.010)	0.213***	(0.010)	0.164***	(0.004)	0.156***	(0.004)
Model fit								
Chi-square test statistic	5,531 (1,885 df)		5,303 (1,885 df)		6,489 (1,890 df)		6,645 (1,890 df)	
TLI	0.898		0.906		0.871		0.870	
RMSEA	0.034		0.030		0.038		0.036	
SRMSR	0.075		0.072		0.125		0.131	
n	1,712		1,959		1,712		1,959	

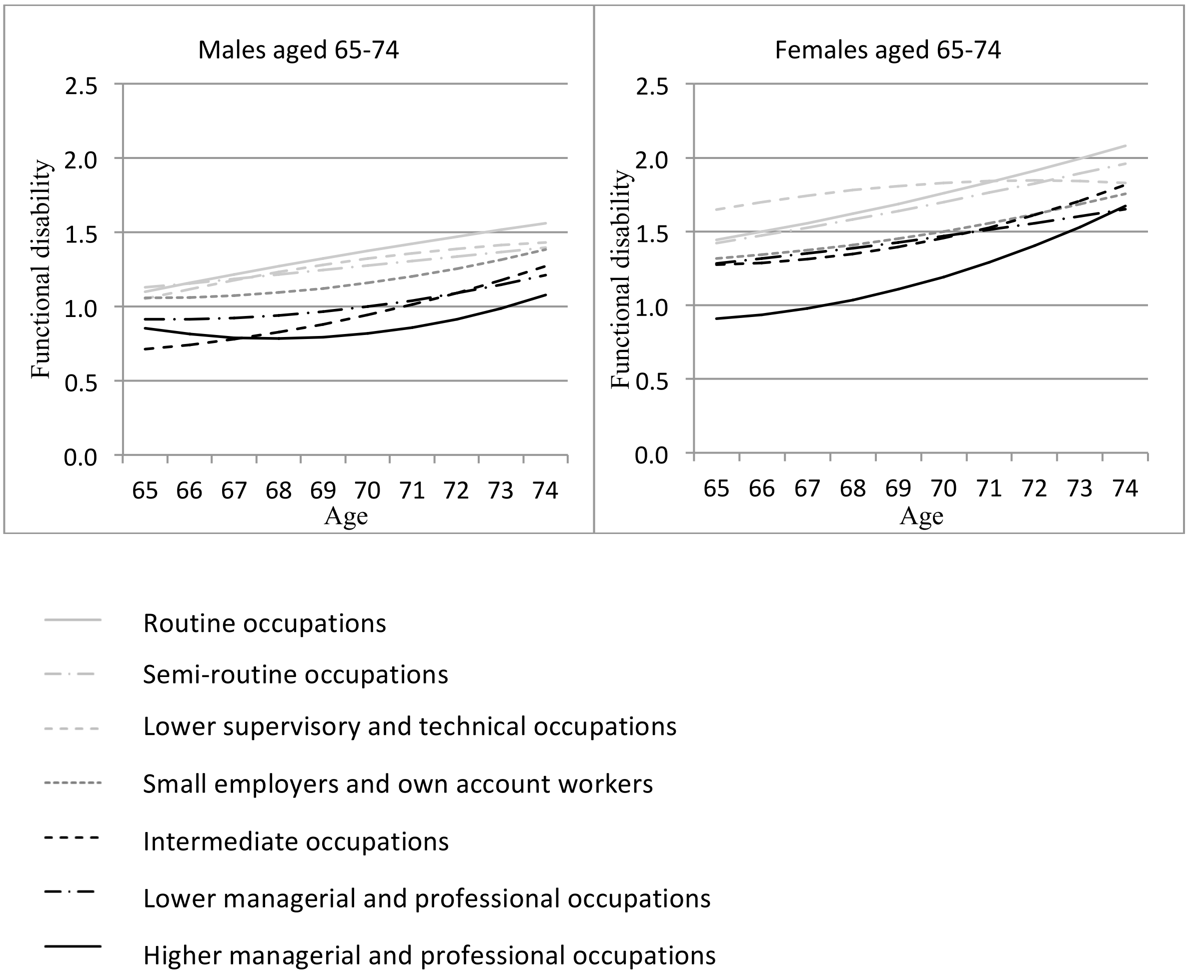
*** p <0.01, ** p<0.05, * p<0.10 . Standard errors in parentheses.

Table 4: Comparison of standardised SES effects for growth model component of Models 4 and 5, SEMs with unequal and equal factor loadings across ADL items.

	Model 4: Growth model from SEM with unequal factor loadings for ADL items				Model 5: Growth model from SEM with equal factor loadings for ADL items			
Age	65	68	71	74	65	68	71	74
Males:								
Routine occupations	0	0	0	0	0	0	0	0
Semi-routine occupations	0.031	-0.059	-0.122	-0.153	0.037	-0.050	-0.083	-0.064
Lower supervisory and technical occupations	-0.052	-0.043	-0.069	-0.122	-0.064	-0.039	-0.060	-0.118
Small employers and own account workers	-0.046	-0.197	-0.233	-0.167	-0.043	-0.198	-0.214	-0.111
Intermediate occupations	-0.431	-0.496	-0.432	-0.273	-0.422	-0.475	-0.400	-0.232
Lower managerial and professional occupations	-0.206	-0.369	-0.404	-0.330	-0.184	-0.348	-0.387	-0.316
Higher managerial and professional occupations	-0.276	-0.546	-0.596	-0.457	-0.260	-0.537	-0.600	-0.476
Females:								
Routine occupations	0	0	0	0	0	0	0	0
Semi-routine occupations	-0.025	-0.040	-0.067	-0.100	-0.024	-0.040	-0.074	-0.118
Lower supervisory and technical occupations	0.219	0.167	0.007	-0.209	0.217	0.162	-0.014	-0.254
Small employers and own account workers	-0.135	-0.220	-0.262	-0.267	-0.129	-0.210	-0.245	-0.242
Intermediate occupations	-0.179	-0.285	-0.289	-0.219	-0.167	-0.277	-0.288	-0.223
Lower managerial and professional occupations	-0.171	-0.242	-0.304	-0.353	-0.173	-0.229	-0.286	-0.337
Higher managerial and professional occupations	-0.571	-0.609	-0.512	-0.337	-0.525	-0.579	-0.493	-0.320

Predicted trajectories for each gender are presented in figure 3 with separate curves for each SES group. These trajectories are calculated using the SES coefficients for functional disability for someone at the mean of the distribution, in other words the individual random effects are set at their means of zero. For all SES groups we estimate a positive linear growth (β_1) in functional disability, and the quadratic growth factor mean (β_2) shows a slight acceleration in growth for females but not for males. The random effect variance associated with the linear age effect, $\mathrm{var}(u_{1i})$, is similar for men and women, though slightly smaller for males. There is a negative covariance (though statistically insignificant for females) between the individual intercepts and slopes suggesting that higher functional disability at baseline is associated with slower increase in functional disability over time. Note that the variance of the random effect for t^2 (u_{2i}) and its covariances with the other random effects were found to be negligible, and were therefore omitted from the structural model.

Figure 3: Functional disability trajectories by socioeconomic status (Model 4)



SES is allowed to affect both the intercept and slope of functional disability. For each gender we find small but significant effects of SES on the intercept (γ_{0m}) compared with the reference category 'routine occupations', though the lower status occupations ('small employers and own account workers', 'lower supervisory and technical occupations', and 'semi-routine occupations') are not statistically significantly different. In terms of the social gradient in the change in functional disability $(\gamma_{1m} \text{ and } \gamma_{2m})$ males show a slight widening of the social gradient in functional disability with age, while females show a slight convergence with age (though from a more divergent baseline), though these relationships are only statistically significant for males and only for the less routine occupations.

In the right-hand side of table 4 we show estimates from a comparison model (model 5) which proxies a growth model fitted to an unweighted sum of ADL scores. Table 5 shows standardised SES effects for models 4 and 5 for the male and female subsamples, calculated for selected ages three years apart. We would not expect the SES effects to be dramatically different given the factor loadings from the measurement part of model 4 shown in table 2 are relatively close to one another. This comparison shows the SES effects would be slightly underestimated when no measurement models is used, with the most noticeable differences for males at the older ages (71 and 74).

Discussion

The general health of the elderly population is typically measured using questions relating to functional ability across a range of dimensions. When using these measures to model trajectories of functional disability as people age, researchers typically use simple methods to combine these indicators, such as the total score. We argue that these approaches are limited since they do not capture the difference in discriminatory power of these different items. We propose supplementing the growth model of functional disability with a measurement model to better capture the underlying latent variable functional disability that we wish to use as the outcome in the growth model.

Another advantage of specifying a measurement model is that it makes explicit and allows testing of the assumption of temporal measurement invariance. We estimated a sequence of three increasingly restricted models in order to test for measurement invariance for the gender sub-samples. Vandenberg and Lance (2000) argue that assessing model fit using only a χ^2 test is limited because it is sensitive to sample size and differences in the covariance structure, and suggest using a suite of fit indices including TLI, RMSEA and SRMR to evaluate the degree of temporal measurement invariance. By recognising the strengths and weaknesses of each of these indices we are able to build a more robust assessment of the temporal measurement invariance assumption.

We estimate SEM of the growth in latent functional disability separately for each gender. Overall we see increasing functional disability, with accelerating growth for males but not for females. For both genders we find evidence of a social gradient in the baseline levels of functional disability between the most routine occupational class (the reference category) and the least routine social classes. The social gradient in the rate of change of functional disability was less clear. Our model predicts that the functional disability of an individual from the lowest SES group at baseline (aged 65) is equivalent to that of an individual from the highest SES group who was ten years older for males and seven years older for females.

Future research which estimates trajectories of functional disability for the elderly could benefit from adopting our approach of using a SEM to incorporate a measurement model which treats functional disability as a latent variable. This includes work using richer datasets, which would allow a wider set of items to measure functional disability and a wider set of controls. For example the English Longitudinal Study of Ageing includes measures of iADL and mobility to supplement the ADL, and has better measures of SES to improve identification of the social gradient in functional disability trajectories (which could be used in a second measurement model for a latent SES measure). To date, these data have only been used to model functional disability cross-sectionally (Gjonça, Tabassum, & Breeze, 2009). Time-varying measures of social status would allow us to explore the relationship between change in SES and change in functional disability, and determine whether the changes in SES effects with age are real or simply a function of increasing time since the measure was taken. We know that there may be reverse causality in this relationship as health status could also impact on social status (Steele, French, & Bartley, 2013). Longitudinal data on both health and SES would allow us to identify the direction of these effects. Residential status is another time-varying characteristic of policy relevance (because of the cost of residential care), which may be included as a determinant of functional disability trajectories. Such a model could be extended to identify the effect of residential status on individuals where care needs (including moves into residential care) are not met (Scott, Evandrou, Falkingham, & Rake, 2001). Finally, studies that incorporate this approach over shorter-term periods would be able to capture aspects of recovery as well as the longer term increase in disability found in this study. Importantly, a shorter time span would also make it easier to satisfy the temporal measurement invariance assumption.

Acknowledgements

This work was supported by the UK Economic and Social Research Council (grant number RES-576-25-0032).

References

Bentler, P. M. (1995). *EQS structural equations program manual* (Vol. 6).

Burchardt, T. (2000). The dynamics of being disabled. *Journal of Social Policy, 29*(04), 645-668. http://dx.doi.org/10.1017/S0047279400006097

Case, A., & Paxson, C. (2005). Sex differences in morbidity and mortality. *Demography, 42*(2), 189-214. http://dx.doi.org/10.1353/dem.2005.0011

Chan, D. (1998). The conceptualization and analysis of change over time: an integrative approach incorporating Longitudinal Mean and Covariance Structures Analysis (LMACS) and Multiple Indicator Latent Growth Modeling (MLGM). *Organizational Research Methods, 1*(4), 421-483. http://dx.doi.org/10.1177/109442819814004

Cheung, G. W., & Rensvold, R. B. (1999). What constitutes significant differences in evaluating measurement invariance? Paper presented at the 1999 conference of the Academy of Management, Chicago.

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research, 38*(4), 529-569. http://dx.doi.org/10.1207/s15327906mbr3804_5

Feng, Q., Zhen, Z., Gu, D., Wu, B., Duncan, P. W., & Purser, J. L. (2013). Trends in ADL and IADL disability in community-dwelling older adults in Shanghai, China, 1998–2008. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 68*(3), 476-485. http://dx.doi.org/10.1093/geronb/gbt012

Fries, J. F. (1980). Aging, natural death, and the compression of morbidity. *New England Journal of Medicine, 303*(3), 130-135. http://dx.doi.org/10.1056/NEJM198007173030304

Gjonça, E., Tabassum, F., & Breeze, E. (2009). Socioeconomic differences in physical disability at older age. *Journal of Epidemiology and Community Health, 63*(11), 928-935. http://dx.doi.org/10.1136/jech.2008.082776

Goldstein, H., & Woodhouse, G. (2001). Modelling repeated measures. In A. H. Leyland & H. Goldstein (Eds.), *Multilevel modelling of health statistics* (pp. 13-26). Chichester: Wiley.

Grundy, E., & Glaser, K. (2000). Socio-demographic differences in the onset and progression of disability in early old age: a longitudinal study. *Age and Ageing, 29*(2), 149-157. http://dx.doi.org/10.1093/ageing/29.2.149

Haas, S. (2008). Trajectories of functional health: The 'long arm' of childhood health and socioeconomic factors. *Social Science & Medicine, 66*(4), 849-861. http://dx.doi.org/10.1016/j.socscimed.2007.11.004

Hancock, G. R., Kuo, W.-L., & Lawrence, F. R. (2001). An illustration of second-order latent growth models. *Structural Equation Modeling: a Multidisciplinary Journal, 8*(3), 470-489. http://dx.doi.org/10.1207/S15328007SEM0803_7

Holstein, B. E., Avlund, K., Due, P., Martinussen, T., & Keiding, N. (2006). The measurement of change in functional ability: dealing with attrition and the floor/ceiling effect. *Archives of Gerontology and Geriatrics, 43*(3), 337-350. http://dx.doi.org/10.1016/j.archger.2005.12.004

House, J. S., Kessler, R. C., & Herzog, A. R. (1990). Age, socioeconomic status, and health. *The Milbank Quarterly, 68*(3), 383-411. http://dx.doi.org/10.2307/3350111

Institute for Social & Economic Research (2010). *British Household Panel Survey: waves 1-18, 1991-2009 [computer file] 7th Edition*. Retrieved from: http://discover.ukdataservice.ac.uk/catalogue?sn=5151

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: a Multidisciplinary Journal, 6*(1), 1-55. http://dx.doi.org/10.1080/10705519909540118

Katz, S., Ford, A., Moskowitz, R., Jackson, B., & Jaffe, M. (1963). Studies of illness in the aged. The index of ADL: a standardized measure of biological and psychosocial function. *Journal of the American Medical Association, 185*, 914 - 919. http://dx.doi.org/10.1001/jama.1963.03060120024016

Kline, R. B. (2005). *Principles and practice of structural equation modelling* (2nd. ed.). New York: The Guilford Press.

LaPlante, M. P. (2010). The classic measure of disability in activities of daily living Is biased by age but an expanded IADL/ADL measure is not. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 65B*(6), 720-732. http://dx.doi.org/10.1093/geronb/gbp129

Lawton, M. P., & Brody, E. M. (1969). Assessment of older people: Self-maintaining and instrumental activities of daily living. *Gerontologist, 9*, 179-186. http://dx.doi.org/10.1093/geront/9.3_Part_1.179

Li, L. W. (2005). Trajectories of ADL disability among community-dwelling frail older persons. *Research on Aging, 27*(1), 56-79. http://dx.doi.org/10.1177/0164027504271348

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. New York: Wiley. http://dx.doi.org/10.1002/9781119013563

Mendes de Leon, C. F., Guralnik, J. M., & Bandeen-Roche, K. (2002). Short-term change in physical function and disability: the Women's Health and Aging Study. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 57*(6), S355-S365. http://dx.doi.org/10.1093/geronb/57.6.S355

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user’s guide. Seventh Edition.* Los Angeles, CA: Muthén & Muthén.

Nathanson, C. A. (1975). Illness and the feminine role: a theoretical review. *Social Science & Medicine (1967), 9*(2), 57-62.

Nilsson, C. J., Avlund, K., & Lund, R. (2010). Social inequality in onset of mobility disability among older Danes: the mediation effect of social relations. *Journal of Aging and Health, 22*(4), 522-541. http://dx.doi.org/10.1177/0898264309359684

Office for National Statistics. (2010). *The National Statistics socio-economic classification: (Rebased on the SOC2010) User Manual* (Vol. 3). Southampton: Palgrave Macmillan.

Park, N. S., Klemmack, D. L., Roff, L. L., Parker, M. W., Koenig, H. G., Sawyer, P., & Allman, R. M. (2008). Religiousness and longitudinal trajectories in elders' functional status. *Research on Aging, 30*(3), 279-298. http://dx.doi.org/10.1177/0164027507313001

Payette, H., Gueye, N. D. R., Gaudreau, P., Morais, J. A., Shatenstein, B., & Gray-Donald, K. (2011). Trajectories of physical function decline and psychological functioning: the Québec Longitudinal Study on Nutrition and Successful Aging (NuAge). *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 66B*(suppl 1), i82-i90. http://dx.doi.org/10.1093/geronb/gbq085

Sacker, A., Clarke, P., Wiggins, R. D., & Bartley, M. (2005). Social dynamics of health inequalities: a growth curve analysis of aging and self assessed health in the British Household Panel Survey 1991–2001. *Journal of Epidemiology and Community Health, 59*(6), 495-501. http://dx.doi.org/10.1136/jech.2004.026278

Sacker, A., Worts, D., & McDonough, P. (2011). Social influences on trajectories of self-rated health: evidence from Britain, Germany, Denmark and the USA. *Journal of Epidemiology and Community Health, 65*(2), 130-136. http://dx.doi.org/10.1136/jech.2009.091199

Scott, A., Evandrou, M., Falkingham, J., & Rake, K. (2001). Moves into residential care amongst older people in Britain. *Paper Presented to the 2001 British Household Panel Survey Research Conference; 5-7 July, 2001, Colchester, UK, 2012* (November). https://www.iser.essex.ac.uk/files/conferences/bhps/2001/docs/pdf/papers/scott.pdf

Steele, F. (2008). Multilevel models for longitudinal data. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 171*(1), 5-19.

doi: 10.1111/j.1467-985X.2007.00509.x

Steele, F., French, R., & Bartley, M. (2013). Adjusting for selection bias in longitudinal analyses using simultaneous equations modeling: the relationship between employment transitions and mental health. *Epidemiology, 24*(5), 703-711. http://dx.doi.org/10.1097/EDE.0b013e31829d2479

Steiger, J. (1990). Structural model evaluation and modification: an interval estimation approach. *Multivariate Behavioral Research, 25*(2), 173-180. http://dx.doi.org/10.1207/s15327906mbr2502_4

Stuck, A. E., Walthert, J. M., Nikolaus, T., Büla, C. J., Hohmann, C., & Beck, J. C. (1999). Risk factors for functional status decline in community-living elderly people: a systematic literature review. *Social Science & Medicine, 48*(4), 445-469. http://dx.doi.org/10.1016/S0277-9536(98)00370-0

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*(1), 1-10. http://dx.doi.org/10.1007/BF02291170

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4-70. http://dx.doi.org/10.1177/109442810031002

Verbrugge, L. M. (1984). Longer life but worsening health? Trends in health and mortality of middle-aged and older persons. *The Milbank Memorial Fund Quarterly. Health and Society, 62*(3), 475-519. http://dx.doi.org/10.2307/3349861

Verbrugge, L. M. (1989). The twain meet: empirical explanations of sex differences in health and mortality. *Journal of Health and Social Behavior, 30*(3), 282-304. http://dx.doi.org/10.2307/2136961

Ware, J. E. J., & Sherbourne, C. D. (1992). The MOS 36-Item Short-Form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care, 30*(6), 473-483. http://dx.doi.org/10.1097/00005650-199206000-00002

Wu, A., Liu, Y., Gadermann, A., & Zumbo, B. (2010). Multiple-indicator multilevel growth model: a solution to multiple methodological challenges in longitudinal studies. *Social Indicators Research, 97*(2), 19. http://dx.doi.org/10.1007/s11205-009-9496-8

Zaninotto, P., Falaschetti, E., & Sacker, A. (2009). Age trajectories of quality of life among older adults: results from the English Longitudinal Study of Ageing. *Quality of Life Research, 18*(10), 1301-1309. http://dx.doi.org/10.1007/s11136-009-9543-6

STUDY PROFILE

Intergenerational transfers and rosters of the extended family: a new substudy of the Panel Study of Income Dynamics

Robert F. Schoeni University of Michigan, Michigan, US
bschoeni@umich.edu
Suzanne M. Bianchi University of California, Los Angeles, US
V. Joseph Hotz Duke University, Durham, US
Judith A. Seltzer University of California, Los Angeles, US
Emily E. Wiemers University of Massachusetts, Boston, US

(Received November 2014 Revised February 2015) http://dx.doi.org/10.14301/llcs.v6i3.332

Abstract

Family members provide support to each other at critical life stages. To better understand the pervasiveness, causes, and consequences of such support, a sub-study of the United States (U.S.) Panel Study of Income Dynamics (PSID) was created. A battery of questions on family relationships and intergenerational transfers was designed, pretested on a U.S. national telephone sample, and then administered in the 2013 wave of the PSID. These new data are available to the public. Given the extensive supporting data available on the respondents and members of their co-resident and non-co-resident family members – many of whom are interviewed themselves – the new sub-study will become a valuable resource to researchers.

Keywords: Intergenerational transfers; longitudinal research; family support; family structure

Scientific motivation

The extended family provides financial resources, time, housing assistance in the form of co-residence, and emotional support to its members to benefit individuals' wellbeing throughout life (Bianchi, Hotz, McGarry & Seltzer, 2008). Parents are a primary source of financial support for their children's post-secondary education (Lovenheim, 2011) and also may provide support to help their children launch careers and purchase homes (Engelhardt & Mayer, 1998; Cox & Stark, 2005). Adult offspring may help elderly parents manage their lives at older ages and provide care at the end of life (Wolf, Soldo & Freedman, 1996; Wolf 1999). Family members may resume living with one another when a family member has a health problem, loses a job, or copes with marital problems or divorce (Kaplan, 2012; Ward, Logan & Spitze, 1992). How family members help each other weather the vagaries of life can have lasting consequences for individuals' health and general wellbeing.

Intergenerational transfers may be especially important at the transition to adulthood and in later life when aging parents become infirm. As children become adults, parents help launch them, in part by giving them money or paying their expenses and by providing housing. Schoeni and Ross (2005) estimate that in the late 1980s an adult child age 18-34 received a total of $50,432 (in 2013 dollars) in combined financial transfers and benefits from co-residence with parents. Transfers varied by parental income, from $30,799 for young adults with parents in the bottom quartile of the income distribution to $93,347 for those with parents in the top quartile. At the other end of life, about 40% of the disabled elderly rely on unpaid help (Spillman & Pezzin, 2000). Adult children are the most common source of informal care, particularly for the unmarried elderly; 44% of primary caregivers are adult children (Center on an Aging Society, 2005). Time help also is an important way that children can help aging parents maintain independence (Kahn, McGill, & Bianchi, 2011). Understanding the decisions families make about the care of the elderly and the family's role in providing care is increasingly important as the population ages and as the costs associated with nonfamily care rise.

Parents may treat each of their offspring differently in young adulthood and even earlier in life, transferring more to some than to other children in their family. The received wisdom is that parents make greater *inter vivos* transfers to their children in greater need (Altonji Hayashi, & Kotlikoff 1997, 2000; McGarry & Schoeni 1995, McGarry, 1997). However, findings by Zissimopoulos and Smith (2009) suggest that this conclusion is premature. Using Health and Retirement Study (HRS) longitudinal data, they find that parents' financial transfers among children in the same family may be more equal than researchers previously thought: when tracked over a 16 year period, 43% of parents who have given money to at least one child also gave money to all of their children, compared to only 11% in a two year period. In addition, amounts tended to become more equal across children over a longer time horizon. Some of the equalization in giving across children over time is due to large transfers, probably for college and buying a home, which occur at different times for different children. Collecting data on both short-term and long-term transfers over the life course is a significant contribution of the new PSID sub-study.

Several demographic and economic trends are likely to have important and far-reaching consequences for the provision and receipt of familial transfers. Family instability, due to divorce and repartnering, affects family members' capacity to transfer time, money, and co-residence across generations, the need for these transfers, and the willingness to participate in family exchanges. High rates of union dissolution and repartnering make stepfamilies and families formed by cohabitation more common now than in the past (Bumpass & Lu, 2000; Kennedy & Bumpass, 2008). Divorce may reduce parents' ability to help children and increase parents' need for help in old age (Furstenberg, Hoffman, & Shrestha, 1995). Remarriage may further weaken the family safety net. Step-children are less likely than biological children to live with or provide help to older parents (Pezzin, Pollak, & Steinberg Schone, 2008; Seltzer, Yahirun, & Bianchi, 2013). Americans feel less obligation to care for step- than biological parents (Coleman & Ganong, 2008) and parents who have both step and biological children receive less help overall than do parents who have only biological children (Eggebeen, 1992; Pezzin & Steinberg Schone, 1999).

