VOLUME 7 | ISSUE 4 | **OCTOBER 2016**

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 current issue and preceding 2016 issues (Volume 7):

Annual Individual rate: **£20 (discounted rate £55 for 3 years)**.

Annual Library **rate: £200** to register any number of library readers.

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

SLLS members have automatic free access to the journal among many other benefits. For further information about SLLS and details of how to become a member, go to http://www.slls.org.uk/#!services/ch6q

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

Print on demand

An attractive printed version of each Issue of the journal, including all back numbers, is available at a price of £10 (plus postage and packaging). If you would like to purchase a full set or individual copies, visit the SLLS Bookshop at http://www.slls.org.uk/#!journal-bookshop/czkc Depending on distance, you will receive the order within two weeks of submitting your order.

Please note

The reselling of personal registration subscriptions and hard copies of the journal is prohibited.

INTRODUCTION

PAPERS

STUDY PROFILE

RESEARCH POLICY DIALOGUE

BOOK REVIEW

LLCS Journal can be accessed online at www.llcsjournal.org

Published by the Society for Longitudinal and Life Course Studies
info@slls.org.uk www.slls.org.uk

SLLS Society for Longitudinal and Life Course Studies

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 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 www.slls.org.uk/#!policy-group/c99m

www.slls.org.uk

Editorial

John Bynner **Executive Editor**

With the completion of the seventh volume and 25th issue of the journal this October, we are rounding off the year by introducing further innovative modes of communicating longitudinal research issues and findings. This is a good time to take stock.

Comment and Debate was launched in the 2015 July issue and featured in four of the subsequent five issues. Each new debate comprised an introductory discussion paper devoted to the topic of interest followed by a number of commentaries from experts in the field and concluding with the lead author's right of reply.

The series started with *Social class differences in early cognitive development* (Leon Feinstein et al.). This was followed by *Population sampling and longitudinal surveys* (Harvey Goldstein et al.); *Life Course and longevity risk* (David Blane et al.) and; *The case for Allostatic Load* (Cyrille Delpierre et al.).

Each debate has been successful in opening up discussion, while also supplying signposts to relevant literature – thus contributing to the resource value of the debate as a whole. More proposals for *Comment and Debate* submissions are warmly welcomed.

The other communications initiative – research to policy – has been more difficult to get off the ground. A key mission of the SLLS, as reflected in the Society's policy group and connections with the think tank Longview, is to strengthen the interactions and dialogue between longitudinal researchers and policy makers. This is with a view to facilitating mutual learning and more effective use of longitudinal research findings in the policy-making process.

Our first venture in this direction came from the SLLS Lausanne conference on early childhood effects of moving home in disadvantaged families in the US (*Fragile Families study*) and UK (*Millennium Cohort Study*). A special section devoted to the symposium was compiled by Mary Clare Lennon, William A.V. Clark and Heather Joshi, comprising interlinked and integrated papers.

The interdependency proved challenging because external blind peer review addresses each paper independently of others, entirely on its own scientific merits. This was contrary to the holistic approach favored for the section – and also tended to squeeze out the policy issues. Through much adjustment in redrafts, the review requirements were met. The final paper by Ruth Lupton then made good the policy gap with an analysis of the different features of the US and UK housing markets and discussion of the policy implications for families and children (LLCS Volume 7 Issue 3, July 2016).

We complete the year with an even more radical approach to the policy-research interface with a paper also derived from a Lausanne conference symposium – this time devoted to the research-policy relations regarding the massive six-cohort *German National Education Panel* run from the University of Bamberg

Symposium presenters included researchers and policy clients who, along with audience participants, were followed up with interviews by the symposium convener, Jutta Von Maurice. A well-rounded picture of the gaps and synergies between research, policy and general observer perspectives emerged, helping to identify key principles of effective communication.

Overall the different approaches contribute, in their different ways, to the richness of the journal's offering, supplying new reporting models, which we hope will be taken up and developed further by journal authors.

The challenge for a scientific journal is to maintain the rigorous standards of blind peer review on which scientific reputation is based while opening discussion to the wider range of contextualising policy perspectives and action principles that inform the policy process. Used effectively, the latter can play a crucial role shaping the direction the scientific program takes.

Many thanks to the conveners who gave their time to pioneering these break-through innovations and to all the contributors for the invaluable insights gained.

MEMBERSHIP OF THE SLLS

SLLS was established in 2009 and has over 350 members worldwide.
Interested in becoming a member? Then read on....

Membership Benefits

- Substantially reduced fees for attendance at the Society's annual conference
- Free, open access to read and publish in our peer reviewed journal, Longitudinal and Life Course Studies
- Access via log-in to the private members area of the SLLS website, where you will find an archive of all SLLS newsletters and be able to flick through copies of the LLCS Journal
- A regular newsletter of global news and events in longitudinal and life course research from our President
- A link with your regional Global Representative, who promotes SLLS in your region and organises local activities
- Collaborative contacts throughout the global longitudinal and life course research community
- An opportunity to be involved with the SLLS Policy Group, which promotes longitudinal research and communication between policy makers and researchers
- An opportunity to join the SLLS Cohort Network
- Access to theory and methods including our annual summer school
- The opportunity to make nominations, be nominated and vote in elections for the future SLLS President and Executive Committee

Fees
Full Annual Membership - £65/ €90/ US$115
Student Annual Membership - £26/ €35/ US$45
Corporate Annual Membership - £300/ €400/ US$540

Become a Member
To become a member or to renew your membership, visit the LLCS Journal site (www.llcsjournal.org) and follow the online instructions. Please note: You will need to log in/register first on the journal site to proceed with membership set up.

Longitudinal and Life Course Studies 2016 Volume 7 Issue 4 Pp 304 – 319 *ISSN 1757-9597*

Vulnerability, scar, or reciprocal risk? Temporal ordering of self-esteem and depressive symptoms over 25 years

Matthew D. Johnson University of Alberta, Canada
matt.johnson@ualberta.ca
Nancy L. Galambos University of Alberta, Canada
Harvey J. Krahn University of Alberta, Canada

(Received November 2015 Revised March 2016) http://dx.doi.org/10.14301/llcs.v7i4.394

Abstract

Three models have been proposed to explain the temporal interrelation between self-esteem and symptoms of depression: vulnerability (self-esteem predicts future depressive symptoms), scar (depressive symptoms predict future self-esteem), and reciprocal risk (self-esteem and depressive symptoms predict each other in the future). This study tested these three models over 25 years in a sample of high school seniors surveyed six times from age 18 to 43 (n = 978). Autoregressive cross-lagged modeling revealed that self-esteem and symptoms of depression prospectively predicted each other at every measurement occasion. Additionally, the cross-lagged association from self-esteem to symptoms of depression and the corresponding link from depressive symptoms to future self-esteem were equally strong. These results provide support for the reciprocal risk model.

Keywords

Depressive symptoms, longitudinal, mental health, reciprocal risk, scar model, self-esteem, vulnerability model

Introduction

Do symptoms of depression erode future perceptions of self-worth or does low self-esteem foretell later depressive symptoms? Or are they both risk factors for worsening mental health over time? Given recognition of depression as a major contributor to disease and disability around the world (World Federation for Mental Health, 2012) and the robust literature demonstrating significant associations between depression and low self-esteem (e.g. Galambos, Barker & Krahn, 2006; Lewinsohn, Hoberman & Rosenbaum, 1988), untangling the temporal ordering between these variables has emerged as a central focus for research. Three models have been proposed: vulnerability (self-esteem prospectively predicts depressive symptoms), scar (depressive symptoms prospectively predict self-esteem), and reciprocal risk (self-esteem and depressive symptoms prospectively predict each other). A burgeoning body of longitudinal research has evaluated these models (Orth & Robins, 2013), but questions remain. Not only has each model received some support, but much of the evidence is based on data gathered over the span of a few years, despite the recognized need for examination over longer time spans such as decades (Orth, Robins & Roberts, 2008). The current work contributes to this literature by testing the three models in a sample of Canadian high school seniors surveyed six times from age 18 to 43 years.

Vulnerability, scar, and reciprocal risk models

Theoretical perspectives on self-esteem and depression have long posited linkages between these constructs. Believing one's self to be of value and worth has been conceptualised as a fundamental human need (e.g. Rosenberg, 1979). Those with low self-esteem suffer a variety of psychological disturbances. Rosenberg and Owens (2001) highlight some characteristics of low self-esteem people that may render them particularly vulnerable to experiencing depression. Those with low self-esteem are hypervigilant for signs of rejection or inadequacy and are, in turn, more likely to internalise such information compared to those with high self-esteem. Unsurprisingly, people with low self-esteem lack self-confidence, ultimately reducing performance in a variety of domains and limiting their ability to enjoy success. Given these dynamics, those with low self-esteem are motivated to protect the little self-worth they possess and are more likely to avoid challenges that may facilitate the development of mastery and achievement of self-respect. A life driven by such tendencies may certainly lead to feelings of depression. Indeed, some theories of depression view negative self-beliefs as playing a causal role in the onset of clinical depression (e.g. Beck, 1967).

In line with these theoretical perspectives, the *vulnerability model* posits that low self-esteem serves as a risk factor for future depression. Poor perceptions of self-worth may increase the likelihood of depression via intrapersonal and interpersonal pathways. For example, self-esteem prospectively predicts the quality of social relations (Johnson & Galambos, 2014; Johnson, Galambos & Krahn, 2015; Marshall, Parker, Ciarrochi & Heaven, 2014) and low self-esteem can impair social ties (Murray, Bellavia, Rose & Griffin, 2003), which may ultimately put one at risk for symptoms of depression (Davila, Bradbury, Cohan & Tochluk, 1997). One intrapersonal mechanism has been found: the self-esteem to depression pathway was partially mediated by an increased tendency in those with low self-esteem to ruminate about negative facets of themselves, which predicted higher levels of depression (Kuster, Orth & Meier, 2012), as expected from prior theory (Rosenberg & Owens, 2001).

The *scar model,* in contrast, suggests that depression precedes self-esteem and other self-focused cognitions (e.g. Lewinsohn, Steinmetz, Larson & Franklin, 1981) by leaving 'scars' in one's life that erode self-esteem over time. The experience of depression may alter self-concept, but, again, depressive symptoms could operate through interpersonal and intrapersonal pathways. Symptoms of depression are intertwined with less adaptive close relationship functioning (Davila, Karney, Hall & Bradbury, 2003) decades in the future (Johnson, Galambos & Krahn, 2014) and self-esteem is shaped by interactions with close others (Mund, Finn, Hagemeyer, Zimmerman & Neyer, 2015). Likewise, the negative feelings inherent in depression may also shape the way individuals process and encode information, such that more negative interpretations of self-relevant information are retained, ultimately lowering self-esteem (Orth et al., 2008).

Of course, the vulnerability and scar models are not mutually exclusive, so a third reciprocal relations model has been proposed, whereby self-esteem and depressive symptoms exhibit bidirectional longitudinal associations over time (Orth & Robins, 2013). We label the third model a *reciprocal risk* model because it proposes that low self-esteem and higher levels of depressive symptoms both pose risks to future mental health (and each other) and that both scar and vulnerability processes could operate simultaneously. From our perspective, a reciprocal risk model assumes that neither self-esteem nor depressive symptoms hold temporal precedence for determining the course of mental health across the life span. Indeed, theories of depression and self-esteem acknowledge a poor sense of self-worth is closely intertwined with feelings of depression (for example, see Beck's (1983) conceptualisation of autonomous depression) and "each strengthens and reinforces the other" (Rosenberg & Owens, 2001, p. 410), supporting the possibility of a reciprocal risk dynamic between these constructs.

Turning to empirical work in this area, a substantial body of literature supports the vulnerability model. Drawing on a variety of large longitudinal datasets with diverse participants from around the world and employing sophisticated statistical methods to tease out effects, recent research found self-esteem precedes symptoms of depression from adolescence

to old age (Orth et al., 2008; Orth, Robins, Trzesniewski, Maes & Schmitt, 2009a; Orth, Robins & Widaman, 2012; Orth, Robins, Widaman & Conger, 2014; Rieger, Göllner, Trautwein & Roberts, 2016). The prospective link from self-esteem levels to future depression is robust, persisting when accounting for self-esteem instability and contingency on external factors (Sowislo, Orth & Meier, 2014), the occurrence of stressful events (Orth, Robins & Meier, 2009b), narcissism (Orth, Robins, Meier & Conger, 2016), and item overlap between measures of self-esteem and depression (Orth et al., 2009b).

Far less longitudinal evidence supports the scar model. Using data from 260 adults diagnosed with a severe mental illness who were participating in an intervention aimed at improving social functioning, baseline depression predicted lower levels of self-esteem four months later (mid-treatment), but there was no association between the constructs from mid-treatment to termination when the majority of gains in social functioning occurred (Shahar & Davidson, 2003). The nature of this sample calls into question the generalisability of the findings. In addition, research specifically investigating the scar model of depression in a community sample found no evidence that those with a history of depression differed from those who never experienced depression on a variety of cognitive measures, including self-esteem (Lewinsohn, et al., 1981).

Although much evidence aligns with the vulnerability model, support for the reciprocal risk model is impressive and relatively unacknowledged. A meta-analysis of longitudinal research with 77 samples found bidirectional linkages between self-esteem and depression (Sowislo & Orth, 2013): the path from self-esteem to future symptoms of depression was significant (β = -.16, p < .05) as was the path between depressive symptoms and future self-esteem (β = -.08, p < .05). Because, compared to the effect from depression to self-esteem, the effect was twice as large for self-esteem to depression, the authors concluded that the evidence best supported the vulnerability model rather than the reciprocal risk model. Additionally, the only study we could locate to test the interrelation between self-esteem and depressive symptoms over decades found reciprocal associations, with self-esteem and depressive symptoms at age 16 years predicting each other at age 45 years (Steiger, Fend & Allemand, 2015); again, the self-esteem to depression effect (β = -.21, p < .001) was significantly stronger than the depression to self-esteem path (β = -.12, p < .05), although both paths were statistically significant. We believe that the results from the meta-analysis and the Steiger et al. (2015) study could be re-interpreted as support for the reciprocal risk model. In our view, casting aside the reciprocal risk model in favour of the vulnerability model implies that little risk is posed by depression for decreasing self-esteem; dismissal of the significant effect of depressive symptoms on self-esteem is not warranted by the evidence.

Finally, some studies have not supported any of the three models. For example, a longitudinal, population-based study found no association between self-esteem at age 16/17 years and the odds of experiencing major depression from age 18 to 31 years, controlling for baseline variables, including psychological distress; the effect of depression on later self-esteem was not examined (Colman et al., 2014). A recent longitudinal study of depression and self-esteem in a sample of older adults also found no cross-lagged associations between the constructs (Gana, Bailly, Saada, Broc & Alaphilippe, 2015). Overall, then, it is difficult to draw conclusions concerning the generalisability of the vulnerability, scar, and reciprocal risk models.

A lifespan developmental perspective

We follow a lifespan developmental perspective, which assumes that (a) development across the life course is influenced by earlier behaviours; (b) different dimensions of development unfold together dynamically across time, making the patterning of interrelationships complex and important to examine; and (c) temporal sequencing of related phenomena is necessary to understand different pathways through life and to pinpoint the direction and determinants of change (e.g. Baltes, 1987; Baltes, Staudinger & Lindenberger, 1999; Lerner, Leonard, Fay & Issac, 2011). The lifespan perspective leads us to consider how self-esteem and depressive symptoms (two dimensions reflecting psychological well-being) are related to one another within and across time over as long a period of the life course as possible, and to explore possible shifts over time in the magnitude and direction of their interrelationship. The dynamic

nature of human development (Lerner et al., 2011) leads to the prediction that, at least across some points in the lifespan, depression and self-esteem will exhibit mutually influential reciprocal relations.

Furthermore, lifespan perspective proponents argue that studies with only two points in time cannot constitute an adequate test of the direction of effects and temporal sequencing of theoretically and empirically intertwined variables. Baltes and Nesselroade (1979, p. 34), for example, provided compelling arguments that "the study of developmental change generally ought to encompass more than two occasions of measurement….multiple occasions of measurement in longitudinal research permit the specification of change functions" (Baltes & Nesselroade, 1979, p. 34). In the current context, the lifespan perspective highlights the need for multiwave longitudinal research that captures and documents the potential complexities in the self-esteem-depression connection.

It is notable that many of the studies in Sowislo and Orth's (2013) meta-analysis included assessments at only two points in time, and only one followed participants longer than a few years (13 years; Schafer, Wickrama, & Keith, 1998). But even that study examined only the effect of self-esteem to future depression and did not consider possible bidirectional associations between the constructs. The only other study spanning decades included assessment at two time points (Steiger et al., 2015), making the current research the only attempt to test the interrelation among self-esteem and depression with varying time intervals in a single sample spanning a good portion of the lifespan. Furthermore, Sowislo and Orth's (2013) meta-analysis included fewer studies estimating the depression to self-esteem path (k = 42), relative to those estimating the self-esteem to depression path (k = 77). It may be premature to reject the reciprocal risk model when there are more studies testing only the vulnerability model, and it is imperative to consider how self-esteem and depressive symptoms are related to one another across varying intervals involving a larger swath of the lifespan than previously considered.

The present study

This study further investigates the interrelation between self-esteem and depressive symptoms as they unfold over decades using a sample of high school seniors surveyed into midlife. The intensive measurement during the transition to adulthood in our study (participants were surveyed five times from age 18 to 25 years) is similar to the time lags (ranging from one to three years between assessments) of most other studies in this area, but has an additional time of measurement at age 43 (an 18 year gap). Thus, we are able to examine whether the magnitude of the longitudinal associations between self-esteem and depression differs over varying time intervals, a question yet to be answered.

Given the stronger effect observed from self-esteem to depressive symptoms than vice versa in the studies documenting reciprocal relations (Sowislo & Orth, 2013; Steiger et al., 2015), we will empirically test the strength of the cross-lagged associations. We address our research questions using autoregressive cross-lagged modeling procedures. This is a particularly potent analytic approach, as the three most likely confounds are controlled in the model: they are prior levels of each construct; within-wave correlations between the constructs; and shared method variance due to reliance on self-report data. We also include baseline (assessed at age 18 years) control variables that have exhibited associations with self-esteem or depressive symptoms, namely: self-rated health (Orth et al., 2012); parental education (Galambos et al., 2006); final year grades (Steiger, Allemand, Robins & Fend, 2014); social support (Marshall et al., 2014); and disagreement with parents (Galambos et al., 2006; Johnson & Galambos, 2014).

Finally, we consider gender as a potential moderator of the association between self-esteem and depression. Across most of the lifespan and around the world, major depression and depressive symptoms are more prevalent in women than in men (Galambos, Leadbeater & Barker, 2004; Raj, Zitko, Jones, Lynch & Araya, 2013). Prior work with the data used in the present research found men had higher self-esteem and fewer symptoms of depression than women at age 18 years, but the gap narrowed by age 25 years (Galambos et al., 2006). Some evidence suggests that gender moderates associations between depression and its correlates. For example,

unemployment and media use are more strongly related to depressive symptoms in young men than in young women (Mossakowski, 2009; Primack, Swanier, Georgiopoulos, Land & Fine, 2009). In contrast, the depression-self-esteem association does not appear to differ for women and men (e.g. Orth et al., 2008; Steiger et al., 2015). Nevertheless, given the reported gender differences in levels of self-esteem and depression and their correlates, we examine the potential moderating influence of gender on the vulnerability, scar, and reciprocal risk models.

Method

Procedures

In spring of 1985, 983 Grade 12 students (age 18 years) completed questionnaires in class in the first wave of this 25-year longitudinal study. Six high schools representing working-class and middle-class neighborhoods in a large western Canadian city were sampled. The baseline sample (47% female, 80% born in Canada, 15% non-White, and 26% with at least one university-educated parent) was representative of western Canadian urban youth born in 1967 (the birth year of our sample) on race, immigration status, and parents' education (McVey & Kalbach, 1995). Follow-up questionnaires were mailed in 1986 (age 19 years; n = 665), 1987 (age 20 years; n = 547), 1989 (age 22 years; n = 503), and 1992 (age 25 years; n = 404) only to previous wave respondents. In each of these mail surveys, data collection involved a slightly modified version of Dillman's (1978) "total design method," with up to five contact attempts made for some participants. In 2010, telephone and web surveys targeted all baseline participants (age 43 years; n = 405; 41% response after 25 years). Half (51%) had participated in all previous waves. Participants were also surveyed in 1999, but data are not included here given the absence of full self-esteem and depressive symptoms measures in that wave. Table 1 contains descriptive information for the sample at age 43 years.

Table 1. Description of sample demographics at age 43 years (n = 405)

Variable	*%*
Female	52.6
White	88.8
Relationship Status	
Married	70.1
Cohabiting	6.2
Divorced or Separated	10.4
Never Married	13.1
Number of Children	
0	22.1
1	13.9
2	42.7
3	16.6
4+	4.7
Education	
High School or Less	15.6
Some Technical School or University	13.6
Technical School Diploma or Associate's Degree	33.8
Bachelor's Degree	24.7
Graduate Degree	12.3
Employed	92.1

Missing data

As expected, the amount of missing data was substantial over the course of this longitudinal study (see table 1 for a variable-level report of missing data at each measurement occasion). We chose to maximise power and generalisability by including every participant who reported self-esteem or depressive symptoms at any time point (n = 978) and by employing full-information maximum likelihood estimation (FIML) to handle missing values. FIML assumes the data are missing at random (MAR), meaning the pattern of missingness is not related to the constructs under investigation (depressive symptoms and self-esteem), but is predicted by other so-called auxiliary variables (see Allison, 2002). We first tested the MAR assumption by conducting a series of t-tests comparing continuing and non-continuing participants at each wave of measurement (ages 19, 20, 22, 25, and 43 years) on prior levels of depressive symptoms and self-esteem measured at baseline (age 18 years). Controlling for family-wise error, there were no significant differences between continuing and non-continuing participants at any wave of measurement. Next we computed a series of logistic regressions to identify auxiliary variables that predicted the pattern of missingness. Male status, parents without a university education, and immigrant status predicted greater odds of attrition. These analyses provide evidence that these data meet the MAR assumption and support the use of FIML to estimate missing values. As a robustness check, we computed our models with listwise deletion to ensure the results obtained using FIML aligned with the results obtained from those who did not drop out of the study. The pattern of results was consistent between the two approaches.

Measures

Means, standard deviations, and Cronbach's α coefficients are presented in table 2.

Table 2. Descriptive statistics for study variables

	Age in years at time of survey					
	1985	1986	1987	1989	1992	2010
	18	19	20	22	25	43
Depressive symptoms						
Mean	2.74	2.73	2.71	2.56	2.46	2.14
SD	.67	.64	.65	.62	.66	.66
Cronbach's α	.68	.68	.71	.69	.73	.72
% Missing	2.2	32.4	44.1	48.8	58.8	58.9
Self-Esteem						
Mean	3.79	3.78	3.86	3.95	3.98	4.32
SD	.70	.69	.69	.70	.72	.64
Cronbach's α	.73	.74	.77	.77	.78	.80
% Missing	0.70	32.4	44.1	48.5	58.8	58.8
Control variables						
Disagreements with parents						
Mean	2.11					
SD	.71					
Self-Reported health						
Mean	3.91					
SD	1.21					
Parent's education						
% No degree	72.30					
% 1 with degree	17.20					
% 2 with degree	10.40					
School grades						
Mean	3.30					
SD	.87					
Social support						
Mean	2.66					
SD	.64					

Note: Range for all variables is 1 – 5, except social support which ranges from 1 – 4 and parent's education ranges from 0 – 2.

Depressive symptoms

Four items from the Center for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977) measured depressive symptoms at ages 18, 19, 20, 22, 25, and 43 years. They asked how often in the past few months respondents 'felt depressed', 'felt lonely', 'talked less than usual' and 'felt like people were unfriendly'. Mean scores were calculated on a scale from 1 = *never* to 5 = *almost always*. Evidence for the validity of this shortened scale was previously provided (Galambos et al., 2006), but we also examined correlations between these items and nineteen of the original twenty CES-D items in the National Longitudinal Study of Adolescent Health (Harris et al., 2009). The shortened scale used in the present study correlated with the almost complete CES-D at .84, providing additional evidence that our measure serves as a reasonable proxy for the full CES-D.

Self-esteem

Five items from the Rosenberg Self-Esteem Scale (Rosenberg, 1989) measured self-esteem at ages 18, 19, 20, 22, 25, and 43 years. Participants reported their level of agreement with each statement: 'On the whole, I am satisfied with myself', 'At times I think I am no good at all (reverse coded)', 'I feel that I have a number of good qualities', 'I certainly feel useless at times (reverse coded)', and 'All in all, I am inclined to feel that I am a failure (reverse coded)'. Responses ranged from 1 = *strongly disagree* to 5 = *strongly agree* and mean scores were calculated. Using data from 1,629 undergraduate students, we examined validity by computing correlations between our five-item measure and the full 10-item Rosenberg Self-Esteem Scale. Our scale correlated with the full measure at .96, strongly supporting its validity.

Control variables

Age 18 years assessments of disagreements with parents, self-rated health, parental education, school grades, and social support were included as control variables. Disagreements with parents were assessed with the mean of 10 items that correspond to potential sources of conflict between parents and adolescents. Participants were asked, 'In the past few months, how often have you had disagreements with your parent or guardian about…'. Items included, 'School (including homework)', 'Your choice of friends', and 'Dating', and were rated on a scale from 1 = *Never* to 5 = *Almost always* (*M* = 2.11, *SD* = .71, *α* = .83). To measure self-rated health, we asked 'In the past few months, how healthy have you felt physically?' Responses were 1 = *Not very healthy,* 2 = *Somewhat healthy,* and 3 = *Very healthy* (*M* = 2.45, *SD* = .60). Parental education referred to the number of parents with a university degree: 0 = *No parents with a university degree,* 1 = *One parent with a university degree,* and 2 = *Two parents with a university degree* (*M* = .38, *SD* = .67). School grades were assessed by asking about participants' final grades, on average, in the past school year. Responses were 1 = *Mainly Fs (under 50%),* 2 = *Mainly Ds (50% to 59%),* 3 = *Mainly Cs (60% to 69%),* 4 = *Mainly Bs (70% to 79%),* and 5 = *Mainly As (80% or above; M* = 3.30, *SD* = .87*).* Social support was measured by asking: 'When you have problems, how much can you rely on each of the following people for help?' Respondents then provided a rating for the following people: 'Mother', 'Father', 'Other family members', 'Friends', 'Teachers' and 'Others'. Mean scores were computed; response categories were 0 = *No such person or not at all,* 1 = *A little,* 2 = *Somewhat,* 3 = *Very Much* (*M* = 2.66, *SD* = .64, *α* = .62).

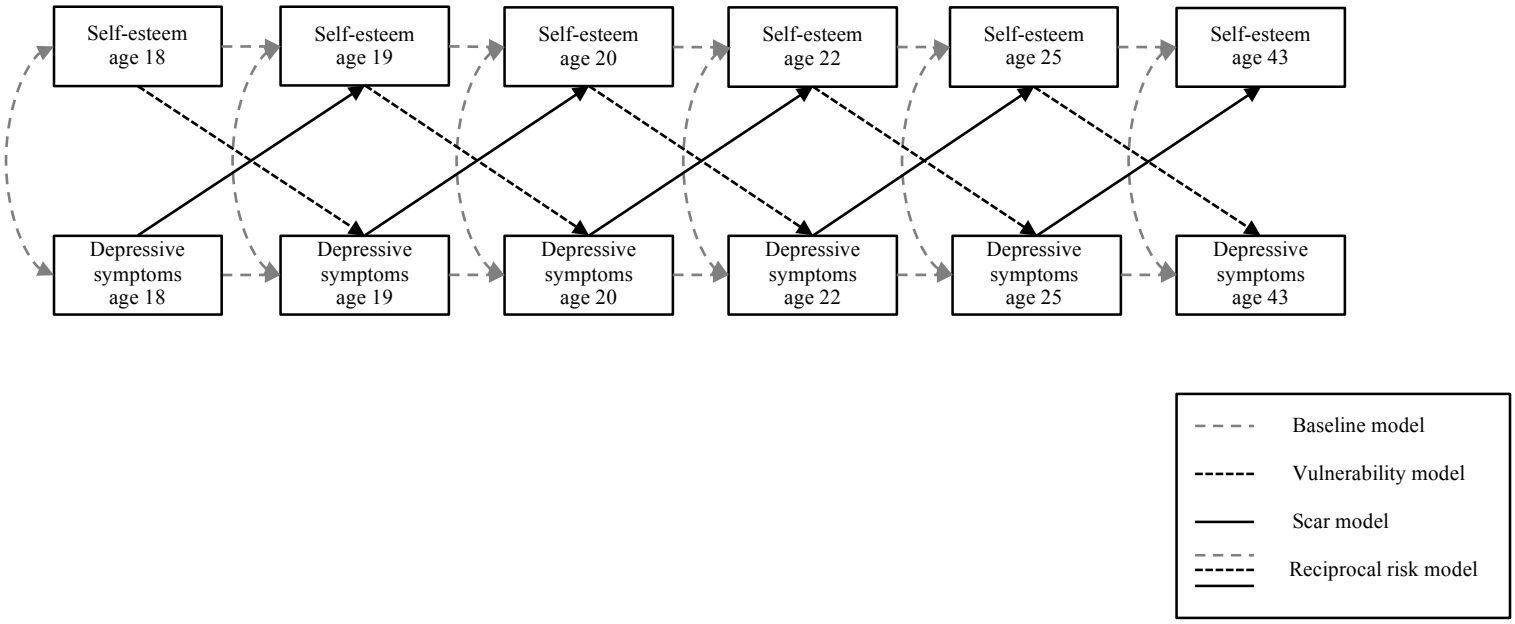
Analytic Plan

The interrelation between depressive symptoms and self-esteem was examined through the use of autoregressive cross-lagged modeling procedures in Mplus 7.11 (Muthén & Muthén, 2012). This approach allows for an empirical comparison of the vulnerability, scar, and reciprocal risk conceptualisations in the literature by computing a series of nested models to determine the most appropriate way to represent their interrelations over time. These nested models are depicted in figure 1. First, a baseline model was computed that included only the autoregressive paths and the within wave covariances between the variables. Two models tested the longitudinal unidirectional cross-lagged associations: the vulnerability model (self-esteem predicting depressive symptoms) and the scar model (depressive symptoms predicting self-esteem). A final model, corresponding to the reciprocal risk model, included the bidirectional cross-lagged linkages. The chi-square difference test (χ^2_{diff}) was used to compare nested models. Degrees of freedom are the same for each unidirectional model, so the one with the smaller chi-square value was considered the better fitting model.

During these initial model comparisons, model fit was poor, so we consulted the modification indices and residuals to diagnose model misspecification. We found that depressive symptoms and self-esteem needed to be regressed on all prior assessments of the construct (e.g., self-esteem at age 43 years was regressed on self-esteem at ages 25, 22, 20, 19, and 18 years), not just the immediately preceding assessment. Such an approach implies one's sense of self-worth and feelings of depression are the result of those constructs at all prior occasions of measurement in this study, not only the temporally adjacent time of measurement. Such a specification is conceptually sound because developmental status at any point represents the accumulation of experiences and continuities in behavior across earlier parts of the

life course, and additional autoregressive paths are often needed to reflect these connections (Little, 2013). With these additional regression paths included, the fit was admissible and we included these paths in all subsequent analyses.

Figure 1. Depiction of nested model comparisons testing the baseline, vulnerability, scar, and reciprocal risk models



After selecting the best fitting model, equality constraints were placed on corresponding cross-lagged paths and a chi-square difference test was computed to determine whether the strength of the associations from self-esteem to depressive symptoms were significantly stronger than those from depressive symptoms to self-esteem. Next, a multiple group analysis was used to test the potential moderating effect of gender on the associations between depressive symptoms and self-esteem. Equality constraints between men and women were placed on corresponding cross-lagged paths and a chi-square difference test was conducted. Finally, control variables were added and final models computed. Overall model fit was evaluated with the chi-square test (χ^2), the root mean square error of approximation (RMSEA), the comparative fit index (CFI), the Tucker-Lewis Index (TLI), and standardised root mean square residual (SRMR). A non-significant chi-square, values greater than .95 for CFI and TLI, and values smaller than .06 and .08 for RMSEA and SRMR suggest good model fit (Hu & Bentler, 1999).

Results

Autoregressive cross-lagged model comparisons

Model fit results from each autoregressive cross-lagged model (baseline, unidirectional, and bidirectional) are shown in table 2. According to chi-square difference tests, the bidirectional or reciprocal risk model (in bold) proved the best fit to the data, signifying depressive symptoms and self-esteem reciprocally influenced each other over time. Next, the application of the equality constraints to corresponding cross-lagged parameter estimates did not worsen the model fit, signifying self-esteem and depressive symptoms exhibited an equally strong effect on each other over time. We retained these more parsimonious models with the cross-lagged equality constraints for all subsequent analyses. With the most appropriate temporal ordering identified, we next tested the potential moderating influence of gender.

Table 3. Model fit indices for the longitudinal associations between depressive symptoms and self-esteem (n = 978)

Model	$\chi^2(df)$	RMSEA (90% CI)	CFI	TLI	SRMR	Model Comparison: $\chi^2_{diff}(df_{diff})$
Baseline	115.654 (30)	.054 (.044, .065)	.966	.927	.104	
Depress to self-esteem	63.188 (25)	.040 (.028, .052)	.985	.961	.061	Baseline: χ^2_{diff} (5) = 52.466, p < .001
Self-esteem to depress	67.265 (25)	.042 (.030, .054)	.983	.957	.053	Baseline: χ^2_{diff} (5) = 49.389, p < .001
Bidirectional	**32.333 (20)**	**.025 (.005, .040)**	**.995**	**.984**	**.027**	Depress to Self-esteem: χ^2_{diff} (5) = 30.855, p < .001

Note: Depress = depressive symptoms. Bolded models have the best fit. Baseline models only include autoregressive paths and within-wave covariances.

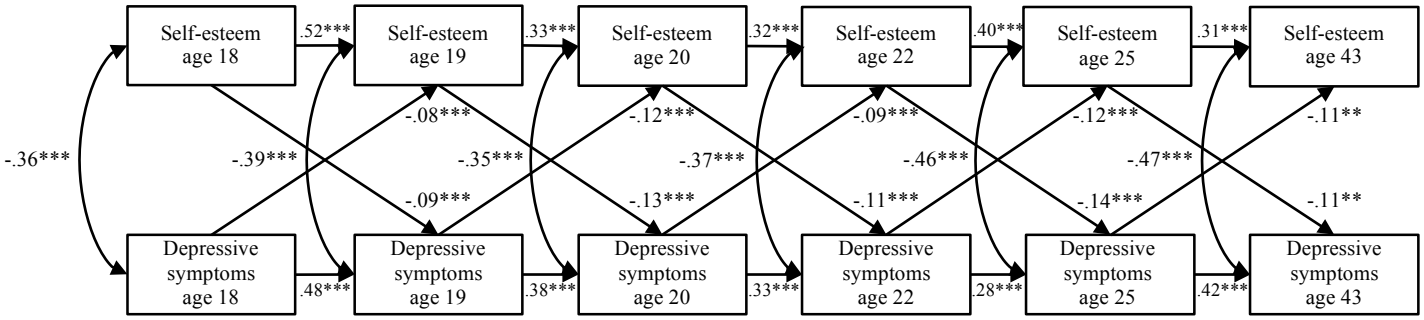
Test of moderation by gender

The multiple group (gender) autoregressive cross-lagged models fit the data well: $\chi^2(40)$ = 66.199; RMSEA = .037 (C.I. = .020, .052); CFI = .990; TLI = .967; SRMR = .035. The application of equality constraints to the cross-lagged paths did not significantly worsen model fit (χ^2_{diff} (10) = 15.959, p = .101), indicating that gender did not moderate the longitudinal associations between symptoms of depression and self-esteem. Thus, we combined men and women in a single model and proceeded to our final analysis, incorporating the control variables (including gender).

Final autoregressive cross-lagged model results

The final autoregressive cross-lagged model proved a good fit to the data (see figure 1). Self-esteem and symptoms of depression at age 18 years were regressed on all control variables (which were also assessed at that age). Being male, reporting better health, experiencing fewer conflicts with parents, earning better grades, and having access to more social support predicted fewer depressive symptoms and higher levels of self-esteem at age 18 years. Additionally, having more highly educated parents was linked with higher age 18 years self-esteem, but was not associated with depressive symptoms.

Figure 2. Standardised autoregressive cross-lagged models depicting bidirectional associations between self-esteem and depressive symptoms in relation to age (n = 978)



Note: Corresponding cross-lagged parameters are constrained equal. Age 18 years assessment of each construct was regressed on gender, disagreements with parents, self-rated health, parental education, school grades, and social support. Self-esteem and depressive symptoms were regressed on all prior assessments of each respective construct. Model fit indices: $\chi^2(85) = 99.839$; RMSEA = .013 (C.I. = .000, .023); CFI = .995; TLI = .992; SRMR = .036.

Turning to the main findings, self-esteem prospectively predicted fewer symptoms of depression at every measurement occasion and symptoms of depression foretold lower self-esteem at each measurement occasion, as well, in support of the reciprocal risk model. Additionally, the effect from self-esteem to depression was equally strong as that from depressive symptoms to self-esteem; some of the corresponding parameter estimates differ slightly in figure 1 because standardised coefficients are presented, but the unstandardised estimates are equivalent.

Discussion

Motivated by the need to examine the interrelation between self-esteem and depressive symptoms using longitudinal data with multiple assessments of both constructs across a large segment of the life span, we tested three conceptual models differing in the temporal ordering of self-esteem and depressive symptoms (vulnerability, scar, and reciprocal risk) in a sample of high school seniors surveyed over a quarter century. Counter to much prior research, we failed to find support for the vulnerability model. Our analyses demonstrated a longitudinal bidirectional relationship between self-esteem and depression, with no differences in the strength of the cross-lagged associations between these constructs. Indeed, as life unfolds from adolescence to midlife, sense of self-worth and depressive symptomatology mutually contribute to each other, reaffirming healthy functioning for some and representing a risk of future problems for others.

Aligning with a lifespan developmental perspective (e.g. Baltes, 1987; Lerner et al., 2011), these findings support our proposition of reciprocal risk, that neither self-esteem nor depressive symptoms hold primacy for determining the course of mental health across the lifespan. One's sense of self-worth and feelings of depression are deeply intertwined in human psychology (e.g. Beck, 1983) and it is likely that the processes by which they become so connected begin early in childhood and adolescence. That is, risks at multiple levels (genetic, personal/individual, parental/family) that are present in childhood are connected to adult depression via pathways that involve internalizing (e.g. low self-esteem, anxiety) and externalising (substance abuse, conduct disorders) problems (Clark, Rodgers, Caldwell, Power & Stansfeld, 2007; Kendler, Gardner & Prescott, 2002, 2006). As an example, lower family socioeconomic status (SES) in childhood is linked with

depression in middle adulthood, but indirectly through children's adjustment problems, young adult depression, and lower SES (e.g. income, occupational status) in adulthood (Elovainio et al., 2012; Gilman, Kawachi, Fitzmaurice & Buka, 2002). Such findings suggest that the associations between self-esteem and depression are set in motion early on by a multiplicity of interacting factors, and they may continue to reinforce each other throughout life. At the same time, diversity in experiences and across transitions at any point in the lifespan may change the course of development of either construct (see e.g. Lerner et al., 2011).

These findings also have direct implications for practitioners working with those suffering from symptoms of depression and/or struggling with a poor self-concept. The bidirectional cross-lagged associations observed in this study indicate that interventions to alleviate either low self-esteem or depressive symptoms may produce gains in both domains. Although either construct can be a target for interventions, treatment protocols for depression are well-articulated (e.g. Power, 2013) with demonstrated efficacy (Ekers, Richards & Gilbody, 2008), and self-esteem has been shown to improve during the course of treatment for depression (Wegener et al., 2015). In contrast, there is little literature on the clinical treatment of low self-esteem (e.g. Hall & Tarrier, 2003). Consequently, it may be more prudent for clinicians to focus on symptoms of depression until similarly effective clinical tools are developed focusing on low self-esteem. Prior work supporting the vulnerability model suggested clinicians boost self-esteem to alleviate symptoms of depression (e.g. Orth et al., 2009b), but this is easier said than done in the absence of a proven treatment protocol.

Two recent studies highlight the difficulty with trying to boost self-esteem, particularly among those with low self-esteem already. First, children who received more frequent praise about who they are as a person rather than in response to their behaviours tended to internalise failures and experience feelings of shame, especially among those with lower self-esteem (Brummelman et al., 2014). Second, giving inflated praise ('that is an incredibly beautiful drawing!') to children with low self-esteem resulted in those children avoiding challenges because they feared failure, thus depriving these children natural opportunities to overcome adversity and build their own sense of self-worth (Brummelman, Thomaes, de Castro, Overbeek & Bushman, 2014). These studies underscore the danger of well-meaning efforts that might perpetuate the very problem they were meant to help and point to a clear need for more investigation into best practices for the clinical treatment of low self-esteem.

Limitations and future directions

This study has limitations that must be considered. First, shortened scales were used to measure depressive symptoms and self-esteem. These shortened measures correlated highly with the full-length scales from which they were drawn (see measures section), however, and the mean levels of each construct exhibited the same pattern over time as documented in prior longitudinal research. Specifically, self-esteem increased (e.g. Orth et al., 2012) and symptoms of depression declined (Elovainio et al., 2012; Sutin et al., 2013) from age 18 to 43 years (see table 1), increasing confidence in the validity of our measures. Nevertheless, investigation of reciprocal risks employing full-length scales administered over decades would provide a useful addition to this literature. A second limitation is that many studies examining the interrelation between self-esteem and depressive symptoms analysed these constructs as latent variables to account for measurement error. We lacked the statistical power to model these data similarly. Once measurement error has been controlled through the use of structural equation modeling, associations between constructs tend to become stronger (e.g. Little, 2013), and so the reciprocal associations observed in this study might be magnified if we had been able to model latent variables. Third, our studies used data from one age cohort in one western Canadian city. Future research in different age groups over varying time spans and diverse geographical areas may indicate to what extent these findings are generalisable. Fourth, our data were best suited to examining the longitudinal interrelation between self-esteem and depressive symptoms, but prior work has documented within-time reciprocal associations between these constructs (e.g. Owens, 1994). While we accounted for within-time covariation between

the constructs, such an approach overlooks potentially meaningful directional relationships between self-esteem at a given point in time. Daily diary studies would be invaluable to shed further light on the interrelation between self-worth and feelings of depression as they unfold on a daily basis.

Conclusion

Untangling the temporal ordering of depressive symptoms and self-esteem has garnered intense interest, but this issue remains unsettled. Drawing on data gathered over 25 years, this study contributes additional evidence in support of a reciprocal risk conceptualisation: self-esteem and depression exhibit bidirectional longitudinal associations over a quarter of a century. Low self-esteem and symptoms of depression both represent risks for future mental health and interpersonal problems, and careful intervention able to alleviate distress in either may produce benefits that cascade into many life domains.