The recent Great Recession and the economy's slow rebound in many parts of the United States may impair the ability of many parents to finance not only their on-going needs, but also their ability to help finance their adult children's education and home ownership. Using PSID data, Lovenheim (2011) found that the rise in housing values during the first part of the 2000-10 decade significantly increased college enrollments, with the largest effects among less wealthy households. These results suggest that the more recent housing bust associated with the recession may reduce college attendance. Parents, particularly those in the hardest hit localities, who might have helped finance their children's college education by drawing on the equity in their homes, now find themselves with far fewer resources than they believed they had just a few years ago.

Disruptions in parents' lives are likely to affect their children's lives as well. As is well documented with PSID data, economic success is correlated across generations (Solon, 1992; Lee & Solon, 2009; Charles & Hurst, 2003). Estimates of the intergenerational correlation of economic attainment range from 0.3 to 0.7 depending on dataset, sample, and measures (Harding, Jencks, Lopoo, & Mayer, 2005). But understanding the mechanisms that produce these correlations is less developed. Theories of intergenerational transfers (Becker and Tomes, 1979) imply that parents invest directly in the human capital of their children and use *inter vivos* transfers of wealth to enhance their offspring's wealth and capacity to produce it. Evidence of nonlinearities in the intergenerational elasticity of earnings suggests that variation in parents' circumstances affects the magnitude and mechanisms that contribute to the intergenerational transmission of economic wellbeing (Bratsberg et al., 2007). However, isolating the influence of these forms of parental transfers is complicated by the lack of information about the incidence and amounts of different types of transfers and the lack of information about exogenous factors which affect parent's ability to provide them.

To summarise, U.S. families are at a unique historical juncture, with the huge Baby Boom cohort facing retirement in uncertain economic times. This is a cohort for whom the family context was radically different than for their parents' generation. They experienced more family disruption, reduced fertility and smaller families but more step-kin, and the increased labour force participation of women. They raised their children during a period of rapidly rising income inequality in the U.S. (Autor, Katz, & Kearney, 2008). This trend, coupled with the growing socioeconomic divide in family stability (McLanahan, 2004), may magnify family effects on the wellbeing of the next generation. The high level of persistence in economic circumstances across generations may make it difficult for those at the bottom to live healthy and secure lives in retirement, and they are unlikely to have children who are able to provide them with additional resources.

Questions that can be addressed

Data from the new PSID sub-study will allow researchers to investigate how parents and offspring use the scarce resources of time and money to alleviate economic distress. The list of scientific questions that can be addressed is long, particularly when these new data are combined with the rich array of social, economic, and health data on individuals and their family members that have been collected during the prior 45 years of the PSID. Because respondents and their family members have been interviewed in the past and will continue to be interviewed into the future, the new sub-study data provide new opportunities to investigate how past circumstances influenced large transfers for schooling and housing and how family characteristics and behaviour reported in the 2013 sub-study influence subsequent social and economic outcomes.

These new data allow researchers to describe the intergenerational structure of American families both within and across households and the transfers that family members make to one another across adult ages — something that, to our knowledge, is not possible with any other nationally representative dataset. These data also allow researchers to investigate how health and economic circumstances of parents and children are correlated with transfers and whether these patterns differ by life stage, demographic characteristics such as education, race, and marital status, and intergenerational family structure. The inclusion of questions on transfers for schooling and housing in an intergenerational study allows for a deeper understanding of how these transfers shape the intergenerational transmission of economic advantage. Furthermore, researchers can compare family transfers in 2013 with data on transfers collected in the 1988 PSID to see if intergenerational support has changed along with changes in family structure and economic circumstances.

PSID as a vehicle for studying families and intergenerational transfers

The PSID is the premier dataset in the United States for studying how life course and intergenerational processes contribute to individual well-being because of its prospective, repeated measures of individuals' economic characteristics, health, living arrangements, its genealogical design, its long life histories of linked family members, and its high wave-to-wave response rates. Launched in 1968, the study follows individuals whether or not they are living in the same dwelling as the original sample household or with the same people. Children who grow up and leave their parents' household become what PSID calls "split-offs," and these children continue to be followed and interviewed as they establish their own households, have children themselves, and even when they sometimes move back in with their parents. Interviews were conducted annually until 1997 when PSID moved to an every other year schedule.

All individuals in households recruited into the PSID in 1968 are said to have the PSID 'gene'. All individuals who are born to or adopted by someone with the PSID gene acquire the gene themselves and therefore are followed and become members of the PSID sample for the rest of their lives. This design feature implies that the study provides, at each wave, data on a sample of extended families, or dynasties. Because it is longitudinal, the PSID provides, across its waves, data on the households and members of multi-generational dynasties at various points in their lives. To facilitate genealogic analyses in which parents and children, siblings, and three-generation families are linked, PSID provides the Family Identification Mapping System (FIMS) tool to allow users to easily create inter- and intra-generational samples.

Americans of all ages are captured by the PSID sampling strategy and then followed throughout their lives. This design allows investigation of intergenerational transfers across the entire life course, including assistance given/received during the years when critical decisions are made about human capital investments, family formation, and homeownership. This feature is particularly important for disadvantaged populations because they leave home earlier, have grandchildren earlier, and experience significant health challenges earlier than more advantaged populations. Surveys that focus just on older populations do not measure these important events.

Despite the undisputed merit of its genealogical design, the fact that non-gened family members, whose needs and resources affect the welfare of gened respondents, are not included in the sample is a disadvantage. For example, if today a 25-year-old woman who is a gened sample member gets married (and her husband does not have the PSID gene), the PSID contains information on the health and economic status of her parents, because they, themselves, are sample members, but not detailed information on the health and economic status of her spouse's parents. Information on both sets of parents' needs and resources is required to understand why an adult child's family may give help to parents but not parents-in-law. The 2013 Family Roster and Family Transfer sub-study addresses this shortcoming of the PSID for studying intergenerational transfers (Bianchi et al., 2008).

When it began in 1968 the PSID had a sample of 18,230 individuals living in 4,802 households. The sample size has changed over time for a number of reasons including mortality, attrition, births, addition of a sample of new immigrants, and dropping of a portion of the sample due to budget cuts. In 2013, interviews were conducted with 9,107 households containing nearly 25,000 individuals. Since 1969, the wave-to-wave response rate has been 91%-98%. The evolution of the sample size is described in detail in the PSID User Manual.

A number of studies have examined the representativeness of the PSID and selective attrition. The findings from these studies were recently summarised in this journal (McGonagle, Schoeni, Sastry, & Freedman, 2012). Studies find that the PSID estimates of a variety of socioeconomic characteristics of individuals are similar to estimates from contemporary cross-sectional gold standard surveys. The correspondence between the PSID and cross-sectional surveys is close even though attrition has been found to be associated with characteristics of respondents; notably there is higher attrition among lower income individuals. Nevertheless the PSID comparisons with other national survey data sources suggest that the Roster and Transfer sub-study will provide estimates that are representative of American families. An indication of its value as a survey representative of the U.S. population is the PSID's inclusion in the Cross-National Equivalent File, a major data source for comparative research (http://cnef.ehe.osu.edu/).

Instrument design, testing, and implementation

The battery of questions in the sub-study included a roster of parents and children age 18 and older. The questions identified biological and step relationships, and asked about transfers to and from these family members regardless of where they lived.

Family roster

We designed a new roster of family members to list parents and parents-in-law, children, step-parents, and step-children of PSID Heads and Wives[1]. As discussed above, the PSID's family-based following rules imply that the PSID already has substantial information about respondents' relatives because they have the PSID gene and are respondents themselves. However, data are not available for relatives of PSID households that do not have the PSID gene or for gened relatives who missed various waves and/or attrited from the study. For example, among married couples in which one spouse does not have the PSID gene, the non-gened spouse's parents and children from prior unions (stepchildren to the PSID-gened respondent) are not interviewed. Therefore, the roster obtained information about all current living adult children, stepchildren, parents, and step-parents of Heads and Wives in the main interview in 2013. Adult children were defined as those age 18 and older. For each adult (step)child and (step)parent , we collected: name, relationship to head/wife, gender, date of birth or age, marital status (including cohabitation for adult children), city and state of residence, educational attainment (for adult children; education of parents already exists in the PSID database), subjective general health status (excellent, very good, good, fair, poor), homeownership, and employment status. For adult offspring we also asked how many children they had. The total numbers of living siblings for the Head and the Wife also were collected because siblings are an important determinant of whether the responding adult child provides support to parents.

The roster includes the names of the full set of adult children and parents. When respondents stated that they had given (received) a transfer, they were asked to (by) whom the transfer was given (received). Interviewers had the family roster displayed on their computers, allowing them to easily select the correct individual; therefore, not only do we know that the respondent gave transfers to a child, we know which child, and the transfer can be linked to the information about that child that is reported in the roster as well as information on the child's PSID family for those children who are living in interviewed PSID family units.

Family transfers

The 2013 transfers module had two parts, with reference to each PSID Head and Wife: part one asked about ***recent transfers*** of time and money (over $100) given to and received from parents and parents in-law and about transfers of time and money given to and received from children and step-children in the last year; part two asked about large, ***life-cycle transfers*** of money that the Head/Wife received from parents since age 18 and about large transfers of money given to children and stepchildren since they were 18 years old. Table 1 outlines the questions.

Recent transfers

Questions were asked about the incidence and amount of transfers of money and time given and received over the last year. Respondents were asked about transfers with parents and children and were allowed to report transfers with both co-resident and non-co-resident parents and children. Similar questions have appeared in the PSID in the past. Table 2 outlines the transfer questions that were asked in a transfer module administered in the 1988 wave of the PSID and indicates when co-resident kin were included. As table 2 shows, respondents were asked about transfers with parents and were asked a general question about transfers with others. In both 1988 and 2013, respondents were asked enough information about the person with whom they engaged in transfers to link these transfers to specific individuals (including children). While maintaining as much comparability as possible with the 1988 PSID, we improved the questions in 2013 by designing specific questions about money and time transfers to children and stepchildren rather than relying on the more general question about transfers with others utilized in 1988 and by uniformly including co-resident family members. (See ftp://ftp.isr.umich.edu/pub/src/psid/questionnaires/q88.pdf for the 1988 questions.) If time transfers were reported in 2013, we also asked which person, the Head or Wife, gave most of the time transfers. While money transfers can be thought of as a transfer from one household to another, time transfers can be assessed for individuals within households.

Long-term, life-cycle transfers

The sub-study includes questions about large transfers that the Head and Wife of a PSID household each may have received from their parents (whether or not the parents are alive in 2013) and/or provided to their children since they/their children were age 18. The bottom panel of table 1 outlines the long-term transfer questions that were asked and when amounts were included. Two specific large life-cycle transfers were assessed—one for post-secondary education and a second for help with the purchase of a home—along with a more general question on large financial transfers between parents and their adult children. These questions capture retrospective information about important and salient types of transfers. For transfers to offspring, both whether assistance was provided and the amount of assistance was assessed. However, for transfers from parents only yes/no was assessed because of the potentially long recall period. Until 2013, the PSID had never asked these types of life-cycle transfer questions.

The roster and transfer module, as well as the entire PSID instrument administered in 2013, are available on the PSID website: ftp://ftp.isr.umich.edu/pub/src/psid/questionnaires/q2013.pdf.

Content that could not be included

The sub-study was constrained to including interview questions that totaled no more than 12 minutes on average across all respondents. This constraint was required by the Directors of the PSID to avoid overburdening respondents, with an average interview length of 94 minutes in total including other topics captured by the survey. Therefore, the project team was forced to make some difficult choices. Two sets of questions were the most difficult to exclude. First, the project team was hoping to include a roster of respondent's siblings, including half and stepsiblings, and their characteristics, as well as questions about transfers to and from siblings. Second, the team would have liked to have included additional measures such as relationship quality, recent episodes of unemployment, health events, intensity of time help, for instance hours of time help in each week, and changes in family structure such as divorce or marriage. We chose to exclude the roster of siblings because transfers of money and time help with siblings are much less common than transfers with parents or adult children (Kahn et al., 2011; Schoeni, 1997). Our choice of characteristics to measure for individuals included in the roster was based on findings from prior literature on the most salient determinants of money and time transfers as well as evidence on which characteristics could be accurately reported by a parent or adult child with only a small number of questions. Another decision the investigators made based on timing estimates from the PSID pretest was to shorten question stems by deleting explanations of types of financial and time transfers to include in responses. The shorter question stems had the advantage of making the module shorter and more conversational. But by excluding examples, the questions may have elicited responses with more underlying variation in the definitions that respondents used to inform their answers (Schaeffer & Presser, 2003). Finally, the investigators chose to exclude amounts from questions about long-term transfers received from parents or spouse's parents because of time constraints on the length of the interview and pretesting for the pilot study suggested that respondents had difficulty reporting amounts of long-term money received, particularly by their spouse. This finding is not surprising given that many respondents had not married their spouse until well after such transfers were received.

Pilot test on national sample

The roster and transfer module was tested on the June 2012 Michigan Survey of Consumer Attitudes (SCA). SCA is a telephone survey of a national probability sample of U.S. adults age 18 and older. The survey collected data from 495 respondents. Approximately two thirds of respondents, 314, had adult children, and these parent respondents provided information on 789 adult children. These data are expected to become available at the University of Michigan’s Inter-University Consortium for Political and Social Research by August 2015.

Survey field effort

The roster and transfer questions were included in the 2013 PSID main interview. The Survey Research Operations group at the Institute for Social Research, University of Michigan, fielded the survey. The 115 interviewers were located throughout the United States, and they were required to successfully complete in-person training prior to beginning fieldwork. At training, interviewers were instructed on PSID specific terminology and procedures, and they were required to pass an examination on this material before beginning fieldwork. Interviewing began in March and finished at the end of December.

Interviews were obtained with 9,107 family units. The re-interview response rate was 95%. Of the completed interviews, 97% were conducted on the telephone and the remainder face-to-face. A cell phone was used by 80% of telephone respondents. The average number of calls to complete an interview was 14.1, with a median of six. Nearly a quarter, 23%, of completed cases required multiple sessions to finish the interview.

Various strategies were used to maximize response rates. Respondents were provided with a $70 incentive, which was mailed to them after completing the interview. To compensate for the use of paid minutes of phone time, respondents using a cell phone were offered an additional $10. Interviewers who worked for PSID previously were assigned to cases they had successfully interviewed in the past. Approximately six months before fieldwork began, sample members were mailed a newsletter that provided short articles describing new research findings based on the PSID. A postcard was also sent to respondents seeking updated contact information. Respondents were sent a $10 check if they returned the postcard with updated information.

Data access, ethical approval, and funding source

PSID data and documentation for all survey years are freely available on the PSID website: www.PSID.org. Users can download all data for a particular year as one large file. Alternatively, the PSID website will create customised cross-year extracts, with users simply choosing the variables they require from each year. Among the thousands of variables are characteristics of the interview process including interview length, number of calls to complete the interview, mode of interview (telephone versus face-to-face), interviewer identification number, date of interview, and a lengthy set of interviewer observations about the interview process for each particular respondent.

An initial release of the 2013 transfer and roster data and documentation, along with a video introduction to the sub-study, is available online at http://simba.isr.umich.edu/Zips/zipSupp.aspx#RAT13. The final and fully edited version of the data is scheduled to be released fall of 2015.

The University of Michigan Health Sciences and Behavioral Sciences Institutional Review Board reviewed and approved the 2013 PSID data collection and distribution protocols and survey instrument to ensure the rights and welfare of research participants were protected.

Funding for this sub-study was provided by the U.S. National Institute on Aging through a program project grant (P01AG029409).

Item non-response

Tables 3 and 4 report un-weighted frequencies of money and time transfers between dyads of parent units and children. We refer to ‘parent units’ because questions about transfers to and from married (or cohabiting) parents did not distinguish between transfers to (from) individual parents. The tables include the percentage of responses where the respondent reported they did not know the answer to the question or refused to answer. The rates of nonresponse are quite low, no higher than 4%. For example, table 3 shows that among all child respondent-parent pairs, time help in the past year was given to parents in 32% of cases, time help was not given in 66% of cases and only 2% of respondent-parent pairs were item nonresponse.

The top panel of table 4 reports the percentage of parent-unit-child pairs in which the parent reported that they gave the child money for schooling, to buy a home, or something else since the child was age 18. The bottom panels show the percentage of Heads and Wives who reported receiving such help from parents since age 18. As for reports about short-term transfers, there are low percentages of item nonresponse. This is especially notable because the respondent reports about his or her own parents as well as the spouse’s parents, and knowledge about long-term transfers to a spouse might be more limited than knowledge about transfers from the respondent’s own parents.

Conclusions

There have been fundamental changes in the family, the economy, and the social safety net in the United States over the last few decades. The new Roster and Transfers sub-study of the PSID will allow researchers to document many of these changes, as well as examine the impact of these changes on families and individuals.

The sub-study highlights the importance of ongoing longitudinal studies. By including the sub-study in the PSID, researchers can draw upon the 45 years of information on respondents and their families to more fully understand the causes and consequences of intergenerational transfers. Moreover, as more and different data are collected in future waves of the PSID, the data collected in the 2013 sub-study can be utilized in conjunction with future data to assess the impact of intergenerational transfers.

The addition of this sub-study to the long-running PSID may motivate the addition of similar modules to the ongoing panel studies in the other countries represented in the Cross-National Equivalence File, including the British Household Panel Study (BHPS), the Australian Household Income and Labor Dynamics Study (HILDA), the Korean Labor and Income Panel Study (KLIPS), the Russia Longitudinal Monitoring Survey (RLMS), the Canadian Survey of Labour and Income Dynamics (SLID), and the German Socio-Economic Panel (SOEP). Inclusion in even a small number of these studies would enable comparative research on how private transfers among family members complement or substitute for public transfers or government support. The different institutional contexts represented by these surveys, particularly variation in arrangements for postsecondary schooling, child care, and old age support, would shed important light on explanations for intra-family transfers of time and money. Even without the potential for new data to support this comparative effort, the roster and transfer sub-study in the PSID is a valuable new public resource for investigating the familial process that may contribute to inequality within and between generations.

Acknowledgements

Funding for this sub-study was provided by the United States National Institute on Aging (P01AG029409). This work was also supported, in part, by the Duke Population Research Institute (DuPRI), which receives core support from grant P30-AG034424 funded by the National Institute on Aging, and the California Center for Population Research at UCLA (CCPR), which receives core support (R24- HD041022) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD).

References

Altonji, J.G., Hayashi, F., & Kotlikoff, L.J. (1997). Parental Altruism and Inter Vivos Transfers: Theory and Evidence. *Journal of Political Economy, 105*(6), 1121-66. http://dx.doi.org/10.1086/516388

Autor, D.H., Katz, L.F., & Kearney, M.S. (2008). Trends in U.S. Wage Inequality: Revising the Revisionists. *Review of Economics and Statistics, 90*(2), 300-323 http://dx.doi.org/10.1162/rest.90.2.300

Becker, G.S. and Tomes, N. (1979). "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility." *Journal of Political Economy*, *87*(6): 1153-1189. http://dx.doi.org/10.1086/260831

Bianchi, S.M., Hotz, V.J., McGarry, K., &, Seltzer, J.A. (2008). Intergenerational Ties: Theories, Trends, and Challenges. In A. Booth, A.C. Crouter, S.M. Bianchi, & J.A. Seltzer (Eds.), *Intergenerational Caregiving* (pp. 3-43). Washington DC: Urban Institute Press.

Bratsberg, B., Røed, K., Raaum, O., Naylor, R., Jäntti, M., Eriksson, T., Österbacka, E., & Björklund, A. (2007). Nonlinearities in Intergenerational Earnings Mobility: Consequences for Cross-Country Comparisons. *Economic Journal,* 117(519), C72–C92. http://dx.doi.org/10.1111/j.1468-0297.2007.02036.x

Bumpass, L., & Lu, H.H. (2000). Trends in Cohabitation and Implications for Children's Family Contexts in the U.S. *Population Studies*, 54, 29-41

Center on an Aging Society. (2005). *Adult Children: The Likelihood of Providing Care for an Older Parent*. Data Profile Number 2. Washington, DC: Georgetown University.

Charles, K.K. & Hurst, E. (2003). The Correlation of Wealth across Generations. *Journal of Political Economy,* 111(6), 1155-1182.
http://dx.doi.org/10.1086/378526

Coleman, M., & Ganong, L. (2008). Normative Beliefs About Sharing Housing with an Older Family Member. *International Journal of Aging and Human Development*, 66, 49-72.
http://dx.doi.org/10.2190/AG.66.1.c

Cox, D., & Stark,O. (2005). On the Demand for Grandchildren: Tied Transfers, Liquidity Constraints and the 'Demonstration Effect'. *Journal of Public Economics*, 89, 665-697.
http://dx.doi.org/10.1016/j.jpubeco.2004.04.005

Eggebeen, D.J. (1992). Family Structure and Intergenerational Exchanges. *Research on Aging*,14, 427-447.
http://dx.doi.org/10.1177/0164027592144001

Engelhardt, G.V., & Mayer, C.J. (1998). Intergenerational Transfers, Borrowing Constraints, and Saving Behavior: Evidence from the Housing Market. *Journal of Urban Economics,* 44(1), 135-157.
http://dx.doi.org/10.1006/juec.1997.2064

Furstenberg, F.F.,Jr., Hoffman, S.D., & Shrestha, L. (1995). The Effect of Divorce on Intergenerational Transfers: New Evidence. *Demography*, 32(3), 319-333.
http://dx.doi.org/10.2307/2061683

Harding, D.J., Jencks, C., Lopoo, L.M., & Mayer, S.E. (2005). The Changing Effect of Family Background on the Incomes of American Adults. In S. Bowles, H. Gintis, & M. Osborne Groves (Eds.), *Unequal Chances: Family Background and Economic Success* (pp. 100-144). Princeton, NJ: Russell Sage Foundation and Princeton University Press.

Kahn, J.R., McGill, B.S., & Bianchi, S.M. (2011). Help to Family and Friends: Are There Gender Differences at Older Ages? *Journal of Marriage and Family*. NIHMS: NIHMS269750

Kaplan, Greg. 2012. "Moving Back Home: Insurance Against Labor Market Risk." *Journal of Political Economy, 120(*3): 446–512.
http://dx.doi.org/10.1086/666588

Kennedy, S., & Bumpass, L. (2008). Cohabitation and Children's Living Arrangements: New Estimates from the United States. *Demographic Research*, *19*, 1663-1692. PMCID: PMC2612998
http://dx.doi.org/10.4054/DemRes.2008.19.47

Lee, C.I., & Solon, G. (2009). Trends in Intergenerational Income Mobility. *Review of Economics and Statistics, 91*(4), 766-772.
http://dx.doi.org/10.1162/rest.91.4.766

Lovenheim, M.F. (2011). The Effect of Liquid Housing Wealth on College Enrollment. *Journal of Labor Economics, 29*(4) (October 2011), pp. 741-771.
http://dx.doi.org/10.1086/660775

McGarry, K., & Schoeni, R. (1995). Transfer Behavior in the Health and Retirement Study. *Journal of Human Resources, 30* (supplement): S184-S226.
http://dx.doi.org/10.2307/146283

McGarry, K. (1997). Transfer Behavior within the Family: Results from the Asset and Health Dynamics Study. *Journals of Gerontology, 52B*: 82-92.
http://dx.doi.org/10.1093/geronb/52B.Special_Issue.82

McLanahan, S. (2004). Diverging Destinies: How Children Are Faring Under the Second Demographic Transition. *Demography, 41*(4), 607-627.
http://dx.doi.org/10.1353/dem.2004.0033

K. McGonagle, R. Schoeni, N. Sastry, V. Freedman. (2012) The Panel Study of Income Dynamics: Overview, recent renovations, and potential for life course research. *Longitudinal and Life Course Studies*, *3,* 268-284.

Pezzin, L.E., & Steinberg Schone, B. (1999). Parental Marital Disruption and Intergenerational Transfers: An Analysis of Lone Elderly Parents and Their Children. *Demography, 36*, 287-297.
http://dx.doi.org/10.2307/2648053

Pezzin, L.E., Pollak, R.A., & Steinberg Schone, B. (2008). Parental Marital Disruption, Family Type, and Transfers to Disabled Elderly Parents. *Journal of Gerontology: Social Sciences*, *63B*(6), S349–S358.
http://dx.doi.org/10.1093/geronb/63.6.S349

Schaeffer, N. C., Presser, S. (2003). The Science of Asking Questions. *Annual Review of Sociology*, *29*, 65-88. http://dx.doi.org/10.1146/annurev.soc.29.110702.110112

Schoeni, R. (1997). Private Interhousehold Transfers of Money and Time: New Empirical Evidence, *Review of Income and Wealth 43*(4), 423-448. http://dx.doi.org/10.1111/j.1475-4991.1997.tb00234.x

Schoeni, R.F., & Ross, K.E. (2005). Material Assistance from Families during the Transition to Adulthood. In R. A. Settersten, Jr., F.F. Furstenberg, Jr., & R.G. Rumbaut (Eds.), *On the Frontier of Adulthood*, (pp. 396-416). Chicago: University of Chicago Press. http://dx.doi.org/10.7208/chicago/9780226748924.003.0012

Seltzer, J. A., Yahirun, J.J., & Bianchi, S.B. 2013. Coresidence and Geographic Proximity of Mothers and Adult Children in Stepfamilies. *Journal of Marriage and Family 75*(5), 1164–80. http://dx.doi.org/10.1111/jomf.12058

Solon, G.R. (1992). Intergenerational Income Mobility in the United States. *American Economic Review*, *82*(3), 393-408.

Spillman, B.C., & Pezzin, L.E. (2000). Potential and Active Family Caregivers: Changing Networks and the 'Sandwich Generation'. *Milbank Quarterly*, *78*(3), 347-374. http://dx.doi.org/10.1111/1468-0009.00177

Ward, R., Logan, J., & Spitze, G. (1992). The Influence of Parent and Child Needs on Coresidence in Middle and Later Life. *Journal of Marriage and Family, 54*(1), 209-221 http://dx.doi.org/10.2307/353288

Wolf, D.A., Soldo, B., & Freedman, V. (1996). The Demography of Family Care for the Elderly. In T. Hareven (Ed.), *Aging and Generational Relations Over the Life Course : A Historical and Cross-Cultural Perspective*, (pp.115-138). Hawthorne, NY: Walter de Gruyter, Los Angeles, CA: Rand Corporation. http://dx.doi.org/10.1177/089826439901100306

Wolf, D.A. (1999). The Family as Provider of Long-Term Care: Efficiency, Equity, and Externalities. *Journal of Aging and Health*, *11*(3), 30-82.

Zissimopoulos, J., & Smith, J.P. (2009). Unequal Giving: Monetary Gifts to Children Across Countries and Over Time. Rand Labor and Population Working Paper No. WR-723, Los Angeles, CA: Rand Corporation.

Endnotes

[1] PSID defines household headship as the Census did when the PSID began. In a husband-wife household the husband is designated the Head, regardless of whether he has the PSID gene.

Table 1. Summary of 2013 transfer questions

Type of Help/Transfer	Amount Collected?
Panel A. Recent Transfers	
Transfers in last year to/from parents/children of PSID Heads and Wives	
Time Given	Y
Time Received	Y
Money Given ($100+)	Y
Money Received ($100+)	Y
Living Arrangements with PSID HH:	
Co-Residence	N/A
Panel B. Long Term, Life-cycle Transfers	
Transfers since age 18 to/from parents/children of PSID Heads and Wives	
Help with post-secondary education given to children	Y
Help with post-secondary education received from parents	N
Help with home purchase given to children	Y
Help with home purchase received from parents	N
Other large transfers given to children	Y
Other large transfers received from parents	N

Table 2. Summary of 1988 transfer questions

Type of Help/Transfer	Separate questions for parents vs. non-parents?	Did the other party reside in or outside the PSID family at the time of the 1988 Interview?	
		Transfer with Parent	Transfer with Non-parent
Time Given	Y	In or Outside	Outside
Time Received	Y	In or Outside	Outside
Money Given ($100+)	N	Outside	Outside
Money Received ($100+)	Y	In or Outside	Outside

Table 3. Types of transfers in the last year and item nonresponse (%)

	Money Given ($100+)	Money Received ($100+)	Time Given	Time Received
Children (N=11511)				
Yes	29%	8%	28%	21%
No	69%	91%	68%	76%
DK/RF	2%	1%	4%	2%
Parent Units (N=12845)				
Yes	14%	17%	32%	26%
No	85%	82%	66%	73%
DK/RF	1%	1%	2%	1%

Table 4. Types of transfers since age 18 and item nonresponse (%)

	Money for School	Money for Home	Money for Other
Children (N=11511)			
Yes	18%	2%	10%
No	79%	97%	88%
DK/RF	4%	0%	2%
Head (N=9107)			
Yes	20%	5%	17%
No	79%	94%	82%
DK/RF	1%	0%	1%
Wife (N=4638)			
Yes	24%	6%	13%
No	76%	94%	87%
DK/RF	1%	0%	0%

COMMENT AND DEBATE
Social class differences in early cognitive development

(Received April 2015 Revised June 2015) http:// dx.doi.org/10.14301/llcs.v6i3.361

John Bynner

Few pieces of recent longitudinal research have had as much influence in United Kingdom policy circles as Leon Feinstein's analysis of 1970 birth cohort study data (reported in 2003) on cognitive development assessed at ages 22 months, 42 months five years and ten years. Breakdown by social class of the test performance data demonstrated that infants of superior cognitive ability in the first assessment from working class backgrounds showed relative decline with age in test performance compared with their middle class counterparts, who while starting from an inferior position, subsequently overtook the working class group. The findings were embodied in what became a famous graph showing this crossover and consequent reversal of predicted life chances. They pointed to substantial obstacles to social mobility and were an important factor in the policy response of major pre-school educational interventions such as the Sure Start programme introduced by the Labour Government to reverse the trend, which attracted support from across the political spectrum.

Subsequent re-analysis by John Jerrim and Anna Vignoles challenged the existence of the crossover as a statistical artefact attributing it to the well-known phenomenon of 'regression to the mean'. The consequence was a cooling off of support for the intervention policy directed at strengthening working class children's early cognitive performance. This shift included termination of the Sure Start programme by the new Coalition Government (Conservative and Liberal Democrat) that took office in 2010. Subsequent research has qualified the picture further raising issues on a number of methodological and substantive fronts – especially the need to give more attention to measurement error in such work and for a more nuanced interpretation of such longitudinal research results. Social class disadvantage in cognitive development is well established but the relative loss of competence developmentally needs to be treated with caution.

In an opening paper Leon Feinstein reviews the methodological criticism of his original research. The points he raises are then debated in commentaries by John Jerrim and Anna Vignoles, Harvey Goldstein and Robert French, Elizabeth Washbrook and RaeHyuck Lee and Ruth Lupton. Leon Feinstein's response to these will be published in the next issue of the journal.

Opening paper by

Leon Feinstein Early Intervention Foundation, UK
leon.feinstein@eif.org.uk

Social class differences in early cognitive development and regression to the mean

Introduction

In April 2011 the then new Coalition Government published its social mobility strategy (HM Government, 2011). As a minor reference within the overall document, figure 2 of Feinstein (2003) was reproduced on page eight as a reference to the claim that "Bright children from poorer families tend to fall back relative to more advantaged peers who have not performed as well."

This claim in the strategy brought an immediate response in a press release from Professor Daniel Read (2003) of Warwick Business School[1] claiming that:

> I am very worried that this graph is being used to shape policy when in fact many statisticians will instantly see that it simply replicates a statistical trap or artefact called 'regression toward the mean'. The apparently shocking pattern of results in the graph is simply what statisticians would expect when you measure extremes of performance in two populations of differing ability. The Feinstein graph is constructed... with undue emphasis on extreme results.

Simultaneously, Jerrim and Vignoles (2011), published a paper on the "use (and misuse) of statistics," undertaking a series of simulations and new analyses based on a model assuming "true abilities" with pre-determined social class mean gaps, which shows that under reasonable assumptions the pattern in the chart could result from regression to the mean.

The 1970 Cohort Study is no longer such a significant source of information on the degree of British inequality in contemporary childhood. We have now much larger and more recent studies such as the Millennium Cohort Study, the first major, public United Kingdom (UK) birth cohort study since the 1970 Cohort and the Life Study at University College London (UCL) with an intended sample of 100,000 babies and a wide range of developmental, health epigenetic and neuro-scientific observations.

But the 1970 Cohort data in general are still of considerable interest and many of those who may have quoted the graph in the past may have been disappointed to learn that they had been so badly misled so I am grateful to the editors of this journal for the opportunity to respond to the critique and to raise a handful of questions for further debate. As the quotations above make clear, there is a wider debate both in learned journals and on the pages in the national newspapers. They do not operate by the same rules. I am grateful to the editors for inviting this paper in a special comment and debate section of the journal that is also about the relationship between research, policy and practice.

Some in the policy world have used the graph without understanding it, though it is perhaps hard to see this an issue specific solely to this chart. Some of those who have used it have understood its weaknesses but found it informative, some have had no idea and are horrified it is all so complicated. Few will have much mind to it. Without wishing to reinstate the graph as a "killer chart"[2] I would like, in this paper, to correct some of the mis-representations that have been suggested and suggest some issues for further discussion.

The graph was based on a relatively small sample from a time and place that is, with due respect to its members, now distant. The Britain of today is transformed in terms of ethnicity and the way inequality is experienced. Much larger datasets are available with much larger samples and more consistent measurement across a broader range of aspects of cognitive development. New methodologies are available. So my argument is not that anyone should return to this chart as the way to model and measure the interaction of a distal measure like class and tests of cognitive ability through childhood, but that there are a range of ways of modelling these data, recognising the importance of the measures used, the age at which children are tested, how different models lead to different tests of this social level interaction and how this plays out in different times and places.

I set out below the basic facts of the graph and then discuss in more detail what I think the graph means and raise some questions about meaning and inference, in particular with regard to the definition of true ability and the difference between average, macro-social phenomena and the lives of individuals. The first section sets out first the charts and then their source. The second section describes some of the challenges for inference that it has raised. The third section concludes with reference to these themes.

The chart

Figure 1: Average rank of test scores at 22, 42, 60 & 120 months, by socioeconomic status (SES) of parents

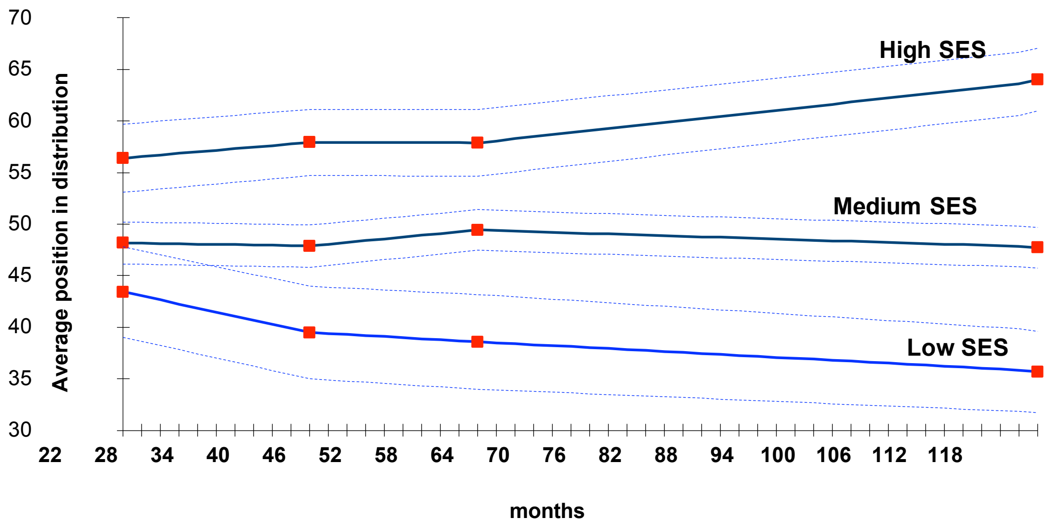


Figure 2: Average rank of test scores at 22, 42, 60 & 120 months, by SES of parents and early rank position

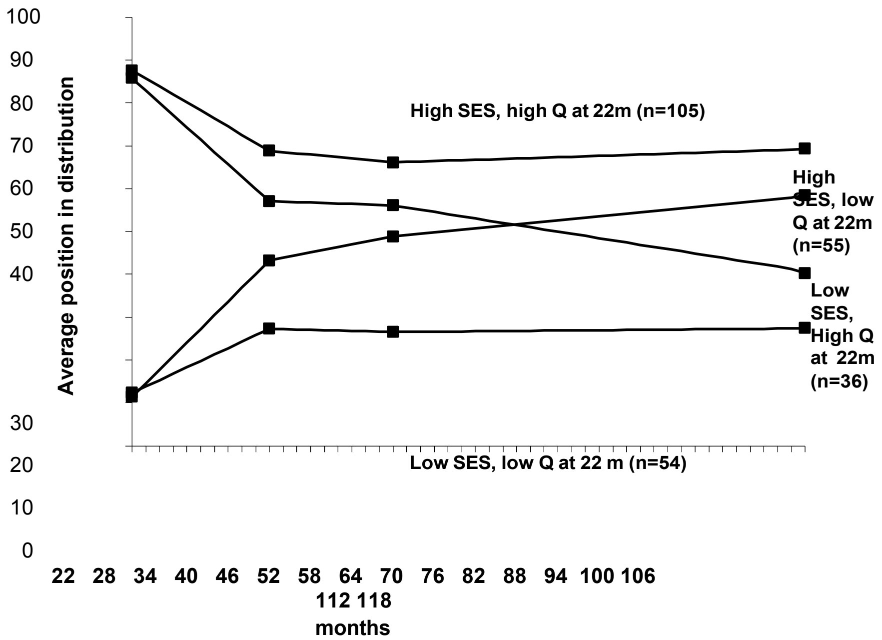


Figure 1 in Feinstein (2003) reports the mean relative positions of children in the 1970 British Cohort Study from different social class groups in age appropriate and hence very different tests of cognitive development at four ages in early and middle childhood (ages 22 months, 42 months, five years and ten years). The explicit rationale of the paper was to test the extent to which average gaps in cognitive development between children from different types of family background were evident before children started school. The difficulty of testing this and of comparing gaps at different ages seemed to centre on the difficulty of finding comparable tests through the complex qualitative developmental changes that occur in early child development. The innovation of the paper was in finding a coherent way to simplify and address the problem. It made possible a cursory, uni-dimensional study of development at the level of average groups of children.

Figure 2 indicates some of the interaction between family background and early ability scores, at particular points of the distribution of early scores.

The main rationale of the paper and of initial discussion of it was of the finding of early socioeconomic status (SES) gaps in figure 1. Figure 2 subsequently became used by me and others in public debate, showing that the children from high SES backgrounds who scored poorly on the age 22 month tests had a higher mean score at 10 years than the low SES children who scored well in the early tests. The shift in relative mean position occurs between the age of five and ten. This always seemed to me remarkable, though recognising that it may be due in untested proportions to measurement error, context, gene*environment interactions, gene*environment correlations and cultural bias. As I said at the time, the graph does not and cannot resolve the question of explanation but it does describe a very common pattern in UK data.

Methods and measurement

Sample

The 1970 Cohort Study was a representative sample based on children born in the first week of April 1970 undertaken initially by the Department of Child Health at Bristol University and taken on to adulthood by the International Centre for Child Studies, City University and then the Institute of Education. Data was collected from members of the cohort studies and from members of their households, through a range of tests and interviews at different ages, as well as through interviews with teachers and medical officers.

The full sample comprising 17,196 children were studied at birth, of whom 13,135 were picked up at age five years and 13,871 at ten years. For the 2003 paper, data were also drawn from the 22 month and 42 month sub-samples of the study. These studies initially comprised 2,457 individuals, of which half were selected from the full sample because at risk of foetal malnutrition and half as a random control group. Attrition and non- response provides a sample of 1,292 children providing test score data and a measure of SES at ages 22 months, 42 months, five years and ten years. This degree of sample loss represents a tremendous achievement by the study team. It is not unproblematic if attrition is non-random such that the remaining sample is no longer typical of the population it is intended to represent. The kind of non-randomness required to cause bias to the general indications of figure 1 and 2 would be one in which children most likely to have been omitted were ones for whom on average the relationship between SES and the development of test scores was different than for the included children.

There is no evidence of this. The paper did have an eye to the differences between children in the control group and those who were selected for the sub-sample because of a distal concern about foetal development. Both issues of sample selection and attrition merit further work. If the paper were to be published today referees would require a greater focus on the handling of missing data than was the case before Little and Rubin (2002) and others had created software and approaches to handling missingness and a greater appreciation of its importance.

Socioeconomic status (SES)

The measures of occupation deposited in the ESRC Data Archive that holds the 1970 Cohort Study were developed by the study team to classify the children by social class using the Registrar-General's Social Classes, which were introduced in 1913 and were renamed in 1990 as Social Class based on Occupation. This was then further aggregated in the 2003 paper into three groups comprising 307 children in a high SES group, 814 in a middle SES group and 171 in a low SES group. There were two further simplifications. Firstly, only one measure of

SES was used for each child, rather than allowing it to change through childhood as family circumstances evolved. The birth measure was used where this was available. Second, the measure was an aggregation of the SES of the male and female adults in the household, in most cases the biological parents. The majority of the mothers of the children in the 1970 Cohort were not working and for these children SES was categorised on the basis of the father's occupation. Where both parents were working and had occupations in different SES groups, high or low SES dominated in the categorisation, such that a child in a household with a high SES father and middle SES mother would be categorised as high SES and vice versa. The very small number of high/lows were categorised as middle.

The intention was to create a simple indicator of occupational skills and access to earnings and status in the economy and society of the day. The broad pattern of results was found to be very similar if groups were constructed on the basis of parents' education levels (Feinstein 2003). The intention was not to indicate that these groupings reflected common genetic inheritances or to indicate anything causal about social class nor to test hypotheses about the nature of social class. Other ways of classifying the children based on different ways of reflecting aspects of family backgrounds such as the distinct contributions of work situation and market situation (see e.g. Erikson 1984) or different ways of reflecting mothers' and fathers' contributions to each would have been possible and equally of interest. These are broad averages based on the available data on the occupation of parents.

Cognitive performance

Comparing measures of cognitive performance at different ages in childhood is made particularly difficult by the qualitative shifts in development that transform the meaning of cognitive capability as children mature. At each age in these data a different set of tests are taken by the children because the tests span the period from early to middle childhood through which considerable shifts in the meaning, nature and measurable manifestation of cognitive development takes place. Piaget (1952), for example distinguishes the sensorimotor stage, from birth to age two in which infants seek and find knowledge through sensory experiences and manipulation of objects; the preoperational stage, from age two to about age seven in which pretending and play are evidently essential to learning; and the concrete operational stage, from age seven to 11 in which logic becomes more routine if at times rigidly applied. In Piaget there is also a formal operational stage, which begins in adolescence and spans into adulthood with an increase in logic, the ability to use deductive reasoning, and an understanding of abstract ideas. There are of course many other models of the nature of developmental change in this period but it is not disputed that very fundamental qualitative change in behaviour and capability occurs through early childhood.

The 22 months tests comprised cube stacking, measures of personal development, measures of language use and a drawing task. The information was collected by a health visitor recruited by the Department of Child Health at the University of Bristol during a visit to the home or other residence of the sample children[3]. The personal development measures were a set of requests from the health visitor to the child such as to point to her nose or her eyes. The cube-stacking test was designed as a test of motor ability, a precursor of later capabilities such as intelligence as well as of physical dexterity. The measures of language use concerned the child's ability reported by the mother to say "ma-ma" and "da-da" and associate the words with the appropriate persons.

The measures as a whole were less well standardised than measures available in the newer UK cohort studies (Chamberlain and Davey, 1976), but merit further study. At age ten years the tests were of maths and reading and the British ability scale test of IQ.

Therefore these measures reflect the transition from early public language use as a 22-month-old child responds to the request from a health visitor to point to her nose to the experience of the age ten child sitting a maths test in a classroom. The relationship between the different features of cognitive and non-cognitive development that lead from the 22 month child pointing to her nose when asked to and the girl sitting the exam are only partially understood but it is clear that there is no simple linear relationship between specific domains of development in early childhood and equivalent domains in adolescence or adulthood.

At 42 months there were tests of what were called in Feinstein (2003) "counting" (use of cubes, such as counting the number of cubes placed on a table) and "speaking" (correctly naming pictures such as of a car) and a copying designs test, all conducted by trained researchers or health visitors in the home. At age five years there was a test of vocabulary, a copying designs test and a human figure-drawing test (Feinstein, 2003). The age 42 month speaking and counting tests are equal in the prediction of the age ten maths and reading scores with little domain continuity spanning across early to middle childhood (Feinstein, 2003). Not surprising as the tasks all involve elements of language, communication, motor skills and attention, amongst other capabilities.

More and better modelling of this issue is possible now in multiple datasets, not least the Avon Longitudinal Study of Parents and Children which has annual measurement across a much better set of measures of cognitive and other development than were possible in the 1970 Cohort Study.

It is important to emphasise that in order to derive the best possible signal from the available data, the measure used in the graph at each age is not any single test but rather the ordinal position of the children at each age in a weighted average of the scores available at that age. The particular form of weighting is the first principal component, chosen to maximise the variance in the weighted index. Particular tests count higher in the weighting if they add more unique information than other tests. The results are robust to the use of other weighting schema such as regression weights taken from regressing the age ten tests on the earlier tests. Therefore, although the measures used change through development the dependent variable itself is always the relative achievement of the sample children at each age in the age appropriate measures of cognitive development available in the 1970 Cohort Study.

The implications of this are that the dependent variable is ordinal position and has meaning only in this relative sense. It does not measure achievement on any specific test but is an estimate of relative cognitive capability in the available age-appropriate measures. It is therefore different to the repeat measures test that Jerrim and Vignoles (2013) use in their analysis of the Millennium Cohort Study.

Descriptive findings

Figure 1 reports the average (mean) relative positions of children classified by the three-fold categorisation of social class on the single measure of relative ability drawn from the range of age appropriate but different tests at each of the four ages.

Figure 2 reports the mean scores of children from different social class groups at the four ages. Critically, it classified children according to their scores on the first set of tests at age 22 months and considers also the mean scores for children depending on their rank in the early scores.

The early scores are particularly unstable. As I show in the original paper they contain sufficient real information about early development to predict final educational qualifications achieved but the correlation is only just statistically significant and weak. Therefore, as I also show in the original paper, it is not surprising that children's scores subsequently move around a great deal.

Subsequent discussion has focussed on the meaningfulness of the crossover in figure 2, regression to the mean, the appropriateness of classifying children to ability groups at 22 months and the focus on arbitrary quartile groups.

Inference and regression to the mean

When the graph was first presented at an econometrics seminar at UCL issues were raised about measurement error and causality. Then and since there have been debates about regression to the mean, a notion whose original use was by Galton (1886) who discussed the issue in terms of the tendency of the individual deviation from the mean in height to be larger in one generation than the next, so that tall parents will tend to have slightly less tall children. The degree of randomness in a variable through measurement error or chance will mean that the deviation from the mean in one period is that much less likely to be replicated in the next.

In this instance regression to the mean refers in part to the statistical property by which because of misclassification bias in the early groupings, those who appeared to do well early on will have a tendency to lower scores in subsequent tests. Conversely, those who score badly early on will have a tendency to better scores. This was shown by Tu and Law (2010) to be a fatal problem for interpretation of the chart as the outcomes for those from different social class groups with different "true" ability.

Based on their resulting modelling Jerrim and

Vignoles (2011; 2013) have shown how regression to the mean plays out in the kind of data shown in the chart and also used the phrase in other ways. Using simulations, they show that the pattern observed in the chart can be substantially reproduced as the result of regression to the mean of various kinds, including both error in measurement and hence classification to high and low groups and differences in what is tested at different ages.

They reference Nick Clegg's use of the phrasing that: "By the age of five, bright children from poorer backgrounds have been overtaken by less bright children from richer ones — and from this point on, the gaps tend to widen still further."

Jerrim and Vignoles wanted to correct the misapprehension that the graph shows that bright working class children in mid-childhood will necessarily fall behind dim middle class children in middle childhood. This misapprehension would be based on the presumptions that: the graph represents a necessary feature of the development of all individuals rather than representing average phenomena; and that it is meaningful and technically possible to identify stable cognitive capabilities at 22 months such that "bright" and "dim" children can meaningfully be identified and classified as such based on tests at 22 months.

Yet, it is not necessary to believe that the groupings are stable, innate or fixed to find the data in figure 2 interesting. The graph shows what happens to the average test scores of different clusters of children in an interaction between an indicator of family origin and average measures of cognitive development, starting very early in childhood. Low social economic status (SES) children in the UK tended (and still tend) on average to fall back relative to middle class children, whatever the early levels of measured ability.

Read (2003) is wrong that the shift between 22 and 42 months was taken by policy makers to be substantive.[1] Much of the chart's role in public debate was as a proxy for a much wider body of research, including more recent analysis of the National Pupil Database and other more recent cohort studies showing how at every stage of education, low income children tend to progress at a slower rate on average than those on higher incomes (Kingdon and Cassen, 2007; Goodman and Gregg, 2010; Magnuson, Waldfogel and Washbrook, 2012). The broad fact is not disputed that the relative access of parents to wealth, income and educational knowledge on average tend to be replicated across generations, in the UK, now and in the past.

It is regrettable that Feinstein (2003) did not include more consideration of the reliabilities of the measures used because differences in reliability at different ages are likely to be responsible for a considerable but unquantified part of the observed pattern of results as children mature. If reliability of measurement increases with age then one might expect the fanning observed in figure 1 and the resulting pattern of figure 2. It is also important that the age ten tests may be more discriminating as tests of cognitive development than the age five tests.