Acknowledgements

This research was funded by grants from the Social Sciences and Humanities Research Council of Canada (SSHRC), Alberta Advanced Education, and the University of Alberta. Data were collected by the Population Research Laboratory, University of Alberta.

References

Allison, P. D. (2002). *Missing data: quantitative applications in the social sciences.* Thousand Oaks, California: Sage Publications. http://dx.doi.org/10.4135/9781412985079

Baltes, P. B. (1987). Theoretical propositions of life-span developmental psychology: on the dynamics between growth and decline. *Developmental Psychology, 23*, 611-626. http://dx.doi.org/10.1037/0012-1649.23.5.611

Baltes, P. B., & Nesselroade, J. R. (1979). History and rationale of longitudinal research. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 1-39). New York: Academic Press.

Baltes, P. B., Staudinger, U. M., & Lindenberger, U. (1999). Lifespan psychology: theory and application to intellectual functioning. *Annual Review of Psychology, 50*, 471-507. http://dx.doi.org/10.1146/annurev.psych.50.1.471

Beck, A. T. (1967). *Depression: clinical, experimental, and theoretical aspects.* New York: Harper & Row.

Beck, A. T. (1983). Cognitive therapy of depression: new perspectives. In P. J. Clayton & J. E. Barrett (Eds.), *Treatment of depression: old controversies and new approaches* (pp. 265 – 290). New York: Raven.

Brummelman, E., Thomaes, S., de Castro, B. O., Overbeek, G., & Bushman, B. J. (2014). "That's not just beautiful – That's incredibly beautiful!": the adverse impact of inflated praise on children with low self-esteem. *Psychological Science, 25,* 728 – 735. http://dx.doi.org/10.1177/0956797613514251

Brummelman, E., Thomaes, S., Overbeek, G., de Castro, B. O., van den Hout, M. A., & Bushman, B. J. (2014). On feeding those hungry for praise: person praise backfires in children with low self-esteem. *Journal of Experimental Psychology: General, 143,* 9-14. http://dx.doi.org/10.1037/a0031917

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

Colman, I., Zeng, Y., McMartin, S. E., Naicker, K., Ataullahjan, A., Weeks, M., Senthilselvan, A., & Galambos, N. L. (2014). Protective factors against depression during the transition from adolescence to adulthood: findings from a national Canadian cohort. *Preventive Medicine, 65*, 28-32. http://dx.doi.org/10.1016/j.ypmed.2014.04.008

Davila, J., Bradbury, T. N., Cohan, C. L., & Tochluk, S. (1997). Marital functioning and depressive symptoms:

evidence of a stress generation model. *Journal of Personality and Social Psychology, 73,* 849-861. http://dx.doi.org/10.1037/0022-3514.73.4.849

Davila, J., Karney, B. R., Hall, T. W., & Bradbury, T. N. (2003). Depressive symptoms and marital satisfaction: within-subject associations and the moderating effects of gender and neuroticism. *Journal of Family Psychology, 17,* 557-570. http://dx.doi.org/10.1037/0893-3200.17.4.557

Dillman, D. A. (1978). *Mail and telephone surveys: the total design method.* New York: Wiley.

Ekers, D., Richards, D., & Gilbody, S. (2008). A meta-analysis of randomized trials of behavioural treatment of depression. *Psychological Medicine, 38,* 611-623. http://dx.doi.org/10.1017/S0033291707001614

Elovainio, M., Pulkki-Råback, I., Jokela, M., Kivimäai, M., Hintsanen, M., Hintsa, T., Viikari, J., Raitakari, O. T., & Keltikangas-Järvinen, L. (2012). Socioeconomic status and the development of depressive symptoms from childhood to adulthood: a longitudinal analysis across 27 years of follow-up in the Young Finns study. *Social Science & Medicine, 74*, 923-929. http://dx.doi.org/10.1016/j.socscimed.2011.12.017

Galambos, N. L., Barker, E. T., & Krahn, H. J. (2006). Depression, self-esteem, and anger in emerging adulthood: seven-year trajectories. *Developmental Psychology, 42,* 350-365. http://dx.doi.org/10.1037/0012-1649.42.2.350

Galambos, N. L., Leadbeater, B. J., & Barker, E. T. (2004). Gender differences in and risk factors for depression in adolescence: a 4-year longitudinal study. *International Journal of Behavioral Development, 28,* 16-25. http://dx.doi.org/10.1080/01650250344000235

Gana, K., Bailly, N., Saada, Y., Broc, G., & Alaphilippe, D. (2015). Relationship between self-esteem and depressive mood in old age: results from a six-year longitudinal study. *Personality and Individual Differences, 82,* 169-174. http://dx.doi.org/10.1016/j.paid.2015.03.021

Gilman, S. E., Kawachi, I., Fitzmaurice, G. M., & Buka, S. L. (2002). Socioeconomic status in childhood and the lifetime risk of major depression. *International Journal of Epidemiology, 31*, 359-367. http://dx.doi.org/10.1093/ije/31.2.359

Hall, P. L., & Tarrier, N. (2003). The cognitive-behavioural treatment of low self-esteem in psychotic patients: a pilot study. *Behaviour Research and Therapy, 41,* 317-332. http://dx.doi.org/10.1016/S0005-7967(02)00013-X

Harris, K. M., Halpern, C. T., Whitsel, E., Hussey, J., Tabor, J., Entzel, P., & Udry, J. R. (2009). *The National Longitudinal Study of Adolescent Health: Research Design.* Retrieved from http://www.cpc.unc.edu/projects/addhealth/design

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

Johnson, M. D., & Galambos, N. L. (2014). Paths to intimate relationship quality from parent-adolescent relations and mental health. *Journal of Marriage and Family, 76,* 145-160. http://dx.doi.org/10.1111/jomf.12074

Johnson, M. D., Galambos, N. L., & Krahn, H. J. (2014). Depression and anger across 25 years: changing vulnerabilities in the VSA model. *Journal of Family Psychology, 28,* 225-235. http://dx.doi.org/10.1037/a0036087

Johnson, M. D., Galambos, N. L., & Krahn, H. J. (2015). Self-esteem trajectories across 25 years and midlife intimate relations. *Personal Relationships, 22,* 635-646. http://dx.doi.org/10.1111/pere.12100

Kendler, K. S., Gardner, C. O., & Prescott, C. A. (2002). Toward a comprehensive developmental model for major depression in women. *American Journal of Psychiatry, 159*, 1133-1145. http://dx.doi.org/10.1176/appi.ajp.159.7.1133

Kendler, K. S., Gardner, C. O., & Prescott, C. A. (2006). Toward a comprehensive developmental model for major depression in men. *American Journal of Psychiatry, 163*, 115-124. http://dx.doi.org/10.1176/appi.ajp.163.1.115

Kuster, F., Orth, U., & Meier, L. L. (2012). Rumination mediates the prospective effect of low self-esteem on depression: a five-wave longitudinal study. *Personality and Social Psychology Bulletin, 38,* 747-759.

http://dx.doi.org/10.1177/0146167212437250

Lerner, R. M., Leonard, K., Fay, K., & Issac, S. S. (2011). Continuity and discontinuity in development across the life span: a developmental systems perspective. In K. L. Fingerman, T. Antonucci, & J. S. Berg (Eds.), *Handbook of lifespan psychology* (pp. 141-160). New York, NY: Springer.

Lewinsohn, P. M., Hoberman, H. M., & Rosenbaum, M. (1988). A prospective study of risk factors for unipolar depression. *Journal of Abnormal Psychology, 97,* 251-264. http://dx.doi.org/10.1037/0021-843X.97.3.251

Lewinsohn, P. M., Steinmetz, J. L., Larson, D. W., & Franklin, J. (1981). Depression-related cognitions: antecedent or consequence? *Journal of Abnormal Psychology, 90,* 213-219. http://dx.doi.org/10.1037/0021-843X.90.3.213

Little, T. D. (2013). *Longitudinal structural equation modeling.* New York: The Guilford Press.

Marshall, S. L., Parker, P. D., Ciarrochi, J., & Heaven, P. C. L. (2014). Is self-esteem a cause or consequence of social support? A 4-year longitudinal study. *Child Development, 85,* 1275-1291. http://dx.doi.org/10.1111/cdev.12176

McVey, W., Jr., & Kalbach, W. (1995). *Canadian population*. Scarborough, Ontario: Nelson Canada.

Mossakowski, K. N. (2009). The influence of past unemployment duration on symptoms of depression among young women and men in the United States. *American Journal of Public Health, 99*, 1826-1832. http://dx.doi.org/10.2105/AJPH.2008.152561

Mund, M., Finn, C., Hagemeyer, B., Zimmerman, J., & Neyer, F. J. (2015). The dynamics of self-esteem in partner relationships. *European Journal of Personality,* http://dx.doi.org/10.1002/per.1984

Murray, S. L., Bellavia, G. M., Rose, P., & Griffin, D. W. (2003). Once hurt, twice hurtful: how perceived regard regulates daily marital interactions. *Journal of Personality and Social Psychology, 84,* 126-147. http://dx.doi.org/10.1037/0022-3514.84.1.126

Muthén, L.K. and Muthén, B.O. (2012). *Mplus User's Guide. Seventh Edition*. Los Angeles, California: Muthén & Muthén.

Orth, U., & Robins, R. W. (2013). Understanding the link between low self-esteem and depression. *Current Directions in Psychological Science, 22,* 455 – 460. http://dx.doi.org/10.1177/0963721413492763

Orth, U., Robins, R. W., & Meier, L. L. (2009b). Disentangling the effects of low self-esteem and stressful events on depression: findings from three longitudinal studies. *Journal of Personality and Social Psychology, 97,* 307-321. http://dx.doi.org/10.1037/a0015645

Orth, U., Robins, R. W., Meier, L. L., & Conger, R. D. (2016). Refining the vulnerability model of low self-esteem and depression: disentangling the effects of genuine self-esteem and narcissism. *Journal of Personality and Social Psychology, 110,* 133-149. http://dx.doi.org/10.1037/pspp0000038

Orth, U., Robins, R. W., & Roberts, B. W. (2008). Low self-esteem prospectively predicts depression in adolescence and young adulthood. *Journal of Personality and Social Psychology, 95,* 695-708. http://dx.doi.org/10.1037/0022-3514.95.3.695

Orth, U., Robins, R. W., Trzesniewski, K. H., Maes, J., & Schmitt, M. (2009a). Low self-esteem is a risk factor for depressive symptoms from young adulthood to old age. *Journal of Abnormal Psychology, 118,* 472-478. http://dx.doi.org/10.1037/a0015922

Orth, U., Robins, R. W., & Widaman, K. F. (2012). Life-span development of self-esteem and its effects on important life outcomes. *Journal of Personality and Social Psychology, 102,* 1271-1288. http://dx.doi.org/10.1037/a0025558

Orth, U., Robins, R. W., Widaman, K. F., & Conger, R. D. (2014). Is low self-esteem a risk factor for depression? Findings from a longitudinal study of Mexican-origin youth. *Developmental Psychology, 50,* 622-633. http://dx.doi.org/10.1037/a0033817

Owens, T. J. (1994). Two dimensions of self-esteem: reciprocal effects of positive self-worth and self-deprecation on adolescent problems. *American Sociological Review, 59,* 391-407. http://dx.doi.org/10.2307/2095940

Power, M. (2013). CBT for depression. In M. Power (Ed.), *The Wiley-Blackwell handbook of mood disorders* (2nd

ed., pp. 173 – 192). Oxford: John Wiley & Sons, Ltd. http://dx.doi.org/10.1002/9781118316153.ch7

Primack, B. A., Swanier, B., Georgiopoulos, A. M., Land, S. R., & Fine, M. J. (2009). Association between media use in adolescence and depression in young adulthood: a longitudinal study. *Archives of General Psychiatry, 66*, 181-188. http://dx.doi.org/10.1001/archgenpsychiatry.2008.532

Radloff, L. S. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement, 1,* 385-401. http://dx.doi.org/10.1177/014662167700100306

Raj, D., Zitko, P., Jones, K., Lynch, J., & Araya, R. (2013). Country- and individual-level socioeconomic determinants of depression: multilevel cross-national comparison. *The British Journal of Psychiatry, 202*, 195-203. http://dx.doi.org/10.1192/bjp.bp.112.112482

Rieger, S., Göllner, R., Trautwein, U., & Roberts, B. W. (2016). Low self-esteem prospectively predicts depression in the transition to young adulthood: a replication of Orth, Robins, & Roberts (2008). *Journal of Personality and Social Psychology, 110,* e16-e22. http://dx.doi.org/10.1037/pspp0000037

Rosenberg, M. (1979). *Conceiving the self.* New York: Basic Books.

Rosenberg, M. (1989). *Society and the adolescent self-image* (Revised ed.). Middletown, Connecticut: Wesleyan University Press.

Rosenberg, M., & Owens, T. J. (2001). Low self-esteem people: a collective portrait. In T. J. Owens, S. Stryker & N. Goodman (Eds.), *Extending self-esteem theory and research: sociological and psychological currents* (400-436). New York: Cambridge University Press. http://dx.doi.org/10.1017/cbo9780511527739.018

Schafer, R. B., Wickrama, K. A. S., & Keith, P. M. (1998). Stress in marital interaction and change in depression: a longitudinal analysis. *Journal of Family Issues, 19,* 578-594. http://dx.doi.org/10.1177/019251398019005005

Shahar, G., & Davidson, L. (2003). Depressive symptoms erode self-esteem in severe mental illness: a three-wave, cross-lagged study. *Journal of Consulting and Clinical Psychology, 71,* 890-900. http://dx.doi.org/10.1037/0022-006X.71.5.890

Sowislo, J. F., & Orth, U. (2013). Does low self-esteem predict depression and anxiety? A meta-analysis of longitudinal studies. *Psychological Bulletin, 139,* 213-240. http://dx.doi.org/10.1037/a0028931

Sowislo, J. F., Orth, U., & Meier, L. L. (2014). What constitutes vulnerable self-esteem? Comparing the prospective effects of low, unstable, and contingent self-esteem on depressive symptoms. *Journal of Abnormal Psychology, 123,* 737-753. http://dx.doi.org/10.1037/a0037770

Steiger, A. E., Allemand, M., Robins, R. W., & Fend, H. A. (2014). Low and decreasing self-esteem during adolescence predict adult depression two decades later. *Journal of Personality and Social Psychology, 106,* 325-338. http://dx.doi.org/10.1037/a0035133

Steiger, A. E., Fend, H. A., & Allemand, M. (2015). Testing the vulnerability and scar models of self-esteem and depressive symptoms http://dx.doi.org/10.1037/a0038478

Sutin, A. R., Terracciano, A., Milaneschi, Y., An, Y., Ferrucci, L., & Zonderman, A. B. (2013). The trajectory of depressive symptoms across the adult life span. *Journal of the American Medical Association Psychiatry, 70*, 803-811 http://dx.doi.org/10.1001/jamapsychiatry.2013.193

Wegener, I., Geiser, F., Alfter, S., Mierke, J., Imbierowicz, K., Kleiman, A., Koch, A. S., Conrad, R. (2015). Changes of explicitly and implicitly measured self-esteem in the treatment of major depression: evidence for implicit self-esteem compensation. *Comprehensive Psychiatry, 58,* 57-67. http://dx.doi.org/10.1016/j.comppsych.2014.12.001

World Federation for Mental Health. (2012). *Depression: a global crisis.* Retrieved from: http://www.who.int/mental_health/management/depression/wfmh_paper_depression_wmhd_2012.pdf?ua=1

Longitudinal and Life Course Studies 2016 Volume 7 Issue 4 Pp 320 – 346 *ISSN 1757-9597*

Order or chaos? Understanding career mobility using sequence analysis and information-theoretic methods

Marc A. Scott New York University, US
marc.scott@nyu.edu
Matthew Zeidenberg Abt Associates, US

(Received June 2015 Revised March 2016) http://dx.doi.org/10.14301/llcs.v7i4.358

Abstract

We examine the careers of a nationally representative U.S. cohort of young adults using sequence analysis and information-theoretic techniques to describe these careers' structure and how this structure might inform differences in wage mobility. We operationalise the career as a sequence of industry-occupation pairs observed quarterly. We investigate how the content of these pairs and their organisation over time relate to future mobility. We perform the analysis across three different mobility groups, one of which is characterised by persistent low-wage work. Contrary to what one might expect, low-wage work is not typified by a lack of structure, even in many of the careers in which the worker is weakly attached to the labour market. Using clustering techniques customised to this problem, we build a typology of careers within three groups of workers defined by their wage mobility. We find significant variation within, as well as similarity across the three groups, enhancing our understanding of careers with different levels of mobility.

Keywords

Sequence analysis; information theory; optimal matching; categorical data clustering; intragenerational mobility; national longitudinal survey.

Introduction

The last quarter of the 20th century witnessed a tremendous growth in the share of the U.S. workforce that has been unable to increase their wages over a substantial portion of their work lives (Andersson, Holzer & Lane, 2004; Bernhardt, Morris, Handcock & Scott, 2001; Boushey, 2005; Strawn & Martinson, 2000). In absolute terms, the proportion stuck in low-wage careers in their late 30s is large, nearly 30% of the workforce.[i] Researchers have explored the theory that early careers are characterised by a somewhat chaotic period of job matching, followed by wage growth that results from a successful match and increased job-specific experience. While wage mobility is by definition limited for workers who are stuck in low-wage jobs, some researchers have examined whether job shopping (and job matching) or increased experience are central to wage growth in this job sector (French, Mazumder & Taber, 2007; Gardecki & Neumark, 1998; Topel, 1993). Other researchers have examined trends in job attachment itself, which has been in decline (Bernhardt, Morris, Handcock & Scott, 1999; Kalleberg, 2009; Neumark, Polsky & Hansen, 2000), and the increased lack of control over work schedules (Clawson & Gerstel, 2014). Some of these stylized characterisations of low-wage careers emphasise their chaotic nature and contend that they are not careers at all, as that term

is commonly used. These characterisations stand somewhat in contrast to case-studies of low-wage work, such as Appelbaum, Bernhardt and Murnane (2006) and earlier work by Becker and Carper (1956), which suggest that low-wage careers are often found within well-defined sectors of the labour market and individuals identify their career with a particular sector.

This apparent lack of consensus in the literature characterising low-wage careers motivates our detailed examination and measurement of the extent of *order* or *disorder* in careers in three different segments of the labour market, with a focus on low-wage, or immobile careers. Specifically, we use a life course analysis approach to characterise low-wage careers and determine which, if any, features distinguish them from their more mobile counterparts. The central research questions we address are:

- Low-wage careers are sometimes characterised as chaotic. We evaluate this by operationalising and evaluating several related but distinct measures of structure on career paths.
- To what extent is the characterisation due to an inappropriate unit of analysis? What do we learn from individual versus aggregated trajectories?
- How do we effectively incorporate time into any assessment of sets of careers?

To address these questions, we utilise techniques developed in life course studies, incorporating them into a framework for comparing and contrasting longitudinal nominal sequences by employing information-theoretic measures, some of which are well established in the life course analysis literature, while others are introduced or customised by us.

In contrast to much of the previous work on low-wage labour (Dickens & Lang, 1985; Mincer, 1958; Mishel, Bernstein & Allegretto, 2007; Osterman, 1975), we examine the individuals' career trajectories as a whole. Specifically, we examine the sequence of industry and occupation pairs that form a career over a portion of the life course (we use a coding scheme that establishes several hundred unique pairs). We measure the coherence or similarity of these pairs to one another across time using information-theoretic and sequence-analytic measures. As suggested by prior research, we examine the role of attachment to the labour force, which we will show is much more *variable* for workers who experience low-wage careers than for other workers. We show that low-wage careers fall into two distinct categories, and we use clustering techniques to tease these out further. Our findings suggest that commonly held beliefs and generalisations about low-wage careers are too simplistic, ignoring substantial groups whose trajectories exhibit clear structure. Moreover, the generalisations obscure or neglect many of these careers' important features, which are revealed through the more nuanced approaches that we develop.

The organisation of this paper is as follows. In the second section, we review the literature on sequence analysis as it relates to the measurement of structure within sequences and between groups of sequences. In the third section, we introduce new measures, which complement or modify existing ones, and establish a time-lagged approach to evaluating structural differences in sequences. We discuss an approach to clustering within mobility groups to establish content-similar sequences and reveal heterogeneity. In the fourth section, we describe the data we will use and discuss the partitioning of subjects into mobility groups. The fifth section applies our techniques to the sequences of labour market entry, movement and exit, and then reassesses our sequence measures in the context of career clusters. The final section concludes with a discussion of the substantive research questions addressed and makes suggestions for future work.

Sequences and life course analysis

In sequence analysis, the outcome is the whole set of nominal states viewed as a discrete time series (Abbott, 1995; Blanchard, Bühlmann & Gauthier, 2014; Cornwell, 2015). Roughly speaking, there are three interrelated methods associated with sequences: visualisation, descriptive summary, and modelling, which we now introduce.

Visualisation of sequences is challenging, as the purpose of a good graphic is to convey just the right amount of information. Each sequence is an ordered list of tokens (nominal states), forming something akin to a sentence. In the usual situation in which

there is no natural ordering to the nominal states, sequences cannot be plotted in the coordinate plane. A workaround is to aggregate the tokens across subjects at each time point, forming a moving window of token or state distributions, and these can be colour-coded to reflect content difference (see, for example, Fasang & Liao, 2013; Piccarreta, 2012). The R library TraMineR (Gabadinho, Ritschard, Müller & Studer, 2011; R Core Team, 2014), which we use extensively in our analysis, has excellent functionality for this purpose and allows for group-specific summaries. While informative, the typical state distribution plot masks one of the key features of a sequence: the dependence between states at different times. One way this could be visualised borrows techniques from network analysis, in which the states are nodes on a network, and edges are movements between states (see Bison, 2014; Hoff, Raftery & Handcock, 2002). A slightly different approach clusters the sequences to establish subgroups that are homogeneous with respect to content and dependence structure. Following this, the medoid or some other summary of each cluster may be compared and contrasted across clusters.

Sequence summaries are an active research topic, so we highlight only a subset of approaches and discuss the need for some extensions or modifications for our research problem. One summary that receives substantial attention is entropy, evaluated cross-sectionally for subgroups (see Billari, 2001; Pierce, 1980). We will make the definition formal in the next section, but the basic idea is that the more uniform the distribution of tokens for a given group or time, the more difficult it is to succinctly describe or make predictions of temporally nearby content. The more concentrated the distribution, the easier it is to predict or describe. Predictive capacity can be operationalised as within or between sequences, and the distinction is important. When we are characterising a mobility group, or a subgroup (cluster) within it, we are referring to the entropy of the ensemble of job types for that group or subgroup. In our example, if low-wage careers consisted entirely of the same industry and occupation pair (token IxO, hereafter) at any given time, they would be low entropy. Entropy could be a measure of structure, but as noted cross-sectional information ignores relationships *within* a sequence; pooled sequence content could be predictable as a whole whereas it could be highly variable within each sequence. To characterise the volatility within a sequence we turn to the notion of turbulence (Elzinga & Liefbroer, 2007; Elzinga, 2010). Turbulence is based on the number of distinct subsequences that may be derived from a given sequence, rescaled to reflect the potential maximal variation in token spell length within it. Intuitively, sequences that do not change tokens frequently tend to be low turbulence while those that have a variety of token spell lengths are higher turbulence. After measuring turbulence for each sequence, we can aggregate the measurements by subgroup. Prior research suggests that the timing of token progression is crucial to career development, thus we measure turbulence on subsequences reflecting different periods (e.g., early, middle and late career).

In the next section, we introduce one additional measure to directly assess the extent to which tokens co-vary within subgroups or sequences. It is a natural analogue to entropy known as mutual information (MI). An important distinguishing characteristic of MI is that it relates one distribution to another. We interpret this in the context of careers as a measure of aggregate similarity, predictability or dependence of one portion of an ensemble of sequences (e.g., all clerical workers) on a latter portion.

Statistical models for sequences are a somewhat controversial topic (for example, see Abbott & Tsay, 2000 and Wu, 2000). Important to this debate is whether sequences derived from the life course are to be viewed through the lens of event history modelling (i.e., time to event models) or processed somehow as a whole sequence, intact. When the state space is large, as it is in our labour market application, models must make simplifying assumptions in order to be tractable, and these assumptions are often questioned. However, event history models can sometimes be customised to address specific research questions. For example, Scott (2011) used a continuous time Markov model to handle a large number of states (tokens) via a Bayesian (random effects) approach, comparing subgroup behaviour and allowing for time heterogeneity. Non-model-based approaches often

use techniques of optimal matching (Abbott, 1995; Abbott & Tsay, 2000; Durbin, Eddy, Krogh & Mitchison, 1998) to compute the distance between pairs of sequences, and then use the distances to construct clusters via hierarchical or agglomerative techniques (Everitt, Landau, Leese & Stahl, 2001; Kauffman & Rousseeuw, 1990). One challenge arises in the assignment of costs to the primitive operations of insertion, deletion and substitution in the OMA algorithm (see Vingron & Waterman, 1994, for example). Recent developments in sequence analysis (Elzinga & Studer, 2015; Halpin, 2014; Studer & Ritschard, 2015) have provided methods that address many of these concerns. Of course, clustering has inherent challenges that remain in the sequence analysis domain (see Hennig & Liao, 2013, for some discussion). We will utilise techniques of sequence analysis, OMA and clustering throughout our evaluation of mobility. As we introduce and use each technique, we will justify the methodological choices in the context of our problem domain.

Controlling heterogeneity

Most of the above measures involve pooled sequences, potentially compared over time. This can be misleading when the heterogeneity between subjects is large and the within-sequence link is obscured. We can ameliorate this substantially through the use of categorical data clustering. Clustering, broadly conceived, is a mechanism for partitioning subjects into homogeneous groups. In fact, goodness of fit measures of a clustering often compare the homogeneity within assigned groups to the heterogeneity between them (Everitt, et al., 2001). Clustering is sometimes used to discover a new typology within a subject, but this usually leaves the researcher open to criticism regarding the choices made in producing the clusters. Hennig and Liao (2013) contend that some subject matter knowledge should be brought to bear in any cluster analysis, even when the goal is to learn about latent subgroups in the population. For the purpose of this study, clustering allows us to re-assess our measures of structure within smaller, more homogeneous subgroups. Given that some of our measures are based on aggregated sequences or subsequences, this reassessment of smaller, more homogeneous groups can be illuminating. The literature suggests that careers are attempts by workers to optimise their match with employers (French, et al., 2007; Gardecki & Neumark, 1998; Topel, 1993), so to the extent that we can identify similar patterns in our career sequences, we will consider the clusters to define a career type.

Clustering sequences presents the researcher with this challenge: what is the distance between two career sequences? One way to address this is by first proposing the dissimilarity between two IxO tokens, or job-types. This is a central challenge, because once we have a proper dissimilarity between tokens, which translates into a substitution cost, methods of optimal matching (OMA) and recent extensions of the core methodology can produce the necessary sequence distances (Elzinga & Studer, 2015; Halpin, 2014; Studer & Ritschard, 2016). In our problem domain, we recognise the need for time-varying distances; the likelihood of moving between different job types changes over time (see Scott, 2011, for empirical support; see Halpin, 2014 and Studer & Ritschard, 2016, for related methodological developments to OMA). Equally important in this domain is the need for sensitivity analysis, which is a comparison of findings made under various choices of metric. The approach we take will be discussed in the next section.

Measuring structure over time and clustering similar trajectories

Entropy

Entropy is a useful one-dimensional summary of the *distribution* of tokens used by a subpopulation (Elzinga, 2010; Pierce, 1980). Entropy measures the difficulty one would have guessing a worker's job type in the absence of additional information (given only the frequency distribution of job types for an ensemble of sequences). Formally, entropy is defined as $H(X) = -\sum_i P(X=i)\log_2 P(X=i)$, where i varies across the different job types, or IxO tokens, captured in variable X.[ii] For categorical data, entropy is zero when only a single type is present and maximised when the distribution is uniform across the set of types.

Choosing the unit of analysis or type of pooling is a key decision one must make when using entropy, as

well as most other measures that we discuss. Given our research questions, we first stratify by mobility group. Within each group, aggregating across individuals, we assess entropy at each time point, providing a moving window of the measure, which might reveal change over time. We use entropy to measure the complexity or variety of job types for a group of workers, rather than complexity within a particular worker's sequence.[iii] For labour market data, there is some advantage to smoothing out short-term shifts in distributions, so while we compute entropy at a single time point (cross-sectionally), we smooth those measures using the running median approach of Tukey (1977).

Turbulence and mutual information

Entropy as we have operationalised it is based on pooled observations. The order or disorder in *individual* career paths cannot be assessed using entropy without significant modification, as the total number of distinct tokens in a single sequence may be several orders of magnitude smaller than the full token alphabet. Fortunately, turbulence is a measure that captures the complexity of single sequences. Turbulence is defined in terms of three functions of a sequence *x*, $\phi(x)$, $c_1(x)$ and $c_2(x)$, where $\phi(x)$ is the number of distinct subsequences, $c_1(x)$ is the maximal variance of spell lengths given the alphabet used and the length of the sequence and $c_2(x)$ is the actual variance of spell lengths. Given these terms, turbulence is given by $T(x) = log_2\left\{\phi(x)\frac{c_1(x)}{c_2(x)}\right\}$ (Elzinga, 2010). If turbulence describes the complexity of a sequence, then evaluating it over time, using moving windows, will reveal changes in career structure as it progresses. To understand how turbulence changes over time, we use an age-based moving window; for example, the first evaluation is based on the subsequence that covers ages 20 through 25; we then shift by one quarter to examine the window from age 20¼ to 25¼, and so forth.

Another way to assess structure within a career is by using a measure called mutual information (MI, hereafter). This is typically defined for two random variables, *X* and *Y*, as follows:

$$I(X,Y) = -\sum_i \sum_j P(X=i, Y=j) \log_2 \frac{P(X=i, Y=j)}{P(X=i)P(Y=j)}$$

(Pierce, 1980). In our analysis of career sequences, *X* and *Y* reflect job content at two points in the *same* career, such as the early and middle stages. While the formulas for MI and entropy are similar, the unit of analysis differs for each, with MI defined within sequences and entropy across them, yielding a negative relationship between these two measures. In our study, we take *P(X,Y)* to be the joint probability of observing job type *X* in the early career and job type *Y* later in the career *of the same individual*. If the likelihood of observing *X* and *Y* together in the same career greatly exceeds expectations based on chance alone (the marginal product in the denominator), MI will be large. Thus MI applied to two portions of a career measures the extent to which the content in an early portion of the career predicts the content in a later portion. We aggregate all of these within-sequence co-occurrence probabilities for a prespecified group of subjects, such as chronically low-wage or 'stuck' workers, and evaluate the MI of the joint distribution. We reiterate, however, that each co-occurrence is derived from an individual sequence.

Addressing the heterogeneity of careers

While the above measures should provide substantial insight into career structure over time, the necessary aggregation masks heterogeneity and homogeneity across the sequences. On the other hand, sequence level measures are hard to summarise due to the large number of individuals in the study, and the concomitant heterogeneity. To explore this, we cluster careers with similar job content together within mobility group, and then make comparisons between mobility groups, at the *cluster level*. We will examine entropy, turbulence and MI for these more homogeneous groupings and may obtain results that are different from the aggregate findings. When the basic trends still obtain, we have simply added a layer of robustness to the prior findings. In what follows, we describe the methodology used to obtain clusters.

Clustering typically begins with measuring the distance between objects, represented as multivariate vectors, but there is no obvious metric

for comparing the components of two categorical sequences, as we have in our application (the simplest metric, exact matches, is too crude a measure). Another issue in comparing career sequences is the timing of the jobs – should the comparison be made at the exact same time point in each of the sequences, or do minor variations in the timing of jobs matter when defining a typical career pathway? It is clear we must consider timing; in the extreme, one could ignore the ordering of these sequences and just compare IxO frequencies across pairs. Without timing, career sequences have reduced meaning—they become mere collections of jobs.

The Optimal Matching approach

To meet the challenges of clustering these sequences, we divide the problem into two subproblems. First, we must decide how 'near' or 'far' each IxO is from the other to construct a pairwise token dissimilarity matrix. Second, we have to decide how to *align* two careers so that the token-by-token pairings may be compared less rigidly. The first subproblem is the topic of substantial research (e.g., Studer & Ritschard, 2016); optimal matching (OMA) is a fairly well established method, which deals effectively with the second subproblem (Abbott, 1995; Durbin et al., 1998; Lesnard, 2014; Sankoff & Kruskal, 1983).

To use OMA, we must specify the cost of each of the primitive operations, which are substitution, insertion, and deletion. Typically, the insertion and deletion operations (sometimes abbreviated 'indel') are given fixed costs (independent of which token is being inserted or deleted), while substitution depends on the token pair.[iv] We set the maximum cost of substitution to be no greater than the sum of the costs for insertion and deletion; otherwise, there would be no benefit to substitution. In our application, we use three different alternative substitution costs, capturing a range of assumptions and emphases. We take the insertion and deletion costs to be fixed at one. The first substitution cost is *fixed* at two, implying no two tokens are similar enough to warrant a substitution less costly than a deletion followed by an insertion. The next choice reduces the substitution cost of token 'a' for token 'b' by the probability of observing a *transition* between temporally adjacent tokens in either direction (we call this a transition based metric). Namely, substitution cost s(a,b)=2-P(a|b)-P(b|a), for arbitrary tokens 'a' and 'b'. This is implemented in the TraMineR function seqsubm by setting method='TRATE'. One concern with this approach is that most transition probabilities are small, as there are large runs of the same token within a sequence, which will set most substitution costs to nearly two.

We base the third of our substitution costs on a *ratio* commonly used in Biological Sequence Analysis (BSA; Durbin et al., 1998), which has the same functional form as the rightmost term in the mutual entropy calculation. We also allow for change over time by adjusting this cost metric as individuals age. We want to compute the similarity of two tokens, measured by how often they occur together. Consider the joint probability, $P(a,b)$, defined to be the probability of witnessing tokens 'a' and 'b' together in the same career sequence. To adjust for the fact that some tokens are extremely frequent while others are rare, we normalise the joint probability by the marginal product of the token component frequencies. That is, if the probability of seeing 'a' and 'b' together is large compared to the product of their marginal probabilities (under an independence assumption), then we have indication that 'a' and 'b' belong together (in MI parlance, one is predictive of the other). This metric may be written

$$s(a,b) \propto \log\left(\frac{P(a,b)}{P(a)P(b)}\right),$$

where s is the similarity and a and b are arbitrary tokens. Our substitution cost is a rescaled version of this similarity oriented so that the most similar pairs have the lowest substitution cost. The difference between this and the kernel of the mutual information measure resides in how we include time in the definition of all of the probabilities involved. We call this the *likelihood ratio* based metric. The use of this metric in the context of career sequences is novel.

We veer slightly from what is commonly done in BSA and base the joint probability on the empirical transition matrix, and to allow for heterogeneity, we compute these probabilities at each time in the analysis period.[v] Thus, $P_t(a,b)$ (substituted in the formula above) is the probability of observing

subsequence {*ab*} beginning at t. The metric $s(a,b)$ spans the entire real line; to transform it to a substitution cost, we refer positive values to a χ^2_1 distribution (large values yield small p-values) and multiply by two so that the least similar substitutions (p=1.0) are as expensive as a single deletion followed by an insertion. Negative values suggest unusually infrequent co-occurrence, so these pairs also receive the maximal substitution cost of two.

Clustering dissimilarities – multiple approaches

Using OMA and the above cost matrices provides us with pairwise distances between sequences. To cluster the sequences, we consider two well-established techniques. The first is a hierarchical clustering approach in which Ward's criterion is used to choose the nearest cluster pair to join (Everitt et al., 2001; Kaufman & Rousseeuw, 1990). The second is a recursive-partitioning algorithm known as partitioning around medoids (PAM; Kaufman & Rousseeuw, 1990). PAM is sometimes called k-medoids clustering, in reference to the k-means algorithm (Hartigan & Wong, 1979).[vi] PAM is similar to k-means in that one pre-specifies the number of clusters (k), and the algorithm (non-hierarchically) finds the optimal partition with that number. The 'centre' of the cluster, the point closest to all others in that cluster, is called the *medoid* and serves as an exemplary 'representative' for the group. PAM algorithms often provide the useful diagnostic 'silhouette width' (SW; a measure between -1 and 1) to assess how well each item belongs in the cluster to which it is assigned (for details, see Kaufman & Rousseeuw, 1990). Roughly speaking, SW is zero or negative for a sequence in which the nearest cluster medoid is not that of the assigned cluster.[vii]

Number of clusters

Using both Ward's method and PAM, crossed with the three different measures of sequence dissimilarity for OMA discussed above (constant, transition based, likelihood ratio based), we form separate clusters within each mobility-based group to identify the common pathways in each. We considered two goodness-of-fit criteria: average silhouette width and the Calinski-Harabasz index. The latter was adapted for dissimilarity matrices by Hennig and Liao (2013). We maximise the fit criterion across different choices for the number of clusters. We view the final set of career clusters as a *typology*, or a way to label similar careers based on their component industry and occupation (along with time out of the labour force) trajectories. The clusters should be relatively *homogeneous within* and *heterogeneous between* each other.

Data and construction of mobility groups

Our data source is the National Longitudinal Survey of Youth (NLSY), which yields a representative sample of non-institutionalised[viii] men and women in the U.S. between the ages of 14 and 21 in 1979 (Bureau of Labor Statistics, 2000). This cohort was interviewed every year from 1979-1994, and then once every two years until the present. The last round in our analysis dataset ended in 2000, when the cohort was between 35 and 42 years old. Blacks, Hispanics and poor whites were oversampled in what are known as 'supplemental samples'. The supplemental sample of poor whites was dropped from this analysis because it was discontinued after 1990, truncating their career sequence prematurely. Poor whites are represented in the common, retained sample, and we have adjusted the weights to accurately reflect their proportionate contribution. A supplemental military sample was dropped from the sample used in our analysis as well. The original sample size, including all supplemental samples, is 12,686. This drops to 9,763 after the two supplemental samples are dropped. After careful evaluation of patterns of missed interviews, we decided to remove individuals who show a gap of more than four years between any two surveys.[ix]

As explained above in the first section, when we refer to the 'career', we mean a sequence of industries and occupations. We build this sequence from quarterly jobs spanning ages 20 to 36 using the individual's work history (the NLSY constructs a *weekly* job history from the questionnaire, and we define the quarterly job to be the job held in the first week of each quarter). We formulate 25 unique industry and 20 unique occupation codes that aggregate the three-digit 1970 Census Classification codes into reasonably homogeneous groups.[x] For example, all durable goods manufacturing industries are collapsed into a single code. In addition,

unemployment, educational enrolment, and time spent out of the labour force are coded into the sequence. In the 16-year age span studied, approximately 450 unique industry and occupation pairings (IxOs) occur. These codes can be understood as ideal types (Weber, 2009), or coarse nominal groupings.[xi]

For our restricted sample, the key variables of industry and occupation for jobs recorded in the work history are missing at about 3%, which is minimal. About 100 individual cases have missing industry or occupation information for more than half of the work history, so these cases are dropped. The sample size becomes 7,816, or 80% of the maximum possible. We reweight the sample so that it is consistent with the demographics of the original baseline sample.

Mobility group definitions

For each respondent, longitudinal wage profiles were constructed using inflation-adjusted, logged hourly wages associated with the CPS job[xii] at the time of each annual interview. Taken as a whole these wages form a profile of growth or stagnation over time. The inflation adjustment is made using the Consumer Price Index research series (CPI-U-RS; see Bureau of Labor Statistics, 2015).

The wage profiles were cleaned of short-term wage fluctuations by substituting, for the original observations, best linear unbiased predictions (BLUPs; see Robinson, 1991) from a longitudinal mixed-effects model. Each predicted profile could be understood as the 'permanent' wage level over a broad time span. Gottschalk and Moffitt (1994) discuss the theory behind such permanent and transient wage decomposition and provide a methodology for their identification. We use a slightly different method to generate permanent wage trajectories, following Bernhardt et al. (2001). Using age as the underlying timeline, we found that a quintic mean structure, quadratic random effects and year-specific variances represented the data best, given the variation observed. The Bayesian Information Criterion (BIC; Kass & Raftery, 1995; Schwarz, 1978) was used to select these model components, which were also guided by and were similar to those used by Murphy and Welch (1990).

We classify each permanent wage as either *low* or *not low* using a poverty line threshold. Wages below 1.25 times the poverty line for a family of four (converted from annual income to an hourly wage) in a given year are considered low.[xiii] At the end of the period examined, 2000, this was about $11.00/hour; for reference, in 2015 this is about $14.50/hour. An additional twenty-five individuals were dropped from this analysis due to inconsistencies in their wage profiles, including severely outlying wages.

We divide workers into three mobility groups based on their permanent wages at age 24 and 38.[xiv] The first age reflects a point in the life course at which most individuals have entered the labour force, and the latter is a point at which most family formation, if it is to occur, has begun. Intragenerational mobility occurs over the life course, within a worker's career; intergenerational mobility occurs from parent to child. Each individual was classified as belonging to one of the following three intragenerational mobility groups:

- *Stuck*: wages are low at age 24 and are still low at age 38
- *Mobile*: wages are low at age 24 but are no longer low at age 38
- *Never low*: wages are never low, at either age.

Seventy-nine workers were found to be downwardly mobile, but this category was dropped due to small sample size. The final sample size is 7,712.

Under our poverty-based definition, fully 28% of the sample is stuck in low-wage jobs over the career. Another 33% begin working in low-wage jobs, but then escape them by mid-career. Fully 39% manage effectively to avoid low-wage jobs altogether, throughout the career. We find these numbers striking, showing both a significant amount of immobility and a significant amount of mobility out of low-wage jobs.

Basic descriptive findings

We have examined many demographic and education measures across career and worker dimensions, and not surprisingly, these differ across the three mobility groups in expected ways. For example, stuck workers are more often female, less-educated, and African American. Never low workers are more often male, white, and college-educated.

This is consistent with the literature on low-wage workers in the U.S. (Andersson et al., 2004; Appelbaum et al., 2006; Boushey, 2005; Danziger, Blank & Schoeni, 2008; Mishel et al., 2007).

We turn to markers of career structure and find that the situation is more complex than one might think. In aggregate, both time spent out of the labour force (OLF) and unemployment spells are a very strong marker of low-wage careers (see also Bernstein & Hartmann, 2000 and Topel, 1993). Stuck workers are OLF (and not enrolled in school) 36.2% of the 16 years examined, while mobile and never low workers average 18.7% and 9.7%, respectively. However, such figures mask tremendous variability. Stuck workers, in particular, have great variation in their attachment to the labour force: at the 75th percentile, nearly 60% of a stuck worker's career is spent OLF, while at the 25th percentile, that figure drops to about 15%, which is close to the average experience of mobile and never low workers.

Another important feature of these careers that distinguishes groups from each other are the job types themselves. While attachment to the labour force, on the whole, is weaker for low-wage workers, the remaining (non-OLF) portions of the career sequences differ as well. We evaluate the similarity of IxO distributions at three different stages of the career using Pearson's χ^2 test of association, pooling the tokens across subjects but within mobility groups. The three stages are early, middle and late career, and are defined as six-year intervals. We do not include all 457 token types; rather, we remove the non-working tokens, and limit our analysis to the top *K* most frequent tokens, with *K*=10 or 50. The latter, being most common, cover the majority of all jobs and the former considers overlap in a smaller subset of common jobs. When all three mobility groups are included, there is clear association between the groups and IxO patterns at each stage ($p<0.001$; simulation-based, using Hope, 1968). When we restrict this to the stuck and mobile groups, it still holds.

While there are clear differences between these groups, many traditional measures that could be used to describe structure are strikingly similar in all three mobility groups. For example, in aggregate, all three have about the same number of industries, occupations, and IxO pairs across ages 20 to 36. That is, individuals accumulate about the same number of unique job types over this portion of their careers. Equally noteworthy, stuck and mobile workers accumulate about the same number of employers, in aggregate. To understand these findings more completely, we need to disaggregate these workers into meaningful subgroups and consider change over time. As an example of the latter point, the aggregate number of employers masks an important change over time: stuck workers begin their careers with fewer employers but end with slightly more than their more mobile counterparts. However, this is only part of the story, because there is great variation in the number of employers among these workers, suggesting that meaningful subgroups of more and less attached low-wage worker careers evolve over time.

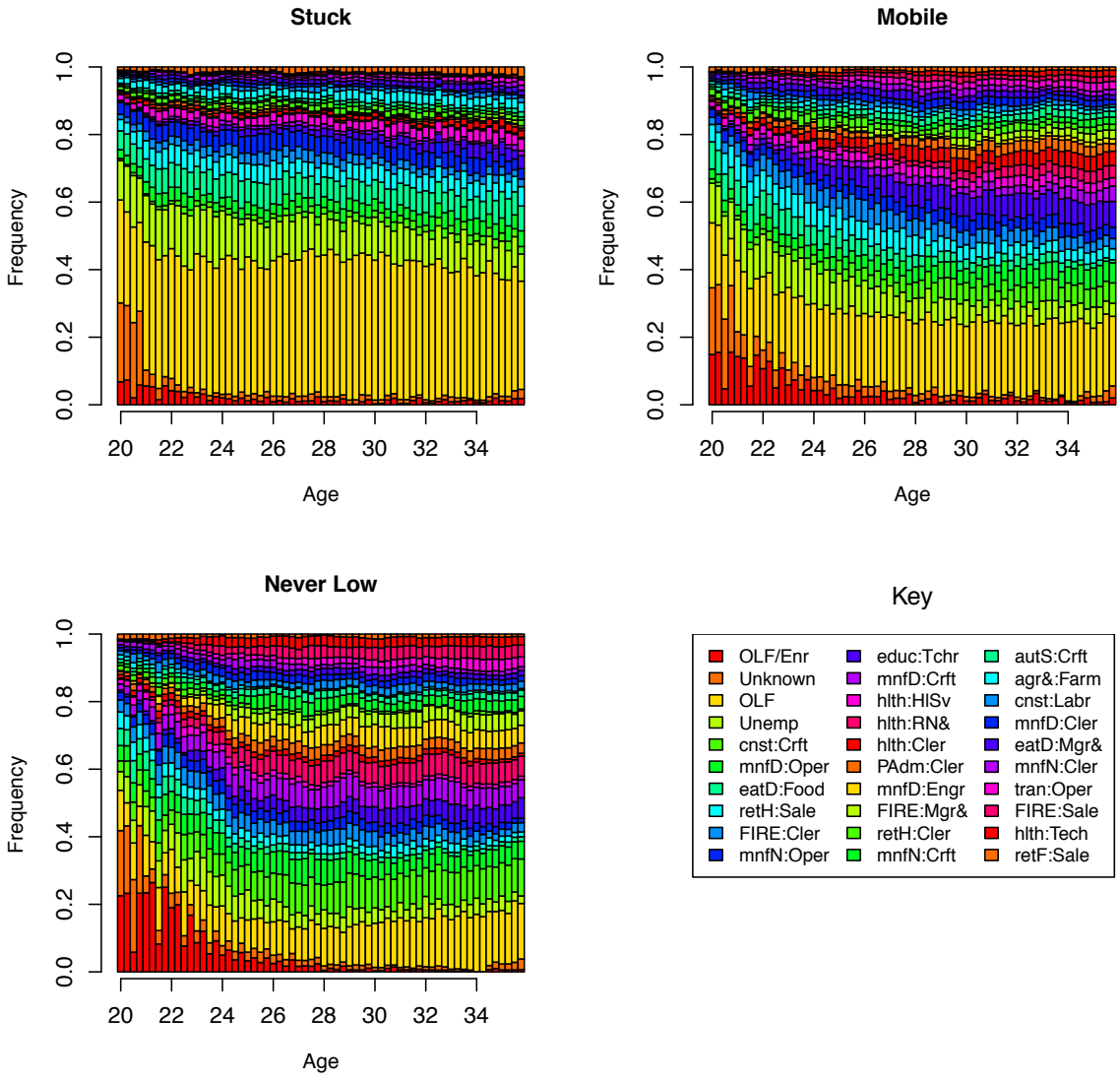
Results

Visualisation

Visualising the content of a large number of sequences relies, commonly, on aggregation, but this can still be revealing. In figure 1, we colour-coded the top 30[xv] of 457 IxO pairs (including labour force status) using fourteen colours (equispaced spectrum-wise, using R's rainbow function), repeated to cycle through those colours, and display the *content* of careers in three mobility groups over time following the approach of Piccarreta (2012) and Fasang and Liao (2013). Any token outside of this prespecified range is removed, so the state distributions can be understood as conditional on work being in the smaller set of tokens. With this smaller token space, we can utilise a legend, which we provide in the bottom right panel (codes are described in Appendix B). Note that three of the first four colours (these are always on the *bottom* of the state distribution plots) represent time spent outside of the labour force (OLF), and are thus non-working periods, with yellow specifically representing OLF periods in which the subject is not in school nor looking for work. As alluded to previously, OLF periods dominate the content of low-wage careers. We re-examined these state distributions after removing non-working tokens (not shown), and we still see differences in content as depicted by colour distributions and patterns. Even in this reduced set of 30 tokens, it can be seen that the

spectrum of colours repeats more fully, with a greater proportion of the lower frequency IxOs present for never low workers, suggesting that they (conditionally) utilise a larger spectrum of job types. This suggests that there is more variety in the types of careers that are obtained. These top 30 tokens account for between 54% and 71% of the full token space, depending on mobility, with the smallest percentage associated with never low workers. This is consistent with our characterisation that these workers explore a wider variety of job types (requiring more than 30 tokens to fully represent them) over their life course.

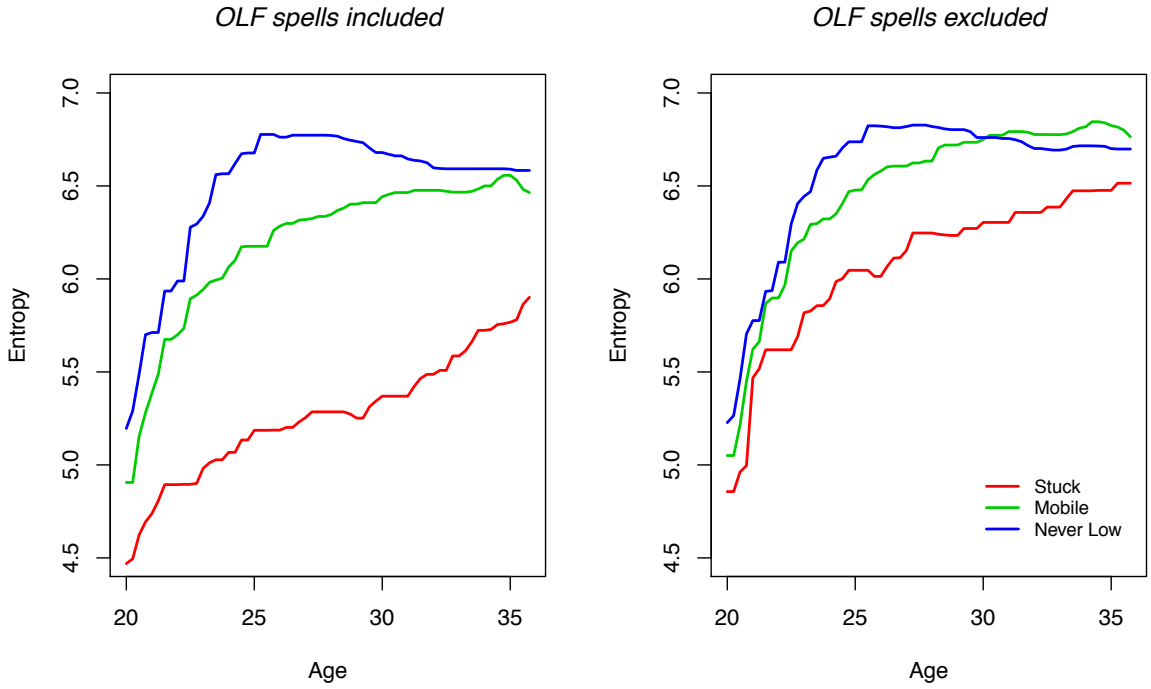
Figure 1. State distributions over time by mobility group



Entropy

The entropy measure is often applied to sequence data cross-sectionally, at every time point. Essentially, we pool job types across workers within mobility group, yielding an empirical distribution of IxOs *for each group* and take the entropy of that distribution. With job IxOs, short-term changes in this distribution can lead to irregular fluctuation in the entropy measure, viewed as a time series, so we smooth it slightly with a running median approach (Tukey, 1977). We measure entropy a second time, after removing quarters of non-work, and plot these in a separate panel.

Figure 2. Entropy over time by mobility group



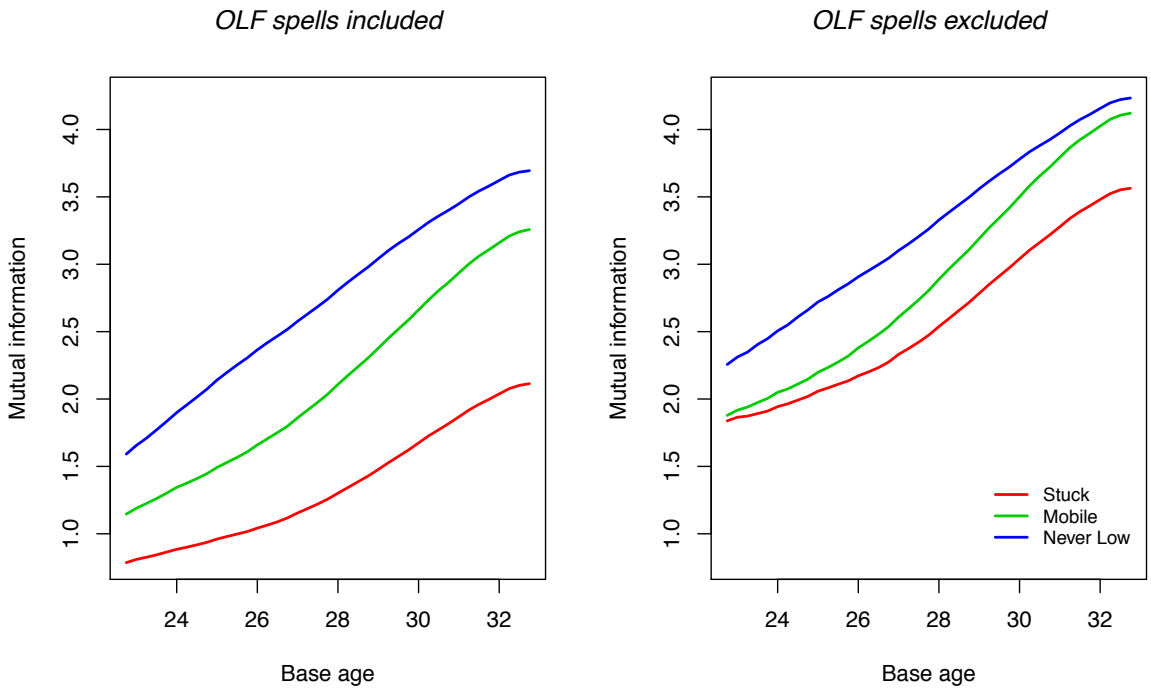
The results are given in figure 2, left and right panels. Surprising, at first, is the *lower* entropy for stuck workers. If lower entropy implies less complexity, then stuck workers appear to have the simplest trajectories. This is consistent with stuck careers being fairly stagnant; the workers in these careers could be failing to 'explore' many of the potential job types in the labour market, and this may be linked to their lack of mobility (note that the sensitivity of our findings to the job classification system is discussed in Appendix A). How the timing and amount of job 'shopping' affects wage growth has been examined in Bernhardt et al. (2001), Murphy and Welch (1990), Topel and Ward (1992) and Gardecki and Neumark (1998), among others. Removing OLF periods of non-work (right panel) increases the stuck career entropy, but relative to more mobile workers, the pattern persists: stuck workers apparently have somewhat less complex careers, in aggregate.

Mutual Information

For the MI analysis, we use a six-year moving window, for a generic subsequence labelled *X*, and we fix the 'target' window for *Y* as the last six years of the career, so *X* gradually moves toward *Y*. Thus, the first evaluation of MI is based on the subsequence from age 20-25 (*X*) compared to 31-36 (*Y*). We then shift by one quarter to compare the windows 20¼-25¼ to 31-36, and so forth, holding the target *Y* fixed. In figure 3, left panel, we show MI as an estimate of the predictability of careers as one gets closer in time, by mobility group, assigning the MI to the midpoint of

each window for *X*, so our plot begins at age 22½. What is immediately striking is the similarity between these results and those for entropy, but the implications are different. Recall that MI is evaluated using joint distributions built up from within-sequence information. Stuck workers have the lowest levels of MI across time. The overall trend of increasing MI over time is consistent with our understanding that early careers consist of more job-shopping[xvi] than later careers, which are more settled (Topel & Ward, 1992). Since stuck workers are least predictable (given a subsequence from the past), well into their 30s, we might conclude that as a group, their careers are more chaotic. Even when we restrict the analysis to time spent working (right panel), the basic differences between mobility groups hold, although they are attenuated.

Figure 3. Mutual information over time by mobility group



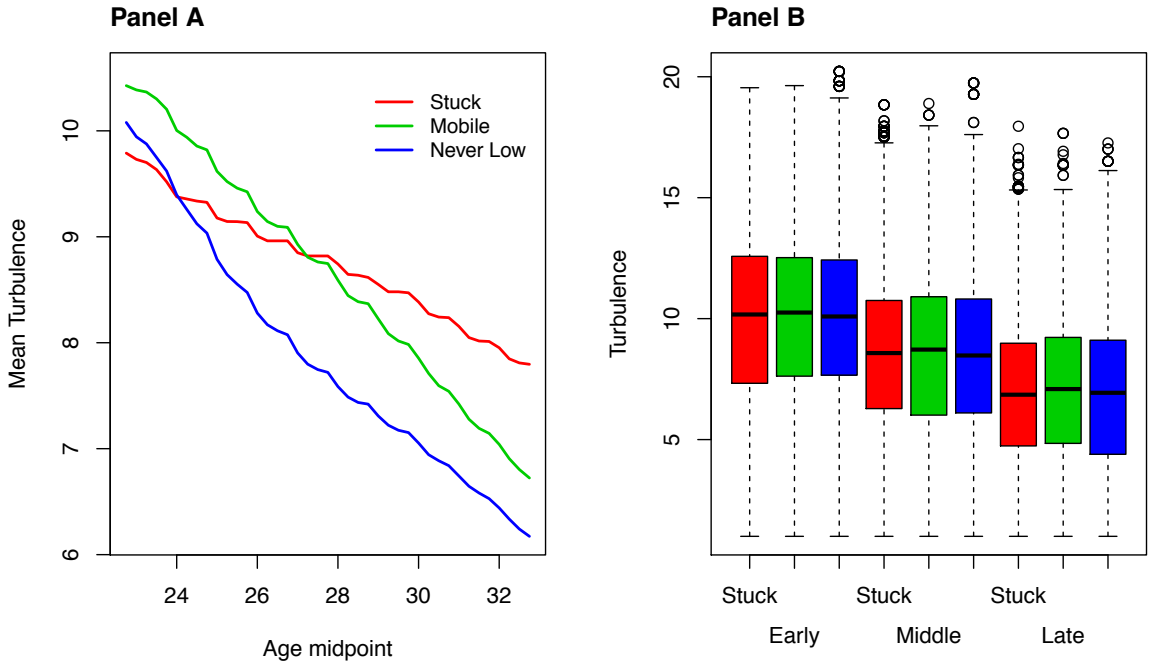
Combining the visualisation with entropy and mutual information measures of structure, we contend that in aggregate, stuck careers explore a smaller spectrum of job types, and that this partially explains their lower entropy, yet this does not translate into predictability. The lack of predictability is consistent with a more chaotic career for stuck workers, partially driven by time in and out of the labour force. Even though the alphabet of IxOs is smaller for these workers, movement back and forth between job types is not forming a consistent pattern or career at this level of analysis. We will learn more from the cluster-based reassessment.

Turbulence

Turbulence is a property of a single sequence, yet we wish to understand differences in turbulence over time and between mobility groups. The latter may be achieved through mean and standard deviation statistics, while the former requires a moving window approach. For this analysis, we use the same longer window of six years so that we witness sufficient sequence complexity.[xvii] We move the window one quarter at a time, generating 41 time points from early to late career, and we median smooth these. For this measure, we do not remove non-working periods for several reasons. First, we wish to consider movement in and out of work and spells of

unemployment as part of the structure (or lack thereof) in the career. Second, the introduction of missing tokens creates completely missing subsequences, and turbulence is not defined on these.[xviii] This could result in stuck careers registering relatively low amounts of turbulence (e.g., if the bulk of the career were spent OLF), and we view this as an empirical question.

Figure 4. Mean and distribution of turbulence over time by mobility group



Judging from figure 4, panel A, mean turbulence is much higher early career for all mobility groups, then it drops, presumably as workers establish the core job types that form their careers. Stuck workers may start out with slightly less turbulent careers, on average, but they do not become less turbulent at the same rate as their more mobile counterparts. We do not conduct significance tests on these comparisons, but any gap of 0.25 units may be considered significant at traditional levels, and such a gap exists for most age and mobility group comparisons. Moreover, we wish to document the heterogeneity of sequence turbulence. Thus, in boxplots given in figure 4, panel B, we see that the interquartile range of turbulence within mobility group is fairly consistent over time (it seems to drop slightly for stuck workers). So while turbulence itself is dropping by age 30, the within-group variance persists. By mid-career, stuck workers' careers appear to be more turbulent judging by the mean trend although not by the more robust median. Furthermore, this characterisation masks tremendous heterogeneity present in each mobility group. This is part of what motivates our cluster analysis, to follow.

We have amassed some evidence that stuck workers utilise a more concentrated subset of job IxOs. They explore a limited set of career types, and spend considerable time outside the labour force, in aggregate. It is harder to predict a stuck worker's career from prior content than to do so for a worker in another group. We must be cautious: aggregate measures have the potential to obscure more subtle differences between and within groups. In what follows, we group careers based on their industry and occupation content to determine whether there are more and less coherent careers in the low-wage sector. If stuck careers do not all contain large spells of OLF, then perhaps there is an underclass with low attachment to the labour force and more chaotic careers, while a solid group of stuck workers' histories

are quite similar to those of mobile and never low workers. This is an empirical question that can be explored using cluster analysis.

An important finding based on our evaluation of marginal OLF distributions (not shown) is that stuck careers are bifurcated into two distinct types (for a general discussion of dual or segmented labour markets, see: Dickens & Lang, 1985; Hudson, 2007; Osterman, 1975; Piore, 1983). In the first third of the career, the distribution of time spent OLF or unemployed exhibits clear bimodality, with a smaller mode near 100% – these are nearly completely unattached, 20-25 year old workers. This suggests that we may gain new insight by first separating careers into those with similar attachment *and* job content. A deeper study, perhaps both qualitative and quantitative, would form matched career types to examine whether or not stuck careers *in retail*, for example, are different from their mobile counterparts *in retail*. We will use our cluster analysis to highlight potential matched types with similar content across mobility groups to highlight similarities and differences that may yield insight into the structure of careers.

Career sequence clustering

We wish to place careers with similar IxO pairs and structure in the same cluster. To review, the career sequences consist of industry-occupation pairs (such as retail hard goods/sales clerk) for the 64 quarters spanning ages 20-36. Quarters spent unemployed, enrolled in school, or out of the labour force are included with unique codes, since they form an important part of the career structure.

As discussed in the third section of this paper, we evaluated three dissimilarity measures and their corresponding substitution costs, two different clustering techniques and two criteria for choosing the number of clusters. In terms of measures, we compare substitution costs of constant (twice the insertion or deletion cost, which are the same), transition rate based, and likelihood ratio based. We found that the raw transition rates were generally low for these data (most transitions are self-loops, indicated by large diagonal entries in the transition matrix), so we included a fourth variant in which rates were conditional on a non-self-loop transition. In other words, we removed the diagonal and renormalised the transition rates before computing substitution costs for the transition matrix based dissimilarity measure. This correction provided the strongest competitor (and most similarity) to the likelihood ratio based approach, which adjusts for the probability of a transition under independence. Within mobility group, we examined a range of 2 to 64 clusters, and used the Calinski-Harabasz index and average silhouette width to identify the most appropriate clustering (Everitt, et al., 2001; Kaufman & Rousseeuw, 1990). While we always identify an optimal number of clusters for each scenario, most approaches selected either very small cluster sizes, or had great disparity in their sizes. For example, a clustering with 2 clusters for stuck workers and 64 for never low workers, even if 'optimal', was deemed a poor fit on *substantive grounds* – it reveals little to nothing about either group.[xix] With the likelihood ratio based metric, the clustering results had much better properties. Whether we used Calinski-Harabasz index or silhouette width to evaluate, and to some extent whether we used PAM or Ward's method to cluster, the number of clusters and the size of the clusters were comparable across clusterings using likelihood ratio based dissimilarity. We chose the PAM, likelihood ratio based, average silhouette width evaluated clustering as our optimal choice, which yielded 12, 13 and 17 clusters for the stuck, mobile and never low mobility groups, respectively. Comparing this choice to our second best choice using the transition-based measure, we find unadjusted Rand Index (Rand, 1971) similarities of 0.73, 0.88, and 0.93, respectively for stuck, mobile and never low workers. This is evidence of content agreement by cluster, but adjusted for chance indices are much lower (10-30%), so there are real differences as well.

Clustering is often viewed as a subjective exploratory approach (Hennig & Liao, 2013), and while there is always some subjective input provided by the researcher in the form of dissimilarity measures and cluster generation, we use available techniques for evaluating the goodness of fit, and note that visualising the clusters often provides additional face validity. We compared graphs of state distributions over time by cluster and mobility group, and found that most had several clearly dominant, or important, IxOs that characterised the cluster. Given the number of clusters and IxO states, we summarise the clustering in table 1, which provides the size (proportion of workers, by mobility group) and the top four IxOs that are

dominant in the state distributions. The four-letter codes for the IxO pairs are given in Appendix B. To facilitate comparison, we name the clusters with a mobility code and number, so that S1 is the first (most frequent) stuck cluster, M3 is the third mobile, N5 is the fifth never low, and so forth.

Table 1. Characterisation of clusters by top four IxO state distributions.

Cluster	*Tot. Freq.*	*Top IxO*	*%*	*2nd IxO*	*%*	*3rd IxO*	*%*	*4th IxO*	*%*
S1	18.6	OLF	31.4	agr&:Farm	7.1	Unemp	3.3	Unknown	1.6
S2	17.4	OLF	21.7	Unemp	8.3	eatD:Food	5.3	OLF/Enr	3.5
S3	17.2	Unemp	9.6	OLF	9.5	Wkg:Msg_IxO	1.9	mnfD:Oper	1.9
S4	13.7	OLF	24.9	Unemp	6.2	eatD:Food	2.3	retH:Sale	2.1
S5	7.5	retH:Sale	18.0	OLF	6.9	Unemp	6.4	retH:Cler	4.9
S6	5.5	eatD:Food	15.5	OLF	13.4	eatD:Mgr&	9.2	Unemp	2.8
S7	4.6	cnst:Crft	25.6	Unemp	4.3	OLF	3.7	cnst:Labr	1.9
S8	4.3	mnfD:Oper	12.6	OLF	7.7	Unemp	3.1	mnfN:Oper	3.0
S9	3.7	mnfN:Oper	23.1	OLF	10.7	Unemp	5.2	mnfN:Crft	2.8
S10	2.7	FIRE:Cler	9.7	OLF	9.6	FIRE:Mgr&	9.1	FIRE:Sale	4.5
S11	2.6	prsl:PrSv	29.9	educ:Tchr	4.8	OLF	3.5	Unemp	3.2
S12	2.3	hlth:HlSv	27.1	OLF	6.4	Unemp	4.6	hlth:RN&	2.1
M1	23.2	OLF	22.0	Unemp	7.3	autS:Crft	4.5	eatD:Food	4.3
M2	14.3	OLF	14.8	OLF/Enr	4.5	Unemp	4.2	Unknown	3.9
M3	13.9	Unemp	22.8	OLF	4.2	tran:Oper	1.5	educ:Cler	1.4
M4	7.0	retH:Sale	5.6	retH:Cler	3.5	OLF	3.0	retH:Mgr&	2.3
M5	6.6	educ:Tchr	20.5	OLF	7.4	OLF/Enr	3.7	prsl:PrSv	3.2
M6	5.7	hlth:HlSv	17.2	hlth:RN&	14.1	OLF	3.5	hlth:Tech	2.1
M7	5.5	FIRE:Cler	12.8	FIRE:Mgr&	10.5	OLF	6.3	FIRE:Sale	5.2
M8	5.4	mnfN:Oper	6.5	Unemp	6.1	mnfN:Crft	3.2	OLF	2.2
M9	4.9	cnst:Crft	14.0	OLF	10.9	cnst:Labr	3.6	cnst:Oper	2.8
M10	4.8	mnfD:Oper	19.7	mnfD:Crft	4.7	Unemp	4.7	OLF	4.6
M11	3.8	hlth:Cler	18.1	PAdm:Cler	3.9	OLF	3.6	Unemp	3.3
M12	2.9	eatD:Food	20.4	eatD:Mgr&	11.8	OLF	4.5	Unemp	2.5
M13	2.0	PAdm:Prot	23.1	Unknown	6.6	Unemp	4.7	OLF	2.8
N1	11.4	OLF	12.6	prfS:Cler	2.9	Unemp	2.8	hlth:Cler	2.4
N2	10.1	mnfD:Oper	18.8	mnfD:Crft	4.9	mnfD:Mgr&	4.4	Unemp	2.6
N3	8.9	cnst:Crft	14.7	cnst:Labr	12.5	Unemp	4.3	cnst:Oper	3.1
N4	8.6	mnfN:Mgr&	3.5	comm:Cler	2.8	OLF/Enr	2.1	Wkg:Msg_IxO	1.9
N5	7.5	FIRE:Cler	15.0	FIRE:Mgr&	10.1	FIRE:Sale	8.1	FIRE:Prof	2.9
N6	7.4	tran:Oper	8.7	Unemp	6.4	OLF	3.8	mnfD:Cler	3.1
N7	7.1	hlth:RN&	20.8	hlth:Tech	6.4	hlth:HlSv	3.4	OLF	2.3
N8	5.5	retH:Sale	3.5	retH:Mgr&	3.1	OLF	2.2	mnfD:Sale	2.1
N9	5.4	educ:Tchr	20.4	OLF/Enr	8.0	OLF	6.7	recr:WrAr	4.4
N10	5.2	mnfD:Engr	7.2	bzSv:Engr	6.5	OLF/Enr	6.1	prfS:Engr	3.5
N11	5.2	mnfN:Oper	22.6	mnfN:Crft	5.5	mnfN:Engr	4.2	OLF	3.8
N12	3.8	PAdm:Cler	19.4	tran:Oper	7.1	util:Cler	3.3	OLF	2.6
N13	3.6	prfS:Prof	13.8	OLF/Enr	9.9	prfS:Cler	6.3	OLF	2.9
N14	3.3	OLF	18.1	retF:Sale	5.0	Unknown	4.7	retF:Mgr&	4.5
N15	3.0	eatD:Food	32.2	eatD:Mgr&	2.2	OLF	2.1	autS:Crft	1.8
N16	2.1	PAdm:Prot	13.4	bldS:Prot	11.3	PAdm:SocW	3.3	Wkg:Msg_IxO	2.9
N17	1.9	util:Crft	12.7	PAdm:Engr	8.6	PAdm:Crft	4.0	OLF/Enr	3.9

In table 1 (first third), we notice that stuck workers have four clusters, S1-S4, in which time spent OLF is paramount. In cluster S3, the most frequent job type is time spent unemployed, but time spent OLF is a very close second. These represent highly unattached workers, but their trajectories seem to differ in subtle ways. For example, the day labourer (e.g., farm worker) job appears somewhat regularly in S1, while food service (e.g., waiter/waitress) jobs appear in S2 and S4. S3 has some connection to durable manufacturing, operative work (e.g., fork lift operator). The remaining eight clusters, S5-S12, accounting for about one-third of the workers, are identified with very specific job types (through the modal IxO): e.g., retail sales, food service, construction, operative work in manufacturing, clerical work, household service (e.g., childcare), health services, respectively. Many of these job types have higher wage equivalents, as we shall see, but they are usually indicated by a different IxO code (e.g., a more professional occupation in the same industry).

Mobile workers' career clusters (second third of table 1) are not as dominated by total time spent OLF, although a solid 51% of careers, as captured by clusters M1-M3, have OLF or unemployment as their most-prevalent token. It is not immediately clear how these three clusters differ from S1-S4 for stuck workers. Visualisation would reveal that these are fairly turbulent careers, with less attachment but still some semblance of IxO content and thus characterisation outside of simply being unattached. Of potential importance is the presence of enrolment in college (code: OLF/Enr), which could provide skills to secure a more stable job in the future. Perhaps more noticeable is the solid dominance of one or two IxO job types in each of the remaining clusters. These clusters are well-described by their modal IxO, which now includes teachers (cluster M5) and medical office managers (cluster M11), in addition to many of the career types in the stuck worker clusters. As an example of how ostensibly similar career types (as determined by dominant IxOs) can differ across mobility groups, we note that in cluster M6, entry-level health service work is eventually replaced, over time, by nursing, technician and administrator jobs, all in healthcare (confirmed via state distributions over time). Such a change is noticeably absent from the comparable stuck worker in healthcare, cluster S12, within which entry level work (hospital attendants, nurses' aides) remains dominant over time.

For never low workers (last third of table 1), perhaps surprisingly, clusters N1 and N14 contain a large proportion of time spent OLF. The variety of types of work that remain (administrative, sales and managerial) seems to characterise this group, despite its weaker attachment. Upon closer inspection using a time-based state distribution plot (not shown), the proportion of OLF periods begins fairly small and grows somewhat over the life course. Given the wage mobility of this group, these are likely intentional exits, perhaps related to family formation (e.g., some of these clusters have a large proportion of females). Also noteworthy is a simple difference in how one might characterise the remaining never low clusters; namely, many have a single industry followed by several occupations within it. For example, N3 are construction workers, N5 work in Finance, Insurance and Real Estate (FIRE), while N2 and N11 are manufacturing clusters. In many of the clusters, occupations shift from sales to manager, with the latter clearly representing a more stable and better-compensated position within a firm.

Alluded to in the above description of the top IxOs for each cluster within mobility group, these clusters inform our understanding of *stuck* careers through the differences in the *variety* of content over time. For example, while the more structured stuck career clusters contain a dominant token describing a job type, there is rarely more than one with any meaningful presence in the distribution. Mobile workers tend to have two or more (working) IxOs with substantial presence. Never low workers often have three or more solid (working) IxOs in a cluster, perhaps most clearly indicating dynamic changes in a career. For example, we witness sales work leading to managerial work in mobile and never low workers, but this progression is not typically observed in a stuck worker's career.

The above highlights several industries and common occupations within them. We note that exploratory analyses, in which clusters are matched – as an amalgam – across mobility groups suggests that nearly all clusters have at least one 'companion' cluster that is more or less mobile. For example, we

find stuck, mobile and never low health care clusters, but each is characterised by different mixtures of occupations, and attachment. The stability and other features of the careers will be addressed next. A complete analysis comparing and contrasting the full set of typical careers, focussing on the timing of job changing and the extent to which later stuck careers begin to resemble portions of their more mobile counterparts is left to future work.

Cluster-based reassessment of sequence characteristics

The cluster analysis reveals that mobility groups can be sorted reasonably well based on the tokens representing job types and their dependence relationships. They reveal increasing complexity, in the sense that career paths share a greater variety of content within a cluster, as we move from stuck to mobile to never low workers. The measures of structure that we have evaluated in the aggregate, entropy, MI and turbulence, can be reassessed through the lens of these more homogeneous clusters to complete the characterisation of career mobility groups.

Figure 5 is our entropy measure evaluated separately for each cluster within mobility group (cluster codes are given near the right margin). In this analysis, OLF spells are included, although some mention of what we may learn from their exclusion will be noted. The thickness of each line is proportional to the size of the cluster. Side by side panels allow for same-scale comparison. Note that in aggregate, never low workers have the highest level of entropy (recall figure 2), and this is true as they age. In contrast to this, within cluster, the 17 clusters in the never low group do not seem terribly different from the 13 clusters in the mobile group, in terms of entropy trajectories over time.[xx] There is roughly a bifurcation into higher and lower entropy paths. The higher entropy clusters are the least attached M1, M2 and M3 (thicker lines, representing a larger portion of the workforce) for mobile workers, whereas only one less-attached cluster, N1, is among those with higher entropy for the never low workers. The remaining high entropy clusters are N4 and N6, in manufacturing and transportation, for which MI and turbulence (to be presented in figures 6 and 7) are also quite high, which is suggestive of predictable, but complex pathways. The stuck workers, however, tend to remain lower entropy as an ensemble, even when their homogeneous clusters are the unit of analysis. Note that there is a single stuck cluster that stands out with larger entropy than most others, consistently over time. It is cluster S3, which has substantial levels of unemployment along with time spent OLF. We know that these workers were placed in the same cluster because of this non-working content. This could imply that the remaining content of these careers lacks structure, at least in the aggregate. This suggests that such careers consist of movement across a wide range of IxOs as opposed to concentrating in a specific sector. In contrast and equally noteworthy, many stuck career clusters are in the same basic entropy range as their more mobile counterparts; perhaps these are comparably structured as well? If so, this stands in stark contrast to a common characterisation of low-wage careers as being chaotic; this cluster analysis has revealed structure in many low-wage sectors of employment. Lastly, when the OLF spells are removed (not shown), the less-attached clusters, M1, M2 and M4, show higher levels of entropy later in the career, confirming that OLF spells suppressed the complexity inherent in the content of the working periods in these careers.

Figure 5. Entropy over time by cluster within mobility group

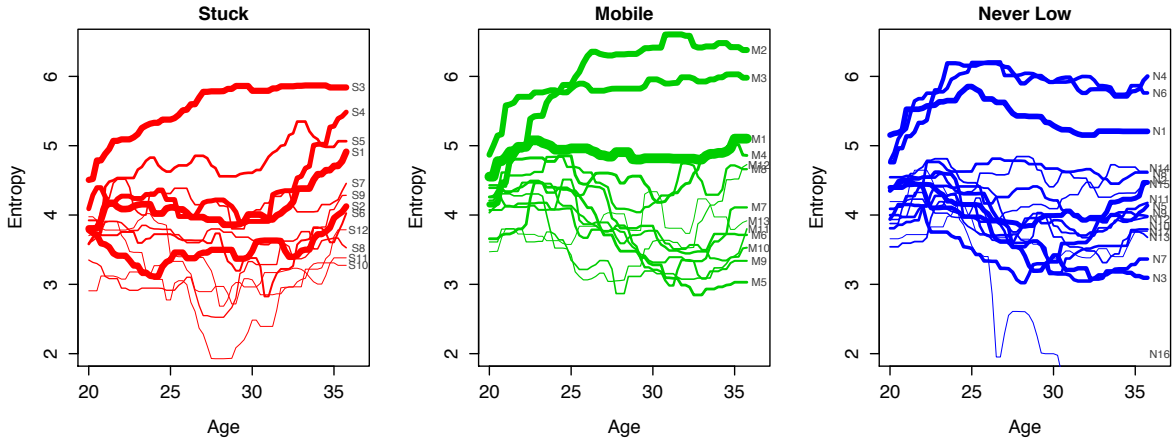


Figure 6 repeats this analysis using the MI measure. Again, we refer primarily to the top panel, in which the content analysis includes OLF spells, and the line width remains proportional to cluster size. Recall that the measure is based on a moving window; when MI increases over time, this implies increasingly similar and predictable content. These curves are much smoother than those for entropy due to the large window used in computing the measure. The most noticeable feature is that while the mean level and trends mimic the aggregate analysis (recall figure 3), with stuck workers at generally lower levels of MI, heterogeneity has emerged with respect to this measure. In particular, never low workers exhibit a wide range of MI, suggesting that some of this group's career types involve much more IxO pair changing than others (closer inspection would reveal some 'steady' career types, such as police officers, are lower MI). An explanation consistent with higher MI clusters is that increased job changing for some never low clusters involves a form of promotion, making their sequence of job types a bit harder to predict throughout the career. It is important to note that clusters are already sorted in a manner that should reflect predictability within elements in the same cluster, so differences in MI reflect differences in predictability, *post-clustering*. In the bottom panel, after removing OLF spells, MI is now much more heterogeneous for the stuck workers. In fact, the mobility groups appear quite similar, when exclusively working periods are considered.

Figure 6. Mutual information over time by cluster within mobility group

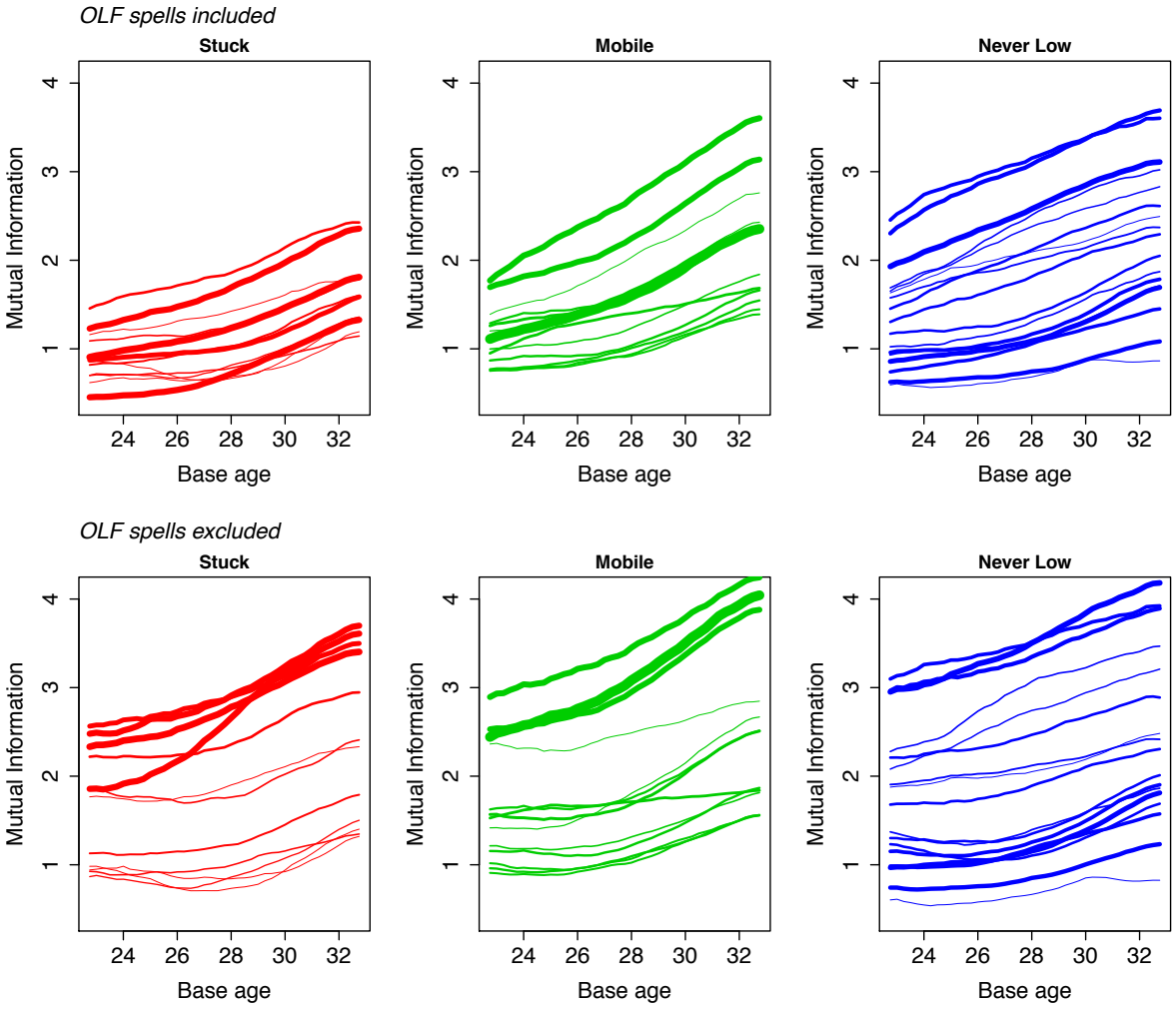
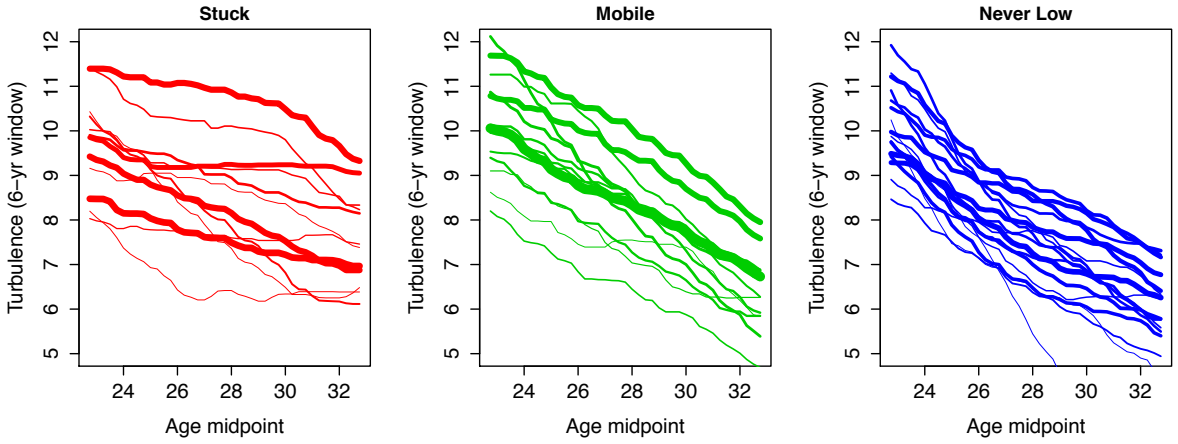


Figure 7 summarises turbulence by cluster within mobility group. The following analysis includes OLF spells.[xxi] While the overall decreasing trends resemble the aggregated analysis given in the third section of this paper (recall figure 4), the variation is again important to note, and it appears to be largest in the stuck group.[xxii] Several larger stuck clusters exhibit very little decline in turbulence over time, suggesting a form of complexity – these are the less attached clusters. At this point, there is mounting evidence that this *subgroup* within stuck careers is more chaotic, but movement in and out of the labour force appears to generate this. For some clusters within the stuck group, there may also be an absence of settling on a more focused path. However, the entropy analysis suggests that there are limitations to the types of jobs in stuck workers' careers – they do not traverse a wide range of job types, which is consistent with dual labour market theory (see Dickens & Lang, 1985; Hudson, 2007; Osterman, 1975; Piore, 1983). This suggests that 'chaotic' is too simplistic a label; less-attached, with reasonably limited exploration, net of this, seems closer to what is observed through this analysis.