The intention in Feinstein (2003) was explicitly descriptive, aiming to offer a sense of scale of the emergence of the gaps in average scores by children classified in very broad groupings in a very raw and single index of cognitive development.

Measurement error was not dealt with. The aim was to present the actual data, of the kind that is used to test children and award them grades and qualifications, suffering as this does from measurement error, rather than to present corrected trajectories based on modelling assumptions.

The change between ages five and ten years

Jerrim and Vignoles show that under reasonable, though not proven, assumptions the misclassification bias in the average score washes out after the second point of measurement. Therefore, the change in relative position between age five and ten years may be substantive. They note it may be due to a difference in the underlying tests, and suggest therefore this should also be seen as a form of regression to the mean.

The required assumption in their model is that the measurement error at the later ages is not correlated with the measurement error in the earlier scores. The age five tests were taken with a different set of instruments than those at 22 or 42 months, as set out above. They were conducted in the home by health visitors. The age ten tests are a different set of tests again, much more scholastic with a strong focus on maths and reading tested through a longer series of questions asked during a test session in the child's school. Therefore, it seems reasonable to suppose that measurement errors of this sort are not correlated across ages.

However, there are other sorts of possible measurement error that may well be persistent across ages, depending on what is meant by true ability. Some have argued (Gillborn & Youdell, 2000) from a more sociological perspective that low SES children will tend to under-perform in tests of cognitive capability because the tests reflect codes, expectations and structures of power that are themselves class-based.

Perhaps authors in this series might comment on the likelihood and implications of the assumption of zero correlation in measurement error across ages. I certainly agree that the difference in the underlying tests is important, but labelling this regression to the mean in a public debate seems to me to confuse the error resulting from misclassification bias in the early tests with the idea of a genetic basis to social class groups. Although this latter shift could technically be described as "regression to the mean," it is of a very different sort to that of the first kind, and is not adequately explained as necessarily a statistical phenomenon. This is an issue on which further clarification would be useful.

Although concerned with measurement, I see the data in figure 2 as evidence that children from working class families in the 1970 Cohort Study who not only scored well at 22 months on fairly raw tests of cognitive capability, but continued on average to do so at ages 42 months and five years, did not on average translate this ability into school success at age ten at anything like the rate of children in middle and upper class families. The shift from more general features of cognitive development at age five to more scholastic tests of reading and maths is important. Working class children in the 1970s appear to have tended to do worse on average on the age ten scholastic tests than they did on the more general age five tests. Some may argue this is because the age ten tests are better measures of true ability and so better indicate the true abilities of children from the different social class groups. My interpretation is more that the working class children tended to translate their earlier capabilities into success in scholastic test scores less well than did their middle class peers. As has been said many times the graph does not resolve this question.

By age ten it is meaningful and possible to conduct long tests of what children have learned in school. The age five tests are much more generic tests of cognitive capability. So it is informative that whereas middle class children who scored well on the age five copying test tended to score well in later tests, working class children did so to much less of an extent. It may be, as some appear to assume, that working class children who did well on the copying test just got lucky. It seems more likely to me that they just didn't achieve their potential in the later tests. The data do not distinguish between these interpretations.

True ability

There are both statistical and political debates being had and much as statisticians might like the rules of political debate to be reduced to the conventions of statistical debate, this is unlikely to happen. The graph has caused confusion in some quarters because of the difficulty of translating accurate and reasonable interpretations for policy audiences. This has also been difficult for Jerrim and Vignoles whose critique of the false interpretation of the chart has been taken by some as proof that social mobility is inevitable (Saunders, 2011 and Guardian 14 April, 2011 "Government social mobility expert under attack."). In other work Jerrim, Vignoles, Lingam and Friend (2013) show the huge gap between the evidence from structural genetics regarding the heritability of intelligence and that from any biological analysis of actual genetic data in explanation of the social class attainment gap. As discussed further below, it is important in the political debate that the Jerrim and Vignoles model is not taken as proof of its own assumptions, that low SES children are innately less cognitively capable, based on confusion about the meaning of "true ability" in their model. The notion of true ability they use is a statistical convenience, not the suggestion that science or social science has shown in any way that the latent ability gap at each age is in any way innate.

Jerrim and Vignoles base their model on the idea that at every age and moment of development each child has a true level of ability by which they can be ranked on a uni-dimensional scale of cognitive capability, as implied by the first principal component measure used in Feinstein (2003). They apply a standard statistical model in which a true, latent construct is hypothesised to be measured with error, which in this case they define as "true ability" – the specific level of ability of the child with some unspecified degree of stability at the time the measurement was taken. There is a particular definition of true ability at the core of their model, but it does not concord with a more usual, popular understanding of the notion of true ability, it is a statistical definition.

A second key assumption of their model is that at all ages and moments of development the true component of their variable is socially stratified, that is to say reflective of the degree of wider structural inequality such that the 22 month differences in rank contain and reflect SES differences. They assume that it is a feature of true ability, as well as of test scores. This follows from their implicit definition of true ability as the latent construct at the time of the test, not from presuming that it is a fixed entity, as the Jerrim Vignoles model allows true ability to vary over time. Furthermore, Jerrim and Vignoles do not assume, as does Saunders (2010; 2011) that a social class gap in true ability is a necessary feature of society, occurring necessarily in all social aggregates in all times and places. However, their use of the phrase "true ability" in their statistical modelling does appear to have been taken by some to imply that their model showed that there are stable, biological foundations to the social class attainment gap.

Crucially, Jerrim and Vignoles (2011), add to this the hypothesis that the degree of misclassification bias will vary by SES. Because low SES children are drawn from a group with a lower average score, children drawn from the low SES group who score well early on are more likely to have had, on their terms, over-estimated ability than the similarly scoring high SES children. They go on to say "Low SES children who get defined as high ability have probably had a particularly large random positive error (i.e. a lot of luck) during the initial test."

This is intended to be a statistical observation but we all need to be cautious in how we phrase attempts to explain statistical assumptions by making statements about people.

People, averages and qualitative change

Those I spoke to about the chart understood that it pertains to average rather than individual phenomena and so is an indicator of society and development in general not individual children. That said, I do particularly regret not being much clearer in public use of the study findings that the data in the two charts are averages. They may not describe the trajectories of any individual children. They are representative of a feature of development in general at the social level not of specific individuals.

Jerrim and Vignoles (2013) have shown the error of interpreting figure 2 as showing that at age six bright children from working class families will be overtaken by dim children from upper middle class households in school achievement. Another misuse in the public debate was the elision from average to individual. The Every Child Matters White Paper (HM Government, 2003, p19) stated that "children from a poor background with a high developmental score at 22 months have fallen behind by the age of 10, compared to children from higher socio-economic groups but with a low developmental score at 22 months". This drafting also conflates the average pattern with a universal phenomenon.

The distinction between averages and people is perhaps obvious to readers of this journal but is very easily blurred when statistics are used in wider public discourse. This causes problems for a public debate in which, at the level of society, it is important to know that family assets and capabilities and contexts may tend to impact on school test scores, but in which it would be false to assert that this makes SES the determining factor in the destiny of any specific child. When politicians today claim that figure 1 shows that "the race is over by age five," this is similarly a confusion of individual and average phenomena as it makes a universal of the average, as well as making an exaggerated claim about the average importance of the early years. The data in the graphs above and in all similar analysis tell us something about the average trends, indicating what tends to happen, the tendency in the time and place of the UK 1970 Cohort, not the true history of any individual case.

As Bronfenbrener (1979) and others have shown any framework for intergenerational change involves actors and action at multiple levels of which biological, individual and social levels are particularly important as distinct domains of change. The graph does not begin to address the breadth and complexity of these issues but it does need to be understood in this context if there is to be any discussion of implications for policy. The multi-level approach to understanding longitudinal data set out in Peck, Feinstein and Eccles (2008) emphasises in particular the importance of recognising qualitative change in modelling life course data. The corrections suggested by Jerrim and Vignoles treat cognitive development in the early years as a time-series of a common constructs rather than the emergence of a complex capability that changes qualitatively through the periods modelled. This is implied by the use of the lines alongside the data points of figure 2, which infuriated many, but it is important not to take this too literally as anything other than the changes in the average scores, that bear a distant relationship to the individual scores and are even more distant from the multi- dimensional and complex development of the individual children. In recognising this, other models might treat these measures in very different ways.

The implication of a literal reading of figure 2 or of the Jerrim and Vignoles correction is that at each age it is unproblematic to compare children in terms of their true ability and stack them up in unique ranks of relative achieved uni-dimensional intelligence. Even at 22 months, their model assumes that the only barrier to achieving this is the technical difficulty of measuring these true ranks. Error results not just from poor measurement but also from the deviation of the "true" distribution of the underlying latent variable from the linearity assumption in the index. So it remains important not to overstate the resulting precision. In their corrections for regression to the mean Crawford, Macmillan and Vignoles (2014) are very careful to label this "high early performance," to distinguish it from anything that might be thought innate, whatever the researchers' intentions. These data in corrected form show a general tendency at a time and a place and between the ages assessed using the specific metrics available, not a fundamental and fixed truth about human beings.

There are a number of different explanations of the facts about cognitive development and social class in the UK. It is conceptually possible that the pattern between 42 months and age ten in figure 2 indicates how capability and context interact to influence outcomes for the children in the 1970 study and hence in general in England, Wales and Scotland in the 1970s. It is also conceptually possible that the pattern is entirely the result of regression to the mean in a very strong sense; that the high scoring working class children were just a group with low true ability with continued luck who eventually got found out as test scores got more accurate. It is true that the data do not discriminate easily between these interpretations. We are left with theory and the wider science to attempt to distinguish them.

Conclusion

The graph shows that children from working class backgrounds in the 1970 cohort with good very early signs of cognitive development were less likely to translate these early signals into good later scores than children from middle class backgrounds. From this graph and many other sources was drawn the line in the strategy: *"Bright children from poorer families tend to fall back relative to more advantaged peers who have not performed as well."*

I wouldn't myself have used the phrase "bright children" but nothing in the Jerrim and Vignoles (2011, 2013) or Read (2003) critiques disprove the statement, as they themselves pointed out (The Guardian 28 April, 2011).

David Willetts MP, at the time Minister for Higher Education in the Department of Business Innovation and Science said subsequently:

> Sometimes over-reliance on one specific piece of evidence can leave you vulnerable. I remember being influenced by Leon Feinstein's very interesting paper for Economica in 2003 called Inequality in the Early Cognitive Development of British Children. He showed that bright poor kids fell behind rich dim kids by the age of 7. I served on Nick Clegg's social mobility group and recommended this powerful evidence to him and he too was impressed and cited it. But Leon's work was challenged by other academics because it was affected by reversion to the mean. The result was that the Guardian ran a piece that the Coalition's social mobility strategy was undermined because the research on which it rested had been disproved. That is not, of course, a reason for giving up on evidence-based policy: but it is a reminder of how careful we have to be in using it.

The question of the age at which supposedly "bright" working class children are overtaken in school performance by supposedly "dim" middle class children is not one that was ever tested or referenced by me. It is regrettable if there was confusion about there being a fixed age of six at which all dim middle class children overtake all bright working class ones. To be clear the crossover in this form is an artefact of the transparent way figure 2 was constructed and a corollary of figure 1. The point that was important for policy and was referenced in the 2010 Social Mobility Strategy was that throughout childhood in the UK children from low SES homes tend on average to fall back in school achievement relative to children from higher SES backgrounds.

The observed pattern between 22 and 42 months has always been understood by me and those with whom I have discussed the graph as mainly a statistical artefact resulting from measurement error. It has also been, in my experience, well understood that you cannot accurately or meaningfully fix children at 22 months on a scale of absolute and fixed ranks of ability. It would be wrong to define children as "bright" or "dim" on the basis of a set of early tests of cognitive development. Indeed, part of the early interest in the paper was because of the instability it showed in early signals of ability.

In an attempt at explaining the data (Feinstein 2003b, p30) I wrote, "so early scores do matter but so does social class after early childhood. The lesson for policy makers is clear. There is mobility (as one would expect) after 22 or 42 months, but upward mobility is mainly for high or medium SES children. Low SES children do not, on average, overcome the hurdle of lower initial attainment, combined with continued low input. Furthermore, social inequalities appear to dominate the apparent early positive signs of academic ability for most of those low SES children who do well early on."

Some would like to argue this is just an inevitable fact of heredity (Lynn, 2011; Saunders 2011). Some have wanted to claim that these patterns of inequality in development demonstrate underlying genetic continuities such that inequality is inevitable, others that the data show the impact of environment. As I stated in the 2003 paper, the graph cannot answer these questions.

However, there is general agreement that intelligence and school achievement have sufficient fluidity and malleability that only in rare cases is school achievement so fixed that there is no role for education and policy. Heckman (2007) puts it very clearly, based on his model of the production of capability:

> The nature versus nurture distinction, although traditional, is obsolete. Abilities are produced and gene expression is governed by environmental conditions. Behaviours and abilities have both a genetic and an acquired character. Measured abilities are the outcome of environmental influences, including in utero experiences, and also have genetic components.

I think this means it is wrong to interpret this type of longitudinal interaction between early scores and late scores (even if corrected for early reversion to the mean) as the later outcomes of dim or bright children, as though these characteristics were easily discernible in early childhood and fixed.

It is helpful that people are reminded that the graph is not simple and should be considered carefully, bearing particularly in mind the strong classification error between 22 and 42 months. We should remember it was a sample of children from the 1970s.

How children perform in tests matters for many reasons, not least as a signal to themselves and others. How this information is interpreted has a very substantial impact on child achievement and life outcomes (e.g. Dweck, 1986) so in the public debate it is always important to make a clear distinction between the meaning of aggregate statistical data and individual lives.

Subject to issues of modelling and measurement, the pattern of emergence of inequality in development tells us about the nature of inequality at the time and place at which the data are gathered. It is my hope that this debate will lead to further comparative work using diverse methods across diverse datasets to establish what differences are due to measurement, what to modelling and what to time and place.

Acknowledgments

I am grateful to those who have offered comment at various workshops, in particular at seminars at the Centre for the Analysis of Social Exclusion (LSE), at the Institute of Social and Economic Research (University of Essex), at the Genomics Forum (University of Edinburgh) and in a debate with John Jerrim at Portcullis House. Others have provided helpful comment on earlier drafts including Ruth Lupton, Kate O'Neill, Bilal Nasim, Kitty Stewart and Judith Dimant. Remaining errors remain mine alone.

References

Björklund, A., Lindahl, M., & Plug, E., (2006). The Origins of Intergenerational Associations: Lessons from Swedish Adoption Data. *The Quarterly Journal of Economics, 121*(3), 999-1028. http://dx.doi.org/10.1162/qjec.121.3.999

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press

Chamberlain, R. & Davey, A. (1976). Cross-sectional study of Developmental Test Items in Children aged 94-97 Weeks: Report of the British Births Child Study. *Developmental Medicine and Child Neurology*, *18,* 54-70. http://dx.doi.org/10.1111/j.1469-8749.1976.tb03605.x

Crawford, C., Macmillan, L., & Vignoles, A.. (2014). *Progress made by high- attaining children from disadvantaged backgrounds*. Social Mobility and Child Poverty Commission. Department for Education, London

Dweck, C.S. (1986). Motivational processes affecting learning. *American Psychologist, 41*(10), 1040-1048. http://dx.doi.org/10.1037/0003-066X.41.10.1040

Erikson, R. (1984) Social Class of Men, Women and Families, *Sociology, 18*, 500- 514. http://dx.doi.org/10.1177/0038038584018004003

Feinstein, L., (2003) Inequality in the Early Cognitive Development of British Children in the 1970 Cohort. *Economica, 70*, 277, 73-98. http://dx.doi.org/10.1111/1468-0335.t01-1-00272

Feinstein, L., (2003b). How early can we predict future educational achievement? *CentrePiece 8* (2). Centre for Economic Performance, London School of Economics.

Galton, M. (1886). Anthropological Miscellanea: Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute, 15*, 246–263

Gillborn, D. & D. Youdell (2000). Intelligence, 'ability' and the rationing of education. In Demaine, J. (Ed) *Sociology of Education Today*, London: Palgrave.

Goodman, A. & Greg, P. (Eds) (2010) *Poorer children's educational attainment: how important are attitudes and behaviour?*. York : Joseph Rowntree Foundation.

Heckman, J., (2007). The economics, technology, and neuroscience of human capability formation. *Proceedings of the National Academy of Sciences, 104*(33), 13250–13255. http://dx.doi.org/10.1073/pnas.0701362104

HM Government (2011). *Opening Doors, Breaking Barriers: A Strategy for Social Mobility*.

HM Government (2003). *Every child matters*. Green Paper, Cm 5860

Jerrim, J., & Vignoles, A. (2011). The use (and misuse) of statistics in understanding social mobility: regression to the mean and the cognitive development of high ability children from disadvantaged homes. *DoQSS Working Paper 11-01*. Institute of Education

Jerrim, J., & Vignoles, A. (2013). Social mobility, regression to the mean and the cognitive development of high ability children from disadvantaged homes. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176* (4), 887–906 http://dx.doi.org/10.1111/j.1467-985X.2012.01072.x

Jerrim, J., Vignoles, A., Lingam, R., & Friend, A. (2013). The socio-economic gradient in children's reading skills and the role of genetics. *DoQSS Working Paper 13-10*

Kingdon, G. & Cassen, R. (2007). *Understanding low achievement in English schools.* CASEpapers, CASE/118.

Centre for Analysis of Social Exclusion, London School of Economics and Political Science.

Little, R.J.A. & Rubin, D.B. (2002). *Statistical Analysis with Missing Data*, 2nd edition, New York: John Wiley. http://dx.doi.org/10.1002/9781119013563

Lynn, R. (2011). *Dysgenics: Genetic Deterioration in Modern Populations*. Praeger Publishers.

Magnuson, K., Waldfogel, J. & Washbrook, E., (2012). SES Gradients in Skills during the School Years. in: Ermisch, J., Jäntti, M. & Smeeding, T. (Eds) *From Parents to Children: The Intergenerational Transmission of Advantage*, 235-261. New York: Russell Sage Foundation.

Peck, S., Feinstein, L. & Eccles, J., (2008). Pathways through education: Why are some kids not succeeding in school and what helps others beat the odds? *Special issue: Journal of Social Issues, 64,*(1), 1–233. http://dx.doi.org/10.1111/j.1540-4560.2008.00545.x

Piaget, J., (1952). *The origins of intelligence in children.* International Universities Press. http://dx.doi.org/10.1037/11494-000

Read, D. (2003). *Researcher warns that Government Strategy for Social Mobility misled by a statistical trap*. Press release, Warwick University. Retrieved from http://www2.warwick.ac.uk/newsandevents/pressreleases/extreme_statistics/

Saunders, P., (2010). *Social mobility myths*. Civitas, London.

Saunders, P., (2011). *Social mobility delusions.* Civitas, London.

Spearman, C. (1904). General Intelligence, Objectively Determined and Measured. *American Journal of Psychology, 15*, 201-293 http://dx.doi.org/10.2307/1412107

Tu, Y.K., & Law, J. (2010). Re-examining the associations between family backgrounds and children's cognitive developments in early ages. *Early Child Development and Care, 180* (10),1243-12 http://dx.doi.org/10.1080/03004430902981363

Endnotes

[1] http://www2.warwick.ac.uk/newsandevents/pressreleases/extreme_statistics/

[2] https://inequalitiesblog.wordpress.com/2011/06/16/the-rise-and-fall-of-a-killer-chart/

[3] It's a shame journalists so routinely state that the author of a recent study is necessarily the author of all the data. So much work goes into the data that never gets recognised by this approach.

Commentary by **John Jerrim** Institute of Education, University College London, UK
j.jerrim@ioe.ac.uk
Anna Vignoles University of Cambridge, UK

Socioeconomic differences in children's test scores: what we do know, what we don't know and what we need to know.

Along with Blanden, Gregg and Machin (2005), Feinstein (2003) is one of the social science papers that has had the biggest impact upon British policymakers since the turn of the Millennium. Despite our subsequent criticism of how these findings have been interpreted (Jerrim & Vignoles, 2013), we encourage readers to remember that Leon Feinstein's 2003 paper offered many important, original and interesting insights. Although blunt in communication, it has always been our intention to engage productively with the important issues raised by Feinstein (2003). With that in mind, we are grateful to both John Bynner (the editor) and Leon Feinstein for their continuing engagement, and the productive platform they have developed for moving the debate forward. Indeed, we hope this Comment and Debate section encourages further work in this area. Our response will be structured as follows. To begin, we will clarify our position on some of the points raised by Feinstein (2015). We then summarise areas where there now seems to be reasonably strong empirical evidence on socioeconomic achievement trajectories and where broad consensus seems to have been reached. This is followed by a discussion of where knowledge is still lacking, or open issues which remain in debate. The conclusion then suggests directions for future research.

Feinstein's contribution to the debate makes the important point that the sample used for his 2003 analysis is a selected subset of the entire cohort, and indeed that there is significant attrition in the British Cohort Study. Undoubtedly this is so and it threatens the external validity of the findings (the extent to which they really are representative of the population as a whole). Further, as attrition is differential, with greater drop out amongst poorer children, the sample of poor students for whom we have complete data may not be representative of the population of poor children. This is an important issue to address when describing the achievement trajectories of poor children. In Jerrim and Vignoles (2013), the Millennium Cohort Study (MCS) data also suffer from sample selection and attrition, albeit at a lower level in the MCS. More work is needed on how sample selection and attrition might impact upon the generalisability of the results. However, it is also important to point out that attrition concerns in and of themselves do not affect the methodological issues of regression to the mean, since in our paper we do a complete case analysis, whereby we examine the trajectories of children in the sample over the entire period. Hence differences between the data sets used by Feinstein (2003) measures and those used by Jerrim and Vignoles (2013) do not explain differences in findings that we attribute to regression to the mean.

Feinstein (2015) also makes the point that the different cognitive achievement tests used across the cohort studies have different scales and meanings, and that the measures used in Feinstein (2003) are ordinal. This is of course correct, and is an issue we consider extensively in Jerrim and Vignoles (2013). Indeed, regression to the mean (RTM) is potentially problematic whether one uses an ordinal scale or not. Nevertheless, it is important that we undertake more work on how best to measure and interpret tests of children's cognitive development at a very early age.

Having clarified these points, we now move on to issues where we believe broad consensus has been reached. First, there are large socioeconomic gaps in children's cognitive skills, and these can be observed from a very early age (indeed, from the very first point measurement of such skills is possible). This is consistent across both Feinstein (2003) and Jerrim and Vignoles (2013), along with a host of other research (e.g. Blanden & Machin, 2007; Crawford, Macmillian & Vignoles, 2014; Cunha, Heckman & Lochner, 2006; Goodman, Sibieta & Washbrook; Jerrim, Vignole, Lingam & Friend, 2014; Jerrim & Choi, 2014; Schoon, 2006). We believe that this represents the main message that policymakers *should* have taken from Feinstein 2003.

Second, the *absolute* difference in average test scores between socioeconomic groups certainly does not seem to decline as children enter school, and in all likelihood continues to grow. (By the 'absolute' skill gap, we are referring to the actual competencies that high and low SES children display at any given age. It is based upon test scores measured in its original scale – one which has a substantive meaning - and has not been standardised or converted into rank position.) This is due, at least in part, to the increase in the variance of the skills children display as they age. For instance, the difference in what three year olds can do in, say, mathematics is a lot smaller than the variability in mathematics skills displayed by 15 year olds. See Magnuson, Waldfogel and Washbrook (2012) for evidence on this issue from the United States.

Third, although measurement issues cannot be completely ruled out, we believe there is now sufficient empirical evidence to assert that socioeconomic differences in *relative* skills do not appreciably narrow during the school years. (By relative skills, we are essentially referring to the rank order of young people in the achievement distribution). See, for instance, evidence from Feinstein (2003), Magnuson, Waldfogel and Washbrook (2012), Jerrim and Choi (2014), Choi and Jerrim (2015) and Schoon (2006). However, evidence on whether and when the relative skill gap grows ('fans out') or remains stable is (in our opinion) still relatively weak, and susceptible to important (yet little discussed) measurement issues. (For instance, if there is random error/noise in children's test scores, but this decreases as children age, this may also produce the 'fanning out' pattern that is so often cited in this literature). This therefore remains an area where further UK evidence, tackling the important issue of measurement of skill, is needed.

Fourth, with regard to the skill trajectories of initially high (low) achieving children from low (high) SES backgrounds, the striking decline between 22 and 42 months reported in Feinstein (2003: Figure 2) and between 36 and 60 months in Jerrim and Vignoles (2013: Figure 5a[1]) is due, at least in part, to a statistical artefact known as '*regression to the mean (RTM)*'. This should therefore *not* be used by academics or policymakers to stress the importance of the early years, that we are failing 'bright' young people from disadvantaged backgrounds, or to highlight the lack of social mobility in the UK. Rather, the fact that early socioeconomic gaps in achievement are so large is by itself highly suggestive of the importance of the earliest years in more general terms.