Figure 7. Turbulence over time by cluster within mobility group



Discussion

Our characterisation of mobility-based labour markets using sequence-analytic and information-theoretic measures reveals that multiple levels of analysis are important to our understanding of the structure of stuck (and more mobile) careers. We evaluated several research questions regarding the mobility prospects of U.S. workers during a period of increasing wage inequality. One of the key questions was whether the careers of stuck workers were more or less chaotic than those of their more mobile counterparts. Our analysis suggests that a substantial subgroup has weak attachment to the labour force, more movement to and from job types *including types of non-work*, and that these careers are less predictable within any given person. Yet cluster analysis also reveals a sizable set of career types that are highly attached, less chaotic, and which have natural, content-driven ties to career paths associated with more mobile workers, consistent with Appelbaum, et al. (2006). Notably, the stuck group utilises a smaller 'alphabet' of job IxOs, making the type of work more predictable in aggregate. This highlights an important methodological contribution of this research; namely, that the level of analysis, choice of descriptive measure, and the temporal aspect of these must be customised to address the research questions of interest.

Examples of customisation include the following. Two very good existing measures of sequence structure, entropy and turbulence, initially provided very different characterisations, and turbulence was much more informative once we evaluated it using a running window approach (this revealed the natural trend of decreasing turbulence over the career). Entropy as a measure of complexity is too coarse an instrument, but when disaggregated to career clusters, it corresponds more closely to the notion of 'career lines' (see Mouw & Kalleberg, 2006; Spilerman, 1977) or movement to related forms of work, particularly for the more mobile workers. We also introduced an information-theoretic measure, MI, and evaluated it within-sequence for a moving time window. This provided a measure of predictability of future IxOs from the current subsequence; MI appears to be underutilised in the literature. Through the lens of MI, under different levels of granularity (aggregate and cluster-specific), our findings suggest that stuck workers' careers are harder to predict at short and long intervals, while this is only the case in some more mobile career clusters.

This last point highlights the often-neglected feature of careers specifically, and sequences more generally; namely, that substantial heterogeneity exists within them. By employing an OMA-based categorical clustering method to each mobility group, we managed to isolate fairly homogeneous clusters within them. The clusters had face validity, in that industries and occupations paired in a familiar manner (e.g., defining police officers, or nurses,

implicitly) in most clusters. Applying the same measures (particularly MI) to the cluster-based subgroups revealed that the never low group has substantially more variability in its (within-cluster) predictability, which was quite unexpected.

By generating clusters, grouping together careers with similar IxO patterns, we have refined our conception of these different labour markets. We have demonstrated many ways in which these mobility groups differ with respect to their IxO progression. We have also found significant similarities between the groups on sequence-analytic measures. This is due in part to the wide variation uncovered through clustering. The clusters, in turn, form many matched trajectory types across mobility groups, providing an avenue for future comparative analysis. We note that we have mentioned demographic differences between mobility groups but not between the career *clusters*. For example, the gender of the worker is likely to be important as we examine mobility differences across matched career types. Further examination of the similarities and differences between these matched trajectory types should lead to deeper understanding of the determinants of mobility in the modern U.S. labour market.

Acknowledgements

The authors wish to thank Annette Bernhardt, Laura Dresser and Joel Rogers for their support and intellectual contribution to this work, as well as the Annie E. Casey Foundation, which provided grant support. The authors also wish to thank the editors and anonymous reviewers for their insights and suggestions.

References

Abbott, A. (1995). Sequence Analysis: New Methods for Old Ideas. *Annual Review of Sociology*, 21, 93-113. http://dx.doi.org/10.1146/annurev.so.21.080195.000521

Abbott, A. & Tsay, A. (2000). Sequence analysis and optimal matching methods in sociology review and prospect. *Sociological Methods & Research*, *29*(1), 3-33. http://dx.doi.org/10.1177/0049124100029001001

Andersson, F., Holzer, H.J. & Lane, J.I. (2004). *Moving Up or Moving On: Who Advances in the Low-Wage Labor Market?* New York: Russell Sage Foundation.

Appelbaum, E., Bernhardt, A.D. & Murnane, R.J. (2006). *Low-Wage America: How Employers Are Reshaping Opportunity in the Workplace*. New York: Russell Sage Foundation.

Becker, H. S. & Carper, J. (1956). The elements of identification with an occupation. *American Sociological Review*, 341-348. http://dx.doi.org/10.2307/2089290

Bernhardt, A., Morris, M., Handcock, M. S. & Scott, M. A. (1999). Trends in Job Instability and Wages for Young Adult Men. *Journal of Labor Economics*, 17(4), S65–S90. http://dx.doi.org/10.1086/209943

Bernhardt, A., Morris, M., Handcock, M. & Scott, M. (2001). *Divergent Paths: Economic Mobility in the New American Labor Market*. New York: Russell Sage Foundation.

Bernstein, J. & Hartmann, H. (2000). Defining and Characterizing the Low-Wage Labor Market. In K. Kaye & D. Smith Nightingale (Eds.), *The Low-Wage Labor Market: Challenges and Opportunities for Economic Self-Sufficiency* (pp. 15-40). Washington, DC: Urban Institute Press

Billari, F. C. (2001). The analysis of early life courses: Complex descriptions of the transition to adulthood. *Journal of Population Research*, 18(2), 119–142. http://dx.doi.org/10.1007/BF03031885

Bison, I. (2014). Sequence as network: An attempt to apply network analysis to sequence analysis. In P. Blanchard, F. Bühlmann & J.A. Gauthier (Eds.), *Advances in Sequence Analysis: Theory, Method, Applications*. (pp. 75-103). New York: Springer. http://dx.doi.org/10.1007/978-3-319-04969-4_12

Blanchard, P., Bühlmann, F. & Gauthier, J.-A. (Eds.). (2014). *Advances in Sequence Analysis: Theory, Method, Applications*. New York: Springer.

Boushey, H. (2005). No Way Out: How Prime-Age Workers Get Trapped in Minimum Wage Jobs. *Briefing Paper.* Washington, DC: Center for Economic and Policy Research.

Bureau of Labor Statistics, U.S. Department of Labor. (2000). National Longitudinal Survey of Youth 1979 cohort, 1979-2000 (rounds 1-19) [computer file]. Produced and distributed by the Center for Human Resource Research, The Ohio State University. Columbus, OH.

Bureau of Labor Statistics, U.S. Dept. of Labor. (2015). Accessed June 18, 2015 from http://www.bls.gov/cpi/cpirsdc.htm.

Clawson, D. & Gerstel, N. (2014). *Unequal Time: Gender, Class, and Family in Employment Schedules*. New York: Russell Sage Foundation.

Cornwell, B. (2015). *Social Sequence Analysis: Methods and Applications.* New York: Cambridge University Press.

Danziger, S.H., Blank, R.M. & Schoeni, R.F. (2008). *Working and Poor: How Economic and Policy Changes Are Affecting Low-Wage Workers*. New York: Russell Sage Foundation. http://dx.doi.org/10.1017/CBO9781316212530

Dickens, W.T. & Lang, K. (1985). A Test of Dual Labor Market Theory. *American Economic Review*, 75(4), 792-805.

Durbin, R., Eddy, S., Krogh, A. & Mitchison, G. (1998). *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. New York: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511790492

Elzinga, C. H. (2010). Complexity of Categorical Time Series. *Sociological Methods & Research*, 38(3), 463–481. http://dx.doi.org/10.1177/0049124109357535

Elzinga, C. H. & Liefbroer, A. C. (2007). De-standardization of Family-Life Trajectories of Young Adults: A Cross-National Comparison Using Sequence Analysis. *European Journal of Population*, 23(3-4), 225–250. http://dx.doi.org/10.1007/s10680-007-9133-7

Elzinga, C. H. & Studer, M. (2015). Spell Sequences, State Proximities, and Distance Metrics. *Sociological Methods & Research*, 44(1), 3–47. http://dx.doi.org/10.1177/0049124114540707

Everitt, B. S., Landau, S., Leese, M. & Stahl, D. (2001). *Cluster Analysis (4th Ed.),* London: Arnold. http://dx.doi.org/10.1201/9781420057492.ch10

Fasang, A. E. & Liao, T. F. (2013). Visualizing Sequences in the Social Sciences Relative Frequency Sequence Plots. *Sociological Methods & Research*, 43(4).

French, E., Mazumder, B. & Taber, C. (2007). The changing pattern of wage growth for low skilled workers. In R.M. Blank, S. Danziger, & R.F. Schoeni (Eds.), *Working and poor: How economic and policy changes are affecting low-wage workers* (pp. 141-72). New York: Russell Sage Foundation.

Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and Visualizing State Sequences in R with TraMineR. *Journal of Statistical Software*, 40(4), 1-37. http://dx.doi.org/10.18637/jss.v040.i04

Gardecki, R., & Neumark, D. (1998). Order from Chaos? The Effects of Early Labor Market Experiences on Adult Labor Market Outcomes. *Industrial and Labor Relations Review*, 51 (2), 299-322. http://dx.doi.org/10.1177/001979399805100209

Gottschalk, P., & Moffitt, R. (1994). The Growth in Earnings Instability in the U.S. Labor Market. *Brookings Papers on Economic Activity,* 2: 217-54. http://dx.doi.org/10.2307/2534657

Halpin, B. (2014). Three narratives of sequence analysis. In P. Blanchard, F. Bühlmann & J.A. Gauthier (Eds.), *Advances in Sequence Analysis: Theory, Method, Applications*. (pp. 75-103). New York: Springer. http://dx.doi.org/10.1007/978-3-319-04969-4_5

Hartigan, J.A. & Wong, M.A. (1979). A K-Means Clustering Algorithm. *Applied Statistics,* 28(1): 100–108. http://dx.doi.org/10.2307/2346830

Hennig, C. & Liao, T. F. (2013). How to find an appropriate clustering for mixed-type variables with application to socio-economic stratification. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62(3), 309–369. http://dx.doi.org/10.1111/j.1467-9876.2012.01066.x

Hoff, P., Raftery, A.E. & Handcock, M. (2002). Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association*, 97, 1090-1098. http://dx.doi.org/10.1198/016214502388618906

Hope, A. C. A. (1968) A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society, Series B,* 30, 582–598.

Hudson, K. (2007). The new labor market segmentation: Labor market dualism in the new economy. *Social Science Research*, 36, 286–312. http://dx.doi.org/10.1016/j.ssresearch.2005.11.005

Kalleberg, A. L. (2009). Precarious work, insecure workers: Employment relations in transition. *American Sociological Review*, 74(1), 1-22. http://dx.doi.org/10.1177/000312240907400101

Kass, R.E., & Raftery, A.E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90(430), 773-795. http://dx.doi.org/10.1080/01621459.1995.10476572

Kaufman, L. & Rousseeuw, P.J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley. http://dx.doi.org/10.1002/9780470316801

Lesnard, L. (2014). Using Optimal Matching Analysis in Sociology: Cost Setting and Sociology of Time. In P. Blanchard, F. Bühlmann, & J.A. Gauthier (Eds.), *Advances in Sequence Analysis: Theory, Method, Applications*. (pp. 39-50). New York: Springer. http://dx.doi.org/10.1007/978-3-319-04969-4_3

Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy*, 66(4), 281-302. http://dx.doi.org/10.1086/258055

Mishel, L., Bernstein, J., & Allegretto, S. (2007). *The State of Working America 2006/2007*. Ithaca, NY: ILR Press.

Mouw, T. & Kalleberg, A. L. (2006). Intragenerational Mobility, Careers, and Task-Specific Human Capital. Paper presented at the annual meeting of the American Sociological Association, Montreal Convention Center, Montreal, Quebec, Canada. Accessed online 29 December 2015, from http://citation.allacademic.com//meta/p_mla_apa_research_citation/1/0/4/7/3/pages104734/p104734-1.php

Murphy, K.M. & Welch F. (1990). Empirical Age-Earnings Profiles. *Journal of Labor Economics* 8(2): 202-229. http://dx.doi.org/10.1086/298220

Neumark, D., Polsky, D., & Hansen, D. (2000). Has Job Stability Declined Yet?: New Evidence for the 1990s. In D. Neumark, ed., *On the Job: Is Long-Term Employment a Thing of the Past?* (pp. 70–110). Russell Sage Foundation.

Osterman, P. (1975). An empirical study of labor market segmentation. *Industrial and Labor Relations Review*, 28(4), 508-523. http://dx.doi.org/10.1177/001979397502800402

Piccarreta, R. (2012). Graphical and Smoothing Techniques for Sequence Analysis. *Sociological Methods & Research*, 41(2). http://dx.doi.org/10.1177/0049124112452394

Pierce, J. R. (1980). *An Introduction to Information Theory: Symbols, Signals and Noise*. New York: Dover Books.

Piore, M.J. (1983). Labor Market Segmentation: To What Paradigm Does It Belong?
The American Economic Review, 73(2), 249-253.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association,* 66 (336): 846–850. http://dx.doi.org/10.1080/01621459.1971.10482356

Robinson, G.K. (1991). That BLUP is a Good Thing: The Estimation of Random Effects. *Statistical Science*, 6(1), 15-32. http://dx.doi.org/10.1214/ss/1177011926

Sankoff, D., & Kruskal, J.B. (1983). Time *Warps, String Edits, and Macromolecules*. Reading MA: Addison Wesley.

Sawhill, E., & McMurrer, D. (1998). *Getting Ahead: Economic and Social Mobility in America.* Washington, DC: Urban Institute Press.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), 461-464. http://dx.doi.org/10.1214/aos/1176344136

Scott, M. A. (2011). Affinity models for career sequences. *Journal of the Royal Statistical Society: Series C (Applied Statistics),* 60(3), 417–436. http://dx.doi.org/10.1111/j.1467-9876.2010.00752.x

Spilerman S. (1977). Careers, Labor Market Structure, and Socio-Economic Achievement. *American Journal of Sociology,* 83(3), 551-593. http://dx.doi.org/10.1086/226595

Strawn, J., & Martinson, K. (2000). Promoting Access to Better Jobs: Lessons for Job Advancement from Welfare Reform. In Kazis, R. & Miller, M. S. (eds.), *Low-Wage Workers in the New Economy*, Washington, D.C.: The Urban Institute Press.

Studer, M. & Ritschard, G. (2016). What matters in differences between life trajectories: a comparative review of sequence dissimilarity measures. *Journal of the Royal Statistical Society: Series A.* Published online: 14 JUL 2015. http://dx.doi.org/10.1111/rssa.12125

Topel, R.H. (1993). What Have We Learned from Empirical Studies of Unemployment and Turnover? *The American Economic Review*, 83 (2), Papers and Proceedings of the 105th Annual Meeting of the American Economic Association, 110-115.

Topel, R.H. & Ward, M.P. (1992). Job Mobility and the Careers of Young Men. *The Quarterly Journal of Economics*, vol. 107(2), 439-79. http://dx.doi.org/10.2307/2118478

Tukey, J. W. (1977). *Exploratory Data Analysis*, Reading Massachusetts: Addison-Wesley.

Vingron, M., & Waterman, M.S. (1994). Sequence Alignment and Penalty Choice. *Journal of Molecular Biology*, 235:1-12. http://dx.doi.org/10.1016/S0022-2836(05)80006-3

Weber, M. (2009). *The theory of social and economic organization*. New York: Simon and Schuster.

Wu, L. L. (2000). Some comments on 'Sequence analysis and optimal matching methods in sociology: Review and prospect.' *Sociological Methods & Research*, *29*(1), 41-64. http://dx.doi.org/10.1177/0049124100029001003

Appendix A. Sensitivity to industry and occupation codes

Entropy and any other metrics, as they are defined, depend on the industry and occupation coding scheme. For example, if a single clerical occupation could be further refined into two clerical sub-occupations, this would change entropy, and the changes could vary by mobility group. We might be concerned that our finding that stuck workers have lower entropy over time is driven, somehow, by the coding scheme itself. In developing our coding scheme, we took care to collapse relatively homogeneous occupations and industries. We used census-based codes, which attempt to organise the market by what is being produced and the skill set involved in the production (see Weber, 2009).

Lower-skilled work tends to be, at least occupationally, less differentiated. In part, this is because there is simply less complexity in the jobs themselves. However, we have direct evidence that this, in and of itself, is not driving our findings of lower IxO entropy for stuck workers. We re-analysed entropy using only our industry codes and then did this again using occupation codes. For stuck workers, entropy quickly grows over time so that by their mid-20s, stuck workers have the second highest levels of industry-based entropy and the highest levels of occupation-based entropy. Rather than being an artefact of the coding scheme, entropy for stuck workers shows great variety, depending on the job characteristic being evaluated.

Appendix B. Abbreviations used in the figures and tables

Labour Force Status	
Abbreviation	Description
OLF/Enr	Enrolled in school
Unknown	Unknown
OLF	Out of the labour force
Unemp	Unemployed
Wkg_Msg_IxO	Working (IxO not available)*

* Excluded from figure 1 state distribution plot

Industries	
Abbreviation	Description
agr&	Agriculture, forestry, fishing, mining
cnst	Construction
mnfD	Manufacturing durable
mnfN	Manufacturing non-durable
tran	Transportation
comm	Communications
util	Utilities, sanitary services
whol	Wholesale trade
retH	Retail hard goods (except automotive)
retF	Retail food
retA	Retail automotive
eatD	Eating & drinking
FIRE	Finance, insurance, real estate
bzSv	High end business services
bldS	Building services
tmps	Temporary agencies
autS	Automotive & repair services
hotl	Hotels & laundry
prsl	Personal services
recr	Entertainment, recreation services
hlth	Healthcare
educ	Education
nonp	Non-profit
prfS	Professional services
PAdm	Public administration

Occupations	
Abbreviation	Description
Prof	Professionals (Drs., lawyers, etc.)
Engr	Engineers, scientists, engineering technicians, non-health technicians
Farm	Farm labourers, farm foremen
RN&	Nurses, dieticians, therapists
Tech	Health technologists, technicians
SocW	Religious, social scientists, social workers
Tchr	Teachers
WrAr	Writers, artists, entertainers
Mgr&	Managers, administrators
Sale	Sales workers
Cler	Clerical, unskilled workers
Crft	Craft workers and mechanics
Oper	Operatives
Labr	Labourers except farm
Clng	Cleaning service workers
Food	Food service workers
HlSv	Health service workers
PrSv	Personal service workers
Prot	Protective service workers
HHWk	Private household workers

Endnotes

[i] By this stage of the career, *two-thirds* of wage growth has occurred for the bulk of workers (Topel & Ward, 1992).

[ii] When estimating entropy from a sample, we drop tokens that never occur.

[iii] Within-sequence entropy could be computed and then summarised at various levels of grouping – this would more closely resemble the turbulence measure that we use, which is sequence-based.

[iv] Some OMA implementations allow for minor variations on these assignments; we use the version implemented in the R package TraMineR.

[v] This mirrors the transition based approach implemented in TraMineR (in fact, we 'smooth' at the endpoints, as they do), allowing the implied substitution costs to vary with time.

[vi] We cannot use the k-means algorithm, as it relies on a multivariate continuous feature set, while we only have a distance matrix as input.

[vii] The cluster assignments are globally optimal, meaning that moving any sequence out of that assignment reduces the overall goodness of fit. Locally, however, a sequence might appear to 'belong' more to a nearby cluster even though moving it there would reduce the fit.

[viii] These are subjects who are free to seek work. Patients restricted to psychiatric hospitals, incarcerated prisoners and military personnel are considered 'institutionalized' in the NLSY.

[ix] The reasoning behind this is based in part on our unit of analysis, which is the sequence of industries and occupations held by each worker. The NLSY carefully reconstructs the work history for any subjects who miss

interviews, but the reconstruction is limited to the five most recent employers. Five employers are sufficient when the time between current and last interval is short. Across wide gaps, however, there is a concern that the occupation reported will mask prior occupations with the same employer. Thus, we adopt a moderately strict criterion for inclusion in the study.

[x] 1970 rather than 1980 Census codes were chosen because these are the only codes collected consistently over the survey span.

[xi] We discuss the sensitivity of the findings to the choices made in building the token alphabet in Appendix A.

[xii] The CPS job is defined as the current or most recent job. For multiple job holders, their CPS job is the job with the employer at which they usually work the most hours. It is so-called because the wording of the question in the NLSY is similar to that used in the Current Population Survey.

[xiii] The poverty line is a need-based assessment adjusted for household size and ages of children/dependents.

[xiv] This approach is well established in mobility analyses (e.g., Andersson et al., 2004; Sawhill & McMurrer, 1998).

[xv] More precisely, we include four labour force status tokens (OLF, unemployed, enrolled, and unknown) regardless of their frequency, and the remaining 26 are the most frequent job types.

[xvi] The term, job *shopping*, is used by labor economists to refer to the worker's search for a well-suited (skill-matched) employer. The related term, job *hopping*, refers to frequently changing employers, with unclear rationale.

[xvii] One or two years would yield relatively few job transitions, and this is a sequence-based measure.

[xviii] For the cluster-based re-assessment, we restrict the subjects to those who spend less than half of their time OLF, and this makes the computation feasible.

[xix] One could view this as a lack of structure in some mobility groups, but we view it more as a function of the methods used.

[xx] Note that the cluster with entropy that goes below two after age 25 represents police officers, who simply do not explore other careers over this time interval.

[xxi] Removal of OLF spells was non-trivial for the MI measure, in part because it created subsequences that were entirely missing. After resolving some of these technical concerns, the findings including and excluding OLF spells were found to be quite similar.

[xxii] The analysis excluding OLF spells revealed slightly greater heterogeneity in all groups, but it was most pronounced in the stuck group.

Understanding older adults' labour market trajectories: a comparative gendered life course perspective

Diana Worts University of Toronto, Canada
dworts@chass.utoronto.ca
Laurie Corna King's College London, UK
Amanda Sacker University College London, UK
Anne McMunn University College London, UK
Peggy McDonough University of Toronto, Canada

(Received November 2015 Revised April 2016) http://dx.doi.org/10.14301/llcs.v7i4.389

Abstract

The recent push to keep older adults in the labour force glosses over who is likely to follow what kind of employment trajectory and why. In this paper, we broaden understandings of later-life labour market involvement by applying a comparative gendered life course perspective. Our data come from the Survey of Health, Ageing and Retirement in Europe and the Health and Retirement Study (US), two representative panel studies of individuals aged 50-plus. Using a unique modeling strategy, we examine employment biographies for older women and men from four nations with diverse policy regimes (Germany, Italy, Sweden, and the US), along with their links to family experiences and earlier attachment to the labour force. We find that, in every nation, women prevail in groups representing a weak(er) attachment to the labour market and men in groups signifying a strong(er) attachment. However, this pattern is much stronger for Germany and Italy than for Sweden and the US. Similarly, both family experiences and prior employment matter more for later-life labour market involvement in Germany and Italy. Our findings demonstrate that older adults' employment trajectories are gendered; moreover, there is evidence that they are influenced by policies related not only to paid work but also to caregiving, and by those affecting not only current decisions but also those made earlier in the life course.

Keywords

Older adults; life course; employment; family; gender; welfare state; optimal matching

Introduction

In the wake of population ageing and the baby boom cohort's impending retirement, analysts have turned their attention to older adults' labour force participation (e.g., OECD, 2006). Yet, the dominant conceptual emphases—state-funded incomes that 'pull,' and workplace barriers that 'push,' older adults out of the labour force—often gloss over life course considerations that illuminate *who* is likely to follow *what kind* of later-life labour market trajectory and *why* (Cooke, 2006).

Studies framed by a gendered life course perspective offer a corrective. They focus on long-term outcomes, the continuity between experiences at younger and older ages, the role of family ties, and

the influence of institutional contexts (Moen & Flood, 2013). These factors especially shape the employment patterns of women and men (Gornick & Meyers, 2006). Yet, studies of older adults' labour market trajectories and the social and institutional processes that shape them are surprisingly rare (Börsch-Supan, Hank, Jürges, & Schröder, 2009). Most longitudinal research on later-life employment focuses on transitions—either withdrawal (e.g., Radl, 2013) or re-entry (e.g., Kail & Warner, 2013). While transitional events are important, they are but one element in a succession of labour market experiences that characterise the later years (Bowlby, 2007). This suggests that older adults' employment patterns are best modeled as extended *biographical sequences*, using methods that take the entire series as the unit of analysis (Billari & Piccarreta, 2005).

In this study, we draw on a gendered life course perspective, and begin to flesh out the *what*, *who*, and *why* of long-term patterns of later-life labour market involvement. Employing a unique modeling strategy that captures biographical sequences over ages 50-69, we examine their gendered correlates across four contrasting institutional contexts: Germany, Italy, Sweden, and the United States (US). We ask:

- How distributions across later-life labour market biographies differ for women and men in each country;
- Whether male-female differences in the distribution of these trajectories are the (gendered) outcome of family and prior work experiences; and
- How both these aspects vary across the four institutional contexts.

We begin with an outline of the gendered life course perspective, existing evidence, and our hypotheses. Following a description of our methods, the results section charts male-female differences in later-life labour market biographies for the four nations, then examines how family experiences and earlier attachment to the labour force contribute to these gender disparities. We conclude by discussing our findings and their policy implications.

Background

The gendered life course

Researchers increasingly argue that older adults' labour market involvement is best understood by drawing on a life course perspective (e.g., Alley, Putney, Rice, & Bengtson, 2010). This entails a focus on individuals' long-term circumstances in a given domain, such as employment (Mayer, 2004). It also encompasses the notions of biographical continuity and linked lives—the ideas that older adults' labour market trajectories follow from their paid work experiences at younger ages (Han & Moen, 1999a), and that those experiences are shaped by ties to family (Moen, 1996). Importantly, the division of labour by sex means that roles will often be gendered, at least earlier in the life course, with men assuming primary responsibility for breadwinning and women providing the bulk of care (Blossfeld & Drobnič, 2001).

The life course perspective also emphasises that biographies are shaped by institutional contexts—most obviously, welfare regimes (Mayer, 2004). These can be distinguished by the relative roles of the state, markets, and families in providing life's necessities (Esping-Andersen, 1990), and the model of family life implied by that balance (Korpi, Ferrarini, & Englund, 2013). The distinctions identified apply to the years when our sample members were in their prime working and childrearing years (Katz, 2010; Persky, 2011; Sundström, 2003). In *corporatist* regimes (typified by Germany), families are the primary providers of social welfare. The state supports a breadwinner-caregiver model of family life: full male employment with fairly generous benefits for periods of non-employment, and primary caregiving for women. Wives and mothers are assumed to work for pay only secondarily, if at all, and childcare spaces are limited in number and hours covered. In *southern European* nations (typified by Italy), breadwinner-caregiver families also play a central role in the provision of social welfare, leading to marked gender differences in employment (Siaroff, 1994). Social spending is generally low, and, at least historically, families have often maintained themselves as multi-generational households in which women care for one or more male breadwinners (Gal, 2010). In *liberal* regimes (typified by the US), social spending is also

low, but the distribution of resources is left to the market. The state passively backs a market-oriented dual-earner model, in which the majority of individuals (male and female) work for pay throughout their lives despite a dearth of public supports for caregiving. *Social democratic* nations (typified by Sweden) are distinguished by the state's prominent role in the distribution of resources and care provision. They actively support a dual-earner model—offering comprehensive publicly-funded caregiving services, fairly generous benefits for periods of non-employment, flexible work arrangements, and better-quality part-time jobs to sustain caregivers' labour force attachment (Halldén, Gallie, & Zhou, 2012). Table 1 (row 1) summarises these distinctions.

Welfare states are also distinguished by the extent to which their labour market policies aim to remove or retain older workers (Blossfeld, Buchholz, & Hofäcker, 2006; Blossfeld, Buchholz, & Kurz, 2011). (Table 1, row 2.) Again, the broad strokes of these approaches have remained quite stable during the years that affect our cohort (Blossfeld et al., 2006). Corporatist and southern regimes favour *employment exit*. High levels of wage replacement through public pensions and early retirement schemes—including, in Italy, relatively lenient access to illness/disability pensions for older workers (Beckstette, Lucchini, & Schizzerotto, 2006)—make early withdrawal attractive. Because these regimes tend not to integrate women into the labour force earlier in the life course, the implications of exit strategies for women's later-life labour market involvement, or for gender differences in that regard, are unclear.

Social democratic and liberal regimes are oriented toward *employment maintenance*. This takes two forms. The *public-induced* strategies of social democratic regimes combine active labour market policies (e.g., retraining and job protection) with generous publicly-funded pensions, to foster labour force attachment until—but not necessarily beyond—state pension age. Nevertheless, relatively easy access to disability benefits has provided some older Swedish workers with an early exit route (Lindquist, 2006). The *market-induced* approaches of liberal regimes blend minimal labour market regulation with meagre public pensions that encourage some older adults to remain employed past state pension age and/or if their work histories are not strong. Moreover, for Americans in this cohort health care coverage has been tied to employment (or the receipt of social assistance/disability benefits). This may motivate many older workers to remain in the labour force until they qualify for Medicare at age 65. Both types of maintenance strategies also foster part-time work in later life. Public-induced approaches can include the option to 'phase out' while receiving a generous partial pension; in market-induced regimes, part-time work may reflect the combined influence of minimally regulated markets and financial need and/or the greater flexibility afforded those with more personal resources.

Table 1: Institutional contexts for breadwinning/caregiving, and later-life labour market involvement

	Germany	Italy	Sweden	US
Breadwinning & caregiving earlier in the life course[a]	• Families are primary providers of social welfare; state supports are strong • Breadwinner-caregiver model of family life • Support for full male employment; fairly generous benefits for periods of non-employment • Little/no direct support for mothers' (full-time) employment	• Families are primary providers of social welfare; state supports are weak • Breadwinner-caregiver model of family life • Fairly ungenerous benefits for periods of (male) non-employment • Little/no support for mothers' employment	• State plays a role in social welfare provision & employment supports • Publicly-oriented dual-earner model of family life[b] • Active commitment to full adult employment • Fairly generous benefits for periods of non-employment • Publicly-funded child care; better-quality part-time jobs	• State's role in both social welfare & employment supports is minimised • Market-oriented dual-earner model of family life • Passive promotion of full adult employment • Ungenerous social safety net • Little/no public support for family responsibilities
Later-life labour market involvement	• Employment exit: high levels of wage replacement through public pensions and attractive early retirement schemes[c] • Official retirement age: 65[d] • Earliest retirement age: 63 for men, 60 for women[d]	• Employment exit: high levels of wage replacement through public pensions and attractive early retirement schemes[c] • Official retirement age: 60 for men, 55 for women[d] • Earliest retirement age: 57 (for men)[d]	• Employment maintenance (public-induced): generous public pensions; only moderate incentives for early retirement[c] • Official retirement age: 65[d] • Earliest retirement age: 61[d]	• Employment maintenance (market-induced): low public pensions; few/moderate incentives for early retirement[c] • Official retirement age: 65[d] • Earliest retirement age: 62[d]

[a] Adapted from Korpi (2000) and Trifiletti (1999: 54, 56).
[b] In a more recent article, Korpi, Ferrarini, and Englund (2013) label Sweden's model of family life "earner-carer," due to the introduction of policies designed to encourage male caregiving. However, these policies would not have applied during the years when our respondents were in their prime child-rearing years.
[c] Adapted from Buccholz, Hofacker, and Blossfeld (2006).
[d] Data sources: Duval (2003: 35) and OECD (2005, 2006, 2009). There will be variation in some of these ages in some countries (e.g., Italy) based on individual work histories. State pension ages listed are those that apply to our birth cohort.

Although broad approaches to older workers differ across welfare regimes, state pension ages for our cohort are generally 65, with uptake possible in the early 60s (table 1, row 2). The exception is Italy, where both normal and early pension ages are lower—in the former case, especially for women.

The framework of Blossfeld and colleagues (2006, 2011) suggests that later-life labour market trajectories will vary cross-nationally despite similar state pension ages—though the extent and nature of gendering is unclear. Esping-Andersen-based typologies imply that institutional contexts may gender paid and family work earlier in the life course; but how this plays out at older ages is less apparent.

Existing research

Cross-sectional and time-series data show variations in older adults' labour force participation in line with welfare regime theories. Rates are lowest in Italy, followed by Germany, then the US, and highest in Sweden (OECD, 2011). In all four countries, participation is higher for older men than for their female counterparts; but the gap is largest in Italy, followed closely by Germany, intermediate in the US, and narrowest in Sweden (Kahn, 2010: 30). Among those in the labour force, women exit earlier than men in all four nations, although the gap is somewhat smaller in the US and Sweden than in Germany and Italy (OECD, 2011). Later working (ages 65–69) is least common in Germany and Italy, more frequent in Sweden, and relatively widespread in the US (OECD, 2011). Part-time work plays a significant role in the lives of older women (but not men) in Germany and Sweden (Morris & Mallier, 2003), is rare for older Italians of both sexes (Morris & Mallier, 2003), and is relatively common among older Americans—especially, but not exclusively, women (US Bureau of Labor Statistics, 2000).

These aggregate-level data cannot, however, reveal *long-term biographical* patterns or uncover their correlates and antecedents. Research using panel data provides clues, but tends to examine single transitions, such as retirement timing (e.g., Radl, 2013; Warner & Hofmeister, 2006) or re-entry (e.g., Kail & Warner, 2013; Pleau, 2010). This emphasis is problematic from a gender perspective, since it excludes anyone without a reasonably steady work history—that is, many women. By contrast, an approach based on extended labour market sequences—regardless of status at a given age—can include all older adults in all countries, and fully capture cross-national differences in the gendering of later-life labour market involvement.

Few studies consider the extended sequences of older adults; and we know of only one that models employment *per se* for any of the countries in our analysis. Han and Moen (1999b) assess work histories from age 30 to retirement, and identify five distinct biographies: delayed-entry, orderly, fast-track, steady part-time, and intermittent. Their study is, however, based on a small non-representative American sample, making their results difficult to generalise. In two loosely related studies, Fasang (2010, 2012) uses nationally representative data to analyse income source sequences of older Germans and Britons. She concludes that, in both countries, men tend to follow institutionalised pathways to retirement, while women predominate in part-time work and non-employment trajectories. Although Fasang speculates that gendered family roles shape women's pathways, she does not test this hypothesis or investigate whether the influence of family experiences differs in the two countries.

Many studies explore the effects of family circumstances on older adults' employment or exit probabilities—typically focusing on current marital status. For Americans, being married decreases older women's attachment to the labour force (e.g., Moen & Flood, 2013), but increases that of older men (e.g., Williamson & McNamara, 2003). Being in a couple has no impact on the odds of employment for German men, but reduces them for German women; makes no difference for Swedish women, but increases employment for Swedish men; and reduces employment for Italian men and, particularly, Italian women (Ogg & Renaut, 2007). However, the life course perspective's emphasis on long-term circumstances suggests that the critical factor may be marital (in)stability, rather than current marital status. Certainly, the loss of a partner ends or compromises the pooling of resources (Wilmoth & Koso, 2002) and the gendering of breadwinning and caregiving, and might reasonably be expected to promote employment, especially for women.

There is some evidence on the influence of parenthood. Being a mother delays labour market

withdrawal for German and American women, as does later age at the time of the first birth (Hank, 2004; Moen & Flood, 2013; Pienta, 1999). Conclusions are mixed regarding the links between labour force participation and having dependent children at older ages. The association may be positive for men and negative for women (Dentinger & Clarkberg, 2002; Pienta & Hayward, 2002); negative for women only (Moen & Flood, 2013; Szinovacz, DeViney, & Davey, 2001); or nonexistent (Denaeghel, Mortelmans, & Borghgraef, 2011). Importantly, though, results are for single countries (or do not test country interactions), and thus cannot reveal cross-national distinctions in effects. In one cross-national study, Hank and Korbmacher (2013) find that the positive association between being a mother and leaving employment in later life is specific to corporatist regimes for our cohort — although their sample was restricted to those with at least some paid work from age 50.

Regardless, the notion of biographical continuity suggests that earlier work histories — themselves linked to family experiences — may be decisive. Among employed married Americans aged 50-75, men (but *not* women) with longer work histories are more likely to leave the labour market (Szinovacz & DeViney, 2000). Older American women (married or not) are more likely to be *in* the labour force if they had fewer employment interruptions earlier in the life course, at least for the cohort born in the 1920s (Pienta, Burr, & Mutchler, 1994).

These studies pertain to the US only. Contrasting institutional contexts mean that prior employment has varied cross-nationally, especially among women — implying that its impact on the gendering of later-life labour market involvement may also vary. In Germany, long maternity leaves, a lack of childcare facilities, and a tax system favouring breadwinner-caregiver families have encouraged women to withdraw or work part-time after marrying or having children (Drobnič, Blossfeld, & Rohwer, 1999; Hofäcker, Stoilova, & Riebling, 2013; Sundstrom, 2003). For Italians, the combination of relatively low education, a lack of part-time work, an underdeveloped service sector, limited childcare, and strong norms around the sexual division of labour often persuaded women with family responsibilities to leave the labour force (Rydell, 2002; Sundström, 2003). In Sweden, where daycare spending has been high, work-family reconciliation policies relatively comprehensive, equal pay legislation long-standing, and the service sector large (Rovny, 2011: 341), women tended to remain employed during the prime childrearing years (Sundström, 2003). How this plays out at older ages is uncertain: Hank and Korbmacher (2013) conclude, based on data from 13 European countries, that length of prior employment is positively associated with the odds of leaving the labour force, but they do not consider gender differences, either within or between nations.

In sum, a number of gaps are evident in existing knowledge of older adults' labour market biographies. The emphasis on transitions underrepresents or excludes women; and the few studies that consider long-term trajectories do not focus on employment *per se*, are not generalisable, lack a comparative dimension, and/or do not take gender into account. Beyond this, comparative evidence on the roles of family and paid work experiences earlier in the life course is sparse, mixed, and not always available for both sexes. In what follows, we begin to flesh out these aspects of older adults' labour market biographies by testing four hypotheses:

Hypothesis 1: Distributions for later-life labour market biographies will be *gendered*—with men concentrated in more 'attached' (mostly full-time; shorter stretches of non-employment), and women in less 'attached' (stretches of part-time; longer periods of non-employment), trajectories.

Hypothesis 2: Male-female distinctions in later-life labour market biographies will be *stronger in Germany and Italy* than in Sweden and the US. This may be especially true for Italy, given the gender gap in pension eligibility.

Hypothesis 3a: Gender differences in older adults' employment trajectories will be partly attributable to *gender-specific influences of family experiences* over the life course. *Hypothesis 3b*: The impact of family experiences will be *stronger in Germany and Italy than in Sweden and the US*.

Hypothesis 4a: The gendering of older adults' labour market trajectories will be at least partly explained by *gendered earlier attachment to the labour force*. *Hypothesis 4b*: The explanatory power of work histories will be *stronger in Germany and Italy than in Sweden and the US*.

Methods

Data

Our data come from two nationally representative panel studies: the Survey of Health, Ageing and Retirement in Europe (SHARE); and the Health and Retirement Study (HRS) (Börsch-Supan & Jürges, 2005; Heeringa & Connor, 1995). Both surveys sample individuals aged 50+, and include detailed employment/activity histories, as well as key socio-demographic variables. SHARE covers more than 85,000 individuals from 21 countries, over a maximum, to date, of four waves: 2004/5, 2006/7, 2008/9, and 2011/12. We restrict analyses to those born in the 1930s or early 1940s (controlling for historical period and matching the HRS sample) and living in one of three countries of interest—(former West) Germany, Italy, and Sweden. HRS has interviewed more than 26,000 Americans aged 50+, biennially since 1992 and most recently in 2012. We use the original cohort, and only those who were age-eligible according to HRS criteria (born 1931-1941). From both studies, we select those with information on at least ten of the 20 possible labour market state variables over ages 50-69 (N = 475 for Germany, 935 for Italy, 654 for Sweden, and 8,832 for the US). And in each case, we utilise current and retrospective information on employment/activity start and end dates for all years in which respondents were aged 50-69 years, along with retrospective information on work, marital, and fertility histories up to age 50, and data on gender and other relevant socio-demographics.

Measures

Later-life labour market biographies

Our study takes a novel approach to modelling later-life labour market involvement—maintaining the integrity of individuals' *long-term* sequences and classifying them using dynamic Hamming distances (Lesnard, 2010) to make analysis feasible. The *Analysis* section provides details on the construction of these employment biographies. The resulting variable comprises seven trajectory 'types' summarising patterns over ages 50-69: 1 = Full-time throughout; 2 = Full-time, exit around 65 (reference); 3 = Full-time, exit around 60; 4 = Full-time, exit around 55; 5 = Non-employed throughout; 6 = Part-time dominant; and 7 = Other (table 2). Note that, because we match each individual to their *closest* model biography, groups are not internally homogeneous; hence, the table and variable labels describe categories as comprising those who were *mostly* employed full-time, part-time or not at all, and/or who exited *around* age 65, age 60, etc.

Gender distinguishes women (yes = 1; no = 0) from men.

Mid-life and prior family experiences comprise two parental measures—*dependent children at age 50* (yes = 1; no = 0) and *age became a parent* (in years)—and two marital measures—*married at age 50* (yes = 1; no = 0) and *marital ending by age 49* (yes = 1; no = 0). In each case, one measure represents mid-life (age 50) circumstances and one captures prior history. Both marital measures are constructed from the starts and ends of current and prior marriages. Dependent children are those under age 19. Age at the birth of the first child is conditionally-coded (Ross & Mirowsky, 1992: 223-224) to permit the retention of non-parents. As such, the measure is centered on its country-specific sample mean and entered into models in conjunction with a dummy for *parent* (yes = 1; no = 0). Thus, individuals without children form the reference group, and those with children differ from them by an average amount (the coefficient for *parent*) plus an amount based on the age at which they entered parenthood (the coefficient for *age became a parent*).

Table 2. Description of the seven later-life labour market trajectories

Later-life labour market trajectory	Employment pattern age 50-69
1. Full-time throughout	Mostly employed full-time ages 50-69
2. Full-time, exit 65	Mostly employed full-time to approximately age 64, exit approximately age 65
3. Full-time, exit 60	Mostly employed full-time to approximately age 59, exit approximately age 60
4. Full-time, exit 55	Mostly employed full-time to approximately age 54, exit approximately age 55
5. Non-employed throughout	Mostly non-employed ages 50-69
6. Part-time dominant	Mostly employed part-time with exits at various ages, or 'downshifting' from full- to part-time
7. Other (residual)	Erratic or unclassified pattern

Prior employment uses starts and ends for employment periods to code total *years employed over ages 15-49*. Distributions are equivalised by cutting at the values nearest each within-country quintile threshold, to produce a five-category ordinal variable.

Controls consist of three variables likely to be associated with gender, family experiences, and labour market involvement: education (low = 1, medium = 2, high (reference) = 3); self-rated health at age 50 (1 = excellent, 2 = very good; 3 = good; 4 = fair; 5 = poor), treated as ordinal in the models; and, for the US, 'minority' status (black/Hispanic = 1; non-Hispanic white = 0).

Table 3 gives weighted means and percentages, by gender and country, for family experiences and prior employment. Of note are the comparatively large proportion of Italians (especially men) with dependent children and/or no marital ending at/by midlife, and the large share of Americans with a marital ending by midlife. These observations are broadly consistent with OECD data for the decades when our samples were under 50, which show low/non-existent divorce rates in Italy, comparatively high rates in the US, and relatively late mean ages at childbirth among Italian women (presumably even later for Italian men, given the gender age gap at first birth) (OECD, 2015a). In addition, while women in all four nations worked less than men in the years before age 50, the size of the gap ranged from less than six years for Sweden, through nearly ten for the US and a little more than 11 for Germany, to 17 for Italy. The relative magnitude of these gaps is in line with population-level data on men's and women's prime-age labour force participation during the years relevant to our cohort (OECD, 2015b).

Table 3. Weighted means and percentages for family and employment history measures, by gender and country[a]

	Germany	Italy	Sweden	United States
Married at age 50 (%)	87.1	88.4	81.0	78.2
Women	87.6	87.3	78.5	75.5
Men	86.6	89.8	83.9	81.1
Dependent child(ren) at age 50 (%)	40.2	53.1	43.2	39.0
Women	32.8	45.4	33.9	33.2
Men	48.8	62.4	53.5	45.2
Ever had a child (%)	91.4	89.0	87.7	82.1
Women	93.6	90.0	87.9	85.9
Men	88.9	87.7	87.5	77.9
Age in years at first child (if parent,)	26.4	27.2	25.9	24.5
Women	24.7	25.6	24.7	22.8
Men	28.6	29.1	27.2	26.4
Ever had a marriage end by age 49 (%)	12.8	7.0	20.8	34.5
Women	13.8	9.0	21.7	36.5
Men	11.6	4.5	19.8	32.4
Years employed ages 15-49	23.6	21.6	27.4	24.0
Women	18.2	13.9	24.8	19.3
Men	29.9	30.8	30.3	29.0

[a] Weighted using the population/design/attrition weights supplied with the surveys.

Analysis

Creating and grouping labour market sequences

To construct the late-life labour market biographies, we first coded, for each of 20 time points (ages 50-69), a set of categorical variables representing whether the respondent was primarily employed full-time, part-time, or not at all. These 20 age-specific variables constitute the later-life labour market sequences. To make analysis feasible while also retaining the integrity of these biographies, we then grouped them using dynamic Hamming distances. Dynamic Hamming distances quantify how alike or distinct individual biographies are, one from another (Abbott and Tsay, 2000), by measuring the 'cost' of converting one person's sequence to another's (MacIndoe and Abbott, 2004). Sequences can then be grouped on that basis. We initially calculated distances relative to a set of 12 model biographies, constructed using existing knowledge of work and retirement among older adults. Because numbers were too small to support further analysis for some groups in the SHARE samples, we subsequently combined the seven biographies involving stretches of part-time work (leaving the remaining five as they were), to create a six-category measure for all countries. We then generated a seventh group by pulling, from the original 12, those cases deemed a less than ideal match to any reference sequence, based on their distance from the one to which they were closest. Initial model sequences were developed separately by all authors, and later inspected for overlap (found to be substantial) and unique patterns of theoretical interest.

The validity of the 12-group classification was checked by examining between-group heterogeneity

and within-group homogeneity — both considerable — using information on individuals' 'own-group' distance measures (available on request). In addition, group-specific sample means for employment states at each age (appendix 1) demonstrate that the collapsed typology represents identifiable longer-term patterns in later-life labour market involvement. Finally, further dynamic Hamming analyses using model reference sequences derived from detailed inspection of the seventh group (not shown) verified that no additional viable later-life employment biography groups could be extracted. Notwithstanding all of the above precautions, it is wise to keep in mind that the later-life labour market categories are not internally homogeneous, but rather, are clusters of individuals with *similar* employment sequences.

Statistical analyses

After assessing the gendering of older adults' labour market trajectories at baseline, we use nested multinomial regression models to investigate the roles of family experiences and prior employment in shaping observed patterns. The models first include controls (education, health at age 50, and, for the US, minority status), then add family circumstances at age 50 (whether married and whether dependent children), followed by family history (whether a parent, age at first child, and whether any marital ending by age 49), and finally include prior employment (years employed ages 15-49). All analyses are stratified by gender and country to allow for expected gender- and country-specific effects of at least the family variables. All models are run in Stata, with procedures designed to handle multiple imputations. Estimates are weighted using the population/design/attrition weights supplied with the surveys.

As multinomial coefficients are difficult to interpret in their raw form — and because we run separate models for women and men — we evaluate the roles of family experiences and prior employment by converting the logits to adjusted probabilities (with confidence intervals) for each trajectory group, by gender and country. Probabilities are assessed at the country-specific sample means for all covariates, and confidence intervals are calculated using Rubin's rules (Rubin, 1987) to accommodate the uncertainty associated with multiple imputations. We then quantify the gendering of later-life labour market biographies using 'gender gradients' or relative risks (adjusted probability among women / adjusted probability among men) for each trajectory type, by country. Where male and female confidence intervals do not overlap, we compare the magnitudes of these gradients within and between countries, and across models.

Imputation

Our method requires complete data on the sequence variables. When data were missing, we imputed values to retain as many cases as possible and minimise bias. Imputation used a two-fold fully conditional multiple imputation specification (Van Buuren, Brand, Groothuis-Oudshoorn, & Rubin, 2006) implemented in Stata -ice-. Two-fold ice was designed specifically for panel data, to impute using all cross-sectional data plus prior and subsequent values for variables with missingness. In brief, the method computes, for each missing value, its posterior distribution conditional on other variables in an imputation model. A value is then sampled from this distribution under the assumption that missingness is random given the values of the other variables in the model. The method uses a Markov chain Monte Carlo algorithm. After double iteration of the algorithm, a complete dataset is created, consisting of a mix of imputed and known values. Enough complete datasets are generated — 20, in our case — to ensure the accuracy of substantive model estimates (Graham, Olchowski, & Gilreath, 2007). All analyses are based on the simultaneous investigation of these 20 data sets, averaging over them and deriving standard errors according to Rubin's rules (Rubin 1987). Following optimal matching, missing covariate values were imputed using chained equations (White, Royston, & Wood, 2011) in Stata -mi impute-.

Results

Baseline distributions and gradients

Table 4. Weighted male and female percentages (CIs) in each later-life labour market trajectory type, along with gender gradients, by country

Later-life labour market trajectory	Germany Men	Germany Women	Germany Female / Male Gradient[a]	Italy Men	Italy Women	Italy Female / Male Gradient[a]
Total	100.0 (sample N=252)	100.0 (sample N=223)		100.0 (sample N=473)	100.0 (sample N=462)	
1 Full-Time Throughout	5.2 (2.3, 8.2)	1.5 (-0.6, 3.7)	0.29	9.9 (6.3, 13.5)	2.7 (0.8, 4.6)	0.28
2 Full-Time, Exit 65	39.6 (33.1, 46.0)	7.5 (3.1, 11.9)	0.19	14.5 (11.2, 17.8)	3.7 (1.8, 5.7)	0.26
3 Full-Time, Exit 60	34.5 (28.2, 40.8)	12.5 (7.2, 17.8)	0.36	32.5 (27.7, 37.3)	10.6 (7.3, 13.9)	0.33
4 Full-Time, Exit 55	12.3 (7.7, 17.0)	4.0 (1.4, 6.6)	0.33	23.7 (19.1, 28.4)	9.4 (6.4, 12.4)	0.40
5 Non-Employed Throughout	2.5 (-0.1, 5.0)	47.8 (40.5, 55.0)	19.26	10.6 (7.5, 13.8)	65.7 (60.7, 70.7)	6.18
6 Part-time Dominant	2.9 (0.3, 5.4)	21.2 (15.5, 26.8)	7.38	3.9 (1.1, 6.8)	5.5 (3.3, 7.8)	1.41

Later-life labour market trajectory	Sweden Men	Sweden Women	Sweden Female / Male Gradient[a]	United States Men	United States Women	United States Female / Male Gradient[a]
Total	100.0 (sample N=316)	100.0 (sample N=338)		100.0 (sample N=4,201)	100.0 (sample N=4,631)	
1 Full-Time Throughout	14.2 (10.0, 18.4)	5.2 (2.7, 7.7)	0.37	18.6 (17.0, 20.1)	9.6 (8.6, 10.6)	0.52†
2 Full-Time, Exit 65	46.2 (40.3, 50.2)	29.5 (24.3, 34.8)	0.64†	21.4 (19.9, 22.8)	13.7 (12.6, 14.8)	0.64†
3 Full-Time, Exit 60	24.3 (19.2, 29.4)	18.3 (13.7, 22.9)	0.75	19.2 (17.8, 20.6)	13.7 (12.6, 14.8)	0.71†
4 Full-Time, Exit 55	5.1 (2.3, 7.8)	5.8 (2.9, 8.8)	1.16	8.2 (7.3, 9.2)	9.8 (8.8, 10.8)	1.19
5 Non-Employed Throughout	1.2 (-0.3, 2.6)	6.6 (3.9, 9.4)	5.64	7.3 (6.4, 8.1)	19.9 (18.7, 21.2)	2.75
6 Part-time Dominant	5.7 (2.9, 8.5)	26.4 (21.3, 31.5)	4.66	20.3 (18.9, 21.7)	24.1 (22.7, 25.6)	1.19†

[a] Gradients are the ratio of the female proportion to the male proportion. As such, they measure the extent to which one gender predominates in a given trajectory group in a given country. Light grey shading identifies larger gradients (> 2.0 for female-dominated or < 0.5 for male-dominated trajectories), for which male and female confidence intervals do not overlap; dark grey shading identifies very large gradients (> 5.0 or < 0.2).

† Confidence intervals do not overlap for male and female estimates, but gradients are smaller than two times. Confidence intervals tend not to overlap for the American sample, simply because of the much larger N; hence, we focus on the magnitudes of (significant) gaps.

Table 4 displays weighted baseline percentages for each trajectory group, by gender and country, along with the resulting gender gradients (female percentage / male percentage). Light grey shading identifies larger gradients (> 2.0 for female-dominated trajectories or < 0.5 for male-dominated)—and dark grey pinpoints very large gradients (> 5.0 or < 0.2)—in which gender-specific confidence intervals do not overlap. Note that we focus on the magnitudes of (significant) gaps as confidence intervals tend not to overlap for the American sample even when gradients are fairly small, simply because of the much larger N. We highlight two features: the amount and type of gendering; and how this varies between countries. We leave aside discussion of the residual category, as its meaning is not consistent across genders or countries — although all models do include this group.

Most trajectory types are gendered, with gradients generally supporting *Hypothesis 1*: Men predominate in the 'attached' biographies (1-3) and women in the less 'attached' (5 and 6). Interestingly, though, gradients are typically largest for female-dominated trajectories. For example, continual non-employment in Germany is almost entirely female; and the part-

time dominant path in Germany, as well as continual non-employment in Italy and Sweden, all have gradients larger than five. Only one male-dominated biography has a gradient of more than five times (i.e., less than 0.20): full-time to exit around 65 in Germany.

The shading in table 4 provides clear support for *Hypothesis 2*: The gendering of later-life employment biographies is strongest in Germany and Italy, where only the very small groups display non-significant male-female differences. Unexpectedly, however, gradients are often larger for Germany than for Italy, despite the gender gap in state pension ages in the latter country. Proportions underlying the strongest gradients show that a far smaller share of German than Italian men followed the female-dominated continual non-employment path, and a far greater proportion of German men are in the male-dominated full-time to exit around 65 group.

Beyond this, the results add nuance to the notion that male-female differences will be weakest in Sweden and the US (*Hypothesis 2*). Instead, gendering is intermediate for Sweden, where one male-dominated and two female-dominated trajectories have steep gradients, and weakest for the US, where only continual non-employment is clearly gendered. Of note is the part-time dominant biography among Swedes, which is both strongly feminized and large (26% of women). This implies that the Swedish dual-earner strategy of incorporating mothers into the labour force partly via better-quality short-hour jobs carries through to older ages. Subsequent investigation (not shown) reveals that Swedish women in the part-time dominant group had spent, on average, nearly a third of their younger years working 'short' hours, versus about a tenth of those years for all other Swedish women combined. Despite this evidence of gendering, we note that the (always male-dominated) full-time to exit around 65 path is actually the most common one taken by Swedish women; and this cannot be said of any other country.

Family and prior work

Table 5. Female/male gender gradients[a] for later-life labour market trajectories, by country

	Model 1 (controls only)	Model 2 (+ family age 50)	Model 3 (+ family history)	Model 4 (+ work history)
		Germany		
1. Full-time throughout[b]				
2. Full-time, exit 65/65+	0.13	0.12	0.13	0.19
3. Full-time, exit 60	0.31	0.29	0.29	0.68
4. Full-time, exit 55	0.32	0.35	0.24	0.89
5. Non-employed throughout	†	†	†	†
6. Part-time dominant	83.42[c]	149.20[c]	204.63[c]	270.84[c]
		Italy		
1. Full-time throughout	0.19	0.07	0.03	0.04
2. Full-time, exit 65	0.19	0.19	0.21	0.35
3. Full-time, exit 60	0.31	0.29	0.30	0.47
4. Full-time, exit 55	0.41	0.38	0.15	0.31
5. Non-employed throughout	6.53	6.62	6.76	3.04
6. Part-time dominant	1.37	1.55	1.69	4.95
		Sweden		
1. Full-time throughout	0.31	0.33	0.07	0.07
2. Full-time, exit 65	0.60	0.59	0.59	0.69
3. Full-time, exit 60	0.73	0.67	0.74	0.78
4. Full-time, exit 55	0.95	0.98	0.98	1.07
5. Non-employed throughout	†	†	†	†
6. Part-time dominant	5.10	8.13	8.56	7.70
		United States		
1. Full-time throughout	0.49	0.49	0.50	0.57
2. Full-time, exit 65	0.64	0.63	0.63	0.68
3. Full-time, exit 60	0.72	0.71	0.73	0.80
4. Full-time, exit 55	1.26	1.23	1.24	1.23
5. Non-employed throughout	2.94	3.03	3.15	1.91
6. Part-time dominant	1.27	1.29	1.27	1.40

[a] In the calculation of the underlying adjusted probabilities, all covariates are set to their within-country sample means. Gradients are highlighted where confidence intervals for male and female adjusted proportions do not overlap. Light grey = male-dominated, dark grey = female-dominated.

[b] Trajectories 1 and 2 were combined for Germany only.

[c] The very large gradients (relative to Table 4) occur because low education dramatically decreases men's, but not women's, odds of having this trajectory. This is largely a function of small cells: There are few German men in in the part-time dominant group and, as a result, the cell for low education is actually empty.

†Virtually all cases are female.

We next investigate the contributions of family experiences and prior employment to the above patterns (table 5). Again, we quantify male-female distinctions using gender gradients. But here the underlying probabilities are calculated from regression coefficients that adjust for the influence of family and earlier work. (See Appendix 2 for the adjusted probabilities and confidence intervals, and Appendix 3 for average marginal covariate effects.) In this table, light-grey shading indicates male-dominated, and dark-grey, female-dominated trajectories. Again, we do not discuss the residual category — although all models do include this group. In addition, for German men we combine full-time to exit around 65 and full-time throughout, to correspond to women's trajectories in that country, where collapsing was necessary because so few were in ongoing full-time employment. For similar reasons, we fold the small number of German and Swedish men in continual non-employment into the residual category. Model 1, which adjusts for controls only, is included as the comparator for Model 2. Models 2 and 3 test *Hypothesis 3a* — that gender differences in later-life labour market biographies are rooted partly in the differential impact of family experiences over the life course. Model 4 tests the hypothesis that the gendering of later-life labour market involvement is at least partly due to differential attachment to the labour force at younger ages (*Hypothesis 4a*). We highlight two features of table 5: how gradients shift with the addition of explanatory variables; and cross-national differences in those shifts.

Relative to the model with controls, some model 2 gender gaps increase — most obviously for the part-time dominant path among Germans and Swedes and ongoing full-time employment among Italians. (For gradients < 1, a larger gap is represented by a *decrease* from one model to the next, meaning a greater preponderance of men.) In addition, women's greater probability of continual non-employment increases for Germany and Sweden, although no gradient is shown because no male comparison group exists. Model 3 (table 5) displays the additional influence of family experiences earlier in the life course. Gender differentials increase for the part-time dominant path among Germans, ongoing full-time employment among Swedes, and the full-time to exit around 55 biography among Italians.

Models 2 and 3 provide support for the notion that the gendering of later-life labour market involvement arises, in part, from the gender-specific influences of family experiences over the life course (*Hypothesis 3a*). They only partially back *Hypothesis 3b* — that effects will be stronger for Germany and Italy than for Sweden and the US. Instead, family experiences appear more important in all three European countries (two breadwinner-caregiver and one dual-earner) than in the US (dual-earner), where they have little or no impact.

Model 4 (table 5) tests *Hypothesis 4a* — that the gendering of older adults' labour market trajectories arises at least partly from gendered earlier attachment to the labour force. Indeed, prior employment partially explains many of the family-adjusted gender gaps (trajectories 2 and 5 in Germany; trajectory 6 and, to some extent, trajectory 2 in Sweden; trajectories 3-5 in Italy; and trajectory 5 in the US), or reduces them to non-significance (trajectory 2 in Italy; trajectory 3 in Germany and the US). The results are also consistent with *Hypothesis 4b*: Male-female differences in earlier employment account for a greater share of the gendering of later-life labour market involvement in Germany and Italy, where most trajectory types are affected, than in Sweden and the US, where effects primarily involve one male-dominated and one female-dominated biography.

Discussion

This study adopted a comparative gendered life course approach, to extend knowledge of older adults' labour market involvement. While the majority of previous work has focused on retirement and thereby underrepresented women, our analysis drew on extended biographical sequences and fully incorporated both genders in four nations with diverse approaches to family and paid work over the life course. This strategy — to our knowledge, unique among studies of older adults' labour market involvement—generated results that both confirmed and expanded existing knowledge.

Consistent with expectations, in every nation women prevailed in groups representing a weak(er)

attachment to the labour market in later life and men in those signifying a strong(er) attachment. Additionally, this pattern was especially pronounced for Germany and Italy, where policies have generally supported men's, but not women's, paid work earlier in the life course. These findings are in line with cross-sectional estimates of older adults' labour force participation (OECD, 2011). However, they also offer novel insights into how gender and institutional contexts play out with respect to *long-term* patterns of labour market involvement. For example, across all nations, gradients tended to be stronger for female- than male-dominated biographies, implying that women were more likely to follow 'male' paths than the reverse. Future research should investigate the institutional and cultural factors behind this finding (e.g., the extent of support for men's family work), along with its relevance to upcoming cohorts where family roles may be less gendered.

We also found support for the idea that male-female differences in later-life labour market involvement arise, in part, from the gendered influence of family experiences over the life course. More importantly, we offered new evidence that family ties matter more in the corporatist and southern welfare regimes, where a breadwinner-caregiver division of labour structures the adult life course, than they do in the liberal nation, where a dual-earner model is (passively) promoted. This highlights the need for approaches to older adults' labour market involvement that consider not only state benefits that 'pull,' and workplace barriers that 'push,' older adults out of the labour force, but also policies related to family circumstances throughout the life course. Future studies might extend these findings by drawing on more detailed marital and parental histories and identifying the specific social, cultural, and policy mechanisms by which they operate in various institutional contexts.

The real significance of family policies undoubtedly lies in their impact on the gendering of earlier attachment to the labour force, which may carry through to older ages. Indeed, prior employment accounted for a significant share of male-female distinctions in later-life labour market involvement. Again, this was most apparent for Germany and Italy, where, historically, women have been encouraged to leave the labour force once married and/or raising children. Conversely, work histories explained less of the gendering at older ages in Sweden, where the state has, for several decades, actively supported women's employment throughout the life course, and in the US, where the state has passively promoted dual-earning. These findings reiterate the need for approaches to older adults' employment that consider the entire life course.

Three key results were not anticipated. First, in some ways the divide was between the three European countries and the US, rather than between those adhering to a breadwinner-caregiver and a dual-earner family model. Family experiences made almost no difference to the gendering of older Americans' labour market biographies. But the part-time dominant biography among Swedes was both highly gendered (female) and common; and the effects of family experiences were particularly strong for that group. These findings, along with the disproportionate levels of earlier part-time work found for Swedish women in the part-time dominant group, suggest that that nation's strategy of incorporating mothers into the labour force partly by means of better-quality short-hour jobs carries through to older ages.

Second, we turned up particularly strong evidence of gendering for Germany — stronger than that found for the other breadwinner-caregiver nation (Italy) and despite the gender gap in the latter nation's pension ages. The much smaller share of German than Italian men in the female-dominated continual non-employment group may follow from Italy's relatively lenient access to illness/disability pensions for older workers; and the far greater proportion of German men with the male-dominated full-time to exit around 65 biography may reflect Germany's pension system, which is closely tied to (male) earnings over the life course (Buchholz et al., 2006).

And finally, the results for work histories were somewhat surprising. We found roughly similar effects for the US and Sweden, as expected; but this was *despite* the larger gender gap in prior employment in the US, which, all else equal, should have meant that work histories accounted for more gendering of later-life labour market involvement in that country. That they did not implies that some American women with a weak(er) prior attachment to the labour force ended up behaving more like men

with respect to their labour market involvement in older adulthood. This may reflect contextual factors — a heavy reliance on accumulated personal resources to fund retirement — requiring some without steady employment records to increase their work effort as they approach (or pass) state pension age. Pending the availability of more detailed data, future research should further investigate cross-national differences in the role of work histories (including full-time/part-time status and interactions with family histories) in shaping older women's employment biographies.

Our contributions should be viewed in light of several limitations. One is the necessary sacrifice of some precision in the construction of the later-life labour market groups. We clustered individuals whose trajectories are similar but not identical, and we cannot say whether this 'muddied' associations with the key explanatory variables. This constraint is, however, inseparable from one of the study's strengths: the ability to include women, even those without steady (or any) employment in their later years, and thus to assess gender differences in older adults' employment trajectories across a range of welfare states. A second consideration is that imputations introduced some uncertainty into the labour market sequences. While this is of concern, we minimised its impact — using a two-fold fully conditional specification, running a large number of imputations, and appropriately adjusting standard errors. Moreover, the vast majority of sequences were complete for all countries except the US, and our sensitivity analysis using a reduced US sample did not alter our major conclusions. A third limitation is the possibility that our measures of family experiences did not capture all relevant aspects, while our prior employment variable could not incorporate, for example, part-time work. These shortcomings reflect a combination of data and methodological constraints—inconsistent information on prior work periods in the HRS, and relatively small sample sizes for the SHARE countries, which restricted the number and type of covariates possible.

Our discussion highlights at least three policy implications. First is the need to go beyond a focus on workplace and economic factors that shape employment decisions in the years surrounding public pension eligibility. We found evidence that older adults' labour market trajectories are influenced by policies related not only to paid work but also to caregiving, and by those affecting not only immediate decisions but also choices made earlier in the life course. Second are possible vulnerabilities associated with the drive to extend working life. The feminization of non-employment in Germany and Italy, and of part-time work in Germany and Sweden, may leave women out in the cold if policies fail to consider the gendering of paid work and caregiving earlier in the life course. Third, the American case hints at concerns related to weakly regulated labour markets and scanty public provisions. That women with weak(er) work histories may end up behaving more like men with respect to later-life labour market involvement, raises questions about the quality of at least some older Americans' work — questions whose urgency will increase if adults are required to work longer. These considerations should be a central focus of ongoing cross-national investigations of older adults' labour market trajectories.

Acknowledgements

This research was funded by the Canadian Institutes of Health Research [grant number MOP 11952, Peggy McDonough PI] and the Social Sciences and Humanities Research Council [grant number 435121267, Peggy McDonough PI]. Anne McMunn was supported by the European Research Council [Starting Grant number ERC-2011-StG_20101124, Anne McMunn PI]; Amanda Sacker by the UK Economic and Social Research Council [grant number ES/J019119/1, Amanda Sacker PI]; and Laurie Corna by the Economic and Social Research Council/Medical Research Council [grant number ES/L002825/1, Karen Glaser PI].

The HRS is sponsored by the National Institute on Aging [grant number NIA U01AG009740] and is conducted by the University of Michigan. We also made occasional use of the RAND HRS Data file, a simplified longitudinal dataset based on the HRS and developed at RAND with funding from the National Institute on Aging and the

Social Security Administration. The SHARE is primarily funded by the European Commission through FP5 [QLK6-CT-2001-00360], FP6 [SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812] and FP7 [SHARE-PREP: N°211909, SHARE-LEAP: N°227822, SHARE M4: N°261982]. Additional funding comes from the German Ministry of Education and Research, the U.S. National Institute on Aging [U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064], and funding sources specific to each national dataset (see http://www.share-project.org/contact-organisation/funding.html).

References

Abbott, A., & Tsay, A. (2000). Sequence Analysis and Optimal Matching Methods in Sociology. *Sociological Methods & Research*, *29*(1), 3-33. http://dx.doi.org/10.1177/0049124100029001001

Alley, D. E., Putney, N. M., Rice, M., & Bengtson, V. L. (2010). The Increasing Use of Theory in Social Gerontology: 1990-2004. *Journals of Gerontology Series B-Psychological Sciences and Social*, *65*(5), 583-590. http://dx.doi.org/10.1093/geronb/gbq053

Beckstette, W., Lucchini, M., & Schizzerotto, A. (2006). Men's Late Careers and Career Exits in Italy. In H.-P.Blossfeld, S. Buchholz, & D. Hofäcker (Eds.), *Globalization, Uncertainty, and Late Careers in Society* (pp. 100-118). New York: Routledge.

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological methods, 9*(1), 30-52. http://dx.doi.org/10.1037/1082-989X.9.1.30

Billari, F. C., & Piccarreta, R. (2005). Analysing demographic life courses through sequence analysis. *Mathematical Population Studies* 12, 81-106. http://dx.doi.org/10.1080/08898480590932287

Blossfeld, H.-P., & Drobnič, S. (2001). *A Cross-National Comparative Approach to Couples' Careers*. Oxford: Oxford University Press.

Blossfeld, H.-P., Buchholz, S., & Hofäcker, D. (2006). *Globalization, Uncertainty and Late Careers in Society*. New York: Taylor & Francis. http://dx.doi.org/10.4337/9781781007495

Blossfeld, H.-P., Buchholz, S., & Kurz, K. (Eds.). (2011). *Aging Populations, Globalization and the Labor Market: Comparing Late Working Life and Retirement in Modern Societies*. Cheltenham ; Northampton, MA: Edward Elgar Pub. http://dx.doi.org/10.4337/9781849805858

Börsch-Supan, A. & Jürges, H. (Eds.). (2005). *The Survey of Health, Ageing, and Retirement in Europe – Methodology*. Munich: Mannheim Research Institute for the Economics of Ageing. Retrieved from http://www.share-project.org/uploads/tx_sharepublications/SHARE_BOOK_METHODOLOGY_Wave1.pdf

Börsch-Supan, A., Hank, K., Jürges, H., & Schröder, M. (2009). Introduction: empirical research on health, ageing and retirement in Europe. *Journal of European Social Policy*, *19*(4), 293-300. http://dx.doi.org/10.1177/1350506809341510

Bowlby, G. (2007). Defining Retirement. *Perspectives*, 8(2), 15-9.

Buchholz, S., & Hofäcker, D., & Blossfeld, H.-P. (2006). Globalization, accelerating economic change and late careers. In H.-P. Blossfeld, S. Buchholz, & D. Hofäcker (Eds.), *Globalization, Uncertainty and Late Careers in Society* (pp. 1-24). New York: Taylor & Francis.

Cooke, M. (2006). Policy Changes and the Labour Force Participation of Older Workers: Evidence from Six Countries. *Canadian Journal on Aging/La Revue Canadienne Du Vieillissement*, *25*(04), 387-400. http://dx.doi.org/10.1353/cja.2007.0015

Denaeghel, K., Mortelmans, D., & Borghgraef, A. (2011). Spousal influence on the retirement decisions of single-earner and dual-earner couples. *Advances in Life Course Research*, *16*(3), 112-123. http://dx.doi.org/10.1016/j.alcr.2011.06.001

Dentinger, E., & Clarkberg, M. (2002). Informal Caregiving and Retirement Timing among Men and Women: Gender and Caregiving Relationships in Late Midlife. *Journal of Family Issues*, *23*(7), 857-879. http://dx.doi.org/10.1177/019251302236598

Drobnic, S., Blossfeld, H. P., & Rohwer, G. (1999). Dynamics of women's employment patterns over the family life course: A comparison of the United States and Germany. *Journal of Marriage and the Family*, *61*(1), 133-146. http://dx.doi.org/10.2307/353889

Duval, R. (2003). *The Retirement Effects of Old-Age Pension and Early Retirement Schemes in OECD Countries*. OECD Publishing. Retrieved from https://ideas.repec.org/p/oec/ecoaaa/370-en.html

Esping-Andersen, G. (1990). *The three worlds of welfare capitalism*. Princeton, N.J.: Princeton University Press.

Fasang, A. E. (2010). Retirement: Institutional Pathways and Individual Trajectories in Britain and Germany. *Sociological Research Online*, *15*(2). http://dx.doi.org/10.5153/sro.2110

------. (2012). Retirement Patterns and Income Inequality. *Social Forces*, *90*(3), 685–711. http://dx.doi.org/10.1093/sf/sor015

Gal, J. (2010). Is there an Extended Family of Mediterranean Welfare States? *Journal of European Social Policy*, 20, 283-300. http://dx.doi.org/10.1177/0958928710374374

Gornick, J. C. and Meyers, M. K. (2006). Welfare Regimes in Relation to Paid Work and Care. *Revue française des affaires sociales,* 5(5), 167–88.

Graham, J., Olchowski, A., & Gilreath, T. (2007). How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory. *Prevention Science, 8*(3), 206-213. http://dx.doi.org/10.1007/s11121-007-0070-9

Halldén, K., Gallie, D., & Zhou, Y. (2012). The skills and autonomy of female part-time work in Britain and Sweden. *Research in Social Stratification and Mobility*, *30*(2), 187-201. http://dx.doi.org/10.1016/j.rssm.2011.07.001

Halpin, B. (2014). SADI: sequence analysis tools for Stata. Working Paper WP2014-03. University of Limerick Department of Sociology Working Paper Series. Limerick, Ireland: University of Limerick. Retrieved from http://www.ul.ie/sociology/wp2014-03.pdf

Han, S.-K., & Moen, P. (1999a). Work and Family Over Time: A Life Course Approach. *The ANNALS of the American Academy of Political and Social Science*, *562*(1), 98-110. http://dx.doi.org/10.1177/0002716299562001007

Han, S-K.., & Moen, P. (1999b). Clocking Out: Temporal Patterning of Retirement. *American Journal of Sociology*, *105*(1), 191-236. http://dx.doi.org/10.1086/210271

Hank, K. (2004). Effects of Early Life Family Events on Women's Late Life Labour Market Behaviour An Analysis of the Relationship between Childbearing and Retirement in Western Germany. *European Sociological Review*, *20*(3), 189-198. http://dx.doi.org/10.1093/esr/jch017

Hank, K., & Korbmacher, J. M. (2013). Parenthood and Retirement: Gender, cohort, and welfare regime differences. *European Societies*, *15*(3), 446-461. http://dx.doi.org/10.1080/14616696.2012.750731
Heeringa, S. G., & Connor, J. H. (1995). Technical Description of the Health and Retirement Survey Sample Design (Public Use Version). Ann Arbor, MI: Institute for Social Research. http://dx.doi.org/10.7826/ISR-UM.06.585031.001.05.0001.1995 Retrieved from http://hrsonline.isr.umich.edu/sitedocs/userg/HRSSAMP.pdf

Hofäcker, D., Stoilova, R., & Riebling, J. R. (2013). The Gendered Division of Paid and Unpaid Work in Different Institutional Regimes: Comparing West Germany, East Germany and Bulgaria. *European Sociological Review*, *29*(2), 192-209. http://dx.doi.org/10.1093/esr/jcr055

Kahn, L. (2010). *Labor Market Policy: A Comparative View on the Costs and Benefits of Labor Market Flexibility* (CESifo Working Paper Series No. 3140). CESifo Group Munich. Retrieved from https://ideas.repec.org/p/ces/ceswps/_3140.html

Kail, B. L., & Warner, D. F. (2013). Leaving Retirement: Age-Graded Relative Risks of Transitioning Back to Work or Dying. *Population Research and Policy Review*, *32*(2), 159-182. http://dx.doi.org/10.1007/s11113-012-9256-3

Katz, M. B. (2010). The American Welfare State and Social Contract in Hard Times. *Journal of Policy History*, *22*(4), 508-529. http://dx.doi.org/10.1017/S0898030610000242

Korpi, W. (2000). Faces of Inequality: Gender, Class, and Patterns of Inequalities in Different Types of Welfare States. *Social Politics: International Studies in Gender, State & Society*, 7(2), 127-91. http://dx.doi.org/10.1093/sp/7.2.127

Korpi, W., Ferrarini, T., & Englund, S. (2013). Women's Opportunities under Different Family Policy Constellations: Gender, Class, and Inequality Trade-offs in Western Countries Re-examined. *Social Politics: International Studies in Gender, State and Society*, *20*(1), 1-40. http://dx.doi.org/10.1093/sp/jxs028

Lesnard, L. (2010). Setting Cost in Optimal Matching to Uncover Contemporaneous Socio-Temporal Patterns. *Sociological Methods & Research*, 38, 389-419. http://dx.doi.org/10.1177/0049124110362526

Lindquist, G. S. (2006). Late Careers and Career Exits in Sweden. In H.-P. Blossfeld, S. Buchholz, & D. Hofäcker (Eds.), *Globalization, Uncertainty, and Late Careers in Society* (pp. 211-233). New York: Routledge.

MacIndoe, H., & Abbott, A. (2004). Sequence Analysis and Optimal Matching Techniques for Social Science Data. In M. A. Hardy, & A. Bryman (Eds.), *Handbook of Data Analysis* (pp. 387-405). London: Sage. http://dx.doi.org/10.4135/9781848608184.n17

Mayer, K. U. (2004). Whose Lives? How History, Societies, and Institutions Define and Shape Life Courses. *Research in Human Development*, 1(3), 161-87. http://dx.doi.org/10.1207/s15427617rhd0103_3

Moen, P. (1996). A life course perspective on retirement, gender, and well-being. *Journal of Occupational Health Psychology*, *1*(2), 131-144. http://dx.doi.org/10.1037/1076-8998.1.2.131

Moen, P., & Flood, S. (2013). Limited Engagements? Women's and Men's Work/Volunteer Time in the Encore Life Course Stage. *Social Problems*, *60*(2), 206-233. http://dx.doi.org/10.1525/sp.2013.60.2.206

Morris, D., & Mallier, T. (2003). Employment of Older People in the European Union. *LABOUR*, *17*(4), 623–648.

OECD. (2005). *Pensions at a Glance 2005*. Paris: Organisation for Economic Co-operation and Development. http://dx.doi.org/10.1111/j.1121-7081.2003.00255.x Retrieved from http://www.oecd-ilibrary.org/finance-and-investment/oecd-pensions-at-a-glance-2005_pension_glance-2005-en

------. (2006). *Ageing and Employment: Live Longer, Work Longer*. Paris: OECD Publishing. Retrieved from https://www.oecd.org/employment/livelongerworklonger.htm

------. (2009). Employment Rate of Older Workers: Percent of Population Aged 55-64. http://dx.doi.org/10.1787/20752342-2009-table6 Retrieved from http://dx.doi.org/10.1787/20752342-2009-table6

------. (2011). *Pensions at a Glance 2011*. Paris: Organisation for Economic Co-operation and Development. Retrieved from http://www.oecd-ilibrary.org/content/book/pension_glance-2011-en

------. (2015a). *OECD Family Database*. Paris: Organisation for Economic Co-operation and Development. Retrieved from www.oecd.org/social/family/database.htm

------. (2015b). OECD.Stat. Tables for "LFS by sex and age – indicators." Retrieved from http://stats.oecd.org/index.aspx

Ogg, J., & Renaut, S. (2007). The influence of living arrangements, marital patterns and family configuration on employment rates among the 1945–1954 birth cohort: evidence from ten European countries. *European Journal of Ageing*, *4*(3), 155-169. http://dx.doi.org/10.1007/s10433-007-0061-5

Persky, L. (2011). *More than a Woman's Issue: The Making of Family Policy in the US and UK Since 1960* (SSRN Scholarly Paper No. ID 1901351). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=1901351

Pienta, A. M. (1999). Early Childbearing Patterns and Women's Labor Force Behavior in Later Life. *Journal of Women & Aging*, *11*, 69-84. http://dx.doi.org/10.1300/J074v11n01_06

Pienta, A. M., Burr, J. A., & Mutchler, J. E. (1994). Women's Labour Force Participation in Later Life: The Effects of Earlier Work and Family Experiences. *Journal of Gerontology*, 49(5), S231-9. http://dx.doi.org/10.1093/geronj/49.5.S231

Pienta, A. M., & Hayward, M. D. (2002). Who Expects to Continue Working After Age 62? The Retirement Plans of Couples. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *57*(4), S199-S208. http://dx.doi.org/10.1093/geronb/57.4.S199

Pleau, R. L. (2010). Gender Differences in Postretirement Employment. *Research on Aging*, *32*(3), 267 -303. http://dx.doi.org/10.1177/0164027509357706

Radl, J. (2013). Labour Market Exit and Social Stratification in Western Europe: The Effects of Social Class and Gender on the Timing of Retirement. *European Sociological Review*, *29*(3), 654-668. http://dx.doi.org/10.1093/esr/jcs045

Ross, C. E., & Mirowsky, J. (1992). Households, Employment, and the Sense of Control. *Social Psychology Quarterly*, *55*(3), 217-235. http://dx.doi.org/10.2307/2786793

Rovny, A. E. (2011). Welfare state policy determinants of fertility level: A comparative analysis. *Journal of European Social Policy*, *21*(4), 335-347. http://dx.doi.org/10.1177/0958928711412221

Rubin, D. B. (1987). *Multiple Imputation for Non-response in Surveys*. Chichester, England: Wiley. http://dx.doi.org/10.1002/9780470316696

Rydell, I. (2002). *Demographic Patterns from the 1960s in France, Italy, Spain and Portugal* (Arbetsrapport No. 2003:2). Institute for Futures Studies. Retrieved from https://ideas.repec.org/p/hhs/ifswps/2003_002.html

Siaroff, A. (1994). Work, Welfare and Gender Equality: A New Typology. In Sainsbury, D. (Ed.), *Gendering Welfare States Gendering welfare states* (pp. 82-100). London: SAGE. http://dx.doi.org/10.4135/9781446250518.n6

Sundström, E. (2003). Gender Regimes, Family Policies and Attitudes to Female Employment : A Comparison of Germany, Italy and Sweden. Doctoral Dissertation in Sociology, Faculty of Social Sciences, Umeå University. Retrieved from www.diva-portal.org/smash/get/diva2:142190/FULLTEXT01.pdf

Szinovacz, M. E., & Deviney, S. (2000). Marital Characteristics and Retirement Decisions. *Research on Aging*, *22*(5), 470-498. http://dx.doi.org/10.1177/0164027500225002

Szinovacz, M. E., DeViney, S., & Davey, A. (2001). Influences of family obligations and relationships on retirement : Variations by gender, race, and marital status. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *56*(1), S20-S27. http://dx.doi.org/10.1093/geronb/56.1.S20

Trifiletti, R. (1999). Southern European Welfare Regimes and the Worsening Position of Women. *Journal of European Social Policy*, 9(1), 49-64. http://dx.doi.org/10.1177/095892879900900103

UNESCO. (2006). *International Standard Classification of Education ISCED 1997*. Retrieved from http://www.uis.unesco.org/Library/Documents/isced97-en.pdf

US Bureau of Labor Statistics. (2000). Table 8. Employed and unemployed full- and part-time workers by age, sex, and race. Retrieved from http://www.bls.gov/cps/aa1999/CPSAAT8.pdf

Van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, C. G. M., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*, 76(12), 1049-1064. http://dx.doi.org/10.1080/10629360600810434

Ware, J. H., & Wu, M. C. (1981). Tracking: Prediction of Future Values from Serial Measurements. *Biometrics*, *37*(3), 427-437. http://dx.doi.org/10.2307/2530556

Warner, D., & Hofmeister, H. (2006). Late Career Transitions among Men and Women in the United States. In H.P. Blossfeld, S. Buchholz, & D. Hofäcker (Eds.), *Globalization, Uncertainty, and Late Careers in Society* (pp. 141-181). New York: Routledge.

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, *30*(4), 377-399. http://dx.doi.org/10.1002/sim.4067

Williamson, J. B., & McNamara, T. K. (2003). Interrupted Trajectories and Labor Force Participation. *Research on Aging*, *25*(2), 87-121. http://dx.doi.org/10.1177/0164027502250015

Wilmoth, J., & Koso, G. (2002). Does Marital History Matter? Marital Status and Wealth Outcomes among Preretirement Adults. *Journal of Marriage and Family*, *64*(1), 254-268. http://dx.doi.org/10.1111/j.1741-3737.2002.00254.x

Longitudinal and Life Course Studies 2016 Volume 7 Issue 4 Pp 368 – 385 *ISSN 1757-9597*

The health impacts of the contemporary manufacturing and service sectors on men and women

Eleni Kampanellou University of York, UK
eleni.kampanellou@york.ac.uk
Donald Houston University of Portsmouth, UK

(Received November 2015 Revised March 2016) http://dx.doi.org/10.14301/llcs.v7i4.392

Abstract

Manufacturing and manual employment and, to a lesser extent, low-grade white-collar work have long been associated with poor health outcomes. This article reports important new findings based on longitudinal micro data that demonstrate important changes and gender-related patterns to this prevailing understanding. Specifically, manufacturing employment now has a protective health effect for men, and women's health is not strongly influenced by occupation. High-paid service sector employment is found to be bad for health, particularly among men. Changing industry within the service sector is linked to a deterioration in health, particularly among women, whereas changing employment from manufacturing to services is found to be bad for men's health. Confirming previous research, shifts from any sector of employment into unemployment and economic inactivity are strongly associated with a deterioration in health. The findings point to four conclusions: i) the emergence of new occupational hazards in the service sector; ii) the improvement of working conditions in manufacturing; iii) changing industry is damaging to an individual's health, possibly due to skills mismatches that may arise, although further research is required to further disentangle the direction of causation; and iv) the impacts on health of different industrial sectors and changes between industrial sectors vary between men and women.