Finally, there remains no robust and consistent evidence that initially high achieving young people from poor backgrounds are overtaken by low achieving children from affluent backgrounds in terms of their cognitive skills. Crawford et al. (2014) have attempted to take account of the problem of regression to the mean when measuring the trajectories of initially high achieving students in secondary school and found that high achieving low SES students do decline relative to high SES students between the ages of 11 and 16. However, for this older age group they did not find support for the "crossover" pattern observed by Feinstein (2003). What is also important to remember is that the evidence base does not suggest that the prospects of high attaining (however defined) young people from poor homes are entirely determined by age ten.

If these now represent what we believe to be consensus views, what are the areas of continuing disagreement, and thus where further research is needed? First, although we know socioeconomic differences in cognitive skills emerge early, it is not clear the extent to which this is due to genetics and 'hereditary' factors, and the extent to which this is environmental (or indeed the interaction between the two). Recently, Krapohl and Plomin (2015:3) have argued that 'half of the phenotypic correlation between children's family SES and their educational achievement is mediated genetically' based upon a genome-wide complex trait analysis of 3,000 unrelated children. This is in contrast to some previous research (e.g. Goldberger, 1979, Gould, 2011, Manski, 2011) which either argue against the strength or relevance of such findings. Nevertheless, recent genome-wide association studies (rather than inferred genetic effects based on comparisons across twins) have also indicated that there may be a high degree of heritability in IQ (Davies et al. 2011). Controversial though this is, it might imply that based upon the empirical evidence alone, it is not currently possible to rule out 'hereditary' (a popular explanation by some – e.g. Saunders, 2012) as an explanation for a significant proportion of the socioeconomic gap in educational test scores. However, it is important to note that the evidence on the importance of hereditary factors is mixed at best.

The picture is undoubtedly complex. Epigenetic studies have suggested that since a child's environment may influence their gene expression, it is by no means straightforward to separate out the effects of hereditary factors and environmental influences, and that the latter influences children's outcomes even in utero (Carey, 2012; Hobcraft, 2012). Further some studies of gene-environment interactions have indicated that whilst differences between socioeconomically advantaged children may be attributable to their genes, environmental factors are more important in socioeconomically deprived environments (Tucker-Drob, Rhemtulla, Paige Harden, Turkheimer & Fask, 2012). Yet the evidence is mixed and partial, with much more research needed. Indeed, on a related note, we must also develop a better understanding of the environmental and genetic mechanisms (and their potential interaction) influencing cognitive development and the growth in absolute socioeconomic skill gaps as children age. Although social scientists typically focus upon the environmental explanations, there is now a growing body of research which suggests that changes in educational attainment over time could be partly due to genetic factors (Haworth, Asbury, Dale & Plomin, 2011).). Rather than shy away from this issue, social scientists should engage more with geneticists and their data – developing a better understanding of how genes may influence cognitive skill growth (including via potential interactions with the environment).

We also still know very little about the educational progress made by initially high-achieving children from disadvantaged backgrounds. Both Feinstein (2003) and Jerrim and Vignoles (2013) have methodological limitations, with the trajectories in both papers subject to a high degree of uncertainty. (As just one example of uncertainty, neither paper presents confidence intervals. But sampling variation is likely to be large, given the small sample sizes of the high/low achieving groups in the data those studies use). Although Crawford, Macmillian and Vignoles (2014) have recently added to the evidence base, further work, using better data and more sophisticated methodology to overcome the RTM problem, is clearly still required.

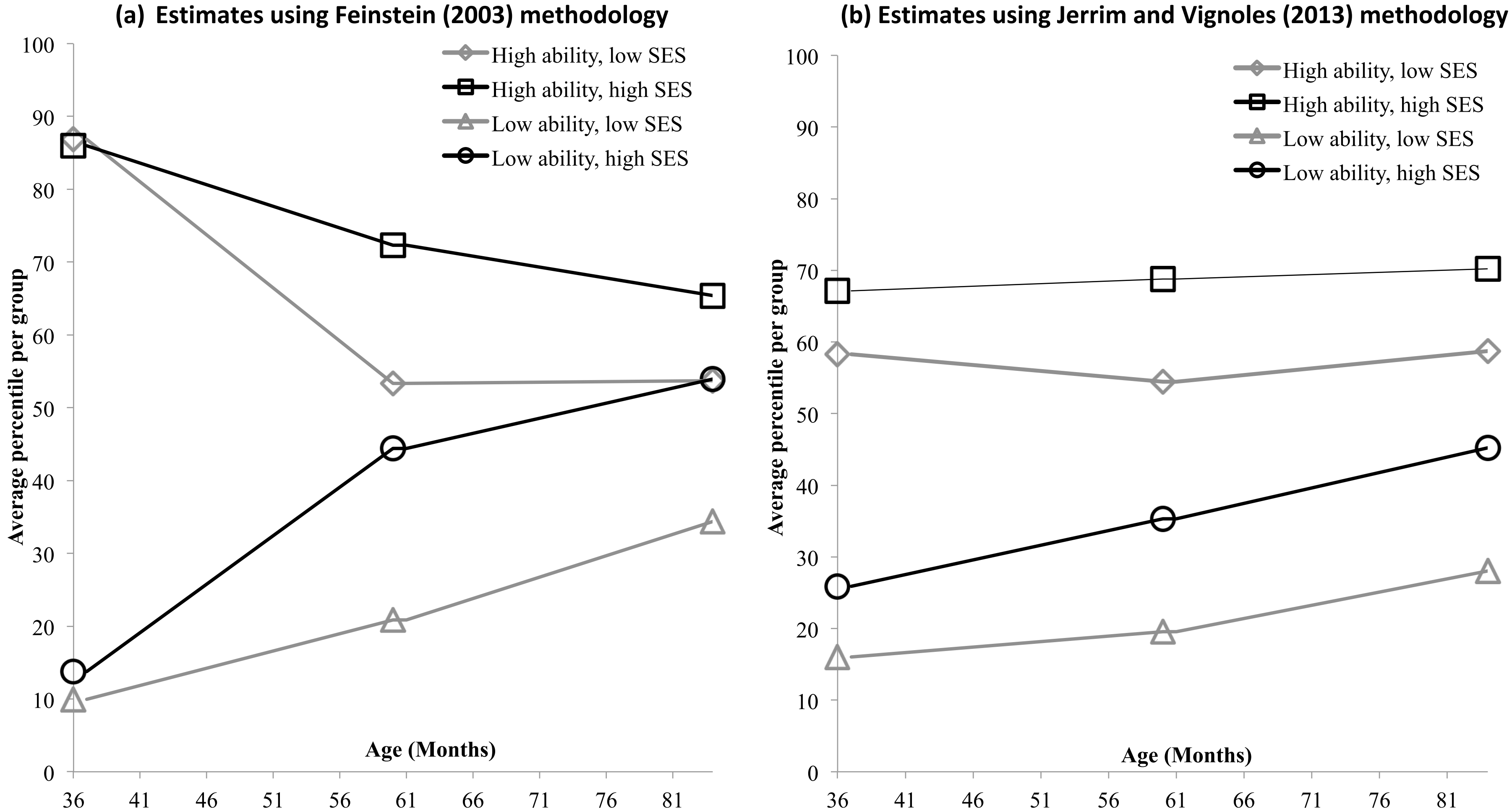
Finally, regarding mean differences in test scores by SES group (not stratified by initial achievement), further detail is needed on the descriptive patterns observed. For instance, if the socioeconomic gap in children's test scores really does increase as children age, is this being driven by the poorest children in society falling behind the rest of the population? Or is it because the most affluent 20% are pulling away from everyone else? These two scenarios would likely warrant quite different policy responses. Our reading of the literature suggests that it is less likely due to the former and more likely to be attributable to the latter (see Goodman & Gregg, 2010; Jerrim & Vignoles, 2015), though again further work is needed.

What then do we suggest are the most pressing issues for those looking to move forward the debate? To begin, more and better data is needed, with greater consideration given to the measurement properties of children's test scores. Despite their many advantages, the UK's 1958, 1970 and 2000 cohort data have some limitations in this respect. Comparable tests have not been conducted at more than two time points for example, with the survey documentation lacking sufficient discussion on possible measurement error and test reliability. The new forthcoming cohort (Life Study) offers potential opportunities to improve the evidence base. It is essential that test scale measurement and psychometric properties are key in any future cohort collecting data on early-life cognition. With regard to existing data sources, the survey organisers and funders should attempt to retrospectively consider such issues. In general, the properties of the achievement data contained within the cohorts need to be more thoroughly investigated and better understood.

Next, more sophisticated methods need to be developed and applied in this area. These should ideally be able to account for the possibility that there is 'systematic' error in the test-score data, including potential measurement bias. In other words, future work needs to move beyond the simplistic assumption made in Jerrim and Vignoles (2013) that any error in test scores is simply random noise. Allowing for systematic error is likely to enable us to get closer to the 'truth'. Although we note the potential for latent variable methods such as growth curve modelling to do this (as per Von

Stumm & Plomin, 2015), this should be accompanied by a clear explanation as to how it overcomes the RTM problem and a simulation study demonstrating the conditions under which it works (and the assumptions being made). After all, the beauty of Feinstein (2003: Figure 2) was its simplicity. (This was one of the key reasons why we proposed the simple adjustment in Jerrim and Vignoles 2013 – reproduced in figure 1 - to maintain this simplicity). Future work should continue with such clarity of presentation to keep policymakers engaged in this matter – while, of course, ensuring the most robust and convincing methods are applied.

Figure 1. Estimated cognitive gradients in MCS when using different methodologies



Note: Reproduced from Jerrim and Vignoles (2013:Figure 5). Estimated cognitive trajectories based upon the MCS. The left hand panel refers to estimates using methodology of Feinstein (2003). The right hand panel is the equivalent figures when applying the methodology proposed by Jerrim and Vignoles (2013

Third, better use needs to be made of existing resources to tackle the issues we have raised. The Twins Early Development Study (TEDS) is a prime example, which contains detailed information on both children's genetic and parental investments for a large sample of UK twins, who have been tested at multiple points throughout childhood (from age two through to age 18). Such data has the potential to provide new descriptive information on SES trajectories. Once the cognitive trajectories are firmly established, this will be the next vital step in this line of research. TEDS is an underutilised dataset by social scientists, and one we believe can potentially address many of the issues described in this paper. We therefore strongly encourage any social scientist looking to conduct further work in this area to consider seeking to use this MRC funded dataset.

Finally, we end on a note of caution. Although there is a desire amongst policymakers to know how policy should respond to counter the deleterious effects of SES on achievement, it is important that we walk before we run. We first need to be certain of the descriptive trajectories regarding how cognitive skills develop differentially across socioeconomic groups. Here, high quality data and robust methodologies are key. It is only once this first stage is complete, and to a satisfactory standard, that we should then attempt to disentangle cause from effect and develop the appropriate policy response.

References

Blanden, J., Gregg, P. & Machin, S. 2005. *Intergenerational mobility in Europe and North America*. The Sutton Trust report. Retreived from http://www.intouniversity.org/sites/all/files/userfiles/files/Sutton%20Trust%20Social%20Mobiliity.pdf

Blanden, J. & Machin, S. (2007). *Recent changes in intergenerational mobility in Britain*. Sutton Trust, London. Retreived from http://www.suttontrust.com/public/documents/summaryintergenerationalmobility.pdf .

Carey, N. (2012). *The Epigenetics Revolution: How Modern Biology is Rewriting Our Understanding of Genetics, Disease, and Inheritance.* New York: Columbia University Press.

Choi, A. & Jerrim, J. (2015). The use (and misuse) of PISA in guiding policy reform: the case of Spain. *IEB working paper series*. Retreived from http://www.ieb.ub.edu/phocadownload/documentostrabajo/doc2015-6.pdf http://dx.doi.org/10.2139/ssrn.2580141

Crawford, C., Macmillan, L., & Vignoles, A. (2014). *Progress made by high attaining children from disadvantaged Backgrounds*. Research report.

Social Mobility and Child Poverty Commission. Department for Education, London

Cunha, F. Heckman, J. & Lochner, L. (2006). Interpreting the Evidence on Life Cycle Skill Formation. In E. Hanushek & F. Welch *(Eds)Handbook of the Economics of Education*. Pp. 697-812. Amsterdam: Holland North.

Davies, G. Tenesa, A., Payton, A., Yang, J., Harris, S. E., Liewald, D. & Deary, I.J. (2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Mol Psychiatry 16*(10), 996-1005. http://dx.doi.org/10.1038/mp.2011.85

Feinstein, L. (2003). Inequality in the early cognitive development of British Children in the 1970 cohort. *Economica 70*, 73-97. http://dx.doi.org/10.1111/1468-0335.t01-1-00272

Feinstein, L. (2015). Social class differences in early cognitive development and regression to the mean. *Longitudinal and Life Course Studies 6*(3)

Goldberger, A. (1979). Heritability. *Economica 46*(184), 327-347. http://dx.doi.org/10.2307/2553675

Goodman A., Sibieta L., & Washbook E. (2009). Inequalities in educational outcomes among children aged 3 to 16. Final report for the National Equality Panel, Institute for Fiscal Studies, London. Retrieved from http://sta.geo.useconnect.co.uk/pdf/Inequalities%20in%20education%20outcomes%20among%20children.pdf.)

Goodman, A. & Gregg, P. (2010). Poorer children's educational attainment: how important are attitudes and behaviour? Report for the Joseph Rowntree Foundation. Retrieved from http://www.jrf.org.uk/system/files/poorer-children-education-full.pdf

Gould, S. *The Measurement of Man* (second edition), W. W. Norton & Company

Haworth, C., Asbury, K., Dale, K. & Plomin, R. (2011). Added value measures in education show genetic as well as environmental influence. *PLoS One 6*(2), e16006. http://dx.doi.org/10.1371/journal.pone.0016006

Hobcraft, J. (2012). Epigenetics and the social sciences: progress, prospects and pitfalls, paper presented at the ESRC International Grand Challenge Symposium, 25-26 June 2012. Retrieved from http://www.esrc.ac.uk/_images/Epigenetics-symposium-background-paper_tcm8-21507.pdf.

Jerrim, J. & Vignoles, A. (2013). Social mobility, regression to the mean and the cognitive development of high ability children from disadvantaged homes. *Journal of the Royal Statistical Society Series A 176*(4), 887 – 906. http://dx.doi.org/10.1111/j.1467-985X.2012.01072.x

Jerrim, J., Vignoles, A., Lingam, R. & Friend, A. (2014). The socio-economic gradient in children's reading skills and the role of genetics. *British Education Research Journal* DOI: 10.1002/berj.3143 http://dx.doi.org/10.1002/berj.3143

Jerrim, J. & Choi, A. (2014). The mathematics skills of school children: how does the UK compare to the high performing East Asian nations? *Journal of Education Policy 29*(3), 349-76. http://dx.doi.org/10.1080/02680939.2013.831950

Jerrim, J. & Vignoles, A. (2015). University access for disadvantaged children: A comparison across English speaking countries. *Higher Education 1007*/s10734-015-9878-6

Krapohl, E. & Plomin, R. (2015). Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Molecular psychiatry*. http://dx.doi.org/10.1038/mp.2015.2

Magnuson, K., Waldfogel, J. & Washbrook, E. (2012).The development of SES gradients in skills during the school years: evidence from the U.S. and U.K. In Ermisch, J., Jantti, M., & Smeeding, T. (Eds) *Inequality from Childhood to Adulthood: A Cross-National Perspective on the Transmission of Advantage*, New York: Russell Sage Foundation.

Manski, C. (2011). Genes, eyeglasses, and social policy. *Journal of Economic Perspectives 25*(4), 83-94. http://dx.doi.org/10.1257/jep.25.4.83

Saunders, P. (2012). *Social Mobility Delusions.* London: Civitas.

Schoon, I. (2006) *Risk and Resilience: Adaptations in Changing Times,* Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511490132

Tucker-Drob, E., Rhemtulla, M., Paige Harden, K., Turkheimer, E. & Fask, D. (2012) Emergence of a gene × socioeconomic status interaction on infant mental ability between 10 months and 2 years, *Psychological Science 83*(2), 743-757.

Von Stumm, S. & Plomin, R. (2015). Socioeconomic status and the growth of intelligence from infancy through adolescence. *Intelligence 48*(1), 30-36. http://dx.doi.org/10.1016/j.intell.2014.10.002

Endnotes

[1]This graph is reproduced in figure 1 below.

Commentary by **Harvey Goldstein** University of Bristol, UK
h.goldstein@bristol.ac.uk
Robert French University of Bristol, UK

Differential educational progress and measurement error

Abstract

The debate around Feinstein's original (2003) analysis is crucially dependent on technical issues associated with definitions of measurement error and how the presence of such error can be adjusted for. In this commentary we explore the ways in which measurement error can be incorporated within regression (and other) models to obtain valid inferences. We suggest that there are flaws in both Feinstein's original analysis and his current response to criticism, as well as problems with some of the existing technical critiques. We conclude that there is reasonable evidence for increasingly divergent educational achievement among social groups as children move through schooling.

Keywords: Social class differences, measurement error, regression to the mean, educational achievement.

Introduction

The publication of Feinstein's original analysis (2003) generated both substantive and methodological concerns with import for social policy. In the latest response to criticisms of his earlier work Feinstein, (2015) broadly defends his earlier findings while at the same time appearing to accept that there are legitimate concerns about the statistical model underpinning his analysis. The main purpose of our paper is to clarify the underlying statistical issues, and in the process of doing this we will comment on Feinstein's conclusions.

We reanalyse the dataset Jerrim and Vignoles (2015) used to critique Feinstein (2003), since this will allow us to illustrate the key technical issues. We then discuss the relevance of our model estimates to the questions originally raised by Feinstein. Before doing this, in the interest of clarity, we need to comment on some of the assertions made by Feinstein (2015).

He claims in section two that his analysis "was explicitly descriptive" as opposed to constituting a statistical model. He says "The aim was to present the actual data…, rather than to present corrected trajectories based upon modelling assumptions". We argue that presenting data, whether based upon a sophisticated statistical model or a simple statistical model such as that which lies behind figure 1 in Feinstein's (2003) paper , is intended to convey an inference about the underlying social process that is generating the data. Feinstein is not presenting 'actual' data, rather he is presenting a summary based upon a particular manipulation (i.e. modelling) of the data with the intention to convey something about how social class differences in achievement are changing over time.

The second issue is Feinstein's use of the terms 'regression to the mean' and 'true scores'. Regression to the mean, which is also used by Jerrim and Vignoles (2012), as introduced by Francis Galton simply occurs when the correlation between two measurements over time is less than one, as is the case with heights of fathers and sons. The notion of measurement error is entirely separate, although if one had perfectly correlated 'true' measures then the addition of random measurement error to these would lead to the same mathematical result. Using standardised measures, so that each measurement is on the same scale with a mean of zero and standard deviation of 1, the mean second occasion score for those with a given high score, *x,* on the first occasion will be smaller than *x*. This raises the issue of what is meant by 'true score'. Feinstein does have a useful discussion on this, but is not entirely clear about the 'statistical' notion of true score. This is a conceptual notion that proceeds from the common observation that the actual score that a child obtains on a test will depend on the actual items chosen for the test plus other factors that might be considered 'transient' such as time of day, test environment etc. Most test constructors (although sadly it appears not the providers of the tests in question) provide estimates of this 'unreliability' or measurement error so that it can be taken account of by data analysts.

The statistical model

We start by presenting the original plot from Jerrim and Vignoles (2012) in figure 1. It illustrates the approach adopted by them and Feinstein (2003). The data consist of four measurements on a sample of children from the 1970 British Births cohort study at the occasions of 22, 42, 60 and 120 months. In this note we will use only two occasions in order to illustrate our points.

Figure 1.

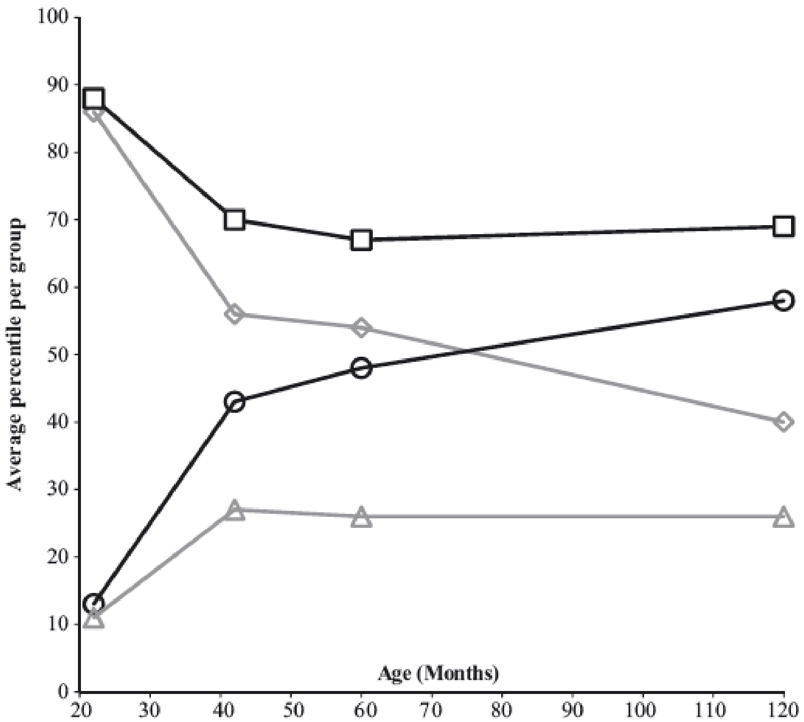


Fig. 1. Development of high and low ability children by socio-economic group—evidence from the existing literature (adapted from Feinstein (2003); based on 1970 British Cohort Study data): ◇, high ability–low SES; □, high ability–high SES; ○, low ability–high SES; △, low ability–low SES

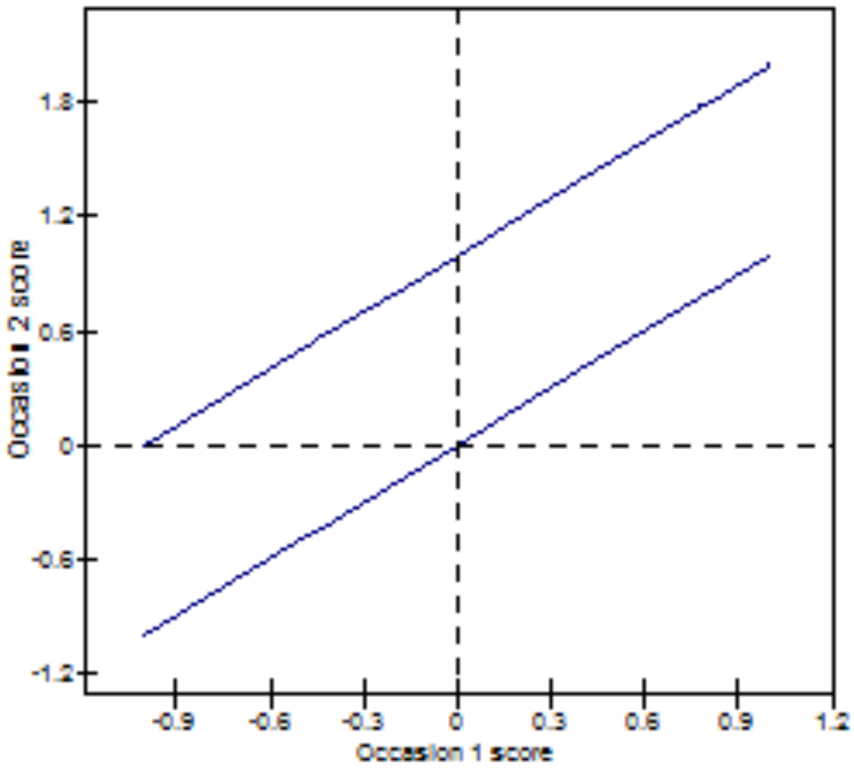
We also use the more extensive Millennium Cohort Study (MCS) dataset, as do Jerrim and Vignoles, in our analysis at ages of approximately three and five years. The number of children available for our analysis, after excluding the small number of cases with missing data (effectively missing at random) is 10,071.

As is clear from figure 1, the previous analyses use the first occasion ability measure by forming two ability groups, the bottom decile and the top decile. Feinstein uses the actual ability measure itself to form these groups whereas Jerrim and Vignoles use a highly correlated 'surrogate' measure. This is because their method requires measurement errors in the grouping variable to be uncorrelated with those in the ability measure itself and they are prepared to make the assumption that this is satisfied by this surrogate measure (see below). In our exposition this is unnecessary, as we shall show. Socioe-conomic status (SES) groups are defined by Jerrim and Vignoles using an income measure with lower and upper quartile thresholds so that about 25% are classified as high SES and 25% low SES. They obtain essentially the same results as those given in figure 1, where the inference is that for children of the same ability at the first occasion those in the high SES group show greater progress than those in the lower SES group, and indeed that the initial high ability low SES children start doing worse than the initial low ability but high SES children by just after the third occasion.