Keywords

Gender, health, inactivity, longitudinal studies, morbidity, occupational hazards, unemployment

Introduction

Manufacturing and construction employment, and to a lesser extent low-grade clerical work, have long been linked with poor health (Hedlund, Järvholm & Lundbäck, 2006; Hein, Stayner, Lehman & Dement, 2007; Jones et al., 2009). The transition of developed economies from manufacturing towards services has been associated with mixed health implications (Ostry et al., 2000, 2002). On the one hand, the decline and automation of hazardous occupations and improved health and safety practices in manufacturing and construction have been hypothesised to reduce the incidence of ill-health and disability in the working-age population (Loomis, Richardson, Bena & Bailer, 2004). On the other hand, work intensification and flexibilisation prevalent in the expanding low-grade service sector may be producing new occupational hazards for health arising from pressure, lack of control and the sedentary nature of non-manual employment (Anderson, Schulte, Sestito, Linn & Nguyen, 2010; Sekine, Chandola, Martikainen, Marmot & Kagamimori , 2009).

Industrial decline and associated job destruction impacts adversely on health through unemployment, economic inactivity or re-employment in lower quality jobs in the service sector. Low-grade service sector jobs are often

characterised by inferior salaries, the health impacts of which may be compounded if individuals coming out of manufacturing are less well suited or equipped to perform the tasks required, for example arising from skills and cultural mismatches (Danson, 2005; Power, 2008; Wessel, 2005). Unemployment and non-participation in the labour market (economic inactivity) are strongly associated with poor health, although the magnitude of causality in each direction remains uncertain – in other words, in some cases job loss comes first and subsequently impacts on health, while in other cases the onset of illness or other conditions subsequently lead to exits from employment (Bambra & Eikemo, 2009; Booker & Sacker, 2013; Korpi, 2001).

Previous research on the health implications for individuals of changing the industrial sector in which they are employed has tended to focus on individuals making the transition from manufacturing to services, particularly men, as a result of the deindustrialisation process producing forced job changes (Beatty & Fothergill, 1996; Morris & Cook, 1991; Rocha, 2001). Little is known about how other transitions between industrial sectors affect health, nor about how the health impacts of changing industry may affect men and women differently. As with job loss, changes in the industrial sector of employment may be either cause or consequence of poor health – or both. Some individuals may change industry in response to a change in their health-related capacity to perform tasks in their previous job. In other circumstances, a change in industrial sector may occur for reasons unrelated to health but have subsequent impacts on health. In both cases, the effectiveness of the labour market and hiring process in matching workers and jobs in terms of skills and preferences is likely to be important in determining subsequent health outcomes. Good matches could be expected to produce better health outcomes.

The effect of an individual changing industry of employment can be hypothesised to impact on that individual's health in two regards. First, the new industry may be associated with different health outcomes (either better or worse) compared to the old industry, arising from differences in pay, job security and physical and psychosocial hazards. Second, the process of change itself may be stressful, and produce a poorer match between the worker and the job in terms of skills, physical and mental capabilities and personal preferences, which may in turn lead to greater physical and mental stresses. Adapting to change, including a possible drop in income and impact on occupational identity, may bring short- and long-term physical and mental stresses.

Involuntary necessity-driven transitions between industries, for example arising from redundancy, could be expected to produce particularly poor 'matching' of workers to jobs because people facing unemployment experience greater pressure to quickly obtain and accept job offers. Because of lower employability, individuals in weaker socioeconomic circumstances and with lower qualifications could be expected to be more likely to experience involuntary changes in job and/or industry and to experience such changes more frequently over the course of their working lives. Repeated transitions and poor worker/job matching may in combination have cumulative adverse effects on career progression and on health over the life course. The rise of precarious or 'non-standard' forms of employment (including zero-hours contracts, agency work and self-employment) may be over time forcing more people into involuntary changes in job and/or industry. Furthermore, the intensification of work, particularly in the low-skilled end of the service sector, may be producing new occupational hazards for health.

This article has three research aims. First, to update evidence on the link between industrial sector employment and health outcomes. Second, to assess the link between a change in industrial sector and a change in health, going beyond existing research that has focussed on men shifting from manufacturing to service-sector employment. Third, to investigate whether industrial sector and a change in industrial sector affect the patterns of poor health among men and women differently.

Industrial sector, gender and health

Industrial sector and health

Traditional manufacturing industries have been associated with specific physical health conditions (respiratory-related diseases, various malignant neoplasms and injuries) for both males and females (Andersson, Persson, Bryngelsson, Magnuson & Westberg, 2010; Jones et al., 2009; Stout, Jenkins & Pizatella, 1996). Some parts of the service sector

are characterised by casualization and low levels of training and career development, which may lead to 'job stress' and poor health outcomes for both men and women (Ferrie, Shipley, Stansfeld & Marmot, 2002). The service sector covers a range of sub-sectors associated with diverse health implications. More precisely, high rates of mortality are observed for employees in the wholesale and retail sector (Anderson et al., 2010). Occurrences of heart disease are more frequent among low-grade civil servants, where employees demonstrate higher blood pressure, smoking and limited physical activity (Van Rossum, Shipley, Van de Mheen, Grobbee & Marmot, 2000).

Additional factors contributing to the prevalence of certain morbidity outcomes are associated with the psychosocial nature of employment. Supplementary elements such as job insecurity and high work-related demands combined with low control and rewards at work are linked with the increase of incidents of coronary heart diseases, sickness absences, emotional exhaustion and overall life dissatisfaction (Aronsson, Gustafsson & Dallner, 2002; Kuper & Marmot, 2003; Sekine et al., 2009; Van der Hulst & Geurts, 2001).

Recent research has revealed that the nature of work in both manufacturing and in the service sector has changed (Murray, Baldwin, Ridgway & Winder, 2005; Ostry et al., 2001). Qualitative evidence suggests that these changes may be altering established employment-related patterns of health – but this has not been quantitatively tested (Ostry et al., 2002). Specifically, automation in manufacturing and construction and improved health and safety practices in hazardous industries may have reduced hazards, and certainly injuries and fatalities reported by industries have fallen (Loomis et al, 2004). Meanwhile, work intensification and flexibilisation prevalent in the service sector may have increased adverse impacts on health (Baumberg, 2014). Work intensification may have increased work-related stress, while aspects of flexibilisation may have led to greater insecurity and an increase in involuntary job changes (Benach, Muntaner & Santana, 2007; Booth, Francesconi & Frank, 2002; Marmot & Bell, 2010).

Deindustrialisation has resulted in high levels of job loss and redundancy, which are strongly linked with poor health (Brown et al*.,* 2012). Unemployed individuals are more likely to demonstrate physical and psychological morbidity (Bambra & Eikemo, 2009; Bartley, 1994; Eliason & Storrie, 2009). On the one hand this association arises because pre-existing ill-health increases the probability of unemployment compared to healthy individuals (Arrow, 1996; Lindholm, Burström & Diderichsen, 2001; Stewart, 2001). On the other hand, unemployment increases unhealthy behaviour driven by anxiety and financial strain (Kessler, Turner & House, 1989; Montgomery, Cook, Bartley & Wadsworth, 1998; Morris , Cook & Shaper, 1994; Rocha, 2001). In addition to unemployment, the sharp decline of manufacturing jobs, pre-existing poor health, the lack of transferable skills of former manufacturing employees, and increased work demands in the service sector in combination led to the acceleration of inactivity rates and claiming of disability-related benefits (Baumberg, 2014; Beatty & Fothergill, 1996; Fieldhouse & Hollywood, 1999), initially mainly among men and later also among women (Beatty, Fothergill & Powell, 2008).

Gender, employment and health

The shift from a largely manufacturing-based to a largely service-based economy has been accompanied by an increase in the labour participation of women – especially in the lower-grade service sector (Marmot et al., 1991), which has been linked with poor health outcomes. In employment, men are more likely to experience higher demands for commitment and responsibility but also higher levels of control; women are more likely to experience higher emotional demands along with greater levels of routinisation, lower salaries and greater insecurity (Campos-Serna, Ronda-Pérez, Artazcoz, Moen & Benavides, 2013; Siegrist et al., 2004).

The gendered division of the labour market has led to an extensive literature looking at the nature and health consequences of the working environment for men and women. Women tend to report higher levels of physical and mental ill-health arising from their job (Menéndez, Benach , Muntaner, Amable & O'Campo ,2007; Vermeulen & Mustard, 2000), particularly musculoskeletal health problems arising from repetitive tasks (Wijnhoven, De Vet & Picavet, 2006). Greater reporting of poor health could be due to the nature of the job, but also due to women more generally reporting poorer health than men, arising from differences in the self-perception of heath and the biological attributes of males and females (Krieger, 2003).

The greater reporting by women of employment-related poor health may be partly the result of the additional burden of domestic work, including mainly household responsibilities and caretaking activities, which can result in emotional burnout in an attempt to balance work and family obligations (Bartley, 1999; Campos-Serna et al., 2013; Madero-Cabib, 2015).

Experiences of working environments are gendered. Men and women are often treated differently in otherwise identical jobs, and men and women tend to place different levels of value on various aspects of the work environment. Specifically, women place more value on good working relationships and the absence of long hours, while men consider remuneration as more important (Bender, Donohue & Heywood, 2005). These gendered dimensions of the workplace can result in men and women experiencing ostensibly the 'same' job somewhat differently. Coupled with biological differences in susceptibility to different health conditions, men and women's health may be impacted differently in similar workplaces (Benach et al., 2007; Kreimer, 2004).

Job changes and health

Most of the literature on the health impacts of industrial sector of employment is quite static and does not consider the impact of switching employment between sectors. Workers are becoming more mobile across and within sectors, making the static approach increasingly problematic.

Research indicates that for men a shift from manufacturing to service-sector employment can lead to skills mismatches and be detrimental in terms of their identity and wellbeing (De Grip, Bosma, Willems & Van Boxtel, 2008). The notion that job shifts between industrial sectors more generally (i.e. not only from manufacturing to services), including at a finer scale within broad groupings such as 'manufacturing' and 'services', has not been assessed in existing literature. In addition, existing literature considering the impact of a shift from manufacturing to service-sector employment has (to some extent justifiably given the dominance of men in manufacturing) focussed on men and largely ignored women (Nixon, 2006; Sayce, Ackers & Green, 2007).

The effectiveness of the labour market and hiring decisions in achieving good matches between workers and jobs in terms of skills, experience and preferences is likely to impact on subsequent health outcomes. Similarly, training provision and support for staff development once in post is likely to impact on skills mismatches, skill deficiencies and job satisfaction (Stier & Yaish, 2014; Zou, 2015). Hiring practices, labour market tightness, credentialism and staff development vary by industry and are gendered (Fernandez & Mors, 2008; Hampson & Junor, 2010; Hoque & Kirkpatrick, 2003; Scherer, 2004). Some existing studies have detected a link between skills mismatch and job satisfaction (Jones, Jones, Latreille & Sloane 2009; Liljegren & Ekberg, 2010), and job satisfaction is likely to subsequently influence health outcomes.

Changing sector of employment may represent a beneficial adjustment mechanism in response to health-damaging or otherwise unsatisfactory employment. A change in the sector of employment may in some circumstances lead to a better match between the skills, preferences and capabilities of an individual and the nature and demands of their job. In other circumstances, however, the decision to change sector of employment may lead to a deterioration, rather than improvement, in the fit between the worker and job. Poor matches between worker and job could be expected to be more likely when the decision to change sector is, at least to some degree, forced upon an individual, for example in response to job loss (Gesthuizen & Dagevos, 2008; Kristensen & Westergaard-Nielsen, 2004).

In the case of re-employment following involuntary job loss, existing evidence points towards mixed health outcomes. Re-employment of former industrial employees to similar jobs contributed to the formation of a positive physical and psychological health experience compared to employees absorbed to less desired jobs (Ostry et al., 2002; Riva, Terashima, Curtis , Shucksmith & Carlebach , 2011). Consequently, the nature of job, in the case of re-employment, combined with elements of job satisfaction and the level of skills mismatch, has significant health impacts. The training and qualification credentials required to enter an industry and the level of training and staff development offered once hired are likely to play a role in minimising 'job stress' and associated adverse health outcomes.

An important distinction between changing sector and exiting employment (the latter has been linked with adverse health outcomes) is the degree

of agency involved. Although many exits from employment involve the exercise of agency on the part of the (former) worker and may even be positive choices or in others necessitated by a deterioration in health, in many cases leaving employment is involuntary, for example as the result of compulsory redundancy. In contrast, changing occupation (almost by definition) requires some agency on the part of the worker to choose and secure employment in an alternative sector – even if the initial trigger was an external factor such as job loss. An individual changing sector of employment may be doing so voluntarily or involuntarily – or making a constrained choice lying somewhere in-between. The degree of voluntariness, the extent to which agency can be exercised and the level of opportunities available in a given local labour market, all influence the likelihood that a change in occupation will lead to an improvement or deterioration in the goodness of fit between worker and job, and the attendant health impacts that may arise from that goodness or badness of fit.

Methods

Research design

Previous studies of the impacts of employment sector and changes in industrial sector of employment are characterised by selection effects and short-term duration of follow-up, and focus mainly on former male manufacturing employees (Morris & Cook, 1991). This is problematic because less healthy workers might not have the same chances of becoming re-employed or entering their favoured alternative occupations compared to their healthy counterparts. In the same context, the short duration of follow-up of those studies does not allow for long-term effects.

This study employs a longitudinal design, with a cohort of employed individuals observed at the beginning and end of a ten-year period, in order to mitigate these limitations in the existing literature. Our research design seeks to explain the probability of an individual reporting poor health at the end of the ten-year period. The key variables of interest on our models relate to industry and change in industry of employment. We control for the existence of a health condition at the beginning of the ten-year period that limits activities, which goes some way to combatting uncertainty in the direction of causation by isolating the effect of employment change on health as opposed to the effect of health in inducing change in employment.

The ideal research design would involve repeated observations at a number of points in time in order to identify the sequencing of changes in sector of employment and health. Unfortunately, such data available in British Household Panel Survey (BHPS) at the time of analysis had insufficient numbers of observations of persons moving between industrial sectors. Therefore, we used the larger number of longitudinally linked Census records available in the Scottish Longitudinal Study (SLS). Future research could make use of the expanded sample size in the UK Household Longitudinal Survey (UKHLS) which replaced the BHPS in 2009, although the SLS, at least for the time being, retains the advantage of spanning a longer time period.

Scotland was chosen as a region that has experienced substantial industrial restructuring, including large numbers coming out of manufacturing throughout the 1990s and growth of a low-paid service sector, and suffers from extensive poor health. Scotland has the added benefit of the availability of the SLS, which has a large sample drawn from Census records (a much larger proportion of the Scottish population is included than in the English Longitudinal Study) and tracks individuals over a long period. At the time of analysis, the 1991 and 2001 Censuses of Population were linked and available within the SLS.

Sample

The Scottish Longitudinal Study represents 5.3% of the Scottish population and information is collected from the Scottish Census and administrative sources (SLS-DSU, 2013). The analysis reported in this article includes two time points, 1991 and 2001. The sample used in this analysis involves the working age employed population (defined as 20-54 in 1991 in order to minimise the inclusion of those still in full-time education and those who retire within the ten-year period of observation) in the secondary and tertiary sectors in the year 1991 (i.e. the primary sector comprising agriculture, forestry and fishing was excluded). The sample selection incorporated only those employed in the year 1991.

Restricting the sample to employed persons only reflects the objective to understand the impact on health of industry and change in industry. Non-employed individuals at the start of the period of

observation (1991) are not exposed to occupational hazards, therefore are not included in our analysis (individuals employed in 1991 but not in 2001 remain in the sample, although results are not reported in the interests of space and because the links between non-employment and poor health are already well documented in existing literature).

Furthermore, individuals who were not in employment in 1991 could not, by definition, display a change in industry – a key 'treatment' in our research design (to use experimental design terminology) – when observed for the second time in 2001. For the same reason, we exclude workers over the age of 54 in 1991 as most will have retired by 2001 and therefore could not display a change in industry. We excluded from our sample individuals aged under 20 years in 1991 on the basis that a high proportion are still in education or have not yet settled on an occupation or industry. Our results, therefore, relate to employed persons aged 20-54 years and not the working-age population as a whole, although include those who have exited employment for unemployment or economic inactivity at the end of the ten-year observation period.

Those who died or left the sample for other reasons were not incorporated into the analysis. Mortality is recorded in the SLS but was not modelled as an adverse health outcome due to insufficient numbers of observations. Employees in full or part-time employment, self-employed with or without employees and those on a government scheme were included in the baseline sample of employed individuals. The sample of the individuals appearing in both the 1991 and 2001 Censuses comprised of 63,150 persons. The number of males was 32,968 and covered 52% of the sample, whereas the number of women was 30,182 (48%).

Health measures (outcome or dependent variable)

Health is captured in the 2001 Scottish Census by self-assessed health over the previous 12 months. The question is: 'Overall the last twelve months would you say your health has on the whole been?' (SLS-DSU, 2013). It has three possible responses: 'good', 'fairly good' or 'not good'. For binary logistic regression analysis, the self-assessed health responses have been transformed into a binary dependent variable: 'good/fairly good' or 'not good'. Self-assessed health is widely used in epidemiological research and longitudinal studies as a good indicator of premature mortality (Burström & Fredlund, 2001) and has been chosen for two reasons. First, it may indicate aspects or symptoms of ill-health not always detectable or yet diagnosed in formal medical procedures (Idler & Benyamini, 1997). Second, it reflects physical and mental conditions related to lifestyle choices and psycho-social influences on health (Eriksson, Unden & Elofsson, 2001; Mackenbach , Van den Bos, Joung, Van de Mheen & Stronks, 1994). Overall, the use of self-assessed health in other studies has shown good reliability (Lahelma, Maritkainen, Laaksonen & Aittomäki, 2004; Lundberg & Manderbacka, 1996). Regarding the Scottish census, a cognitive question testing was conducted by Ipsos MORI Scotland and it was concluded that this question is effective (Scotland's Census, 2015).

Also captured is liming long-term illness (LLTI), which is based on binary (yes/no) responses to the question, 'Do you have any long-term illness, health problem or disability which limits your daily activities or the work you can do?' (SLS-DSU, 2013). We modelled outcomes based on both self-assessed health and LLTI, which produced very similar results. We report the results for self-assessed health in the main body of the article and the results for LLTI in annexes.

Sectoral stability and change (key explanatory variable of interest)

The manufacturing sector was sub-divided into 'mining and manufacturing' and 'energy and construction' (table 1). The service sector was sub-divided into high and low-paid occupations (see Annex 3 and Holdsworth, 2010). Employed individuals in the year 1991 were followed up in 2001 and this formed an indicator based on stability or change between four employed states (two manufacturing sectors and two service sectors). The influences of leaving employment for unemployment or economic inactivity were also modelled, but these have not been reported in the interests of clarity and length and because the health impact of employment exits are well documented. The indicator of stability or change is based on four possible states in 1991 (the four occupational groups) and six possible states in 2001 (the four occupational groups plus the two non-employed states of unemployed or inactive). This produces a 4x6 matrix of possibilities, i.e. 24 possible combinations of states across the two

years of observation. All groups are mutually exclusive.

For moves in and out of manufacturing, the two manufacturing groups have been combined in order to maintain an adequate sample size. Again because of sample size constraints we have not differentiated between upward and downward mobility within manufacturing or within services (although our sub-division of manufacturing is not hierarchal in any case), but rather created categories that indicate whether there has been a change in occupation within the broad manufacturing and services categories. These amalgamations of cells in the 24-cell matrix reduce the total number of stability or change categories to 16. Six of these are exits from employment which have not been reported, leaving the ten occupational stability and change groups reported in each model in the results section. The models include a number of control variables relating to demographic, socioeconomic and pre-existing (1991) health (see next section), but these have not been reported in the interests of clarity and space.

Table 1. Population groups

Groups	*Description*	*Total*
Employed in Mining/Manufacturing both years	Same sector both in 1991 & 2001	6162
Employed in Construction/Energy both years	Same sector both in 1991 & 2001	3246
Changed Industry	From/To: Mining/Manufacturing 1991/2001 To/From: Construction/Energy 2001/1991	1059
Employed in High-Paid Services in 2001 (From Industry 1991)	From: Industry 1991 To: High Paid Services in 2001	3017
Employed in Low- Paid Services in 2001 (From Industry 1991)	From: Industry 1991 To: Low-Paid Services in 2001	2195
Employed in Industry in 2001 (From High-Paid Services 1991)	From: High Paid Services 1991 To: Industry in 2001	1294
Employed in Industry in 2001 (From Low-Paid Services 1991)	From: Low- Paid Services 1991 To: Industry in 2001	1542
Employed in High-Paid Services both years	Same sector both in 1991 & 2001	19449
Employed in Low-Paid Services both years	Same sector both in 1991 & 2001	7670
Changed Services	From/To: High-Paid Services 1991/2001 To/From: Low-Paid Services 2001/1991	5383

Source: Scottish Longitudinal Study

Control variables

The analysis controlled for socioeconomic and demographic indicators that acted as confounding factors when examining the influence on health of sector and change in sector of employment. The determinants of health cover a range of factors that encompasses the social and economic environment, the physical environment as well as the individual level characteristics and behaviour. Age and gender are well-known determinants of health. The younger the age of an individual, the higher the probability of him/her being healthy and being part of the labour market. Moreover, gender plays an additional part, since men and women tend to suffer from different morbidity events, have diverse health expectations and be frequently involved in different occupational tasks (WHO, 2016).

An important socioeconomic factor across the life course is education. Low levels of education are linked to poor health, fewer career prospects and therefore challenging economic circumstances. The lack of adequate educational qualifications can hinder the accessibility to additional necessities such as housing (Bambra, Gibson, Sowden, Wright, Whitehead & Petticrew, 2010). Housing assets, especially building type, number of rooms and ownership have been frequently used as indicators of wealth, instead of income, and have been linked to reduction in injuries, increase in overall wellbeing and social participation (Macintyre, Ellaway, Hiscock, Kearns, Der & McKay, 2003).

As with high living standards, stable and supportive personal relationships, proxied by marital status, can positively influence health. Being

married can act as a source of health-promotion, where individuals tend to adopt healthier lifestyle choices and through social support improve their overall wellbeing (Kiecolt-Glaser & Newton, 2001). Moreover, marriage can enhance household income and therefore provide accessibility to additional material resources and protection against individual level poverty (Lahelma et al., 2004).

More precisely, as well as gender, age was divided in five year intervals, and marital status contained three categories (single, married/remarried, divorced/widowed). Additional indicators such as educational qualifications and car ownership controlled for differences between individuals in socioeconomic status. Educational qualifications were split in three categories: First/Higher, other and no qualifications. The first category includes all the individuals with professional or university degrees/diplomas and those with Higher National Certificates (HNC/HND/SVQ level 4 or 5). Other category refers to those with a Certificate of Sixth Year Studies (CSYS), A-levels together with vocational qualifications (GSVQ/SVQ) and Ordinary National Certificates (ONC/OND Level 3). The final category involves all those without any qualifications, those with General Certificate of Secondary Education ('O' Grade) and those with vocational qualifications (Level 1).

Likewise, housing conditions such as tenure, number of rooms and heating facilities were supplementary indicators of economic status (Annex 4 includes additional descriptive information). Finally, some models controlled for pre-existing limiting long-term illness in 1991, on the basis that ill-health that limits activities could accelerate the transition towards inactivity and induce health-related changes in occupation (table 2). Unfortunately, the self-assessed health question was not introduced in the 1991 census. In addition to initial models with gender as a control variable, separate models were also run for men and women.

Table 2. Control variables

Variables	*Categories*	*Labels*
Age	20-24 25-29 30-34 35-39 40-44 45-49 50-54	Age 1991
Marital Status	Single Married-Remarried Widowed-Divorced	Marital Status 1991
Education	First or Higher Education Other Non-Degree No Qualifications	Qualifications 1991
Housing Tenure	Owner/Private rented/ Social rented	Housing Tenure 1991
Rooms	Two or less / Three /Four or more	Number of Rooms 1991
Cars	Owns a car /Does not own a car	Car ownership 1991
Central Heating	Has central heating / No central heating	Central Heating 1991
LLTI	No/Yes	Limiting long-term illness 1991

Source: Scottish Longitudinal Study

Analysis

Logistic regression was used for the analysis. The dependent variable is coded as 1 if health is reported as 'not good' and 0 if 'good' or 'fairly good'. Therefore, odds ratios reported relate to the probability of reporting poor health – the higher the odds ratio, the more likely that health is poor. The reference category in relation to the sector and sectoral change categories is 'remained in manufacturing' – odds ratios less than 1 mean poor health is less likely than for those who remained employed in manufacturing over the ten-year period, while odds ratios greater than 1 mean poor health is more likely than for those who remained employed in manufacturing over the ten-year period.

Table 3 reports three models. Model 1 controls for demographic measures (age, gender and marital status). Model 2 also controls for the socioeconomic indicators mentioned in the previous section. Model 3 adds a control for pre-existing limiting long-term illness. Coefficients and odds ratios are stable across the three models, suggesting the model is robust. Table 4 reports

separate models for men and women, which included the full suite of control variables (i.e. as model 3 in table 3, excluding the gender variable of course).

Results

The results indicate that manufacturing employment has a protective effect on health compared to employment in any part of the service sector. With all controls in place, those who remained in high-paid service sector employment over the ten-year period are most likely of all sectoral stability and change groups to report 'not good' health, and are 1.45 times more likely to report poor health than those who remained in the manufacturing sector (model 3, table 3). Even only controlling for demographic factors and not socioeconomic characteristics, high-paid service workers remain 1.20 times more likely to report poor health than manufacturing workers (model 1, table 3). After manufacturing, stable employment in low-paid services offers the second-best protective effect for health, with an odds ratio not statistically significantly different to stable manufacturing (model 3, table 3). However, without controlling for socioeconomic factors (model 1, table 3), low-paid service workers are 1.22 times more likely to report poor health than manufacturing workers and this is statistically significant at the 95% level.

Although not shown in the table, men have a higher odds ratio (1.51 $p<0.05$, model 3, table 3) of reporting poor health compared to women. Descriptive statistics without any controls indicate that women report slightly more poor health than men (8.31% of women in our sample compared to 8.18% of men reported their health was 'not good' in 2001).

Table 3. Influence of industry and change in industry on self-assessed health

Activity 2001	*Model 1*	*Model 2*	*Model 3*
Employed in Mining/Manufacturing both years (base)			
Employed in Construction/Energy both years	0.85(0.65-1.11)	0.83(0.64-1.09)	0.85(0.65-1.11)
Changed Industry	0.49(0.28-0.83)**	0.46(0.27-0.79)**	0.47(0.27-0.80)**
Employed in High-Paid Services 2001 (from Industry 1991)	0.92(0.69-1.21)	0.97(0.73-1.28)	0.96(0.72-1.26)
Employed in Low-Paid Services 2001 (from Industry 1991)	1.31(0.99-1.73)*	1.26(0.95-1.66)	1.24(0.94-1.64)
Employed in Industry 2001 (from High-Paid Services 1991)	1.01(0.69-1.46)	1.07 (0.74-1.55)	1.06(0.73-1.53)
Employed in Industry 2001 (from Low-Paid Services 1991)	1.00(0.70-1.42)	0.94(0.66-1.34)	0.94(0.66-1.34)
Employed in High-Paid Services both years	1.20(1.01-1.43)*	1.46(1.23-1.74) **	1.45(1.22-1.73)**
Employed in Low-Paid Services both years	1.22(0.99-1.49)*	1.17(0.96-1.43)	1.15(0.94-1.41)
Changed Services	1.39(1.12-1.72)**	1.40(1.13-1.74)**	1.38(1.11-1.71)**
Diagnostics (for full models including control variables and non-employed outcomes not reported)			
Pseudo R-squared	0.2381		
N	63,150		
LR Chi-Square(33)	8,096.70		
Prob > Chi-Square	<0.0000		

***.01<p<.05 **p<.01. Model 1: age and sex controls; Model 2: age, sex and socioeconomic controls; Model 3: age, sex, socioeconomic and LLTI controls. Results: odds ratios and confidence intervals. Source: Scottish Longitudinal Study**

In order to investigate whether the patterns of poor health among men and women are affected differently by industry or change in industry, we ran separate models for men and women. These are reported in table 4. The models identify patterns of poor health among men and women across industry and change in industry groupings. This allows us to conclude, for example, that industrial differences in male health are wider than for women. It does not, however, allow us to conclude that industry X is better or worse for men's health than for women's health. In other words, odds ratios for men and women cannot be directly compared along the rows in table 4 but the patterns up and down the columns can be compared. This is because odds ratios calibrate relative rather than absolute differences between the industry/change in industry groups.

For women, there are no statistically significant differences between industrial groups in likelihood of reporting poor health. Nevertheless, despite not being statistically significant, low-paid services appear to be the best type of employment for women's health (on the basis of having lowest odds ratio (0.92) of all the stable occupational groups for women in table 4). This is in contrast to manufacturing or construction which appears to be best for men's health (on the basis of both stable service sector groups for men having odds ratios greater than 1.0 compared to stable manufacturing – 1.40 and 1.13 for high- and low-paid services, respectively in table 4).

The R-squared for the model for women in table 4 is lower than that for men (0.197 for women compared to 0.288 for men) – in other words, sector of employment explains less health variation among employed women than it does among men. Men and women who change sector within manufacturing are less likely to report 'not good' health than those who remain within the same manufacturing sub-sector (odds ratio of 0.47 for men and 0.65 for women – although only statistically significant for men – table 4). The opposite is true in the service sector, where changing sector is associated with a 1.38-times increase in the probability of reporting poor health (model 3, table 3).

In terms of moves between manufacturing and services, there are again differences between the two broad sectors. A move from manufacturing to low-paid services is associated with a deterioration in health (odds ratio of 1.24 in model 3, table 3 – not statistically significant in model 3 but significant in model 1 when socioeconomic controls are removed), and this is true for men and women (table 4). In contrast, a move in the opposite direction from low-paid services to manufacturing appears to have little or no impact on health, although a small but not statistically significant benefit for men is detectable (odds ratio for men in table 4 is 0.82 compared to 1.13 for men remaining in low-paid services). Although not statistically significant, a move from high-paid services to manufacturing appears to be beneficial for men's health (odds ratio 0.92 compared to 1.40 if remaining in high-paid services) but damaging to women's health (odds ratio of 1.29 compared to 1.15 if remaining in high-paid services).

Table 4. Influence of industry and change in industry on self-assessed health by Gender

Activity 2001	*Men*	*Women*
Employed in Mining/Manufacturing both years (base)		
Employed in Construction/Energy both years	0.90(0.68-1.20)	0.96(0.37-2.48)
Changed Industry	0.47(0.26-0.85)**	0.65(0.15-2.77)
Employed in High-Paid Services 2001 (from Industry 1991)	0.94(0.67-1.32)	0.93(0.56-1.54)
Employed in Low-Paid Services 2001 (from Industry 1991)	1.22(0.88-1.70)	1.11(0.67-1.85)
Employed in Industry 2001 (from High-Paid Services 1991)	0.92(0.57-1.47)	1.29(0.69-2.40)
Employed in Industry 2001 (from Low-Paid Services 1991)	0.82(0.52-1.31)	0.99 (0.56-1.76)
Employed in High-Paid Services both years	1.40(1.13-1.75)**	1.15(0.82-1.59)
Employed in Low-Paid Services both years	1.13(0.88-1.47)	0.92(0.64-1.31)
Changed Services	1.15(0.84-1.58)	1.16(0.81-1.67)
Diagnostics (for full models including control variables and non-employed outcomes not reported)		
Pseudo R-squared	0.2877	0.1967
N	32,968	30,182
LR Chi-Square(32)	5,080.68	3,213.83
Prob > Chi-Square	<0.0000	<0.0000

***.01<p<.05 **p<.01. Model includes age, socioeconomic and LLTI controls. Results: odds ratios and confidence intervals. Source: Scottish Longitudinal Study**

Discussion

This study has identified changes in established employment-related patterns of health and trends across different sectoral categories. By adopting a longitudinal approach, it has revealed that in the case of self-assessed morbidity, which acts as a proxy of premature mortality (Bentham, Eimermann, Haynes , Lovett & Brainard , 1995), employment in the manufacturing sector has a protective effect , even after socioeconomic controls are in place – but this only applies to men. This could be due to the fact that this study covers a relatively contemporary time period (the 1990s) which captures changes to industrial sub-sectoral composition and improved safety practices across the manufacturing sector. More technologically advanced industries can possibly provide safer working environments and the time frame included in the analysis captures this change. In addition, industrial working environments include a more highly educated labour composition than in the past. More precisely, contemporary manufacturing industries, in an effort to be more innovative and remain internationally competitive, employ highly qualified individuals in posts where manual effort may be present but is minimal and sits alongside an increased premium on problem-solving and communication skills. This trend demonstrates a transition from traditional manual heavy engineering towards lighter and less labour-intensive manufacturing industries that lead to a reduction in physical risks. In construction, there has also been automation of some heavily physical tasks, but here improvements in health and safety practices may have had a transformative effect in reducing occupational hazards.

Sectoral and gender differentiation of the results have implications for our understanding of the health impacts of employment. High levels of customer pressure and employer productivity demands, leading to imbalance of work and personal life as well as physical impacts of long periods of sitting in sedentary jobs, such as musculoskeletal disorders and increased levels of obesity (Biddle et al., 2010; Lundberg, 1999; Sekine et al., 2009; Stansfeld, North, White & Marmot,

1995) can explain some of the health impacts in the service sector. Consequently, daily and long-term exposure in a sedentary working environment combined with an overall sedentary lifestyle is linked to cardiovascular mortality, diabetes, cancer and metabolic dysfunction (Biddle et al., 2010). The high-paid service sector appears to be particularly harmful to men's health.

There are important differences between the manufacturing and service sectors in the impact on health of changing industry within each of these sectors. Specifically, changing industrial sector within manufacturing is associated with lowered probability of an individual reporting poor health, whereas the opposite is the case in the service sector, where those changing sector within services are more likely to report poor health. This may reflect effective labour market matching taking place within the manufacturing sector, i.e. job changes result in better matches between worker and job. It may also be that training for new entrants to an industry is more effective in manufacturing than in services.

The association of transition within manufacturing with reduced likelihood of morbidity is true for men and women (although the results for women are not statistically significant, the odds ratios point in the same direction as for men). Nevertheless men still have higher odds of morbidity compared to women, especially for those employed in the service sector. This finding, that men are more likely to report poor health, is in contrast to numerous studies which have repeatedly shown that women report greater levels of poor health than men. Our finding, however, arises after controlling for industry of employment, indicating that men are more likely than women to work in industrial sectors that are protective of health. Furthermore, our sample includes the employed population, thus the sample size of women in some groups, especially manufacturing, may not be sufficient to produce significant results. Moreover, domestic rather than occupational aspects of women's lives may have a greater influence on their health compared to men, as suggested in existing literature.

There are two possible interpretations of these outcomes. One is that job matching and training are better in manufacturing, which minimises job stress that can result from poor hiring decisions or inadequate training perhaps more prevalent in the service sector. Therefore, the health impact of changing sector within services may reflect poor labour market matching within the service sector, for example arising from labour shortages and/or poor screening and hiring practices. It may also be the result of low levels of training producing 'job stress' and subsequent adverse health outcomes. An alternative interpretation, however, is that in manufacturing there are barriers to entry which make it more difficult for those with poor health to change industry of employment, for example arising from physical demands of jobs and/or health and safety or productivity concerns with hiring people with poor health – whereas there may be more opportunities to change sub-sector of employment in response to deteriorating health within the service sector.

Conclusion

The results suggest that employment in the manufacturing sector has a protective impact on health, especially for men, whereas employees in the service sector are more likely to report morbidity. This is a significant break with the past, when manufacturing and manual employment has been linked with poorer health. Likewise, the movement of an individual from manufacturing to low-paid service sector is associated with health deterioration. The ongoing decline of employment in manufacturing and growth in the low-paid service sector, therefore, suggest that sectoral change in the labour market is having an adverse impact on health. The gendered variations of occupational transitions imply that male employees are more susceptible to the health implications of these changes. Overall, there is a shift in morbidity risk from conventionally hazardous sectors of employment, such as manufacturing, to services. This implies that contemporary labour market conditions are accompanied with new occupational hazards related to the nature of the job and the working environment in the service sector.

As a result, individual level health implications of the labour market appear to operate at two levels: first, exposure to stressful working environments in the service secto; and second, difficult transitions between sectors in the labour market. Consequently, labour market policies should consider these two elements, since the requirements for policy implications may vary. The current labour market is characterised by a rise in

precarious forms of employment and intensification of work, especially for the low-skilled service sector jobs. Therefore, policies should focus on improving conditions related to the nature of the job and the working environment. For example, hiring practices incorporating training provision and support for staff development can create more satisfactory employment settings and promote career progression.

When transition occurs to another type of job within the same or in a different sector, effective labour markets (e.g. in providing good information about jobs and individuals) should help produce a good fit between worker and job, especially for low-skilled employees, who are more vulnerable to frequent and involuntary transitions. On the whole, transitions have become an integral part of an individual's employment history, as the 'job for life' has declined, therefore policies should integrate appropriate planning that will take into account the importance of skills and retraining, leading to more productive and less stressful adaptions to new working environments.

Consequently, future research should reflect upon the complexities of different types of employment, especially types of precarious jobs, among various sub-sectors of the economy and their connection with morbidity outcomes. Research is required in particular to understand the health impacts of new occupational hazards that appear to be emerging in the low-skilled service sector. In parallel to this, exploring detailed employment histories will shed some light on the mechanisms of association between occupational transitions and health. In addition, investigating the scope for re-employment in response to changes in health in different industries is an important area for further research, including the role of employers' occupational health practices.

Acknowledgements

The help provided by staff of the Longitudinal Studies Centre – Scotland (LSCS) is acknowledged. The LSCS is supported by the ESRC/JISC, the Scottish Funding Council, the Chief Scientist's Office and the Scottish Government. The authors alone are responsible for the interpretation of the data. Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

References

Anderson, V., Schulte, P., Sestito, J., Linn, H., & Nguyen, L. (2010). Occupational fatalities, injuries, illnesses, and related economic loss in the wholesale and retail trade sector. *American Journal of Industrial Medicine*, 53(7),673–685. http://dx.doi.org/10.1002/ajim.20813

Andersson, E., Persson, B., Bryngelsson, I.L., Magnuson, A., & Westberg, H. (2010). Cancer mortality in a Swedish cohort of pulp and paper mill workers. *International Archives of Occupational and Environmental Health*, 83(2), 123–132. http://dx.doi.org/10.1007/s00420-009-0446-1

Aronsson, G., Gustafsson, K., & Dallner, M. (2002). Work environment and health in different types of temporary jobs. *European Journal of Work and Organizational Psychology*, 11(2),151–175. http://dx.doi.org/10.1080/13594320143000898

Arrow, J.O. (1996). Estimating the influence of health as a risk factor on unemployment: a survival analysis of employment durations for workers surveyed in the German Socio-Economic Panel (1984-1990). *Social Science and Medicine*, 42(12),1651–1659. http://dx.doi.org/10.1016/0277-9536(95)00329-0

Bambra, C., & Eikemo, T.A. (2009). Welfare state regimes, unemployment and health: a comparative study of the relationship between unemployment and self- reported health in 23 European countries. *Journal of Epidemiology and Community Health*, 63(2),92–98. http://dx.doi.org/10.1136/jech.2008.077354

Bambra, C., Gibson, M., Sowden, A., Wright, K., Whitehead, M., & Petticrew, M. (2010). Tackling the wider social determinants of health and health inequalities: evidence from systematic reviews. *Journal of Epidemiology and Community Health*, 64(4), 284-291. http://dx.doi.org/10.1136/jech.2008.082743

Bartley, M. (1994). Unemployment and ill health: understanding the relationship. *Journal of Epidemiology and Community Health*, 48(4),333–337. http://dx.doi.org/10.1136/jech.48.4.333

Bartley, M. (1999). Measuring women's social position: The importance of theory. *Journal of Epidemiology and Community Health*, 53(10),601–602. http://dx.doi.org/10.1136/jech.53.10.601

Baumberg, B. (2014). Fit-for-Work – or Work Fit for Disabled People? The Changing Role of Job Demands and Control in Incapacity Claims. *Journal of Social Policy,* 43(2),289-310. http://dx.doi.org/10.1017/S0047279413000810

Beatty, C., & Fothergill, S. (1996). Labour Market Adjustment in Areas of Chronic Industrial Decline: The Case of the UK Coalfields. *Regional Studies*, 30(7),627– 640. http://dx.doi.org/10.1080/00343409612331349928

Beatty, C., Fothergill, S., & Powell, R. (2008). *Women on Incapacity Benefits: New Survey Evidence*. Sheffield: CRESR.

Benach, J., Muntaner, C., & Santana, V. (2007). *Employment conditions and health inequalities. Final Report to the WHO Commission on Social Determinants of Health*. Geneva: WHO.

Benach, J., Vives, A., Amable, M., Vanroelen, C., Tarafa, G., & Muntaner, C. (2014). Precarious employment: understanding an emerging social determinant of health. *Annual Review of Public Health*. 35, 229-253. http://dx.doi.org/10.1146/annurev-publhealth-032013-182500

Bender, K.A., Donohue, S.M., & Heywood, J.S. (2005). Job satisfaction and gender segregation, *Oxford Economic Papers*, 57(3),479-496. http://dx.doi.org/10.1093/oep/gpi015

Bentham, G., Eimermann ,J., Haynes, R., Lovet,t A., & Brainard, J. (1995). Limiting long-term illness and its associations with mortality and indicators of social deprivation. *Journal of Epidemiology and Community Health*, 49 (Supplement 2), S57–S64. http://dx.doi.org/10.1136/jech.49.Suppl_2.S57

Biddle, S., Cavill, N., Ekelund, U., Gorely ,T., Griffiths, M. D. & Jago, R. (2010). *Sedentary behaviour and obesity: review of the current scientific evidence*. London: Department of Health / Department For Children, Schools and Families.

Booker, C.L., & Sacker, A. (2013). Labour force sequences, unemployment spells and their effect on subjective well-being set points. *Longitudinal and Life Course Studies*, 4(2), 88-104

Booth, A.L., Francesconi, M. & Frank,J. (2002). Temporary Jobs: Stepping Stones or Dead Ends. *The Economic Journal*, 112(480), F189-F213. http://dx.doi.org/10.1111/1468-0297.00043

Brown, J., Demou ,E., Tristram, M.A., Gilmour, H., Sanati, K.A., & MacDonald, E. (2012). Employment status and health: understanding the health of the economically inactive population in Scotland. *BMC Public Health,* 12,327. http://dx.doi.org/10.1186/1471-2458-12-327

Burström, B., & Fredlund, P. (2001). Self-rated health: is it as good a predictor of subsequent mortality among adults in lower as well as in higher social classes? *Journal of Epidemiology and Community Health*, 55(11), 836-840. http://dx.doi.org/10.1136/jech.55.11.836

Campos-Serna, J., Ronda-Pérez, E., Artazcoz, L., Moen, B.E. & Benavides, F.G. (2013). Gender inequalities in occupational health related to the unequal distribution of working and employment conditions: a systematic review. *International Journal for Equity in Health,* 12,57. http://dx.doi.org/10.1186/1475-9276-12-57

Danson, M. (2005). Old Industrial Regions and Employability. *Urban Studies*, 42(2),285–300. http://dx.doi.org/10.1080/0042098042000316155

De Grip, A., Bosma, H., Willems, D. & Van Boxtel, M. (2008). Job-worker Mismatch and Cognitive Decline, *Oxford Economic Papers*, 60(2), 237-253. http://dx.doi.org/10.1093/oep/gpm023

Eliason ,M. & Storrie, D. (2009). Does job loss shorten life? *Journal of Human Resources*, 44(2),277–302. http://dx.doi.org/10.3368/jhr.44.2.277

Eriksson, I., Unden, A.L. & Elofsson, S. (2001) Self-rated health. Comparisons between three different measures. Results from a population study. *International Journal of Epidemiology*. 30(2),326–333. http://dx.doi.org/10.1093/ije/30.2.326

Fernandez, R.M., & Mors, L.M. (2008). Competing for Jobs: Labor Queues and Gender Sorting in the Hiring Process. *Social Science Research,* 37,1061-1080. http://dx.doi.org/10.1016/j.ssresearch.2007.10.003

Ferrie, J.E., Shipley, M.J., Stansfeld, S.A. & Marmot, M.G. (2002). Effects of chronic job insecurity and change in job security on self-reported health, minor psychiatric morbidity, physiological measures, and

health related behaviours in British civil servants: the Whitehall II study. *Journal of Epidemiology and Community Health*, 56(6), 450-454. http://dx.doi.org/10.1136/jech.56.6.450

Fieldhouse, E. & Hollywood, E. (1999) Life After Mining: Hidden Unemployment and Changing Patterns of Economic Activity amongst Miners in England and Wales, 1981-1991. *Work, Employment & Society*, 13(3),483–502.

Gesthuizen, M. & Dagevos, J. (2008). Mismatching of persons and jobs in the Netherlands: consequences for the returns to mobility. *Work, Employment & Society,* 22(3),485-506. http://dx.doi.org/10.1177/0950017008093482

Hampson, I., & Junor, A. (2010). Putting the process back in: Rethinking service sector skill. *Work, Employment & Society* 24(3),526–545. http://dx.doi.org/10.1177/0950017010371664

Hedlund, U., Järvholm, B., & Lundbäck, B. (2006). Persistence of respiratory symptoms in ex-underground iron ore miners. *Occupational Medicine*, 56(6),380–385. http://dx.doi.org/10.1093/occmed/kql035

Hein, M.J., Stayner, L.T., Lehman, E., & Dement, J.M. (2007). Follow-up study of chrysotile textile workers: cohort mortality and exposure-response. *Occupational and Environmental Medicine*, 64(9),616–625. http://dx.doi.org/10.1136/oem.2006.031005

Holdsworth, C. (2010). Patterns of pay: results of the Annual Survey of Hours and Earnings 1997 to 2009. *Economic and Labour Market Review*, 4(3),59–70. http://dx.doi.org/10.1057/elmr.2010.38

Hoque, K. & Kirkpatrick, I. (2003). Non-standard employment in the management and professional workforce: Training, consultation, and gender implications. *Work, Employment & Society*, 17(4),67–689. http://dx.doi.org/10.1177/0950017003174004

Idler, E.L. & Benyamini ,Y. (1997). Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behavior,* 38(1), 21-37. http://dx.doi.org/10.2307/2955359

Jones, M.K., Jones, R.J., Latreille, P.L. & Sloane, P.J. (2009). Training, Job Satisfaction, and Workplace Performance in Britain: Evidence from WERS. *Labour*, 23(s1), 139-175. http://dx.doi.org/10.1111/j.1467-9914.2008.00434.x

Jones, D.R., Sutton, A.J., Abrams, K.R., Fenty, J., Warren, F.,& Rushton, L. (2009). Systematic review and meta-analysis of mortality in crop protection product manufacturing workers. *Occupational and Environmental Medicine*, 66(1),7–15.http://dx.doi.org/10.1136/oem.2007.035733

Kessler, R.C., Turner, J.B. & House, J.S. (1989). Unemployment, reemployment and emotional functioning in a community sample. *American Sociological Review*, 54(4),648–657. http://dx.doi.org/10.2307/2095885

Kiecolt-Glaser, J. K. &Newton, T. L. (2001). Marriage and health: His and hers. *Psychological Bulletin*, 127(4), 472–503. http://dx.doi.org/10.1037/0033-2909.127.4.472

Korpi, T. (2001). Accumulating Disadvantage. Longitudinal Analyses of Unemployment and Physical Health in Representative Samples of the Swedish Population. *European Sociological Review*, 17(3),255–273. http://dx.doi.org/10.1093/esr/17.3.255

Kreimer, M. (2004). Labour market segregation and gender-based division of labour. *European Journal of Women's Studies*, 11(2),223-246. http://dx.doi.org/10.1177/1350506804042097

Krieger, N. (2003). Genders, sexes, and health: what are the connections – and why does it matter? *International Journal of Epidemiology*, 32(4),652-657. http://dx.doi.org/10.1093/ije/dyg156

Kristensen, N., & Westergaard-Nielsen, N. (2004). *Does Low Job Satisfaction Lead to Job Mobility?* IZA Discussion Paper No.1026. Accessed 25th September: http://ssrn.com/abstract=511722

Kuper, H., & Marmot, M. (2003). Job strain, job demands, decision latitude, and risk of coronary heart disease within the Whitehall II study. *Journal of Epidemiology and Community Health*, 57(2),147–153. http://dx.doi.org/10.1136/jech.57.2.147

Lahelma, E., Maritkainen, P., Laaksonen, M., & Aittomäki, A. (2004). Pathways between socioeconomic determinants of health. *Journal of Epidemiology and Community Health,* 58 (4),327–332. http://dx.doi.org/10.1136/jech.2003.011148

Liljegren , M., & Ekberg, K. (2010). Job mobility as predictor of health and burnout, *Journal of Occupational and Organizational Psychology*, 82(2),317-329. http://dx.doi.org/10.1348/096317908X332919

Lindholm, C., Burström, B., & Diderichsen, F., (2001). Does chronic illness cause adverse social and economic consequences among Swedes? *Scandinavian Journal of Public Health*, 29(1),63–70. http://dx.doi.org/10.1177/14034948010290011201

Loomis, D., Richardson, D.B., Bena, J.F. & Bailer, A.J. (2004). Deindustrialisation and the long term decline in fatal occupational injuries. *Occupational and Environmental Medicine*, 61(7),616–621. http://dx.doi.org/10.1136/oem.2003.009571

Lundberg, U. (1999). Stress Responses in Low-Status Jobs and Their Relationship to Health Risks: Musculoskeletal Disorders. *Annals of the New York Academy of Sciences,* 896(1),162–172. http://dx.doi.org/10.1111/j.1749-6632.1999.tb08113.x

Lundberg, O., & Manderbacka, K. (1996). Assessing the reliability of a measure of self-rated health. *Scandinavian Journal of Public Health* 24(3),218-224. http://dx.doi.org/10.1177/140349489602400314

Macintyre, S., Ellaway, A., Hiscock, R., Kearns, A., Der, G., & McKay, L. (2003). What features of the home and the area might help to explain observed relationships between housing tenure and health? Evidence from west Scotland. *Health and Place,* 9(3), 207–218. http://dx.doi.org/10.1016/S1353-8292(02)00040-0

Mackenbach, J.P., Van den Bos ,J., Joung, I.M., Van de Mheen, H. & Stronks, K. (1994). The determinants of excellent health: different from the determinants of ill-health? *International Journal of Epidemiology*, 23(6),1273–1281. http://dx.doi.org/10.1093/ije/23.6.1273

Madero-Cabib, I. (2015). The life course determinants of vulnerability in late careers. *Longitudinal and Life Course Studies*, 6(1),88-106. http://dx.doi.org/10.14301/llcs.v6i1.299

Marmot, M., & Bell, R. (2010). Challenging health inequalities-implications for the workplace. *Occupational Medicine*, 60 (3),162 – 164. http://dx.doi.org/10.1093/occmed/kqq008

Marmot, M.G., Smith, G.D., Stansfeld, S., Patel,C., North, F., Head, J., White, I., Brunner, E., & Feeney, A. (1991). Health inequalities among British civil servants: the Whitehall II study. *Lancet,* 337 (8754), 1387–93. http://dx.doi.org/10.1016/0140-6736(91)93068-K

Menéndez. M., Benach, J., Muntaner, C., Amable, M., & O'Campo, P. (2007). Is precarious employment more damaging to women's health than men's? *Social Science and Medicine*, 64(4), 776–781. http://dx.doi.org/10.1016/j.socscimed.2006.10.035

Montgomery, S.M., Cook, D.G., Bartley, M.J. & Wadsworth, M.E.J. (1998). Unemployment, cigarette smoking, alcohol consumption and body weight in young British men. *European Journal of Public Health*, 8(1),21–27. http://dx.doi.org/10.1093/eurpub/8.1.21

Morris, J.K. & Cook, D.G. (1991). A critical review of the effect of factory closures on health. *British Journal of Industrial Medicine*, 48(1),1–8. http://dx.doi.org/10.1136/oem.48.1.1

Morris, J.K., Cook, D.G. & Shaper, A.G. (1994). Loss of employment and mortality. *BMJ*, 308(6937),1135–1139. http://dx.doi.org/10.1136/bmj.308.6937.1135

Murray, R., Baldwin, J., Ridgway, K. & Winde, B. (2005). Socio-economic Decline and Adaptation: South Yorkshire's Former Coalfields. *Local Economy*, 20(4),344–359. http://dx.doi.org/10.1080/02690940500286552

Nixon, D. (2006). I just like working with my hands: employment aspirations and the meaning of work for low-skilled unemployed men in Britain's service economy. *Journal of Education and Work*, 19(2),201–217. http://dx.doi.org/10.1080/13639080600668051

Ostry, A.S., Barroetavena, M., Hershler, R., Kelly, S., Demers, P.A., Teschke, K. & Hertzman, D. (2002). Effect of de-industrialisation on working conditions and self-reported health in a sample of manufacturing workers. *Journal of Epidemiology and Community Health*, 56(7),506–509. http://dx.doi.org/10.1136/jech.56.7.506

Ostry, A.S., Hershler, R., Kelly, S., Demers, P., Teschke, K., & Hertzman, C. (2001). Effects of de-industrialization on unemployment, re-employment, and work conditions in a manufacturing workforce. *BMC Public Health*, 1(1),15. http://dx.doi.org/10.1186/1471-2458-1-15

Ostry, A.S., Marion, S., Green, L., Teschke, K., Hershler, R., Kelly, S., & Hertzman, C. (2000). The relationship between unemployment, technological change and psychosocial work conditions in British Columbia sawmills. *Critical Public Health*, 10(2),179–192. http://dx.doi.org/10.1080/713658238

Power, R. (2008). After the Black Gold: A View of Mining Heritage from Coalfield Areas in Britain. *Folklore*, 119(2),160–181. http://dx.doi.org/10.1080/00155870802056977

Riva, M., Terashima, M., Curtis, S., Shucksmith, J. & Carlebach, S. (2011). Coalfield health effects: variation in health across former coalfield areas in England. *Health and Place*, 17(2),588–597. http://dx.doi.org/10.1016/j.healthplace.2010.12.016

Rocha, C. (2001). From plant closure to reemployment in the new economy: Risks to workers dislocated from the declining garment manufacturing industry. *Journal of Sociology and Social Welfare*, XXVIII(2),53–75.

Sayce, S., Ackers, P., & Green, A.M. (2007). Work Restructuring and Changing Craft Identity: The Tale of the Disaffected Weavers (or what happens when the rug is pulled from under your feet). *Work, Employment & Society*, 21(1),85-101. http://dx.doi.org/10.1177/0950017007073618

Scherer, S. (2004). Stepping-stones or Traps? The Consequences of Labour Market Entry Positions on Future Careers in West Germany, Great Britain and Italy. *Work, Employment & Society*, 18(2), 369-94. http://dx.doi.org/10.1177/09500172004042774

Scotland's Census (2015). *Scotland's Census. Shaping our future.* Accessed 18th February from: http://www.scotlandscensus.gov.uk/variables-classification/health

Sekine, M., Chandola, T., Martikainen, P., Marmot, M., & Kagamimori, S. (2009). Socioeconomic inequalities in physical and mental functioning of British, Finnish and Japanese civil servants: role of job demand, control and work hours. *Social Science and Medicine*, 69(10),1417–1425. http://dx.doi.org/10.1016/j.socscimed.2009.08.022

Siegrist., J, Starke, D., Chandola, T., Godin, I., Marmot, M., Niedhammer, I., & Peter, R. (2004). The measurement of effort-reward imbalance at work: European comparisons. Social Science and Medicine, 58(8),1483-1499. http://dx.doi.org/10.1016/S0277-9536(03)00351-4

SLS-DSU (2013) *Scottish Longitudinal Study Development and Support Unit* (SLS-DSU). Accessed 2nd November from: http://sls.lscs.ac.uk/

Stansfeld, S.A., North, F.M., White, I. & Marmot, M.G. (1995). Work characteristics and psychiatric disorder in civil servants in London. *Journal of Epidemiology and Community Health*, 49(1),48–53. http://dx.doi.org/10.1136/jech.49.1.48

Stewart, J.M. (2001). The impact of health status on the duration of unemployment spells and the implications for studies of the impact of unemployment on health status. *Journal of Health Economics*, 20(5),781–796. http://dx.doi.org/10.1016/S0167-6296(01)00087-X

Stier, H. & Yaish, M. (2014).Occupational segregation and gender inequality in job quality: a multi-level approach. *Work Employment & Society*, 28(2), 225–246. http://dx.doi.org/10.1177/0950017013510758

Stout, N.A., Jenkins, E.L. & Pizatella, T.J. (1996). Occupational injury mortality rates in the United States: changes from 1980 to 1989. *American Journal of Public Health*, 86(1),73–77. http://dx.doi.org/10.2105/AJPH.86.1.73

Van der Hulst, M. & Geurts, S. (2001). Associations between overtime and psychological health in high and low reward jobs. *Work & Stress*, 15(3),227–240. http://dx.doi.org/10.1080/026783701110.1080/02678370110066580

Van Rossum, C.T.M., Shipley, M., Van de Mheen, H., Grobbee, D.E. & Marmot, M. (2000). Employment grade differences in cause specific mortality. A 25 year follow-up of civil servants from the first Whitehall study. *Journal of Epidemiology and Community Health*, 54(3),178–184. http://dx.doi.org/10.1136/jech.54.3.178

Vermeulen, M. & Mustard, C. (2000). Gender differences in job strain, social support at work, and psychological distress*, Journal Occupational Health Psychology*, 5(4),428-440. http://dx.doi.org/10.1037/1076-8998.5.4.428

Wessel, T. (2005) Industrial Shift, Skill Mismatch and Income Inequality: A Decomposition Analysis of Changing Distributions in the Oslo Region. *Urban Studies*, 42(9),1549–1568. http://dx.doi.org/10.1080/00420980500185421

WHO (2016) Health Impact Assessment (HIA). World Health Organization. Accessed 16 February from: http://www.who.int/hia/evidence/doh/en/

Wijnhoven, H.A., De Vet, H.C., Picavet, H.S., (2006). Explaining sex differences in chronic musculoskeletal pain in a general population. *Pain*, 124(1–2),158-166. http://dx.doi.org/10.1016/j.pain.2006.04.012

Zou, M. (2015). Gender, work orientations and job satisfaction. *Work, Employment & Society,* 29(1),3-22, http://dx.doi.org/10.1177/0950017014559267

"You are our eyes and ears": A new tool for observing parent-child interactions in large samples

Elaine Reese University of Auckland and University of Otago, New Zealand
Amy L Bird University of Auckland, New Zealand
a.bird@auckland.ac.nz
Mele Taumoepeau University of Otago, New Zealand
Joanna M Schmidt University of Waikato, New Zealand
Jatender Mohal University of Auckland, New Zealand
Cameron C Grant University of Auckland and Auckland District Health Board New Zealand
Polly E Atatoa Carr University of Auckland and University of Waikato, New Zealand
Susan M B Morton University of Auckland, New Zealand

Received September 2015 Revised May 2016) http://dx.doi.org/10.14301/llcs.v7i4.381

Keywords
Parent-child interaction, observation, longitudinal research

Abstract
Differences in parent-child interactions have implications for a range of developmental outcomes that are of interest to large longitudinal cohort studies. We describe a new method for observing parent-child conversations specifically designed to be a component of a more comprehensive collection of data about child health and development. Participants were mothers and their two-year-old children who were part of the Growing Up in New Zealand study. During a series of brief, prompted parent-child conversations, observers were trained to rate mothers' warmth, use of open-ended questions, talk about emotions and 'linking' talk, children's emotional expression, and mothers' overall use of discipline. Reliability was established before and reviewed mid-way through the one-year data collection wave. We observed differences in parent-child interaction ratings as a function of socio-demographic variables, ethnicity, and child gender that were in agreement with published research. Inter-scale correlations and correlations between observer ratings and maternal self-report measures provide preliminary evidence of convergent and discriminant validity. Specifically, higher maternal self-reported affiliation and more frequent book reading were significantly correlated with observer ratings of maternal warmth, maternal language style, and children's emotional expression; and negatively correlated with observer ratings of maternal discipline. Higher maternal self-reported parenting hostility was negatively correlated with observed maternal warmth and language; and positively correlated with observed maternal discipline. This observational method is a potentially useful technique for obtaining independent measures of parent-child conversational interactions during the preschool years within large cohort studies.

Introduction

Large, transdisciplinary longitudinal studies improve our understanding of development across the lifespan (Wadsworth & Dezateux, 2013). Collaboration across disciplines in the design and development of such studies is necessary if they are to increase our understanding of child development (Lu, 2014). This type of transdisciplinary research encourages researchers to incorporate knowledge and skills from other fields, and hence requires the development of new methodologies. The capacity to span multiple domains is one of the strengths of large cohort studies, yet inevitably limits the depth to which each individual construct can be measured. We outline here a new method specifically developed for observing parent-child interactions within a large child cohort study.

Observing parent-child interactions in large cohort studies

Developmental psychologists and child development researchers have long been interested in how parents' day-to-day behaviour with their children impacts on social, emotional, and behavioural development. The quality of parent-child interactions has been implicated in developmental outcomes ranging from psychopathology, to externalising behaviour, to educational attainment (Locke & Prinz, 2002). Parents' reports of their own parenting behaviours can be extremely informative, but for some behaviours, unbiased observations are needed (Dunn & Kendrick, 1980). For instance, it may be difficult for parents to estimate how often they smile at their children, because this behaviour is largely outside their awareness, or highly influenced by their own emotional state, their beliefs about the child, and about expectations for that behaviour in general (Gardner, 2000; Margolin et al., 1998).

Recognising the importance and value of parent-child interaction, a few large cohort studies have successfully incorporated direct observations. The Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) has incorporated parent-child play ('Two Bags Task') and book-reading tasks during the preschool and kindergarten data collection waves. Over 6,000 Two Bags Tasks interactions were coded from recordings at both timepoints (Najarian, Snow, Lennon, Kinsey & Mulligan, 2010), while a subsample (700) of the book-reading tasks were coded from recordings (Hindman, Skibbe & Foster, 2014). Having detailed coding from recordings of parent-child interactions for such large samples is ideal, but may not be financially feasible for all cohort studies.

The Home Observation Measure of Environment scale (HOME) (Bradley & Caldwell, 1984) has also been pivotal in demonstrating the feasibility and value of observing children's home environments within large cohort studies. The HOME scale is a well-validated measure of the quality of the home environment, in particular the learning environment, environmental stimulation, and parent-child interaction. The HOME scale has been instrumental in identifying associations between aspects of the home environment and developmental outcomes for both typical and atypical populations (Totsika & Sylva, 2004). The quality of the home environment – as measured by the HOME scale – has been associated with a range of important developmental outcomes such as attachment security (NICHD Early Child Care Research Network, 2001) and obesity (Strauss & Knight, 1999). The strongest associations across time have been found with cognitive and academic outcomes (Bradley et al., 1989). The development of a short-form has enabled the HOME scale to be more widely used by large cohort studies, such as the National Longitudinal Surveys of Youth. The HOME scale is succinctly described by Mott (2004, p. 260) as: "a psychometric hybrid. It represents an attempt – in my opinion, a very successful one – to integrate a psychological assessment into a large-scale data collection."

The huge contribution and strong psychometric properties of the HOME scale therefore begs the question: do large cohort studies *need* any other measure of the home environment? We present a task developed here for use in the *Growing Up in New Zealand* cohort when children were age two that uses a prompted task to elicit parent-child conversation, and focuses more specifically on the content of the verbal interaction, as well as the non-verbal context. Based on extensive research of adult-child book-reading (e.g., Fletcher & Reese, 2005; Haden, Reese & Fivush, 1996; Reese & Cox, 1999) and past-event conversations (e.g., Bird & Reese, 2006; Farrant &

Reese, 2000; Reese, Haden & Fivush, 1993; Reese & Newcombe, 2007), we use photo prompts to elicit parent-child conversation and observe the verbal and non-verbal quality of communication.

Why a semi-structured task*?*

One of the key advantages to the HOME scale is that no specific tasks are required and interviewers can code behaviours and materials as they occur in their natural environment. This has clear benefits for multi-domain home assessments, but some limitations, particularly with preschoolers. Young children may be present for only some, or none, of the home visit with parent(s) and consequently completion rates and reliability estimates are lower for children under three years compared with older children (Mott, 2004).

In contrast, semi-structured methods require parents and children to engage in a particular task designed to tap into underlying constructs more efficiently. Examples from child development research range from providing the dyad with a standard set of toys or a book to elicit play or reading style (e.g., Fuligni & Brooks-Gunn, 2013; Hindman, Skibbe & Foster, 2014), to inviting the dyad to discuss specific topics (e.g., Fivush, Berlin, Sales, Menutti-Washburn & Cassidy, 2003; cf. Margolin et al., 1998), to lab procedures such as the Strange Situation (e.g., Ainsworth, Blehar, Waters, & Wall, 1978). Shorter observations are not necessarily inferior. For example, Lovejoy, Graczyk, O'Hare and Neuman (2000) showed that, in comparison with longer observations, brief (< 10-min) observations provided larger effect sizes for differences in the positive behaviours shown by depressed and non-depressed mothers' in their interactions with their children. Both unstructured and semi-structured observational methods can provide reliable measures of parent-child interaction with concurrent long-term and predictive validity of children's development (e.g., Bird & Reese, 2006; Dunn, Brown, & Beardsall, 1991; Sroufe, Egeland, Carlson, & Collins, 2005; Taumoepeau & Ruffman, 2006). We hoped that by using a semi-structured picture discussion task we would be able to engage a higher proportion of our two-year-old cohort, while also eliciting important maternal language and parent-child interaction indicators.

Why focus in more depth on the quality of parent-child conversation?

Talking with one another is an integral part of human nature; it is also a key medium through which children develop relationships and learn about language, other people, their culture, the world, and themselves (Bowlby, 1969; Brockmeier & Carbaugh, 2001; Miller, Mintz, Hoogstra, Fug & Potts, 1992). Exposure to language through adult dialogue, narration of activities, and book reading is recognised as critical for children's language and cognitive development (Bornstein & Haynes, 1998; Huttenlocher, Haight, Bryk, Seltzer & Lyons, 1991; Huttenlocher, 1998). The preschool HOME scale measures the general home language environment well; it taps into the frequency (e.g., spontaneously talks to child at least twice) and type (e.g., parent talks with interviewers, names an object or person for child) of parental talk across the home visit.

Yet, while total language exposure is clearly important, not all communication is created equal: interactive conversations between parents and children seem particularly important. For example, Zimmerman and colleagues examined associations of total adult language exposure, adult-child interactive conversations, and total television viewing with two-four year-old children's language development. While all were associated, only adult-child interactive conversations were uniquely predictive of children's language 18 months later (Zimmerman, Gilkerson, Richards, Christakis, Xu, Gray & Yapanel, 2009). These findings highlight the importance of more detailed measurement of parent-child conversations, beyond global indicators of parental speech.

During infancy most parent talk occurs in the 'here and now', but by the age of two children are engaging in discussions that extend both temporally and contextually: talk about the past, the future, other people, and the hypothetical (Snow, 1991). Examples of such decontextualized talk could include: talking about a past trip to the zoo while reading a book about animals; building a plane out of Lego and talking about a family holiday you would like to take one day; or even talking about the day at childcare while driving home in the car.

Naturalistic observations indicate that mothers of two year olds talk about up to six past-events per hour, and two and a half year olds themselves

contribute to, on average, two past-event discussions per hour (Miller & Sperry, 1988). This 'there-and-then' talk appears particularly important for children's cognitive, social and emotional development. For example, parents who guide their children to consider new information by using open-ended questions and discuss emotional or evaluative content have children with more advanced autobiographical memory (Farrant & Reese, 2000; Reese, Haden & Fivush, 1993; Reese & Newcombe, 2007). When children are engaged in conversations with their parents about past events that highlight and explain the child's internal states and emotions, they are more likely to display a coherent self-concept and higher self-esteem (Bird & Reese, 2006; Reese, Bird & Tripp, 2007; Welch-Ross, 1997).

Book-reading is another key medium through which children are exposed to varying levels of decontextualised talk. Overall, parents who describe pictures, discuss story meaning, and use the book as a link to other comments or conversations have children with more advanced language and literacy outcomes than parents who focus more on 'reading the words' (Haden, Reese & Fivush, 1996; Reese & Cox, 1999). Interestingly, mothers' talk about internal states during a picture description task at age 15 months was associated with children's social and emotional understanding at two years of age (Taumoepeau & Ruffman, 2006).

The broader context of parent-child interactions

While the specific verbal content of parent-child interactions is increasingly seen as important, the broader non-verbal context of the parent-child interaction is also critical. Parental discipline and warmth – both key constructs measured by the HOME scale – appear central. Differences in parental discipline are associated with children's externalising behaviour difficulties and academic achievement difficulties (see Locke & Prinz, 2002, for a review). At a broad level, discipline might encompass many different practices (e.g., from verbal correction, to use of behavioural techniques such as 'time-out', to shouting or smacking); or be related to the degree of control in parenting styles (e.g., an authoritarian style characterised by high control and low warmth; Baumrind, 1967). Similar to the HOME scale, we focus here on specific unhelpful discipline behaviours that relate negatively with child outcomes, such as shouting, visible hostility or physical hitting (Bradley, Corwyn, Burchinal, McAdoo & Garcia Coll, 2001).

Parental warmth – defined as "the expression of positive affect, affection, and admiration towards the child" – is theoretically and empirically related to several other key aspects of parent-child relatedness, such as parental sensitivity and responsivity, which have their origins in attachment theory (Ainsworth et al., 1978; Bowlby, 1969). Measurement of parental warmth typically includes visible positive affect and animation when talking with the child, as well as physical affection and closeness. Parental warmth is an important moderator or protective factor for a range of developmental outcomes (Rutter, 2013). For example, maternal warmth has been found to moderate the relationship between low birth weight and ADHD (Tully, Arseneault, Caspi, Moffitt & Morgan, 2004), and between peer bullying and internalising and externalising problems (Bowes, Maughan, Caspi, Moffitt & Arseneault, 2010).

From the child's perspective, understanding and expressing emotion is a crucial developmental task, and underlies healthy functioning across multiple domains (Gross, 1998). The parent-child relationship is a fundamental context through which children develop emotional expression (Grusec, 2011): first through basic cries, and later through more complex facial, behavioural and verbal communication (Malatesta & Wilson, 1988). Children who are able to express social emotions such as empathy tend to exhibit more prosocial behaviour (Roberts & Strayer, 1996) and may elicit more positive social responses from both parents and peers (Findlay, Girardi & Coplan, 2006). Conversely, low empathy has been linked with children's externalising behaviour and social difficulties (de Wied, Gispen-de Wied & van Boxtel, 2010); and deficits in emotional understanding and expression underlie most forms of psychopathology (Aldao, Nolen-Hoeksema & Schweizer, 2010).

Across all of these domains of parent-child interaction – parental discipline, warmth, and verbal communication – reliable differences have been observed in parent-child interactions as a function of parents' ethnicity, educational achievement, and socioeconomic status. Parents from cultures with an independent orientation (e.g., most European

cultures) tend to engage in more distal (talking) and fewer proximal (touching, smiling) behaviours compared to parents from interdependent cultures (e.g., most non-European cultures; Bornstein et al.,1992). Parents with higher educational achievement and socioeconomic status also tend to talk more, and to be less harsh in their disciplinary practices (Bornstein & Bradley, 2014; Hart & Risley, 1995; Jansen et al., 2012).

Establishing Inter-Observer Agreement in Very Large Samples

A crucial issue for cohort studies that employ a large team of interviewers is the establishment of inter-observer agreement or reliability. Most developmental research studies involve, at most, several hundred participants. The 'gold standard' of reliability measurement for parent-child interaction was developed in this context: two independent observers evaluate the same behaviour from recordings of the original interaction, and inter-rater reliability is established on a subset of the sample. This gold-standard inter-rater reliability procedure is not always pragmatically possible in the context of a large cohort study.

Motivated by a desire to observe parent-child interactions for our full sample without the means to record and then code, we sought to identify other methods of establishing inter-rater reliability. We identified these from research conducted in large educational settings, where similar challenges are encountered but for a different reason – namely, the busyness of the classroom environment which results in video recordings failing to capture the behaviours of interest. To overcome this issue in the classroom setting, Coffman, Ornstein, McCall, and Curran (2008) trained all observers prior to going into the field to look for specific teacher behaviours within 30-second intervals. Inter-rater reliability was established by watching pre-prepared video clips of teacher behaviour and comparing ratings with those of an expert coder.

Similar time-sampling techniques were used in research conducted prior to the 1970s, before filming techniques had become widespread in child development research (see Lytton, 1971 for a historical review). For example, Rheingold (1960) used this method to observe maternal care of very young infants. Observers were given a checklist of 42 maternal and infant behaviours to observe and tally during set time periods. Again, inter-observer reliability was established prior to the commencement of the study: observers were paired and asked to simultaneously rate maternal and infant behaviours on seven dyads who were not part of the larger study. This procedure was repeated several times during the study to prevent observer drift.

Development of a new observational measure within a child cohort study

We developed a new observational measure of parent-child interaction, the Parent-Child Interaction task, ideally suited for inclusion in studies with larger sample sizes. This measure was designed for the *Growing Up in New Zealand* study and completed with two-year-old children and their mothers (Morton et al., 2013). Ideally, we would have repeated the observation with fathers, but for this first step we focused only on mothers as the primary caregivers for most of the children. Given the practical limitations of recording and observing thousands of parent-child interactions, observer reliability was established prior to going into the field (Coffman et al., 2008).

We asked mothers and children to engage in a semi-structured shared discussion task. The observers were the group of interviewers employed to complete the face-to-face interview with the mother of each cohort child when the child was two years old. We used time-sampling techniques to train the observers to observe one behaviour at a time for 30-second intervals. The constructs were chosen for their prominence in the child development literature as outlined above: maternal discipline, maternal warmth, maternal verbal communication (open-ended questions, linking, and emotion talk), and children's emotional expression.

Sample diversity was an essential design feature of the *Growing Up in New Zealand* cohort (Morton et al., 2013). We were interested in how parent-child interaction constructs differed across our diverse sample as a function of ethnicity, child gender, maternal education and socio-demographic status. In order to examine construct validity, we examined associations between *observations* of parent-child interactions and parent *self-report* of parenting

warmth and hostility, parent-child affiliation, and frequency of book reading and oral story-telling. We predicted that observed maternal warmth and language constructs would be positively associated with self-reported maternal warmth and affiliation, and negatively associated with maternal hostility. Similarly, we predicted that observed maternal discipline would be positively associated with maternal hostility, and negatively associated with observed maternal warmth, affiliation, and language variables. We also predicted that mothers who reported more frequent oral story-telling and book reading interactions with their children would be observed to use more of all aspects of verbal communication: more open-ended questions, linking talk, and emotion references.

Method

Cohort study participants

Participants were members of New Zealand's longitudinal pre-birth cohort study: *Growing Up in New Zealand*. Analyses here refer to 5,536 two-year-old children and their mothers who completed the Parent-Child Interaction task. This represented 88% of the 6,327 children who took part in the two-year data collection wave and 81% of the original 6,853 children in the cohort (Morton et al., 2013). Women were recruited during pregnancy from a geographically defined region of New Zealand chosen for its population diversity. All pregnant women who resided within this region and who had an estimated due date between 25th April 2009 and 25th March 2010 were eligible. A multi-faceted recruitment strategy was utilised with the goal of recruiting a sample broadly generalisable to the contemporary New Zealand national birth cohort (Morton et al., 2014a). Alignment of the enrolled cohort with the national birth cohort was confirmed (Morton et al., 2014b). Ethical approval was obtained from the Ministry of Health Ethics Committee. Written informed consent was obtained from all participating women.

Data collection procedure with the cohort

The two-year computer-assisted interview was conducted face-to-face in the child's home. Mothers were asked a range of questions across multiple domains (health, psychosocial and cognitive development, family and whanau, education, culture and identity, and neighbourhood and societal context; see Morton et al., 2013). A series of child observation tasks were completed part-way through the interview, including the Parent-Child Interaction task[i].

Parent-Child Interaction task description

Mothers were given a series of five photographs in the same order. These photographs were specifically selected to elicit the parent-child behaviours of interest (see table 1). Mothers were asked to describe each picture to their child as if they were telling a story. After 30 seconds the interviewer retrieved the picture, recorded his or her rating of the target behaviour, and handed the parent the next picture. Each picture and 30-second interval corresponded to a specific construct (e.g., the picture of a mother and child walking in a forest was coded for maternal warmth), with the exception of maternal discipline, which was rated across all five pictures. If children did not engage in the task, interviewers recorded the reason from the following response options: child was asleep; child was unwell; child not in the home; child did not engage; mother refused; child has a physical injury; child has a physical disability; child has a developmental delay; the interview was interpreted; or other.

Table 1. Parent-child interaction task: pictures, constructs and coding instructions

Picture	Photo	Parent-child Interaction Construct	Focus	Behaviour Rating
Mother and child in forest		Maternal warmth	Look	*1 = No emotional expression* Mother may be engaged with child and talking but is not smiling or laughing during the coding period. *2 = Smile only* Mother gives a smile, however slight, at any point during the coding period. *3 = Laugh OR cuddle OR kiss* Mother laughs at any point during the coding period, or cuddles / embraces the child with one or both arms, or she kisses the child.
Children washing the car		Maternal questions	Listen	*1 = No questions* *2 = One question only* *3 = Two or more questions* Once two open-ended questions have been counted, record response as 3 and stop coding. Note, an open-ended question requires more than a "yes" or "no" and contains a *What, Where, Who, How, Why or When*.
Child reaching up for Dad		Maternal desire or emotion words	Listen	*1 = No desire or emotion words* *2 = One desire or emotion word* *3 = Two or more desire or emotion words* If two desire/emotion words are counted, record response as 3 and stop coding. Desire or emotion words include *want, like, don't like, don't want, hate, happy, sad, scared, grumpy, excited*. Not included are words such as crying, hungry or tired because these are about physiological (or physical) rather than emotional states.
Child crying		Children's emotional expression (empathy)	Look	*1 = Absence of concern or empathy* *2 = Presence of concern or empathy (chid appears even mildly distressed, sad or concerned. For example, furrowed brow, raised eyebrows, downturned mouth)* Once empathy or concern is observed, record response as 2 and stop coding. The most reliable indicator of concern/empathy is a furrowed brow, which may or may not be accompanied by a downturned mouth. Note, the target is empathy so does not include other emotional expressions (e.g., smile) or words.
Child stacks blocks		Maternal linking	Listen	*1 = Description of picture, but no link to child's own experience/world.* *2 = Link to child's own experience/world.* If mother makes a link to child's experience straight away, record response as 2 and stop coding. Maternal linking occurs when the mother connects or links the task to the child's own experiences or world. One example of linking includes talking about the child's own block play (*You built a really tall tower like that yesterday*). Note, children had completed a Stack and Topple task earlier in the *Growing Up* interview, so there was an opportunity for all parents to link to recent block play. Another example could include a link to a sibling or relative (*That boy looks like your cousin).*
Across all 5 photos		Maternal discipline	Listen and look	*1 = No discipline/behavioural correction* *2 = One instance of mild behavioural correction (Don't sit like that; Stop it; gentle shove; slightly stern look)* *3 = Harsh discipline (yelling, smacking, hard shove/pull, evil eye, cursing at child) OR more than one mild behavioural correction* If even one instance of harsh discipline occurs, record as 3 and stop coding. Discipline is defined here as any instance of correcting a child's *behaviour*, either verbal or nonverbal (i.e. *Don't sit like that; Stop that*; or pulling/shoving child into place; or stern look). Corrections of a child's response (*No, I don't think that's a potato; it's a kumara*) were not included. Note that this only relates to discipline during the interaction exercise, not to anything that occurs outside of this exercise.

Observer training in administration of the Parent-Child Interaction task

Because observers would be coding in the field rather than from video recordings, inter-rater reliability needed to be established *before* the two-year interviews were conducted. Face-to-face training events were organised for the observers to establish reliability[ii].

Two expert coders (ER and MT) prepared video clips for reliability training from an observational study of toddlers' emotional development with New Zealand parents (Taumoepeau & Ruffman, 2006). The mothers in the clips were either European or Pacific; all had a two-year-old child. Videos showed mothers interacting with their toddlers in a parent-child interaction task[iii].

All 33 interviewers attended a training day (there were two separate training days for interviewers in different geographical locations). The expert coders first introduced the task to interviewers using a power-point presentation. We explained the reasons for observation, and the way that we were adapting traditional methods for the *Growing Up in New Zealand* study. We told the interviewers "you are our eyes and ears" because we would not be able to directly observe the mother-child interactions. We then introduced each target dimension to the interviewers, noting the decision points (table 1) and emphasising that for each dimension, the interviewer was only going to be looking or listening for a single type of behaviour. We believe that this focus on a single behaviour at a time is one reason we were able to succeed in establishing reliability.

Interviewers then watched the videos of mothers interacting with their toddler children and practiced coding for the constructs of interest. Feedback was given and discrepancies explained. Interviewers were then shown six video clips corresponding to the six constructs of interest. This process was repeated four more times with different videos, resulting in interviewers viewing a total of 30 different video clips of parent-child interaction (five for each of the six constructs). It should be noted that the training from video may have actually required a higher threshold than in vivo, as many of the disagreements in coding were due to difficulties in hearing or seeing important cues on the video. Interviewer agreement was calculated by dividing the number of agreements across constructs by the total number of observations. Mean interviewer agreement was 86%, range 73.3%-96.7%.

Reliability check at midpoint in data collection wave

Six months into the 12-month long data collection wave when children were aged two, measurement of the reliability of the interviewer coding was repeated[iv]. Mothers and their preschool children were recruited from a database of primarily European families. Interviewers were shown eight video clips for each of the six constructs, resulting in a total of 48 different video clips. Mean interviewer agreement was 86%, range 73.3%-90%.

Maternal self-report measures of interactions with their child

Mothers were asked a series of questions about their interactions with, and feelings about, their children. The Time Spent with Child Scale (Davies, Harold, Goeke-Morey & Cummings, 2002) was administered as a measure of parent-child affiliation (e.g., 'I enjoy having my child around me', 'I tell my child how proud I am of them when he/she is good'). Parents indicated their responses to each question on a 1-4 scale. Item responses were totalled to give an overall parental affiliation score. The Warmth/Hostility Scale (derived from the Iowa Family Interaction Rating Scale; Melby et al., 1989-1993) contained nine questions with a 1-7 response scale. Responses to five of the items were added to give a parenting warmth total, and responses to the remaining four items were added to give a parenting hostility total. Mothers were also asked to indicate on a 1-5 scale how frequently they read books with their child and how often they told stories with their child. These last two items were designed specifically for the current study and were derived from research demonstrating links between the home literacy environment and children's language development (Payne, Whitehurst & Angell, 1994). Full questionnaires with scales and items can be found at www.growingup.co.nz.

Socio-demographic measures and ethnicity classification

Women were asked a range of standard demographic questions at the antenatal interview. Area-level

socioeconomic deprivation was measured using the NZ Index of Deprivation (Salmond, Crampton & Atkinson, 2007). Maternal education was grouped into the following categories based on highest qualifications: no formal qualifications; secondary school; trade or university qualification. Women were asked to self-prioritise their ethnicity and responses were grouped into the following categories: European, Māori, Pacific, Asian and Other.

Data Analysis

Ordinal regression models were developed to test associations between socio-demographic variables and three-level parent-child interaction variables (maternal warmth, maternal open-ended questions, maternal use of emotion language and maternal discipline). Logistic regression models were developed to test associations between socio-demographic variables and categorical parent-child interaction variables (maternal linking language and children's empathy expression). Spearman correlation coefficients were used to examine associations between observed parent-child interactions and self-report measures.

Results

Descriptive statistics

Descriptive statistics were calculated for each parent-child interaction construct and are shown in table 2. There was a good spread of scores across the scales for all but the discipline dimension, for which most parents were scored as administering no discipline across all five photos.

Table 2. Number and proportion of sample receiving each Parent-Child Interaction construct rating

Parent-Child Interaction Constructs	Number (%) parents N = 5536
Maternal warmth	
1 = No emotional expression	425 (8)
2 = Smile only	1518 (29)
3 = Laugh OR cuddle OR kiss	3378 (63)
4 = Did not engage in task	17 (0)
Maternal (open-ended) questions	
1 = No questions	1031 (19)
2 = One question only	1172 (22)
3 = Two or more questions	3092 (58)
4 = Did not engage in task	43 (1)
Maternal emotion or desire words	
1 = No desire or emotion words	1628 (31)
2 = One desire or emotion word	1622 (30)
3 = Two or more desire or emotion words	1992 (37)
4 = Did not engage in task	96 (2)
Child's emotional expression / empathy	
1 = Absence of concern or empathy	2995 (56)
2 = Presence of concern or empathy	2276 (43)
4 = Did not engage in task	67 (1)
Maternal linking	
1 = Description of picture, but no link to child's own experience/world	1503 (28)
2 = Link to child's own experience/world	3774 (71)
4 = Did not engage	61 (1)
Maternal discipline	
1 = No discipline/behavioural correction	4564 (85)
2 = One instance of mild behavioural correction	634 (12)
3 = Harsh discipline OR more than one mild behavioural correction	135 (3)
4 = Did not engage in tasks	<10 (<1)

Comparing children who completed the task with those who did not

Among the 791 children for whom the Parent-Child interaction task was not completed (12.5% of the two-year cohort), the reasons for non-completion were that the child was asleep (167, 21%), the interview was completed via phone or skype (140, 18%), there were language/translation difficulties (137, 17%), the child was not at home (125, 16%) or did not co-operate (127, 16%), the child was unwell (18, 2%), or there was a developmental or other physical reason (10, 1%).