At this point it is worth remarking that some care has to be taken with this form of adjustment for initial achievement, and this is also pointed out by Jerrim and Vignoles. Since the low SES children at the first occasion on average have a markedly lower ability – about 0.5 standard deviations below the mean and high SES children have a mean ability about 0.3 standard deviations above the mean, the mean abilities for the SES groups may also be expected to differ when we study just the low ability or high ability children. In fact, for the MCS data the mean abilities are about the same for the high and low SES groups within the group of high ability children, but for the low ability children the low SES children have an average ability about 0.3 standard deviations below the average ability for the high SES children. This result, of course, is based upon the observed scores but in general, for unimodal distributions, we would expect a similar result for true scores, that is with no measurement errors. We discuss the definition of true score and measurement error in equations (2) & (3) below. In such a case, with known true scores and an overall average difference between upper and lower quartile SES groups of 0.8, the average difference between high and low SES groups is about 0.2 for low ability children and also 0.2 for high ability children. The more extreme the ability groupings that are used the less this difference will be. Thus in general, even if there were no measurement errors, and no changes in ability over time, we would expect the low SES low ability children on average to have lower average ability over time than the low ability high SES children, and the same principle would apply to the high ability children, simply because they initially have different average abilities because of the way they were selected. In other words, grouping on initial ability in this fashion, whether using a surrogate for true score, or just the observed score, is not generally an adequate adjustment for initial ability, even when we are dealing with true scores without measurement error. This suggests that this kind of approach to studying changing social gaps over time is inherently flawed and is not to be recommended.

To adjust properly for initial ability differences, in order to study progress over time, we will adopt a modification of what is a commonly used approach whereby we model the second occasion ability score as a suitable function of the first occasion score, together with SES and possible interactions. Figure 2 shows a hypothetical example of the results of such a model where for any given true occasion one score the high SES group has a higher than expected true score at occasion two.

Figure 2. Hypothetical relationship for two SES groups



The parallel lines in this graph imply that the mean group difference is the same for all individuals at each occasion one score and we note that this adjusts for the initial ability without any ability grouping involved. This approach is able to show the SES differences for the full dataset rather than just two extreme groups. Interest lies in whether in reality the slopes of the lines in fact differ or whether the relationship may be non-linear. In practice, of course, where we have 'observed' scores that include measurement error, rather than 'true' test scores we will need to adjust for this, which may well change the inferences we make. In the following analysis we shall look at the effects of SES on progress before and after adjusting for measurement error. We do not here present details of the measurement error adjustment, since these form the basis of another paper in preparation. A basic reference, however, is Richardson and Gilks (1993) who outline a Bayesian modelling approach of which ours is a further extension, most notably by allowing interaction terms that include measurement errors. The presence of interaction (and power) terms as in model (1) below is needed in our analysis. In fact, consistent procedures for standard regression models that allow adjustment for measurement error have been known about for over 40 years and Goldstein (1979) uses these in an analysis of 1958 cohort data to study precisely the question of differential progress for different social groups.

A model for measuring differential group progress

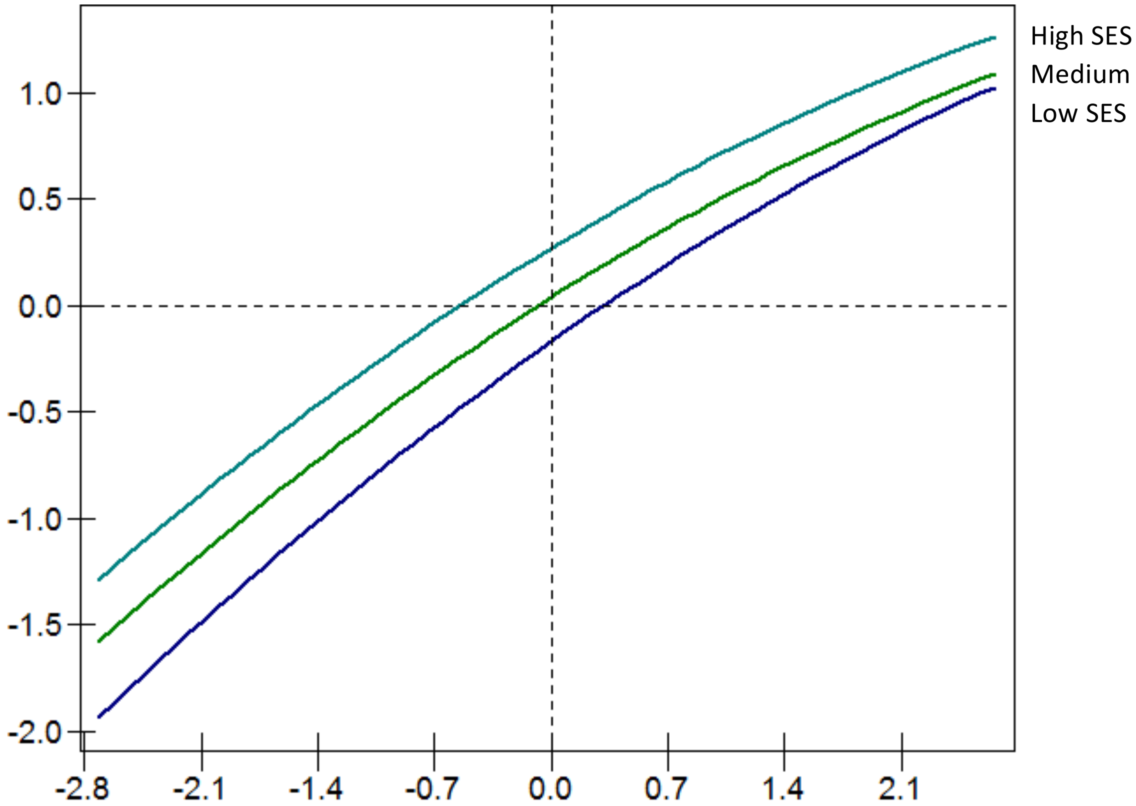
Our basic model can be written as

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_{2i} z_{1i} + \beta_{3i} z_{2i} + \beta_{4i} x_{1i}^2 + \beta_5 x_{1i} z_{1i} + \beta_{6i} x_{1i} z_{2i} + e_i \qquad (1)$$

Where x_{1i} is the occasion one ability measure, in this case a reading (vocabulary) score, and for which we have used a quadratic term also to reflect the nonlinear nature of the relationship as seen in figure 3. The terms z_{1i}, z_{2i} are two dummy variables denoting the lower and upper SES quartile groupings with the middle group as the base category. In this case SES groupings are based upon a measure of income, for further details of the variables used see Jerrim and Vignoles (2012). Note that we have also allowed for a possible interaction between initial ability and SES group. We have chosen the SES thresholds such that they contain 26%, 48% and 26% of the sample for the low, medium and high groups respectively.

Column 1 (**σ_m^2=0)** of table 1 and figure 3 show the results from fitting this model without any adjustment for measurement error, σ_m^2 being the measurement error variance as explained in (3) below.

Figure 3. Predicted age five test score by age three test score for three SES groups. Standardised scores.



We see that for every age three score those in the highest SES group are predicted to have the highest age five score, followed by the middle SES group followed by the lowest SES group. This is consistent with Feinstein's argument for an increasing gap emerging between SES groups. Note also the interaction indicating that the greatest gap between progress made by different groups is for the initial low achievers. In effect these results are similar to those of Feinstein. Thus a low achiever at approximately the lowest quartile position in the low SES group has an expected age five score about 0.7 below a high SES low achiever, whereas a low SES high achiever at approximately the highest quartile position has an expected score about 0.3 below the corresponding high SES child. A high SES child at about the lowest quartile position on age three ability has a predicted age five score of about -0.5 which is what would be predicted for a low SES child a little below the mean.

We now extend the approach used above to adjust for measurement error. The actual size of the measurement error, its variance, is unknown. Jerrim and Vignoles in their simulation use a value of 0.25, equivalent to a reliability of 0.75 for these data. The reliability is defined as the variance of the true scores divided by the variance of the observed scores, where the true score variance is simply the observed score variance minus the measurement error variance. For simplicity we assume that the measurement error variance in the predictor and the response, the reading score at five years, are the same. Now, the observed correlation between the three and five year scores is 0.55 so that this provides the lower limit for the reliability, *R,* since the adjusted true score correlation is estimated by $0.55/R \leq 1$. In fact the correlation varies from 0.56 in the low SES group to 0.44 in the high SES group, and the measurement error variances may well differ also, but we do not explore this further. For the reliability values 0.75 and 0.65 the estimated true score correlations are 0.73 and 0.85 and the latter would seem to be a reasonable upper estimate in practice. We will carry out analyses for these two values. We shall explore the effects of using these reliabilities to judge the sensitivity of the results to the value chosen.

Model (1) is thus augmented by adding the standard measurement error assumption

$$x_{1i} = X_{1i} + m_i, \quad \sigma_m^2 = 0.25 \tag{2}$$

and

$$y_i = Y_i + q_i, \quad \sigma_q^2 = 0.25 \tag{3}$$

Where in (2) X_{1i} is the 'true score' and m_i is the measurement error assumed to have a normal distribution with zero mean. Likewise for the response *Y* in (3). Additionally we assume that the measurement errors at occasion one and two are independent, since these are well separated in time. The test scores are all standardised to have zero means and standard deviations of 1.

In addition to the uncertainty surrounding the term 'true score' a further complication is that the size of any measurement error may vary according to individual characteristics. Thus, for example it may be higher for some SES groups than others. We shall not pursue such matters here, and assume that we have a common measurement error variance.

Table 1 shows the results from fitting models (1) + (2), using Markov Chain Monte Carlo estimation with default priors (see Richardson and Gilks, 1993, for further details).

Table 1. Age five reading score related to age three reading score and SES. Different amounts of measurement error variance, σ_m^2 and reliability

Parameter	σ_m^2=0, $R = 1.0$	$\sigma_m^2 = 0.25, R = 0.75$	σ_m^2=0.35, $R = 0.65$
β_0	0.038 (0.013)	0.052 (0.013)	0.042 (0.014)
β_1	0.494 (0.012)	0.712 (0.016)	0.874 (0.019)
β_2	-0.207 (0.021)	-0.092 (0.022)	-0.006 (0.025)
β_3	0.232 (0.021)	0.168 (0.022)	0.125 (0.024)
β_4	-0.037 (0.006)	-0.090 (0.011)	-0.112 (0.014)
β_5	0.054 (0.021)	0.030 (0.029)	0.024 (0.032)
β_6	-0.022 (0.022)	0.0016 (0.0300)	0.021 (0.029)
σ_e^2	0.668 (0.009)	0.515 (0.009)	0.431 (0.011)

Estimation by MCMC: burn in = 500, iterations = 1000.

We see that for the upper and lower SES groups, with a reliability of 0.75 the mean difference $(\beta_3 - \beta_2)$, that is for students with the mean occasion one ability of zero, is reduced from 0.44 to 0.26 SD units and this changes little across the age three reading score since the interaction terms are small, and in fact not significant at the 5% level. When the reliability drops to 0.65 the SES difference becomes 0.13.

Finally, we make a brief comment on the method used by Jerrim and Vignoles where they use an instrumental variable approach. The instrument they use is an 'auxiliary' or 'instrumental variable' taken at the same time as the test of interest at age three and they assume that this is uncorrelated with measurement error in the test of interest. This does seem to us a very strong assumption, especially since the tests were taken on the same day. Furthermore, for instrumental variable methods where it is likely that the instrument is uncorrelated with measurement errors in the test of interest, the instrument itself will tend not to be a very good predictor of the true score. This implies that it will often tend to lack statistical power. In addition there is a substantive problem in that using the instrument as a measure of 'ability' assumes that it is the same measure of 'ability' that is being measured by the test of interest, in other words it is what is known as a parallel test. This does, however, seem questionable, and Feinstein (2015) also picks up on this point. For these reasons we have adopted the above approach, but we do need to emphasise that a sensitivity analysis, using more than one estimate for the measurement error variance is important.

Concluding remarks

The debate about how to study differential progress of children from different socio-economic backgrounds is clearly important and has illustrated the crucial nature of the modelling assumptions that need to be made. Much depends upon knowledge of the quality of the measures used to define educational or other performance, and it is just such information that is typically absent. In particular the reliability of the tests used needs to be available, or at least a plausible range for such values. Such information should ideally be provided by the constructors and suppliers of the tests and users of the data need to be provided with such information, or at the very least made aware of the need for it. The contribution by Jerrim and Vignoles is therefore important in raising this issue, and in this short note we have suggested a comprehensive approach to studying the issue, using a method that allows quite general models, including multilevel ones, to be fitted. Even without such an extension, however, simple moment based estimators using observed variances and covariances of the observed variables corrected for reliability, are available that will generally provide insight into the extent to which inferences are changed when measurement errors are allowed for. Goldstein (1979) describes this approach in the analysis of data from the 1958 British birth cohort in a similar analysis of differential progress. He showed that reasonable amounts of measurement error, when adjusted for, reduced the size of the estimates for differential social class progress and in the case of change from 11 to 16 years in Mathematics reduced the unadjusted difference to a negligible amount. Nevertheless for change between seven and 11 years for both reading and mathematics attainment, there were still differences after adjustment. Unfortunately this evidence was ignored by Feinstein, as was any procedure for handling measurement error.

As far as the substantive issue goes, our exploration of the data shows support for the proposition that there is indeed a widening performance gap especially for children of high SES parents compared to the remainder. Thus, we do not concur with the claim by Jerrim and Vignoles that there is no convincing evidence to support this. To this extent Feinstein's original conclusions are broadly supported, but the actual extent of the widening gap is still an open question, although likely to be much less and more nuanced than he has claimed. Nevertheless, even with our low estimate of reliability (0.65) we still estimate that those from the high SES group are on average 0.13 of a standard deviation ahead of the remainder at age five given the same achievement at age three.

To be fair, Feinstein is far from alone in ignoring these methodological issues. In a recent highly quoted report looking at the progress between 11 and 16 years of 'bright but disadvantaged students', Sammons, Toth and Sylva (2015) also fail to recognise the problems associated with conditioning on a high achieving group, in their case the top third of students, and they also fail to take account of measurement error.

The debate engendered by Feinstein's original paper and various critiques, especially that of Jerrim and Vignoles, has clearly been a difficult one for policymakers, turning as it does on a rather poorly understood set of technicalities. In our view all of this suggests that a more cautious, long term attitude should be taken towards such research findings. Social research is a highly contested area, whether published in a 'reputable' journal or as a non peer-reviewed report to a sponsor.

Policymakers would do well to promote a wide debate about any findings that appear important, where technical and interpretational issues are debated in terms that are widely accessible, and where other relevant research can be discussed. This would be in everyone's interests, not least that of the policymakers themselves who would more often be seen as interested in pursuing useful knowledge rather than advancing their own predilections.

References

Feinstein, L. (2003). Inequality in the early cognitive development of British children in the 1970 cohort. *Economica, 70*, 73-97. http://dx.doi.org/10.1111/1468-0335.t01-1-00272

Goldstein H. (1979). Some Models for Analysing Longitudinal Data on Educational Attainment (with discussions). *Journal of the Royal Statistical Society Series A 142* (4) 407-442 http://dx.doi.org/10.2307/2982551

Jerrim, J. & Vignoles, A. (2013). Social mobility, regression to the mean and the cognitive development of high ability children from disadvantaged homes. *Journal of the Royal Statistical Society Series A 176*(4), 887 – 906. http://dx.doi.org/10.1111/j.1467-985X.2012.01072.x

Richardson,S. & Gilks,W. (1993). Conditional independence models for Epidemiological studies with covariate Measurement error. *Statistics in medicine, vol. 12,* 1703-1722. http://dx.doi.org/10.1002/sim.4780121806

Sammons, P., Toth, K. & Sylva, K. (2015). *Subject to background: What promotes better achievement for bright but disadvantaged students?* Oxford University Department of Education. Retrieved from http://www.suttontrust.com/researcharchive/subject-to-background/

Commentary by **Elizabeth Washbrook** University of Bristol, UK
Liz.Washbrook@bristol.ac.uk
RaeHyuck Lee Columbia University, US

Beyond the Feinstein chart: Investigating differential achievement trajectories in a US cohort

Introduction

The policy and media attention attracted by Feinstein's (2003) paper revolved chiefly around a single "killer chart", depicting trajectories of the average test scores over the course of childhood for four distinct groups of children – "high-" and "low-" scoring children at 22 months of age from high- and low-socioeconomic status (SES) backgrounds respectively. As will be the case in any killer chart, the compelling telling of a visual story can only be achieved by a great deal of selection and simplification of the underlying messy data. This is of course precisely what the analyst sets out to do – to understand and communicate the relationships that are meaningful in a substantive sense. Feinstein's chart was a hugely valuable contribution in the way that it captured imaginations and drew attention to the topic of the developmental trajectories of children from disadvantaged backgrounds. But having kick-started the debate, it is unfortunate that certain features of the chart seem now to be taken as inherent features of the problem at hand, rather than particular choices made for their convenience or visual impact. In this paper we echo Feinstein's (2015) call for more developed analyses of SES differentials in children's trajectories: analyses that build on his original chart but that consider the full range of children's achievement levels, draw on more sophisticated statistical methods, and that take particular care in the interpretations placed on the data. The ideas are illustrated briefly with a longitudinal analysis of the reading skills of a recent cohort of children from the United States.

Our starting point is that the real question of interest that underlies Feinstein's chart is whether trajectories of development tend to diverge, on average, between children who begin with an identical level of baseline achievement but come from different social backgrounds. This is essentially a question about the *timing* of the evolution of SES gaps, about how far inequality observed at the end-point is "locked in" at the time of baseline assessment versus how far it develops subsequently. Jerrim and Vignoles (2013) highlight one important statistical reason why Feinstein's chart might give inaccurate answers to this question: measurement error in baseline test scores will tend to exaggerate the extent to which low SES children fall behind higher SES children who begin in the same initial achievement group. Although this point is undoubtedly valid, it is unfortunate that Jerrim and Vignoles' critique has been interpreted as evidence that the relationships depicted in Feinstein's chart are entirely spurious, as their analysis in no way proves this to be the case. In fact, another feature of Feinstein's analysis – the use of cut-points to classify children into discrete "low" or "high" achieving groups at baseline – also hinders our ability to define identical starting points for children from different backgrounds. There are statistical problems with this approach, plus it may have obscured the fact that the issue considered by Jerrim and Vignoles is in fact a classic measurement error problem, for which a highly sophisticated toolkit of methods for correction and assessment of sensitivity has long been available.

Technical issues aside, we also welcome the opportunity to comment on some of the interpretations that have been applied to Feinstein's findings in general. We question the legitimacy and possible implications of labelling a child as being of high or low ability on the basis of a narrow set of tests at a particular age, and also the disproportionate focus that some have placed on the outcomes of "high ability" low SES children as compared to other low SES children. We begin with this discussion of the messages that have been taken from Feinstein's work in the world of policy and practice. We then go on to address a number of statistical issues in the way that the chart, and some related research, has been constructed, before presenting our analysis of trajectories of reading achievement among US children born in the early 1990s. We show that low SES children systematically failed to keep pace with higher SES children who began with identical reading skills in kindergarten, and that these inequalities cannot be explained by errors in test score measurement. We then offer some concluding remarks.

Issues of interpretation

The question of whether it is either meaningful or ethical to label some young children and not others as "highly able" on the basis a test taken at one particular age is one discussed in detail by Feinstein (2015), and we strongly welcome the problematizing of this issue. We believe that the uncritical use of this term, as well as of the even more emotive adjectives "clever", "bright" and "dim" in this context is unscientific on the part of academics, and politically loaded on the part of any commentator. We agree that, when characterising the starting point of whatever trajectories are being measured, terms like "initially high performing" or, as we use here, "initially high achieving" should be used as much as possible, even if they are less elegant than the phrases used in common discourse.

The word "initially" is important because it emphasizes that children are being differentiated on the basis of something measured at a single point in time rather than something fixed. The words "performing" or "achieving" emphasize that any test measures the ability to do something specific, and does not even aim to capture all mental capacities. As Feinstein (2015) and Goldstein (1979) among others have argued, it is neither necessary nor appropriate to insist that a single uni-dimensional construct be used to characterise children's development over long periods of time. Nor is it necessary, we would add, to seek to measure the abstract concept of "cognitive ability" via specific tests of vocabulary, or general intelligence or Key Stage mathematics skills. All of these measure something different and substantively important in terms of children's capacities, and can simply be discussed in terms of what they are. To say that a child has a poorly-developed vocabulary may be a narrower statement, but it embodies fewer assumptions and is more precise, than the claim that the child is "low ability" or "dim" (particularly if that child is from a foreign language background, for example).

The labelling by researchers of some children as able or not able implicitly supports the idea that this is a useful way to think about human development. One needs only to think about the debates that led to the abolition of the selective "11-plus" exam in large parts of England in the 1960s to see that such a view has long been politically contested. The idea that certain individuals inherently have more potential than others is one that plays a key role in fundamental debates over social and educational inequality (Dorling, 2015), and the incautious use of language by researchers can give the erroneous impression that the evidence they present is supportive a particular view.

This point is perhaps even more relevant in the light of the disproportionate focus on "highly able" low SES children, relative to the majority who are presumably "not highly able", in much of the comment and analysis that has followed from Feinstein's chart. The key message that provoked a sense of injustice in numerous policymakers and commentators was that, in the words of Jerrim and Vignoles "highly able children from disadvantaged homes are overtaken by their rich (but less able) peers before the age of 10 in terms of their cognitive skill" (e.g. Baumberg, 2011; Harford, 2012; HM Government, 2003, p.19; statement by the Deputy Prime Minister, HC Deb 5 April 2011). Several pieces of related research have then focused specifically on the outcomes of low SES children classed as "high-attaining" (Crawford, Macmillan & Vignoles, 2014), "talent[ed]" (Education Datalab, 2015), and "bright" (Sammons, Toth & Sylva, 2015) at some baseline assessment age. It is surprisingly difficult to find an explicit rationale for why the highly able sub-set of low SES children is singled out in this way, nor (where it occurs) for the purpose of comparing their outcomes with a low-ability high-SES group.

The idea that it is even meaningful to talk about the "overtaking" of an initially high-achieving group by a different low-achieving group is problematic, as we discuss in the following section, but even accepting that this phenomenon can be objectively identified it is somewhat mysterious as to why it matters. As stated in the introduction, the key question from a social equality perspective seems to us to be whether low SES children systematically underperform relative to higher-SES children with identical initial capacities. It is not clear why the significance of the underperformance of a particular group of high-achieving low-SES children should depend on the relative over-performance of a very different group of low-achieving high-SES children, as seems implicit in so many interpretations of Feinstein's chart. It seems possible that the exclusion of any results for children with middling achievement levels at baseline from the chart may have unintentionally skewed reactions to Feinstein's findings. It is interesting to note that the US literature on the evolution of black-white achievement gaps, which addresses similar social and identical methodological issues, places no particular emphasis on high-achieving black children (e.g. Phillips, Crouse & Ralph, 1998; Reardon, 2008; McDonough, 2015). This suggests that a preoccupation with the initially high-achieving among the disadvantaged is not in some sense "natural". It is likely, of course, that there are many legitimate reasons for the slant taken in reactions to Feinstein, but when these are not articulated it risks giving the impression that the relative underperformance of the vast majority of low-SES children is viewed as matter of lesser social concern than the outcomes of an exceptional few. Ultimately, our main contention is that there is a continuum of both achievement and disadvantage that needs to be explored to inform policy, and our empirical example in this paper provides one example of how this might be done.

Statistical issues

The issue of potential measurement error bias in any estimation of cross-group differences in trajectories is one that, as Jerrim and Vignoles (2013) rightly emphasise, must be addressed. Their analysis traces out the consequences of the fact that there is a difference between the true score of an individual that a test attempts to measure, and the observed score that actually results from the test, which contains some component of random measurement error. The characterisation of this issue made by them, and a number of authors, as one of "regression to the mean" (RTM) has, as Feinstein (2015) notes, led to a lack of clarity in a number of areas. As Jerrim and Vignoles fully recognise, the correlation of outcomes over time that produces RTM is affected by many factors that are nothing to do with mismeasurement of the underlying constructs: factors such as transitory influences on development and shifts in the underlying skills that are relevant at different ages. It seems confusing, therefore, to refer to all these processes as RTM, and even more to describe RTM as a "spurious statistical artefact".

The real advantage of making this distinction and of framing the problem as one of measurement error is not just greater clarity, however. In this context it immediately makes transparent the links between this problem and a vast body of statistical results on how to tackle measurement error in a range of different forms (see e.g. Fuller, 2009; Carroll, Ruppert, Stefanski & Crainiceanu, 2006). To give an illustration, Jerrim and Vignoles propose the use of an auxiliary test score to categorize children as "high" or "low" ability on the first measurement occasion as a way to correct for measurement error. Although not acknowledged as such, this method can be viewed as the basis of a rudimentary instrumental variables (IV) estimator, an approach that has long been used as a standard solution in this context (e.g. Blackburn and Neumark, 1992; Ecob and Goldstein; 1983). The IV method is presented formally in the online supplementary material and illustrated in practice in our application in the next section. The advantage of drawing on the classic IV framework here is that it allows one to harness all the associated statistical results and software that have been developed over many years. As a second example of how results from the measurement error literature can inform our analysis of trajectories, we also employ an alternative method in the next section based on extensions to the textbook exposition of attenuation bias (e.g. Wooldridge, 2010). This method provides explicit expressions for the degree of bias in the estimates induced by a given amount of measurement error, and so enables systematic testing of the sensitivity of the results.

A statistical issue with Feinstein's chart that has received much less attention than measurement error is the use of cut-points to classify children into discrete high and low ability groups. On a simple level this specification is wasteful in terms of data because information on each child's individual test score at baseline is discarded. More problematically, it makes it impossible to identify the difference in later outcomes between children from different SES groups who were *identical* at baseline which, as argued, we believe is the ultimate quantity of interest. The problem is that if higher SES children have higher baseline achievement on average than lower SES children, even in the absence of measurement error the achievement levels of those who make it into the "high achievement" group will be systematically greater than those of the low SES children classified in the same group. Comparisons of the later outcomes of the two groups of children will not be comparing like with like, and differences in subsequent rates of progress will be confounded by differences in initial conditions. These initial differences are potentially very large when cut-points such as the 75th percentile are imposed on the data. If we take the distributions of true baseline achievement in the population used in Jerrim and Vignoles' simulations as an example, the score of the average high SES child in the top quartile must exceed the score of the average low SES child in the top quartile by over a third of a standard deviation[1], and note that in this example there is no measurement error at all in the data and hence no misclassification.

The use of cut-points also has implications for the question of whether initially high-scoring low SES children are "overtaken" by initially low-scoring high SES children, which we highlighted previously. Visually this is represented in Feinstein's chart by the "crossing" of the relevant trajectories between the ages of five and ten. Quite apart from the question of why this crossing is worthy of particular attention, whether it occurs or not is highly likely to depend on arbitrary definitions of the high and low achievement categories. As we show in the empirical example in this paper, slight modifications to the presentation can generate or eliminate the appearance of overtaking in predictions from a single underlying model. Hence without greater precision in terms of definition, the statement that this phenomenon does, or does not, occur is essentially meaningless.

A framework that relates continuous measures of baseline achievement to later outcomes overcomes these problems, by allowing the baseline measures to be "matched" exactly across groups, and by forcing the analyst to be transparent about which comparisons have been selected for presentation and why. In contrast, one advantage of using raw group-specific means is that the way in which the results have been generated is immediately intuitive for a non-technical audience. It is possible there are other advantages and it would be helpful if these were articulated in applications where cut-offs are imposed.

Estimating SES differentials in trajectories in a US cohort

In this section we present some evidence on trajectories of relative reading achievement by SES that illustrates the potential of several analytical methods. The data used are taken from the Early Childhood Longitudinal Study - Kindergarten cohort (ECLS-K), a nationally representative longitudinal study of US children who entered kindergarten in 1998. An initial sample of around 19,000 children, along with their parents and teachers, were surveyed on six occasions between entry to kindergarten (average 5.7 years of age) and eighth grade (average 14.2 years of age). This analysis draws on the sample of 7,340 children with valid data in all six waves[2]. Longitudinal survey weights and design variables are provided by the ECLS-K to allow inferences about the underlying national population on the basis of this sample; these are used in all analyses. Reading outcomes are captured on each of the six occasions by the ECLS-K's reading theta score, derived from a suite of tests designed to measure achievement in six dimensions of reading skill, from basic letter recognition through understanding and inference to demonstrating a critical stance (see Tourangea, Nord, Sorongon, Najarian & Germino, 2009, for further details). Test scores at each age were adjusted for age-in-months at assessment and standardized to mean zero unit variance z-scores using the survey weights. We measure parental SES using the highest qualification of a parent resident with the child during the kindergarten year. We distinguish a high SES group corresponding to a parent with a bachelor's college degree or more (30% of the sample); a low SES group corresponding to no parental education beyond high school graduation (37%); and a residual medium SES group (33%).

A flexible model for investigating the extent to which trajectories diverge by SES is

$$A_{i2} = \alpha_0 + \alpha_L L_i + (\beta_0 + \beta_L L_i)A_{i1} + (\gamma_0 + \gamma_L L_i)A_{i1}^2 + u_{i2} \qquad (1)$$

Where A_{it} is the "true" or perfectly-measured achievement of child i on measurement occasion t; L_i is a dummy variable indicating membership of the low (relative to high) SES group, which can easily be extended to a vector distinguishing multiple SES categories; and u_{i2} is a mean zero uncorrelated residual term. The inclusion of the quadratic term, A_{i1}^2, allows the strength of the association between initial and final outcomes to vary with the level of baseline achievement (so that, for example, low initial scores can be less predictive than high ones of future outcomes). The conditional SES gap at occasion two – that is, the difference in test scores predicted to open up between low and high SES children with a truly identical achievement level at time one – is given by

$$E(A_{i2}|L_i = 1, A_{i1}) - E(A_{i2}|L_i = 0, A_{i1}) = \alpha_L + \beta_L A_{i1} + \gamma_L A_{i1}^2 \qquad (2)$$

Tests of the parameters β_L and γ_L can be conducted to assess formally whether developing inequalities are more severe among those who were initially higher or lower achievers. For example, if $\beta_L < 0$, it tells us that (at least over some range) SES gaps open up more between high, rather than lower, scoring children at baseline; if $\beta_L > 0$ it is the weaker-performing low SES children who fall relatively further behind. The inclusion of the quadratic term again allows for greater flexibility, this time in where the largest gaps are estimated to appear. In our analysis of the reading trajectories of US children presented below, the hypothesis that β_L and γ_L were jointly zero could not be rejected in the vast majority of models. In order to simplify the presentation we proceed here with a more parsimonious model that imposes the constraint that the SES differential in progress is simply a constant, α_L (that is, it is the same regardless of initial achievement level) [3].

$$A_{i2} = \alpha_0 + \alpha_L L_i + \beta A_{i1} + \gamma A_{i1}^2 + u_{i2} \qquad (3)$$

The key problem highlighted by Jerrim and Vignoles (2013) is that A_{i1} and A_{i2} are not observed directly. Instead we observe imperfect test score measures, Y_{i1} and Y_{i2} that contain random measurement error components for each individual. A standard result from the statistical literature tells us that ordinary regression estimates of equation (3) will be biased when we substitute the observed error-prone variable, Y_{i1}, for the true baseline achievement measure A_{i1} [4]. One method for correcting for this bias is to use an auxiliary variable, known as an instrumental variable, to statistically "purge" Y_{i1} of its measurement error component. This "corrected" measure of baseline achievement is then used in place of the observed value in a procedure known as two-stage least squares (2SLS).

The first key requirement for this 2SLS procedure to yield correct estimates of the equation of interest is that the error components of the test scores be uncorrelated with the chosen auxiliary variable or instrument. This assumption is likely to be violated if the instrument is an alternative test taken on the same day and under the same conditions as the baseline assessment, because random environmental factors are likely to affect the two observed scores in the same way [5]. Even if this condition is satisfied, however, perhaps a more stringent requirement is that the instrument must be "redundant" in the equation of interest, that is, it contains no information about A_{i2} once true achievement A_{i1} (plus its square and L_i) are conditioned on. If the instrument has some predictive power for A_{i2} independently of the other factors included in the statistical model, then the 2SLS methods will not fully correct for measurement error biases, and may even make matters worse.

In our application we define the first measurement occasion as the spring of kindergarten, roughly half way through the child's first year of compulsory schooling (average age 6.2 years). This allows us to use a prior score on the same test taken just after kindergarten entry (about six months previously, at age 5.7 years on average) as the instrument. The outcome variables in four separate models are then reading achievement in the spring of first, third, fifth and eighth grade, or around ages seven, nine, 11 and 14 respectively. For the instrument to be valid, therefore, we must assume that measurement errors in each of the tests are independent of one another, and also that when realised (true) achievement towards the end of kindergarten is known, the observed reading test score from the start of that school year contains no further information about subsequent achievement from first grade onwards.

"Corrected" (2SLS) and "uncorrected" ordinary least squares (OLS) results for the model (3) are presented in table 1. The key parameters of interest are the coefficients corresponding to α_L, the top two rows that give estimates of the gaps predicted to appear between children who began with an identical level of reading achievement at age six, but who were from low- and medium-SES backgrounds respectively (relative to the reference high-SES group). Provided the instrumental variable assumptions are satisfied, the results in the top row show that a (marginally) significant gap of .07 standard deviations is predicted to open up by age seven between the lowest and highest SES children who had identical reading achievement in their first year of formal schooling (12 months previously). That gap is predicted to widen steadily to .53 standard deviations by age 14. Comparison of the OLS and 2SLS results suggests that while measurement error in the test score at occasion one does indeed tend to bias naïve estimates of the size of the gaps upwards, it accounts only for around 20% of the estimated high-low SES gap in eighth grade.

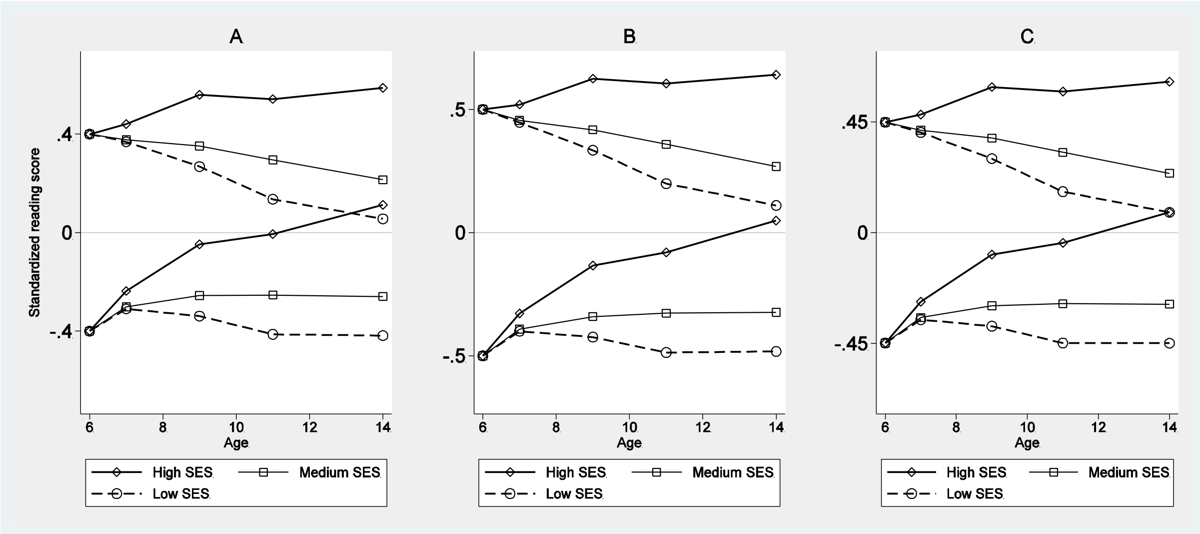
Table 1. The relationships between SES, reading achievement at age six and later reading achievement

	1st grade (age 7)	3rd grade (age 9)	5th grade (age 11)	8th grade (age 14)
A. "Corrected" 2SLS estimates				
Low SES	-.073 (.037)	-.291 (.045)	-.407 (.049)	-.531 (.055)
Medium SES	-.064 (.029)	-.208 (.040)	-.247 (.041)	-.372 (.045)
Age 6 test score	.848 (.023)	.759 (.029)	.686 (.024)	.593 (.028)
Age 6 test score squared	-.067 (.014)	-.105 (.016)	-.058 (.016)	-.056 (.015)
Constant	.113 (.024)	.272 (.031)	.278 (.032)	.359 (.036)
B. "Uncorrected" OLS estimates				
Low SES	-.117 (.035)	-.443 (.044)	-.525 (.049)	-.656 (.050)
Medium SES	-.104 (.029)	-.263 (.038)	-.294 (.040)	-.417 (.044)
Age 6 test score	.735 (.015)	.595 (.021)	.557 (.019)	.458 (.020)
Age 6 test score squared	-.034 (.009)	-.043 (.011)	-.025 (.011)	-.004 (.009)
Constant	.128 (.024)	.280 (.029)	.299 (.029)	.364 (.033)

Standard errors in parentheses. High SES is the omitted reference category. N = 7340.

The results in table 1 allow a predicted outcome to be generated for any combination of SES and kindergarten test score, which can be used to illustrate the consequences of different definitions of what constitutes "high" and "low" achievement on the first measurement occasion. Panel A of figure 1 plots the trajectories predicted by the 2SLS estimates of children from different SES groups who started with achievement either .4 standard deviations above or below the mean in kindergarten. The choice of .4 as an initial benchmark generates a pattern of "crossing" trajectories: a low SES child starting from the higher achievement score is expected, on average, to have poorer outcomes by age 14 than a high SES child whose achievement was .8 standard deviations lower in the first year of schooling. Panel B plots trajectories for slightly more extreme initial achievement levels – .5 standard deviations above and below the mean respectively – and shows that with this minor difference the crossing property is eliminated. Although there is considerable convergence, a low SES child starting at the higher of the two scores is in this case predicted to continue to outperform a child who began with a one standard deviation relative disadvantage in kindergarten, even if that child is from a high SES background. To us, this illustrates the arbitrary nature of comparisons of that single out specific trajectories associated with different starting points for different socioeconomic groups. The compelling fact here is that if we take two children with identical reading achievement at age six – one from a high-SES background and one from a low-SES background – by age 14 the high-SES child is predicted to out-perform the low-SES one by over half a standard deviation. This is equally true and, we argue, equally important, for all low-SES children, including those whose reading achievement was not exceptional when they started school.

Figure 1. Expected trajectories of reading achievement by SES, from selected starting points at age six



Note: Charts plot outcomes predicted on the basis of the 2SLS estimates in panel A of Table 1, for the selected values of the initial age 6 score marked on the y-axis.

If, however, there remains an interest in comparing initially higher-achieving low-SES children with their mirror image lower-achieving high-SES peers, then the use of continuous achievement measures, rather than arbitrary ability groupings, allows for a much more systematic analysis. For example, one could trace out the frontier of exactly how much higher the initial achievement of a low SES child than a high SES child must be in order for them to reach the same level of performance in eighth grade. This distance can be calculated for every possible outcome level in eighth grade, allowing a detailed characterisation of where "overtaking" occurs in the joint baseline achievement distribution of the two groups. Panel C of figure 1 provides just one example, showing that a high SES child scoring .45 standard deviations below the mean in kindergarten (the 33rd percentile) is predicted to attain the same reading level in eighth grade (a score of .08) as a low SES child who began .45 standard deviations above the mean (the 67th percentile).

The instrumental variables strategy outlined here is only valid if the chosen instrument satisfies the necessary statistical requirements. In our application, it also constrained our analysis of trajectories to begin with baseline achievement at age six (the second measurement occasion in the survey), because the first age five score had to serve as the auxiliary variable to correct for measurement error. In datasets where there is a long lag between the first and second measurement occasions, this approach could prevent the analysis of trajectories over important periods earlier in childhood. An alternative approach to adjusting for measurement error – the moment-based estimation method – makes a different set of assumptions, which may be more or less valid than the IV assumptions, but has the advantage that it dispenses with the need to employ an auxiliary variable at all. Instead, this approach requires that we make additional assumptions about the distributions of both unobserved true achievement and measurement error[6], and crucially that we know the degree of error in, or reliability of, the observed test score.

The reliability, denoted r, is the proportion of variance in the observed test score generated by variation in true underlying achievement. (r lies between 0 and 1, with higher values corresponding to more accurate measurement.) When r is known, correction factors can be derived and applied directly to the OLS estimates (see Goldstein, 1979 and Phillips et al., 1998 for examples of similar applications, and the supplementary material for the derivation of the correction factors applied here). Test developers often provide estimates of reliability based on the internal consistency of the individual test items, but it seems unlikely that this will capture all potential sources of noise in all contexts. In cases where the reliability is uncertain or unknown, estimates for a range of values of r can be computed to assess sensitivity. In addition, we might usefully ask how low the reliability would need to be for divergence by SES to be purely a statistical artefact.

With regard to the current application, the ECLS-K test developers provide an estimate of .95 for the internal reliability of the age six reading test score (Tourangeau et al., 2009, table 3-10). Corrected estimates using this value of the reliability are provided in table 2. Given the very high value assumed for r, it is unsurprising that these estimates are very close to the unadjusted OLS estimates shown in table 1. Alongside the coefficients on the SES indicators, table 2 also shows the minimum reliability needed to generate a negative SES gap at each age, r^*. More developed analyses could provide estimates of the reliabilities needed to generate statistically significant, rather than just non-zero, SES gaps, but it is clear that, at least from third grade onwards, observed age six test scores would have to contain an implausibly large proportion of measurement error for the finding of divergence by SES to be entirely spurious.

Table 2. Measurement-error corrected estimates of trajectories in reading achievement from age 6 onwards, *r* = .95

	1st grade (age 7)		3rd grade (age 9)		5th grade (age 11)		8th grade (age 14)	
	Coef (SE)	r^*	Coef (SE)	r^*	Coef (SE)	r^*	Coef (SE)	r^*
Low SES	-.142 (.035)	.79	-.415 (.044)	.55	-.499 (.050)	.49	-.635 (.051)	.39
Medium SES	-.088 (.029)	.75	-.249 (.038)	.50	-.282 (.040)	.46	-.407 (.044)	.32
Age 6 test score	.775 (.016)		.628 (.023)		.588 (.020)		.482 (.021)	
Age 6 test score sq	-.037 (.010)		-.048 (.012)		-.027 (.012)		-.004 (.010)	
Constant	.114 (.024)		.270 (.029)		.289 (.029)		.353 (.034)	

r^* denotes the minimum value of r consistent with a negative, non-zero estimate of the associated coefficient in a measurement-error-corrected model.

Conclusion

Feinstein's (2003) chart attracted a great deal of attention, perhaps because its strong visual image resonated with people's intuition about the way social class differences become embedded over the course of childhood. The interest generated suggests there is great value in digging deeper into precisely when and for whom trajectories diverge, and our analysis of the ECLS-K suggests a number of avenues for research that could refine our understanding in the British context.

The question of the timing of the evolution of SES gaps is important for thinking about when to target policy interventions, particularly given the recent emphasis on the preschool period as advocated by James Heckman and others. The trajectory analysis presented here suggests that educational inequality in the US strengthens considerably in the eight years after school entry, a finding that would be missed from inspection of cross-sectional achievement gaps, which change little with age[7]. The question of whether the degree to which inequalities widen varies with baseline achievement level is also one with important implications for policy. The interventions that are likely to prevent gaps emerging among initially higher-achieving children may be very different from those needed to promote equality among those struggling with basic skills. In our US cohort we find that divergence is a common problem across the full range of the initial achievement distribution, but this is unlikely to be the case in all countries and time periods. A full characterization requires that we move beyond a preoccupation with children predefined as "low" and "high" achievers and give equal consideration to the potential of all lower SES children.

It is important that the issue of measurement error be dealt with if we are to provide convincing answers to these important questions, but we believe it is a mistake to assume any evidence of diverging trajectories must be spurious until proven otherwise. A range of methods for assessing sensitivity to and adjusting for measurement error are available. We explore only a few of these here and our results suggest that, although measurement error bias is present, it plays only a minor role in what is a much larger story about the evolution of social inequality.

Acknowledgements

This paper draws on work supported by the Russell Sage Foundation and the Australian Research Council (grant DP130103440). The authors would also like to thank Bruce Bradbury, Miles Corak, Jane Waldfogel and Anna Zhu for their help in developing the ideas in this project.

References

Baumberg, B. (2011, June 16). *The rise and fall of a killer chart* (Web log comment) Retrieved from http://inequalitiesblog.wordpress.com/2011/06/16/the-rise-and-fall-of-a-killer-chart

Blackburn, M. & Neumark. D. (1992). Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials. *The Quarterly Journal of Economics*, *107*(4), 1421-1436. http://dx.doi.org/10.2307/2118394

Bradbury, B, Corak, M, Waldfogel, J & Washbrook, E. (2015). *Too Many Children Left Behind: The US Achievement Gap in Comparative Perspective.* Russell Sage Foundation: New York (in press)

Carrol, R.J., Ruppert, D., Stefanski, L.A. & Crainiceanu, C. (2006). *Measurement error in nonlinear models: A modern perspective, Second Edition*. London: Chapman and Hall. http://dx.doi.org/10.1201/9781420010138

Crawford, C., Macmillan, L. & Vignoles, A. (2014). *Progress made by high-attaining children from disadvantaged backgrounds.* Research report. June 2014. Social Mobility and Child Poverty Commission.

Dorling, D. (2015). *Injustice: Why social inequality still persists* (revised edition). Bristol: Policy Press

Ecob, R., & Goldstein, H. (1983). Instrumental variable methods for the estimation of test score reliability. *Journal of Educational and Behavioral Statistics*, *8*(3), 223-241. http://dx.doi.org/10.3102/10769986008003223

Education Datalab (2015). *Missing Talent*. Research brief. 5 June 2015. The Sutton Trust. http://www.suttontrust.com/wp-content/uploads/2015/06/Missing-Talent-final-june.pdf

Feinstein, L. (2003). Inequality in the early cognitive development of British children in the 1970 cohort. *Economica*, *70*, 73-97. http://dx.doi.org/10.1111/1468-0335.t01-1-00272

Feinstein (2015). Social class differences in early cognitive development and regression to the mean. *Longitudinal and Life Course Studies 6*(3)

Fuller, W. A. (2009). *Measurement error models.* John Wiley & Sons.

Goldstein, H. (1979). Some models for analysing longitudinal data on educational attainment. *Journal of the Royal Statistical Society. Series A (General)*, 407-442. http://dx.doi.org/10.2307/2982551

Harford, T. (2010, December 11). *Why education fails the poor* (Web log comment) Retrieved from http://timharford.com/2010/12/why-education-fails-the-poor/

HC Deb (2011). House of Commons Debate, 5th April 2011, c875.

HM Government (2003). *Every Child Matters*. Green Paper, September 2003, Cm 5860.

Jerrim, J. & Vignoles, A. (2013). Social mobility, regression to the mean and the cognitive development of high ability children from disadvantaged homes. *Journal of the Royal Statistical Society series A 176*(4), 887 – 906. http://dx.doi.org/10.1111/j.1467-985X.2012.01072.x

McDonough, I. K. (2015). Dynamics of the black–white gap in academic achievement. *Economics of Education Review*, *47*, 17-33.

http://dx.doi.org/10.1016/j.econedurev.2015.03.007
Phillips, M., Crouse, J. & Ralph, J. (1998). Does the black-white test score gap widen after children enter school? Pp. 229-272. In C. Jencks, C. & Phillips, M. (Eds) *The Black-White Test Score Gap*, Washington, DC: Brookings Institution Press.
Reardon, S. F. (2008). *Thirteen ways of looking at the black-white test score gap.* Stanford Institute for Research on Education Policy & Practice, Working Paper, 8.
Sammons, P., Toth, K. & Sylva, K. (2015). *Subject to background: What promotes better achievement for bright but disadvantaged students?* Oxford University Department of Education. http://www.suttontrust.com/researcharchive/subject-to-background/
Tourangeau, K., Nord, C., Le T., Sorongon, A.G., Najarian, M. & Germino Hausken, E. (2009). *Combined User's Manual for the ECLS-K Eight-grade and K-8 Full Sample Data Files and Electronic Codebooks*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
Wooldridge, J.M. (2010). *Econometric Analysis of Cross Section and Panel Data, 2nd Edition*. MIT Press.

Endnotes

[1] In this example, baseline achievement follows a standard normal distribution and two equal-sized groups have means of -0.7 (labelled low SES) and 0.7 (labelled high SES). Average scores of children in the top quartile overall can therefore be calculated as 1.13 for the low SES group and 1.48 for the high SES group.
[2] In accordance with NCES reporting rules, all sample sizes are rounded to the nearest 10.
[3] We note, however, that the interaction terms are significant when achievement in maths, rather than reading, is the outcome of interest. See Bradbury, Corak, Waldfogel & Washbrook (2015), Ch. 6, for related analyses of these data, which include fully interacted models and estimates for both types of outcome.
[4] See supplementary material for further details.
[5] As they acknowledge, this problem is likely to affect the corrected results presented by Jerrim and Vignoles (2013).
[6] Specifically that they are normally distributed with constant variances – see the supplementary material for a formal treatment
[7] See Bradbury et al. (2015), Ch. 6 for more discussion on this point.

Commentary by Ruth Lupton University of Manchester, UK
ruth.lupton@manchester.ac.uk

The practice of policy-related research

Introduction

There are many fascinating and important aspects of the debate over the 'Feinstein graph' which is the subject of this section. Here I address just one – what the case suggests about processes of research production, dissemination and use in public policy fields. For those reading the paper in isolation from the rest of the collection, I first offer a brief descriptive account of the case. I then explore policy implications of the findings of the Feinstein and Jerrim/Vignoles papers, both as they were put and as they were taken up. I suggest that avoidance of the wrong policy implications being drawn from this debate could be helped, among other things, by reference to research in other disciplines, which speaks to the same questions, and conclude with some questions on how such interdisciplinarity might be achieved in practice.

The debate over the Feinstein graph

In 2003, the journal *Economica* published a paper by Leon Feinstein based on analysis of the 1970 British Birth Cohort Feinstein (2003) looked at children's performance in various tests of functioning and cognitive attainment at ages 22 months, 42 months, five years and ten years. He described the average relative position in the overall rankings, at each age, of children from different social economic backgrounds (hereafter SES). This showed that at 22 months, there was already a socioeconomic gradient. This pattern persisted at 42 months, five and ten years and became more pronounced between the years of five and ten.

Feinstein then went on to group the low and high SES children according to their performance on the first test, and to look at their subsequent average rankings in later tests. On average the ranking of high attaining low SES children declined over time, while that of low attaining high SES children increased. Indeed, this latter group had overtaken the former group by age ten (i.e. during the primary school years). The visualisation of this finding (figure 2 in the original paper) is what became familiarly known as 'the Feinstein graph'. Feinstein also ran the analysis classifying groups on the basis of the second test rather than the first. The same pattern was also evident, but was less marked. The ranks of the high attaining low SES children and the low attaining high SES children were converging by age ten but the latter had not overtaken the former. A further important point, often overlooked in the debate about the converging lines, is that low SES children in the lowest quartile at 42 months made no relative gains by age ten. Their primary school experiences did not enable them to catch up with other children. Similar findings were reported in Schoon (2006) using data both from the British Cohort Study (BCS) and the 1958 National Child Development Study, and also in Goodman and Gregg (2010).

In 2011, John Jerrim and Anna Vignoles published a working paper entitled *The use (and misuse) of statistics in understanding social mobility: regression to the mean and the cognitive development of high ability children from disadvantaged homes* (Jerrim & Vignoles, 2011a) a version of which subsequently appeared in the Journal of the Royal Statistical Society with a less provocative title (Jerrim & Vignoles, 2013). This paper argued that the phenomenon of the converging trajectories of initially high and low attainers could be caused by regression to the mean. One way to correct for this is to use a different baseline measure to classify the children than the one used as the starting point of the trajectory. This is not possible in the British Cohort Study, so the authors used data from the Millennium Cohort study, with a baseline at age three. They found some convergence between groups between ages three and five (but less than found with the conventional method). Between age five and seven, the initially low attaining high SES children continued to improve their ranking, but the high attaining low SES children maintained theirs, so there was no crossover effect. For both initial high attainers and low attainers, there was some widening of the gaps between social class groups between ages three and seven.

Policy implications and policy reactions

As Feinstein recounts in his paper in this issue, the Jerrim and Vignoles working paper set off a public debate over the Conservative/Liberal Democrat Coalition Government's Social Mobility Strategy. Writing in the Guardian newspaper, Allegra Stratton suggested that Conservatives were circulating the working paper, having renamed it 'Forget Feinstein' and seeing it as a way to challenge the Social Mobility Strategy so strongly associated with the Liberal Democrat Deputy Prime Minister Nick Clegg. Quoting the paper's conclusions, she stated "in other words, the entire basis for the government's Social Mobility Strategy is wrong" (Stratton, 2011). This line was also taken by Read (2011) and by Saunders (2012). Saunders used Jerrim and Vignoles findings as one plank in an argument that Britain does not have a social mobility problem nor need a Social Mobility Strategy.

The statement that "the entire basis for the government's social mobility strategy is wrong" was a misinterpretation of Jerrim and Vignoles message, as they pointed out in a published riposte (Jerrim & Vignoles, 2011b). First of all, a careful look at the ways in which the Feinstein graph was being cited in policy documents shows that it was his evidence on the wide gaps early in life that was being used in support of policy, not the disputed point about whether and at what age initially high attaining low SES children are overtaken by initially low attaining high SES children. In the key document, the Social Mobility Strategy, the graph is reproduced in a section suggesting the need for very early intervention (Cabinet Office and Deputy Prime Minister's Office, 2011). This is also the context in which it is cited in two other influential documents: the 'Field Review', a government commissioned report on Poverty and Life Chances (Field, 2010); and the 'Marmot Review', a government commissioned report on health inequalities, where the conclusion drawn from the graph is that "even greater priority must be given to ensuring expenditure early in the developmental life cycle" (Marmot, 2010 p.22). The need for investment in the early years is supported by both papers, as Jerrim and Vignoles state explicitly in their conclusion. Their findings did not demonstrate that this element of the Social Mobility Strategy was wrong. The Coalition Government has been criticised for failing to follow through on its early years pledges, cutting funding and increasing poverty among families with the youngest children (Stewart & Obolenskaya, 2015). But that this was a result of the underlying evidence apparently being shaken is implausible, since the effective policy decisions were made prior to this debate.

So what about the issue really in dispute in the papers - whether initially high attaining children from low SES backgrounds really fall behind during the school years? Feinstein's findings on this would tend to suggest the need for policies continuing to target children from low income families during schooling and beyond. This was indeed an explicit policy of the Coalition Government, particularly through its establishment of the Pupil Premium, a new per capita funding amount for schools to spend specifically on raising the attainment of children from low-income homes. The previous Labour government had also increased the amount of funding for disadvantaged pupils and introduced a range of targeted initiatives to support their attainment (see Lupton & Obolenskaya, 2013; Lupton & Thomson, 2015 for a more extended discussion). Notably, Jerrim and Vignoles do not suggest that their work implies disinvestment in redistributive efforts after children have started school. Such policies would also be justified by the (undisputed) evidence on the persistence of wide gaps throughout childhood, unless there was evidence that there is no point (because end-of-school gaps merely reflect fixed differences evident at the start) or that there are no interventions that are effective. Feinstein points to the need for more evidence on these points in the conclusion to his paper. However this implication was taken up by Saunders (2012) who argued that if there is no evidence of wasted working class talent, there is no rationale for policies to break down barriers (for example in university admissions). Instead, Saunders argued that differences in socioeconomic outcomes are largely genetically determined and that policy makers should target the 'underclass problem', i.e. that children's lives are blighted by poor parenting in welfare-dependent households.

Again, there is no evidence that this debate actually changed policy. The government continued with its Pupil Premium and Social Mobility Strategy. However, it is clear that although not its intention, the Jerrim and Vignoles paper was used in support of arguments made in and around the Conservative Party against efforts to create more equitable institutional processes and outcomes in the interests of social mobility, and in favour of efforts to increase the personal responsibility, resilience and skills of the poor.[i] Those arguments have arguably gained some traction. In the manifesto of the Conservative government elected in 2015 there is no mention of social mobility, nor of early intervention, early childhood disadvantage, an early years strategy, or socioeconomic inequalities in education, although the Pupil Premium policy is maintained. Those wishing to resurrect such priorities and strategies now face another round of gathering and interpreting evidence that they are needed.

How can the wrong policy implications be avoided?

Cases like this where something goes wrong in the public and political interpretation of research provide a useful if painful reminder of the difficulties academics face in working at the interface with policy. Writing from the position of someone whose findings from longitudinal research have also previously been taken up by the political right to undermine equitable policies, and as someone who works on the substantive issues discussed in these papers, I want to try to draw out some of the lessons that might be learned.

Two things appear to have gone wrong in this case. One is that public commentators failed to identify the specific challenge being made by Jerrim and Vignoles (the extent of convergence of trajectories of initially high and low attainers from different socioeconomic groups) and the policies to which it related, instead understanding that all the underlying evidence for the Social Mobility Strategy was flawed. The other is that they read Feinstein's research as the <u>only</u> evidence relevant to social mobility policies, thus over-stating the policy implications of a challenge to that research.

In relation to the former, the terminology of the Jerrim/Vignoles paper would appear to be implicated. Bearing in mind that the working paper and to a lesser extent the journal article are comprised principally of a dense statistical text which few people have the training to understand even if they were so minded, the choice of a controversial title to the working paper and an abstract announcing 'dramatically different results' exposing 'serious methodological problems' due to a 'spurious statistical artefact' may well have encouraged the impression that all that had gone before was now disproven. The use of terms such as 'bright' and the introduction of the notion of 'true ability', with the assumption that this is lower in working class children, may have unintentionally provided material to those who believe that intellectual capabilities are genetically fixed. Not too many years ago, such terminology used in a methodological working paper would have been inconsequential. However working papers are now published online where they are more accessible than in academic journals, and often with accompanying press releases highlighting key messages. Wide accessibility to people who need only read very few lines to see the main things that are being said substantially increases the risk that the complex analyses forming the bulk of these papers will be misinterpreted. This situation perhaps requires more than usual caution over terminology and sensitivities to the political environment in which it is taken up (not something in which academics are trained).

Both the working paper and the journal article also made a very clear link between Feinstein's work, which it challenged, and the Social Mobility Strategy. According to the article, figure 2 (the converging lines) and similar findings in later studies "has had a significant influence on both academic research and public policy in Britain" (Jerrim & Vignoles, 2013, p. 889), and it drew attention to the citations in major policy documents, and to the fact that the Deputy Prime Minister specifically cited this graph in a House of Commons Debate launching the latter strategy. Following this, the question is asked: "but is the statistical methodology lying behind this result robust?" The implication (intended or not) is that the policies were based on findings that were wrong, so it is unsurprising that this was the conclusion drawn. I make this point at length simply because this is exactly the kind of practice which is now encouraged in an environment in which researchers are measured on their ability to produce policy impact, and in which universities are keen to find newsworthy stories to promote their activities. Academic caveats, preamble and understatement are easily lost in efforts to make clear links to policy. However, it is the second of the misinterpretations that seems to me to provide the more challenging lessons. In the conclusion of their working paper, Jerrim and Vignoles sounded "a clear warning to academics and policymakers not to place too much emphasis on one single result" (emphasis added). This was echoed by Saunders (2012, p.18), who stated that that until the publication of the Jerrim and Vignoles paper there had been "one striking and compelling piece of research" that supported politicians in their thinking that differences in social outcomes must be the result of unfair disadvantages growing up, and by David Willetts (cited in Feinstein, this issue) who suggested that policy-makers had placed too much emphasis on one result (the Feinstein graph). But how did this come to be the case, or to be seen to be the case, by a senior and well-informed government minister? In the sociology of education alone there is a vast body of work pointing to the ways in which 'environmental factors', broadly put, hold back the achievement of low SES children relative to high SES children. These include (and this is by no means a complete list), evidence on:

- differential access to high quality schools (Gewirtz, Ball, & Bowe, 1995; Reay and Lucey, 2003)
- processes of setting and streaming which serve to disproportionately allocate more disadvantaged children to lower classes (Gillborn & Youdell, 2000), and evidence that low attaining pupils are more likely to be demotivated and less likely to attain well if placed in low attaining groups (Ireson & Hallam, 2001).
- the limited pedagogies and narrowed curriculum that can arise in such classes (Thrupp, 1999; Lingard, 2007; Lupton & Hempel-Jorgensen, 2012).
- low SES children not feeling valued at school (Reay, 2006; Bright, 2011)and being demotivated by messages that they are failing.
- the effects of material poverty, housing conditions, and the social and emotional consequences of disadvantage (Ridge, 2002; Horgan, 2007).
- the effects of disadvantaged contexts on school organisation and processes, including teacher recruitment and retention (Lupton, 2006).
- social and cultural capital and the practices of middle class parents to support educational progress (Ball, 2003; Brantlinger, 2003; Lareau, 2011; Reay, Crozier, and James, 2011)

Feinstein also noted in his paper that the results would not surprise those working in the fields of developmental psychology, psychometrics and behavioural genetics. Studies in economics have also documented some of these processes and their effects, particularly school choice (e.g. Burgess, Briggs, McConnell & Slater, 2006; Allen & West, 2011). Cooper and Stewart's recent (2013) systematic review of quantitative studies finds clear evidence of the effect of material poverty on children's cognitive outcomes.

Admittedly, a more systematic review would be needed to differentiate studies that point specifically to the experiences and trajectories of initially higher and lower attaining children, and to differentiate processes in early years, primary and secondary school. Many studies are small scale. We need to know more about their generalisability. They have also been conducted in different time periods and we need to know whether things have changed over time. It may be the case that there are no other specific studies that show the exact pattern described in figure 2 of Feinstein's paper. But readers of this large body of literature would find it no surprise that low SES children who show early signs of high cognitive attainment are less able to translate that into later academic success than their higher SES peers, nor that schools fail to transform the trajectories of initially lower attaining SES pupils. Jerrim and Vignoles' results with the Millennium Cohort Study also show the accelerating performance of the higher SES initial lower attainers, compared to lower SES peers. As Francis and Mills put it (2012, p. 254) "To observe that schools reproduce social inequality is by no means novel".

Given all this, the conclusions that should be drawn are that there are substantial gaps in early attainments which are not, on average, reversed and indeed widen during the school years, and thus that the kinds of policies that need to be explored are not only early years interventions but (inter alia) the reduction of child poverty, less social stratification in access to schools, less setting and streaming, greater funding for schools serving disadvantaged areas, the development of pedagogies and curriculum which secure ongoing engagement of marginalised learners, and efforts to support families and build social and cultural capital. Jerrim and Vignoles' paper is a valuable contribution to the measurement of educational trajectories, but it is not one that should lead to policies being overthrown or even specific policy conclusions being drawn, and they do not claim this. So how can we ensure that policy debates are not conducted at the crude level that followed the production of this paper, but in a more sophisticated way in which multiple evidence sources are utilised?

One answer would be to be clearer about the contributions of different kinds of papers and about what is expected of academics in relation to policy issues. Researchers exploring methodological issues need not necessarily be cognisant of work in other disciplines nor should they be required to reproduce all these findings in their working papers. We need disciplinary specialists and focused enquiries on specific problems. But we should not then expect or encourage disciplinary specialists to pronounce upon policy issues, which demand a wide spectrum of knowledge across disciplines. This is exactly the situation that is developing in the UK, with 'impact' playing a substantial part in the regular university research quality assessment exercises that partly determine university funding and rankings. These imperatives create a real danger that academics working within their own disciplines will be encouraged to find meaningful policy implications to add on to their scholarly articles in order to further careers and promote university reputations, although most operate without detailed knowledge of the policies upon which they are commenting or the wider evidence base that should inform these. This is a combination which will not help evidence-based policy-making.

On the other hand, if academics are expected to comment on policy issues in a particular field, for example education or housing, we might reasonably expect them to know about work in other disciplines and to be able at least to comment on the concurrences, tensions and disagreements. This has multiple implications for academic training, traditions of publication, and career advancement, all of which privilege the individual discipline, and do not incentivise this wider knowledge to develop. In particular it has implications for the ways we organise academic knowledge exchange through conferences and publication. How should an economist working with cohort study data know how to interpret her findings alongside those of findings from pyscho-social studies, organisational sociology or psychometrics? How should an ethnographically trained sociologist know what to make of longitudinal data analyses? If we expect academics to operate as policy experts, then we must find ways of creating dialogue between researchers and a shared body of knowledge, and ways to synthesise and communicate findings across disciplines.

This case also questions the degree of effectiveness of communication of research findings to the civil service. It is noticeable that the last government's Social Mobility Strategy cited not a single one of the wider evidence sources referred to in this paper – preferring to rely on a narrow range of sources (almost exclusively quantitative and mainly economic). This may reflect the disciplinary knowledge or prejudices of the people concerned, perhaps a view (mistaken in my opinion) that the only valid kind of evidence is quantitative, and with a preference for studies that establish causality (Spicker, 2011). It may also be because other disciplines are less effective in communicating to policy audiences. Working papers and policy briefings are much less common in predominantly qualitative disciplines than they are in quantitative ones, and little academic priority is given to the synthesis of existing small-scale studies, published in books and journals, so that policy-makers can see what they collectively add up to.

I do not claim to have the answers to these questions, and this short paper does not offer room to cover them, even if I did. I hope merely to have raised some of issues and highlighted the need to address them collectively. The debate over the Feinstein graph has illuminated not only the complexities of measurement of educational trajectories but also the complexities of academic practice in a changing environment.

References

Allen, R. & West, A. (2011). Why Do Faith Secondary Schools Have Advantaged Intakes? The Relative Importance of Neighbourhood Characteristics, Social Background and Religious Identification amongst Parents. *British Educational Research Journal 37* (4): 691–712.
http://dx.doi.org/10.1080/01411926.2010.489145

Ball, S.J. (2003). *Class Strategies and the Education Market the Middle Classes and Social Advantage.* London: RoutledgeFalmer.
http://dx.doi.org/10.4324/9780203218952

Brantlinger, E. (2003). *Dividing Classes: How the Middle Class Negotiates and Rationalizes School Advantage*. Taylor & Francis.
http://dx.doi.org/10.4324/9780203465479

Bright, G. N. (2011). 'Off The Model': Resistant spaces, school disaffection and 'aspiration' in a former coal-mining community. *Children's Geographies 9* (1)
http://dx.doi.org/10.1080/14733285.2011.540440

Burgess, S., Briggs, A., McConnell, B., & Slater, H. (2006). *School Choice in England: Background Facts. Working Paper 06/159*. Bristol: Centre for Market and Public Organisation.

Cabinet Office and Deputy Prime Minister's Office (2011). *Opening Doors, Breaking Barriers: A Strategy for Social Mobility*. 26. London: HM Government. https://www.gov.uk/government/publications/opening-doors-breaking-barriers-a-strategy-for-social-mobility.

Centre for Social Justice. (2007). *Breakthrough Britain: Ending the Costs of Social Breakdown*. London: Centre for Social Justice.

Cooper, K. & Stewart, K. (2013). D*oes Money Affect Children's Outcomes: A Systematic Review*. York: Joseph Rowntree Foundation.

DFE& DWP. (2012). *Measuring Child Poverty: A Consultation on Better Measures of Child Poverty*. London: Department for Education.

Feinstein, L. (2003). Inequality in the Early Cognitive Development of British Children in the 1970 Cohort. *Economica 70*, 73–97.
http://dx.doi.org/10.1111/1468-0335.t01-1-00272

Field, F. (2010). *The Foundation Years: Preventing Poor Children Becoming Poor Adults: The Report of the Independent Review on Poverty and Life Chances.* London: HM Government.

Francis, B. & Mills, M. (2012). Schools as Damaging Organisations: Instigating a Dialogue Concerning Alternative Models of Schooling. *Pedagogy, Culture & Society 20* (2), 251-71.
http://dx.doi.org/10.1080/14681366.2012.688765

Gewirtz, S., Ball, S.J. & Bowe, R.. (1995). *Markets, Choice and Equity in Education*. Buckingham ; Philadelphia: Open University Press.

Gillborn, D., & Youdell, D. (20000. *Rationing Education: Policy, Practice, Reform and Equity*. Buckingham England: Philadelphia: Open University Press.

Goodman, A., & Gregg, P. (2010). *Poorer Children's Educational Attainment: how Important Are Attitudes and Behaviour?*. York: Joseph Rowntree Foundation.

Horgan, G. (2007). *The Impact of Poverty on Young Children's Experience of School.* York: Joseph Rowntree Foundation.

Ireson, J., & Hallam, S. (2001). *Ability Grouping in Education*. London: Sage.

Jerrim, J, & Vignoles, A. (2011a). The Use (and Misuse) of Statistics in Understanding Social Mobility: Regression to the Mean and the Cognitive Development of High Ability Children from Disadvantaged Homes. *DoQSS Working Paper 11-01*. Department of Quantitative Social Science - UCL Institute of Education, University College London. Retrieved from https://ideas.repec.org/p/qss/dqsswp/1101.html.

Jerrim, J, & Vignoles, A. (2011b, April 28). Response: Our Early-Years Research Does Not Contradict the Government. *The Guardian*. Retrieved from http://www.theguardian.com/commentisfree/2011/apr/28/social-mobility-early-years.

Jerrim, J. & Vignoles, A. (2013). Social mobility, regression to the mean and the cognitive development of high ability children from disadvantaged homes. *Journal of the Royal Statistical Society series A 176*(4), 887 – 906.
http://dx.doi.org/10.1111/j.1467-985X.2012.01072.x
Lareau, A. (2011). *Unequal Childhoods*. 2nd ed. Oakland: University of California Press. http://www.ucpress.edu/book.php?isbn=9780520271425.
Lingard, B. 2007. Pedagogies of Indifference. *International Journal of Inclusive Education 11* (3), 245–66.
http://dx.doi.org/10.1080/13603110701237498
Lupton, R. (2006). Schools in Disadvantaged Areas: Low Attainment and a Contextualised Policy Response. In. Lauder, H., Dillabough, J. & Halsey, A.H. (Eds) *Education, Globalization and Social Change.* 654–72. Oxford: Oxford University Press.
Lupton, R., & Hempel-Jorgensen, A. (2012). The Importance of Teaching: Pedagogical Constraints and Possibilities in Working-Class Schools. *Journal of Education Policy 27* (5), 601–20.
http://dx.doi.org/10.1080/02680939.2012.710016
Lupton, R., & Obolenskaya, P. (2013). Labour's Record on Education: Policy, Spending and Outcomes 1997-2010. *Social Policy in a Cold Climate Working Paper WP03.* London: CASE, LSE.
Lupton, R., & Thomson, S. (2015). The Coalition's Record on Schools: Policy, Spending and Outcomes 2010-2015. S*ocial Policy in a Cold Climate Working Paper 13*. London: Centre for Analysis of Social Exclusion,.
Marmot, M. G. (2010). *Fair Society, Healthy Lives: The Marmot Review ; Strategic Review of Health Inequalities in England Post-2010*. [London].
Read, D. (2011, April 11). *Regression to the Mean: Researcher warns that Government Strategy for Social Mobility misled by a statistical trap*. Warwick University press release. Retrieved from http://www2.warwick.ac.uk/newsandevents/pressreleases/extreme_statistics/detail/.
Reay, D. (2006). The Zombie Stalking English Schools: Social Class and Educational Inequality. *British Journal of Educational Studies 54* (3), 288-307.
http://dx.doi.org/10.1111/j.1467-8527.2006.00351.x
Reay, D., Crozier, G. & James, D. (2011). *White Middle Class Identities and Urban Schooling*. Basingstoke: Palgrave Macmillan. http://www.palgraveconnect.com/pc/doifinder/10.1057/9780230302501.
http://dx.doi.org/10.1057/9780230302501
Reay, D., & Lucey, H. (2003). The Limits of "Choice": Children and Inner City Schooling. *Sociology: The Journal of the British Sociological Association 37* (1),121–42.
http://dx.doi.org/10.1177/00380385030370011389
Ridge, T. (2002). *Childhood Poverty and Social Exclusion: From a Child's Perspective*. Bristol: Policy Press.
http://opus.bath.ac.uk/1294/.
Saunders, P. (2012). *Social Mobility Delusions*. London: Civitas.
Schoon, I. (2006). *Risk and Resilience: Adaptations in Changing Times*. Cambridge: Cambridge University Press.
http://dx.doi.org/10.1017/CBO9780511490132
Spicker, P. (2011). Generalisation and Phronesis: Rethinking the Methodology of Social Policy. *Journal of Social Policy 40* (01), 1-19.
http://dx.doi.org/10.1017/S0047279410000334
doi:10.1017/S0047279410000334.
Stewart, K. & Polina Obolenskaya, P. (2015). The Coalition's Record on the Under Fives: Policy, Spending and Outcomes 2010-2015. *Social Policy in a Cold Climate Working Paper WP12.* London: Centre for Analysis of Social Exclusion, LSE.
Stratton, A. (2011, April 14). David Davis Takes up Challenge to Prepare next Round of Tory Policies'. *The Guardian*. Retrieved from http://www.theguardian.com/politics/2011/apr/14/david-davis-challenge-tory-policies.
Thrupp, M. (1999). *Schools Making a Difference: Lets Be Realistic!*. First Edition. Buckingham England ; Philadelphia: Open University Press.

Endnotes

[i] These arguments are also prominent in Conservative documents on the causes of child poverty (DFE and DWP, 2012) , and in the thinking of the Conservative think-tank the Centre for Social Justice, set up by the Work and Pensions Minister Iain Duncan-Smith (Centre for Social Justice, 2007).

Referencing

The debate should be referenced as: Feinstein, L., Jerrim, J. & Vignoles, A., Goldstein, H. & French, R., Washbrook, E. & Lee, R. & Lupton, R. (2015). Social class differences in early cognitive development debate. *Longitudinal and Life Course Studies, 6,* 331-376

The individual contributions may be referenced as:
[Authors(s) names], [title of contribution], in Feinstein, L., Jerrim, J. & Vignoles, A., Goldstein, H. & French, R., Washbrook, E. & Lee, R. & Lupton, R. (2015). Social class differences in early cognitive development debate. *Longitudinal and Life Course Studies, 6, 331-376*

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 5,000 words**, with longer papers also accepted up to a maximum of 7,000 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 bibliography – 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 three 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.

INTRODUCTION

PAPERS

STUDY PROFILE

COMMENT AND DEBATE

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