Participation in the task differed by ethnicity: 94% of European dyads, 85% of Māori dyads; 82% of Pacific dyads; 72% of Asian dyads; and 79% of Other ethnicity dyads participated. Mothers with a trade or university qualification (89%), and those with no formal qualifications (87%), were more like to engage

in the task than dyads where mothers whose highest qualification was secondary school (84%). Dyads living in the least deprived areas (93%) were more likely to engage in the interaction task than those living in medium deprivation (88%) and high deprivation (84%) areas.

Associations between Parent-Child Interaction ratings and socio-demographic variables

Distribution of parent-child interaction ratings by socio-demographic variables are given in table 3. Ordinal and logistic regression models were then developed to test associations. Each model tested maternal ethnicity, child gender, area deprivation and maternal education as predictors of each parent-child interaction variable (table 4).

Compared with European mothers, Māori mothers (OR = 0.79) were rated as displaying less warmth, as were Asian mothers (OR = 0.74). Asian mothers were also rated as using fewer open-ended questions (OR = 0.66), fewer emotion words (OR = 0.79), more discipline (OR = 1.82) and less linking language (OR = 0.74). Children of Asian mothers were rated as displaying less empathy (OR = 0.52). Compared with European mothers, Pacific mothers were rated as using fewer emotion words (OR = 0.76) and using more discipline (OR = 1.62) and their children as displaying less empathy (OR = 0.82).

Compared with mothers of daughters, mothers of sons were rated as displaying more warmth (OR = 1.17), more discipline (OR = 1.51), and sons were rated as displaying less empathy (OR = 0.80).

Compared with families living in low deprivation areas, living in high deprivation was associated with mothers asking fewer open-ended questions (OR = 0.84) and with lower child empathy ratings (OR = 0.85). Similarly, medium deprivation was associated with lower use of maternal emotion words (OR = 0.85).

Compared with mothers with no formal educational qualifications, mothers with a trade or university qualification were more likely to use open-ended questions (OR = 1.52), emotion words (OR = 1.45; as were mothers with high school qualifications, OR = 1.46), and less likely to be rated as using discipline (OR = 0.69).

Table 3. Distributions of Parent-Child Interaction scores by maternal ethnicity and education, area deprivation and child gender: n(%)

	Maternal warmth			Maternal (open-ended) questions			Maternal emotion or desire words			Maternal discipline			Maternal linking		Children's emotional expression / empathy	
	Low	Smile only	Physical affection	0	1	2+	0	1	2+	0	1 mild	Harsh or >1 mild	No links	Link to child's world	Absence	Presence
Ethnicity																
European	254 (8)	881 (27)	2175 (66)	604 (18)	688 (21)	2010 (61)	979 (30)	991 (30)	1305 (40)	2901 (87)	336 (10)	79 (2)	897 (27)	2399(72)	1796(55)	1496 (45)
Māori	72 (10)	210 (30)	411 (59)	127 (19)	172 (25)	386 (56)	210 (31)	228 (33)	244 (36)	614 (88)	66 (9)	17 (2)	200 (29)	487 (71)	360 (53)	325 (47)
Pacific	46 (7)	195 (32)	377 (61)	131 (21)	150 (24)	332 (54)	217 (36)	196 (32)	193 (32)	481 (78)	119 (19)	18 (3)	182 (30)	427 (70)	380 (62)	233 (38)
Asian	38 (7)	179 (35)	291 (57)	129 (26)	120 (24)	256 (51)	175 (35)	148 (30)	173 (35)	405 (79)	90 (18)	16 (3)	167 (34)	331 (66)	348 (70)	149 (30)
Other	12 (7)	45 (27)	107 (65)	35 (22)	34 (21)	93 (57)	41 (26)	53 (34)	62 (40)	141 (87)	18 (11)	<10 (2)	50 (31)	111 (69)	94 (60)	63 (40)
Child gender																
Male	197 (7)	758 (28)	1789 (65)	529 (19)	599 (22)	1599 (59)	849 (32)	835 (31)	1007 (37)	2287 (83)	376 (14)	84 (3)	757 (28)	1950 (72)	1608 (59)	1102 (41)
Female	228 (9)	760 (29)	1589 (62)	502 (20)	573 (22)	1493 (58)	779 (31)	787 (31)	985 (39)	2277 (88)	258 (10)	51 (2)	746 (29)	1824 (71)	1387 (54)	1174 (46)
Area deprivation																
Low	103 (7)	426 (29)	957 (64)	265 (18)	294 (20)	923 (62)	418 (29)	437 (30)	614 (42)	1303 (88)	144 (10)	42 (3)	411 (28)	1066 (72)	800 (54)	676 (46)
Medium	158 (8)	543 (27)	1318 (65)	293 (19)	441 (22)	1182 (59)	648 (32)	604(30)	743 (37)	1755 (87)	221 (11)	48 (2)	570 (28)	1434 (72)	1128 (56)	872 (44)
High	162 (9)	542 (30)	1097 (61)	370 (21)	433 (24)	979 (55)	558 (32)	578 (33)	627(36)	1496 (83)	264 (15)	45 (2)	521 (29)	1260 (71)	1058 (59)	722 (41)
Maternal education																
No formal	39 (11)	91 (27)	206 (62)	84 (25)	86 (26)	166 (49)	133 (40)	101 (30)	99 (30)	275 (82)	52 (15)	10 (3)	101 (30)	235 (70)	192 (57)	144 (43)
High school	98 (9)	348 (30)	705 (61)	242 (21)	283 (25)	617 (54)	353 (31)	348 (31)	434 (38)	957 (83)	168 (15)	31 (3)	319 (28)	823 (72)	659 (58)	477 (42)
Trade or degree	284 (7)	1067 (28)	2453 (64)	698 (18)	796 (21)	2294 (61)	1131 (30)	1168 (31)	1447 (39)	3310 (87)	409 (11)	93 (2)	1079 (29)	2694 (71)	2124 (56)	1647 (44)

Table 4. Ordinal and logistic regressions: Differences in Parent-Child Interaction variables by maternal ethnicity and education, area deprivation and child gender

	Maternal warmth			Maternal (open-ended) questions			Maternal emotion or desire words			Maternal discipline			Maternal linking (at least one link made vs none)	Children's emotional expression / empathy (displayed vs not)
	Low	Smile only	Physical affection	0	1	2+	0	1	2+	0	1 mild	Harsh or >1 mild		
	OR (95% CI)			OR (95% CI)			OR (95% CI)			OR (95% CI)			OR (95% CI)	OR (95% CI)
Ethnicity														
European	1.00			1.00			1.00			1.00			1.00	1.00
Māori	**0.79 (0.67, 0.94), *p = .009***			0.99 (0.84, 1.18)			0.96 (0.81, 1.12)			0.81 (0.62, 1.05)			0.92 (0.76, 1.12)	1.17 (0.98, 1.40)
Pacific	0.91 (0.75, 1.11)			0.91 (0.76, 1.10)			**0.76 (0.64, 0.91), *p = .003***			**1.62 (1.27, 2.07), *p = .0001***			0.88 (0.72,1.09)	**0.82 (0.67,0.99), *p* = .0007**
Asian	**0.74 (0.61, 0.89), *p = .001***			**0.66 (0.55, 0.79), *p < .0001***			**0.79 (0.67, 0.95), *p = .01***			**1.82 (1.43, 2.32), *p < .0001***			**0.74 (0.61, 0.91), *p* = .004**	**0.52 (0.42, 0.64), *p* < .0001**
Other	1.01 (0.73, 1.40)			0.89 (0.65, 1.21)			1.10 (0.81, 1.48)			1.07 (0.67, 1.70)			0.83 (0.59, 1.17)	0.83 (0.59, 1.15)
Child gender														
Male	**1.17 (1.05, 1.31), *p = .005***			1.01 (0.91, 1.13)			0.95 (0.86, 1.05)			**1.51 (1.29, 1.77), *p < .0001***			1.05 (0.93, 1.18)	**0.80 (0.72,0.90), *p* = .0002**
Female	1.00			1.00			1.00			1.00			1.00	1.00
Area deprivation														
Low	1.00			1.00			1.00			1.00			1.00	1.00
Medium	1.05 (0.92, 1.21)			0.89 (0.77, 1.01)			**0.85 (0.75, 0.96), *p = .009***			1.02 (0.84, 1.25)			0.99 (0.85, 1.15)	0.93 (0.81, 1.06)
High	0.94 (0.80, 1.10)			**0.84 (0.72, 0.97), *p = .02***			0.91 (0.79, 1.05)			1.20 (0.96, 1.49)			0.98 (0.83, 1.16)	**0.85 (0.73,0.99), *p* = .003**
Maternal education														
No formal	1.00			1.00			1.00			1.00			1.00	1.00
High school	0.98 (0.77. 1.26)			1.21 (0.96, 1.53)			**1.46 (1.16, 1.84), *p = .001***			0.89 (0.64, 1.22)			1.10 (0.84, 1.44)	1.00 (0.78, 1.28)
Trade or degree	1.09 (0.87, 1.38)			**1.52 (1.22, 1.90), *p = .0002***			**1.45 (1.17, 1.80), *p = .0007***			**0.69 (0.50, 0.93), *p = .02***			1.07 (0.82, 1.37)	1.07 (0.84, 1.35)

Parent-Child Interaction task inter-correlations

Because some of the Parent-Child Interaction scales contained only two ordinal response options, Spearman correlation coefficients were conducted (Cliff, 2014). Parent-Child Interaction inter-correlations are shown in table 5. Small but significant positive correlations were present between maternal warmth, maternal verbal communication variables, and child emotional expression. Observed maternal discipline was negatively correlated with maternal verbal communication variables and child emotional expression.

Table 5. Correlations among Parent-child Interaction task constructs

	Maternal warmth	Maternal (open-ended) questions	Maternal emotion or desire words	Maternal linking	Child's emotional expression	Maternal discipline
Maternal warmth (n = 5321)	1.00	0.09***	0.06***	-0.00	0.02	0.00
Maternal (open-ended) questions (n = 5295)		1.00	0.06***	-0.03	0.05***	-0.06***
Maternal emotion or desire words (n = 5242)			1.00	0.13***	0.09***	-0.10***
Maternal linking (n = 5277)				1.00	0.06***	-0.04***
Child's emotional expression (n = 5271)					1.00	-0.07***
Maternal discipline (n = 5333)						1.00

*** p <.0001

Convergent and discriminant construct validity

Parent-Child Interaction scale scores were correlated with maternal self-reported enjoyment of parenting, warmth and hostility towards their child, and frequency of story-telling and book reading with their child (table 6). Maternal self-report of the frequency of oral story-telling and book-reading was positively correlated with most of the positive Parent-Child Interaction constructs, and negatively correlated with observed maternal discipline. Similarly, mothers' self-reported affiliation with her child was positively correlated with all of the positive Parent-Child Interaction constructs, and negatively correlated with observed maternal discipline. Maternal self-reported warmth had few significant associations with observed Parent-Child Interaction constructs. Maternal self-reported parenting hostility was negatively correlated with the positive Parent-Child Interaction scale constructs, and positively correlated with observed discipline.

Table 6. Correlations between Parent-Child Interaction task observation ratings and maternal self-report of parent-child interactions

Observer Ratings	Maternal Self-Report				
	Frequency of oral story-telling with child	Frequency of reading books with child	Self-reported parenting warmth	Self-reported parenting hostility	Parent-child affiliation (Time Spent with Child Scale)
Maternal warmth (n = 5321)	0.01	0.09 ***	0.02	-0.06 ***	0.03 *
Maternal (open-ended) questions (n = 5295)	0.07 ***	0.09 ***	-0.01	-0.04 *	0.04 **
Maternal emotion or desire words (n = 5242)	0.03 *	0.05 ***	0.03 *	-0.06 ***	0.06 ***
Maternal linking (n = 5277)	0.01	0.03 *	0.00	-0.02 *	0.03 *
Child's emotional expression (n = 5271)	0.05 ***	0.09 ***	0.01	-0.05 **	0.06 ***
Maternal discipline (n = 5333)	-0.05 **	-0.10 ***	-0.01	0.08 ***	-0.07 ***

*** p <.0001, ** p <.001, *p <.05

Discussion

We have demonstrated the inter-observer reliability and provided preliminary evidence of the convergent and discriminant validity of a new observational tool specifically designed to measure parent-child conversations in large interdisciplinary cohort studies. Significant correlations between observed variables were in the directions predicted (e.g., maternal discipline was negatively correlated with maternal verbal communication variables and children's empathy; maternal warmth was positively correlated with maternal open-ended questions and emotion talk). The pattern of significant correlations between observer Parent-Child Interaction ratings and mothers' self-reported parenting variables were in the directions predicted, and provide preliminary evidence of convergent and discriminant validity. For example, observed discipline was positively correlated with maternal reports of parenting hostility. Observer ratings of positive Parent-Child Interaction constructs were positively correlated with maternal self-reports of parenting affiliation and book-reading and story-telling with their children; and negatively correlated with maternal self-reports of parenting hostility.

The small effect sizes warrant further discussion, particularly given statistical significance is more likely to be achieved with large samples. Small correlations between different Parent-Child Interaction constructs do indicate that we are in fact measuring discrete and specific constructs, rather than observers relying on a global impression of 'positive', or 'negative' parenting.

Associations between observed and self-report parent-child interaction variables, although generally in the direction expected, were small also. We must

acknowledge the possibility that these small effect sizes reflect larger than desired error (e.g., from observer ratings), particularly given that observer reliability was established prior to going into the field. However, observer agreement ratings were good: both at the initial training and mid-stream check.

Another explanation is simply that we are tapping into related, but meaningfully different constructs in these two modes of measurement: observation and self-report. On the one hand, this demonstrates one of the key reasons for observing parent-child interaction in the first place. Self-report parenting measures are designed to tap into parents' internal experience of their relationship with their child (e.g., "I enjoy having my child around me"; "I get angry at my child"), but parents may not always be aware of some aspects of their behaviour. In the example we gave earlier, parents may not be aware of how often they smile as an indicator of parental warmth. Indeed, there was no significant correlation between parents' observed smiling behaviour and their self-reported warmth. On the other hand, this highlights a limitation whereby self-report measures have been used to validate observational measures. Ideally we would have included another observational measure of the quality of parent-child interaction (e.g., the HOME scale) against which to compare our new task, but this simply was not possible give the interview time and overall cost constraints of the two-year data collection wave.

The current findings demonstrate that observers can be reliably trained to code the quality of parent-child conversations and interactions before going into the field, and that this reliability can be maintained across the data collection period. This is a particular strength given that while observers were experienced in data collection, none had prior experience with behavioural observation of parent-child interactions. We believe this was possible because of our emphasis during training on explaining a clear rationale for why we measure parent-child interactions, and simplifying the observation task to focus on a single behaviour during each 30-second time interval (i.e., to utilise *either* what you see *or* hear, not both). Certainly with greater personnel and financial resource, inter-observer agreement could always be improved. Ideally observers would demonstrate reliability on around 15% of their sample: for our interviewers that would equate to between thirty and forty different dyads. It was simply not possible to record this number of different interactions before going into the field: a sample of this magnitude would represent an entirely separate study. Despite this, our findings demonstrate that reliability can be achieved within the practical constraints of a large cohort study.

Parent-child conversations differed as a function of maternal ethnicity. These findings were generally consistent with past research: Western parents tend to have more elaborative discussions (of which open-ended questions is a key coding target), for example, when compared with Asian parents (Wang, 2001; Wang & Fivush, 2005; Wang, Leichtman & Davies, 2000). There is some evidence that Pacific parents (Schluter, Sundborn, Abbott & Paterson, 2007), and some groups of Asian parents (Lau, Takeuchi & Alegria, 2006) more commonly use physical discipline. The current findings highlight slightly higher levels of observed physical discipline as well as higher rates of mild behavioural corrections for Pacific and Asian mothers.

We must consider the possibility, however, that ratings for ethnic minority dyads may be partly impacted by the nature of the task. None of the pictures specifically included Māori, Pacific or Asian children. Our data for these populations may also be impacted by higher rates of non-participation, particularly for Asian families. It may also be that the nature of the task is less culturally relevant. For example, the use of prompting pictures may be less appropriate or necessary within Māori culture with its strong tradition of oral story-telling (Reese, Hayne, & MacDonald, 2008).

Gender differences were also consistent with the existing research pointing to differences in how sons and daughters are socialised. Girls are consistently found to display greater empathy than boys (see Chaplan & Aldao, 2013, for a recent review). This is thought to be due to a combination of genetically-based temperament differences and socialisation influences that encourage girls to be more socially driven and express positive rather than negative emotions (Chaplan & Aldao, 2013; Zahn-Waxler, 2000). The finding of higher maternal discipline ratings with boys is also consistent with the existing literature. Interestingly, mothers were also rated as displaying greater warmth with sons than daughters.

While this is somewhat surprising, it does concur with theory and research that emphasises warmth and discipline as two distinct parenting constructs. It also highlights the need to consider mediating and moderating factors – something that large longitudinal studies are ideally placed to do. In particular, further research is needed to examine interactions between child gender, child temperament, self-reported parenting practices, and observations of parent-child interactions.

Findings related to maternal warmth warrant further discussion. While maternal warmth is theoretically (Ainsworth et al., 1978; Bowlby, 1969) and empirically (NICHD Early Child Care Research Network, 2001) *related* to attachment security, these are distinct constructs. A secure attachment bond is characterised by responsive, sensitive care-giving which allows a child to use their caregiver as a secure base from which to explore, and to return to seek comfort. In contrast, insecure attachment is more likely to be associated with inconsistent, rejecting or avoidant caregiving, and children in turn display avoidant, anxious, or mixed patterns of exploration and responding (Ainsworth et al., 1978). Attachment security is a dyadic relationship (not a specific maternal behaviour) that encompasses cognitive internal working models as well as observed behaviours. Valid measurement of differences in attachment security is based on observations of young children under mild stress, for example, during separation (Ainsworth et al., 1978). This is very different to the observation of maternal warmth during a typical parent-child interaction used here. Similarly, our measure of child empathy is just one example of emotional expression, which may not necessarily reflect children's responding across more salient situations (e.g., if a sibling or parent hurts themselves). This highlights the importance of considering observed behaviours in combination with parental self-report.

Our findings also highlight areas where, in hindsight, a more differentiated coding scheme may be warranted. For example, more than half of our mothers used two or more open-ended questions. There may be variability in the complexity and quality of these questions that could be further separated. For example, a mother asking 'what's happening here?' and 'what else can you see?' could receive the same rating as a mother who asked 'why do you think they are washing the car?' and 'what sort of reward will they get for being so helpful?' Any future researchers using this task may want to consider modifying in this way.

Observational methods, of course, are not without criticism. The very presence of an observer may create bias, combined with the somewhat artificial nature of engaging parents and children in a semi-structured interaction task (Gardner, 2000). For example, research indicates that parents talk more and laugh less with their toddlers when they think they are being videotaped (Field & Ignatoff, 1981). However, one could argue that longitudinal cohorts may be less likely to display this type of reactivity: by the time our families were visited at age two, they had already met with *Growing Up in New Zealand* researchers on two occasions face-to-face in their own homes, and had completed at least two telephone interviews.

In conclusion, we believe that we have created a promising tool for the direct observation of parent-child behaviour in very large samples. Valid measurement requires multiple informants and multiple forms of measurement (Dunn & Kendrick, 1980). It is our opinion that including behavioural observations of parent-child interactions, as well as parent report, will only strengthen our understanding of their contribution to children's emotional, behavioural and social development across time. We look forward to extensions and refinements of the tool and training procedures as other researchers adapt it for their specific purposes.

Acknowledgements

We acknowledge the children and families who are part of the *Growing Up in New Zealand* study and contribute their immensely valuable time and knowledge. We also acknowledge all members of the *Growing Up in New Zealand* team, as well as our Kaitiaki and Scientific Advisory Groups.

Growing Up in New Zealand has received ethical approval from the Northern Y Regional Ethics Committee for data collection with participants and for specific data linkages. Reference NTY/08/06/055.

We acknowledge the key support and funding of New Zealand's Social Policy Evaluation and Research Unit (Superu) and the Ministry of Social Development. Alongside the Ministry of Social Development and Superu (previously the Families Commission), the Ministry of Health and The University of Auckland (with Auckland UniServices Limited) have contributed the most significant funding and support to the cohort. Other agencies have also contributed funding including the Ministries of Education, Justice, Science and Innovation, Women's Affairs, and Pacific Island Affairs; the Departments of Labour, and Corrections; Te Puni Kōkiri (Ministry of Māori Affairs); New Zealand Police; Sport and Recreation New Zealand; Housing New Zealand; and the Mental Health Commission. Treasury and the Health Research Council also provided support in the development phase of the study, and the Office of Ethnic Affairs, Statistics New Zealand and the Children's Commission provided consultation.

Elaine Reese contributed to the conception and design of the study, developed the data collection instruments, and completed the first and final drafts of the manuscript.

Amy Bird analysed and interpreted the data, and completed the first and final drafts of the manuscript.

Mele Taumoepeau developed the data collection instruments, revised the manuscript, and approved the final manuscript as submitted.

Joanna Schmidt assisted with the data collection instruments, revised the manuscript, and approved the final manuscript as submitted.

Jatender Mohal assisted with the data analysis approach, revised the manuscript, and approved the final manuscript as submitted.

Cameron Grant contributed to the conception and design of the study, revised the manuscript, and approved the final manuscript as submitted.

Polly Atatoa Carr revised the manuscript and approved the final manuscript as submitted.

Susan Morton conceived and designed the study, revised the manuscript, and approved the final manuscript as submitted.

References

Ainsworth, M. D. S., Blehar, M. C., Waters, E. & Wall, S. (1978). *Patterns of Attachment: A Psychological Study of the Strange Situation.* Hillsdale, N.J.: Erlbaum, 1978. http://dx.doi.org/10.1016/j.cpr.2009.11.004

Aldao, A., Nolen-Hoeksema, S. & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review*, *30*(2), 217-237.

Baumrind, D. (1967). Child care practices anteceding three patterns of preschool behavior. *Genetic Psychology Monographs*, *75*(1), 43-88

Bird, A., & Reese, E. (2006). Emotional reminiscing and the development of an autobiographical self. *Developmental Psychology*, *42*(4), 613. http://dx.doi.org/10.1037/0012-1649.42.4.613

Bornstein, M. H. & Bradley, R. H. (Eds.). (2014). *Socioeconomic Status, Parenting, and Child Development*. Routledge.

Bornstein, M. H. & Haynes, O. M. (1998). Vocabulary competence in early childhood: Measurement, latent construct, and predictive validity. *Child Development*, *69*(3), 654-671. http://dx.doi.org/10.1111/j.1467-8624.1998.tb06235.x

Bornstein, M. H., Tal, J., Rahn, C., Galperin, C. Z., Pêcheux, M. G., Lamour, M., Toda, S., Azuma, H., Ogino, M. & Tamis-LeMonda, C. S. (1992). Functional analysis of the contents of maternal speech to infants of 5 and 13 months in four cultures: Argentina, France, Japan, and the United States. *Developmental Psychology*, *28*(4), 593-603. http://dx.doi.org/10.1037/0012-1649.28.4.593

Bowes, L., Maughan, B., Caspi, A., Moffitt, T. E. & Arseneault, L. (2010). Families promote emotional and behavioural resilience to bullying: evidence of an environmental effect. *Journal of Child Psychology and Psychiatry, 51*, 809-817. http://dx.doi.org/10.1111/j.1469-7610.2010.02216.x

Bowlby, J. (1969). *Attachment and Loss: Vol. 1. Attachment*. London: Hogarth Press.

Bradley, R. H. & Caldwell, B. M. (1984). The HOME Inventory and family demographics. *Developmental Psychology*, *20*(2), 315. http://dx.doi.org/10.1037/0012-1649.20.2.315

Bradley, R. H., Caldwell, B. M., Rock, S. L., Ramey, C. T., Barnard, K. E., Gray, C., Hammond, M. A, Mitchell, S., Gottfried, A. W., Siegel, L., & Johnson, D. L. (1989). Home environment and cognitive development in the first 3 years of life: A collaborative study involving six sites and three ethnic groups in North America. *Developmental Psychology*, *25*(2), 217. http://dx.doi.org/10.1037/0012-1649.25.2.217

Bradley, R. H., Corwyn, R. F., Burchinal, M., McAdoo, H. P., & García Coll, C. (2001). The home environments of children in the United States Part II: Relations with behavioral development through age thirteen. *Child Development*, *72*(6), 1868-1886. http://dx.doi.org/10.1111/1467-8624.t01-1-00383

Brockmeier, J. & Carbaugh, D. A. (Eds.) (2001). *Narrative and Identity: Studies in Autobiography, Self and Culture* (Vol. 1). John Benjamins Publishing. http://dx.doi.org/10.1075/sin.1

Brown, J. R. & Dunn, J. (1996). Continuities in emotion understanding from three to six years. *Child Development, 67(3)*, 789-802. http://dx.doi.org/10.2307/1131861

Chaplan, T. M. & Aldao, A. (2013). Gender Differences in Emotion Expression in Children: A Meta-Analytic Review. *Psychological Bulletin, 139(4)*, 735-765. http://dx.doi.org/10.1037/a0030737

Cliff, N. (2014). *Ordinal Methods for Behavioural Data Analysis*. Psychology Press.

Coffman, J. L., Ornstein, P. A., McCall, L. E. & Curran, P. J. (2008). Linking teachers' memory-relevant language and the development of children's memory skills. *Developmental Psychology*, *44*(6), 1640. http://dx.doi.org/10.1037/a0013859

Davies, P. T., Harold, G. T., Goeke-Morey, M. C. & Cummings, E. M. (2002). Child emotional security and interparental conflict. *Monographs of the Society for Research on Child Development, 270,* 67(3).

de Wied, M., Gispen-de Wied, C. & van Boxtel, A. (2010). Empathy dysfunction in children and adolescents with disruptive behavior disorders. *European Journal of Pharmacology, 626,* 97-103. http://dx.doi.org/10.1016/j.ejphar.2009.10.016

Dunn, J., Brown, J. & Beardsall, L. (1991). Family talk about feeling states and children's later understanding of others' emotions. *Developmental Psychology*, *27*, 448-455. http://dx.doi.org/10.1037/0012-1649.27.3.448

Dunn, J. & Kendrick, C. (1980). Studying temperament and parent-child interaction: comparison of interview and direct observation. *Developmental Medicine & Child Neurology, 22(4)*, 484-496. http://dx.doi.org/10.1111/j.1469-8749.1980.tb04353.x

Farrant, K. & Reese, E. (2000). Maternal style and children's participation in reminiscing: stepping stones in children's autobiographical memory development. *Journal of Cognition and Development, 1*, 193-225. http://dx.doi.org/10.1207/S15327647JCD010203

Field, T. & Ignatoff, E. (1981). Videotaping effects on the behaviors of low income mothers and their infants during floor-play interactions. *Journal of Applied Developmental Psychology*, *2*, 227-235. http://dx.doi.org/10.1016/0193-3973(81)90003-4

Findlay, L. C., Girardi, A. & Coplan, R. J. (2006). Links between empathy, social behavior, and social understanding in early childhood. *Early Childhood Research Quarterly*, *21*(3), 347-359. http://dx.doi.org/10.1016/j.ecresq.2006.07.009

Fivush, R., Berlin, L., Sales, J. M., Menutti-Washburn, J. & Cassidy, J. (2003). Functions of parent–child

reminiscing about emotionally negative events. *Memory, 11*, 179–192. http://dx.doi.org/10.1080/741938209

Fletcher, K. L. & Reese, E. (2005). Picture book reading with young children: A conceptual framework. *Developmental Review*, *25*(1), 64-103. http://dx.doi.org/10.1016/j.dr.2004.08.009

Fuligni, A. S., & Brooks-Gunn, J. (2013). Mother–child interactions in Early Head Start: Age and ethnic differences in low-income dyads. *Parenting*, *13*(1), 1-26. http://dx.doi.org/10.1080/15295192.2013.732422

Gardner, F. (2000). Methodological issues in the direct observation of parent-child interaction: do observational findings reflect the natural behavior of participants? *Clinical Child and Family Psychology Review, 3*, 185-198. http://dx.doi.org/10.1023/A:1009503409699

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*, 271-299. http://dx.doi.org/10.1037/1089-2680.2.3.271

Grusec, J. E. (2011). Socialization processes in the family: Social and emotional development. *Annual Review of Psychology, 62*, 243-269. http://dx.doi.org/10.1146/annurev.psych.121208.131650

Haden, C. A., Reese, E. & Fivush, R. (1996). Mothers' extra-textual comments during storybook reading: Stylistic differences over time and across texts. *Discourse Processes*, *21*(2), 135-169. http://dx.doi.org/10.1080/01638539609544953

Hart, B. & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing.

Hindman, A. H., Skibbe, L. E. & Foster, T. D. (2014). Exploring the variety of parental talk during shared book reading and its contributions to preschool language and literacy: evidence from the Early Childhood Longitudinal Study-Birth Cohort. *Reading and Writing, 27,* 287-313. http://dx.doi.org/10.1007/s11145-013-9445-4

Huttenlocher, J. (1998). Language input and language growth. *Preventive Medicine*, *27*(2), 195-199. http://dx.doi.org/10.1006/pmed.1998.0301

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, *27*(2), 236. http://dx.doi.org/10.1037/0012-1649.27.2.236

Jansen, P. W., Raat, H., Mackenbach, J. P., Hofman, A., Jaddoe, V. W. V., Bakermans-Kranenburg, M. J., Van IJzendoorn, M. H., Verhulst, F. C. & Tiemeier, H. (2012). Early determinants of maternal and paternal harsh discipline: The Generation R study. *Family Relations, 61(2),* 253-270. http://dx.doi.org/10.1111/j.1741-3729.2011.00691.x

Lau, A. S., Takeuchi, D. T. & Alegria, M. (2006). Parent-to-child aggression among Asian American parents: culture, context, and vulnerability. *Journal of Marriage and Family*, *68*(5), 1261-1275. http://dx.doi.org/10.1111/j.1741-3737.2006.00327.x

Locke, L. M. & Prinz, R. J. (2002). Measurement of parental discipline and nurturance: A review. *Clinical Psychology Review, 22*, 895 – 929. http://dx.doi.org/10.1016/S0272-7358(02)00133-2

Lovejoy, M. C., Graczyk, P. A., O'Hare, E. & Neuman, G. (2000). Maternal depression and parenting behavior: A meta-analytic review. *Clinical Psychology Review, 20*, 561-592. http://dx.doi.org/10.1016/S0272-7358(98)00100-7

Lu, M. C. (2014). Improving maternal and child health across the life course: where do we go from here? *Maternal and Child Health Journal*, *18*(2), 339-343.

Lytton, H. (1971). Observation studies of parent-child interaction: A methodological review. *Child Development*, 651-684. http://dx.doi.org/10.2307/1127439

Malatesta, C. Z. & Wilson, A. (1988). Emotion cognition interaction in personality development: A discrete emotions, functionalist analysis. *British Journal of Social Psychology*, *27*(1), 91-112. http://dx.doi.org/10.1111/j.2044-8309.1988.tb00807.x

Margolin, G., Oliver, P. H., Gordis, E. B., O'Hearn, H. G., Medina, A. M., Ghosh, C. M. & Morland, L. (1998). The nuts and bolt of behavioural observation of marital and family interaction. *Clinical Child and Family Psychology Review, 1*(4)*,* 195-213. http://dx.doi.org/10.1023/A:1022608117322

Melby, J. N., Conger, R. D., Book, R., Rueter, M., Lucy, L., Repinski, D., Ahrens, K., Black, D., Brown, D., Huck, S., Mutchler, L., Rogers, S., Ross, J. & Stavros, T. (1989-1993). *The Iowa Family Interaction Rating Scales (editions 1-4). Unpublished coding manual*. Iowa State University, Institute for Social and Behavioral Research, Ames.

Miller, P. J., Mintz, J., Hoogstra, L., Fung, H. & Potts, R. (1992). The narrated self: Young children's construction of self in relation to others in conversational stories of personal experience. *Merrill-Palmer Quarterly (1982-)*, 45-67.

Miller, P. J., & Sperry, L. L. (1988). Early talk about the past: The origins of conversational stories of personal experience. *Journal of Child Language*, *15*(02), 293-315. http://dx.doi.org/10.1017/S0305000900012381

Morton, S. M. B., Atatoa Carr, P. E., Grant, C. C., Robinson, E. M., Bandara, D. K., Bird, A., Ivory, V. C., Kingi, T. K., Liang, R., Marks, E. J., Perese, L. M., Peterson, E. R., Pryor, J. E., Reese, E., Schmidt, J. M., Waldie, K. E. & Wall, C. (2013). Cohort Profile: *Growing Up in New Zealand*. *International Journal of Epidemiology, 42*(1)*,* 65-75. http://dx.doi.org/10.1093/ije/dyr206

Morton, S. M. B., Grant, C. C., Atatoa Carr, P. E., Robinson, E. M., Kinloch, J. M., Fleming, C. J., Kingi, T. R., Perese, L. M. & Liang, R. (2014a). How do you recruit and retain a prebirth cohort? Lessons learnt from *Growing Up in New Zealand*. *Evaluation and the Health Professions, 37*(4), 411-433. http://dx.doi.org/10.1177/0163278712462717

Morton, S. M. B., Ramke, J., Kinloch, J., Grant, C. C., Atatoa Carr, P. E., Leeson, H., Lee A. C. & Robinson, E. (2014b). *Growing Up in New Zealand* cohort alignment with all New Zealand births. *Australian and New Zealand Journal of Public Health, 39*(1)*,* 82-87. http://dx.doi.org/10.1111/1753-6405.12220

Mott, F. L. (2004). The utility of the HOME-SF Scale for child development research in a Large nationa longitudinal survey: The National Longitudinal Survey of Youth 1979 Cohort. *Parenting*, *4*(2-3), 259-270. http://dx.doi.org/10.1080/15295192.2004.9681273

Najarian, M., Snow, K., Lennon, J., Kinsey, S. & Mulligan, G. (2010). Early Childhood Longitudinal Study, Birth Cohort (ECLS-B). *Preschool-Kindergarten 2007 Psychometric Report (NCES* 2010-009). National Center for Education Statistics, Institute of Education Sciences, U. S. Department of Education, Washington, DC.

NICHD Early Child Care Research Network. (2001). Child-care and family predictors of preschool attachment and stability from infancy. *Developmental Psychology*, *37*(6), 847-862. http://dx.doi.org/10.1037/0012-1649.37.6.847

Payne, A. C., Whitehurst, G. J. & Angell, A. L. (1994). The role of home literacy environment in the development of language abilities in preschool children from low income families. *Early Childhood Research Quarterly, 9*, 427-440. http://dx.doi.org/10.1016/0885-2006(94)90018-3

Reese, E., Bird, A. & Tripp, G. (2007). Children's self-esteem and moral self: Links to parent–child conversations regarding emotion. *Social Development*, *16*(3), 460-478. http://dx.doi.org/10.1111/j.1467-9507.2007.00393.x

Reese, E. & Cox, A. (1999). Quality of adult book reading affects children's emergent literacy. *Developmental Psychology*, *35*(1), 20-28. http://dx.doi.org/10.1037/0012-1649.35.1.20

Reese, E., Haden, C. A. & Fivush, R. (1993). Mother-child conversations about the past: Relationships of style and memory over time. *Cognitive Development*, *8*(4), 403-430. http://dx.doi.org/10.1016/S0885-2014(05)80002-4

Reese, E., Hayne, H. & MacDonald, S. (2008). Looking back to the future: Māori and Pakeha mother–child birth stories. *Child Development*, *79*(1), 114-125. http://dx.doi.org/10.1111/j.1467-8624.2007.01114.x

Reese, E. & Newcombe, R. (2007). Training mothers in elaborative reminiscing enhances children's autobiographical memory and narrative. *Child Development*, *78*(4), 1153-1170. http://dx.doi.org/10.1111/j.1467-8624.2007.01058.x

Rheingold, H. L. (1960). The measurement of maternal care. *Child Development*, *31*(3), 565-575. http://dx.doi.org/10.2307/1126050

Roberts, W. & Strayer, J. (1996). Empathy, emotional expressiveness, and prosocial behavior. *Child Development*, 449-470. http://dx.doi.org/10.2307/1131826

Rutter, M. (2013). Annual research review: Resilience – clinical implications. *The Journal of Child Psychology and Psychiatry, 54*, 474-487. http://dx.doi.org/10.1111/j.1469-7610.2012.02615.x

Salmond, C., Crampton, P. & Atkinson, J. (2007). *NZDep2006 Index of Deprivation*. Wellington: Department of Public Health, University of Otago.

Schluter, P. J., Sundborn, G., Abbott, M. & Paterson, J. (2007). Smacking—are we being too heavy-handed? Findings from the Pacific Islands Families Study. *New Zealand Medical Journal*, *14*.

Snow, C. E. (1991). The theoretical basis for relationships between language and literacy in development. *Journal of Research in Childhood Education*, *6*(1), 5-10. http://dx.doi.org/10.1080/02568549109594817

Sroufe, L. A., Egeland, B., Carlson, E. A. & Collins, W. A. (2005). *The development of the person: The Minnesota Study of Risk and Adaptation from Birth to Adulthood*. New York: Guilford Press.

Strauss, R. S. & Knight, J. (1999). Influence of the home environment on the development of obesity in children. *Pediatrics*, *103*(6), e85-e85. http://dx.doi.org/10.1542/peds.103.6.e85

Taumoepeau, M. & Ruffman, T. (2006). Mother and infant talk about mental states relates to desire language and emotion understanding. *Child Development*, *77*(2), 465-481. http://dx.doi.org/10.1111/j.1467-8624.2006.00882.x

Totsika, V. & Sylva, K. (2004). The Home Observation for Measurement of the Environment revisited. *Child and Adolescent Mental Health*, *9*(1), 25-35. http://dx.doi.org/10.1046/j.1475-357X.2003.00073.x

Tully, L. A., Arseneault, L., Caspi, A., Moffitt, T. E. & Morgan, J. (2004). Does maternal warmth moderate the effects of birth weight on twins' Attention-Deficit/Hyperactivity Disorder (ADHD) symptoms and IQ? *Journal of Consulting and Clinical Psychology, 72,* 218-226.

Wadsworth, M. & Dezateux, C. (2013). The growing inter-disciplinary field of longitudinal studies beginning in early life: this journal's response. *Longitudinal and Life Course Studies, 4*(1)*,* 1-3.

Wang, Q. (2001). "Did you have fun?": American and Chinese mother–child conversations about shared emotional experiences. *Cognitive Development*, *16*(2), 693-715. http://dx.doi.org/10.1016/S0885-2014(01)00055-7

Wang, Q. & Fivush, R. (2005). Mother–child conversations of emotionally salient events: exploring the functions of emotional reminiscing in European-American and Chinese families. *Social Development*, *14*(3), 473-495. http://dx.doi.org/10.1111/j.1467-9507.2005.00312.x

Wang, Q., Leichtman, M. D. & Davies, K. I. (2000). Sharing memories and telling stories: American and Chinese mothers and their 3-year-olds. *Memory*, *8*(3), 159-177. http://dx.doi.org/10.1080/096582100387588

Welch-Ross, M. K. (1997). Mother–child participation in conversation about the past: Relationships to preschoolers' theory of mind. *Developmental Psychology*, *33*(4), 618-629. http://dx.doi.org/10.1037/0012-1649.33.4.618

Zahn-Waxler, C. (2000). The development of empathy, guilt, and internalization of distress: Implications for gender differences in internalizing and externalizing problems. *Anxiety, Depression, and Emotion*, 222-265. http://dx.doi.org/10.1093/acprof:oso/9780195133585.003.0011

Zimmerman, F. J., Gilkerson, J., Richards, J. A., Christakis, D. A., Xu, D., Gray, S. & Yapanel, U. (2009). Teaching by listening: The importance of adult-child conversations to language development. *Pediatrics*, *124*(1), 342-349.

Endnotes

[i] There was occasional variability depending on chid availability and cooperation, but the Parent-Child Interaction task always followed the consent process and mother interview, and was typically administered part-way through the other child observations (it always followed the Stack and Topple motor play task, and was usually before the child's height and weight measurements).

[ii] Observers were also trained to establish coding reliability for other child observation tasks not included in this paper (e.g., Stack and Topple).

[iii] The initial training videos also used a mother-child picture description task, although the dyads in these videos used different prompting pictures to those used in the main study. The mid-stream reliability training videos, however, were specifically developed for this project and used the same pictures as the main study.

[iv] Different clips were used to the initial training. The parent-child interaction task was the same as that used for the main study.

STUDY PROFILE
The UK Millennium Cohort Study: the making of a multi-purpose resource for social science and policy in the UK

Heather Joshi University College London Institute of Education, UK
h.joshi@ucl.ac.uk
Emla Fitzsimons University College London Institute of Education, UK

Received February 2016 Revised May 2016) http://dx.doi.org/10.14301/llcs.v7i4.416

Abstract

This paper gives an account of the origins, objectives and structure of the Millennium Cohort Study (MCS) – some 19,000 individuals born in the UK in 2000-2001 – and its use in a wide range of research on many aspects of their lives in childhood years. We highlight some of the mass of output on the first five surveys to age 11 in 2012. Topics discussed are social inequalities in child development; comparisons with other cohorts; areas not well covered by previous national cohorts: season of birth, fathers, ethnicity and childcare; parental behaviour; intergenerational links; social ecology and differences between and within UK countries. We also discuss the challenges faced by the National Evaluation of Sure Start (NESS) in drawing controls from the MCS. As the cohort marches to its seventh survey in 2018, and beyond, the potential for research across life course domains will only continue to grow.

Keywords

Millennium Cohort Survey, life course inequalities, Sure Start, neighbourhood effects, UK regions

Introduction

The UK inaugurated a new national cohort study to greet the new millennium with an investment in a multi-purpose data resource. The study has now run for six sweeps – at ages nine months, three, five, seven, 11 and 14 years – and is preparing for a seventh follow-up at age 17. The scale of the resource and its variety means that a complete story would extend well beyond the scope of a single article. Indeed a profile of the Millennium Cohort (MCS) up to its fifth sweep at age 11 has already been published, focussing on its potential use for epidemiological research (Connelly & Platt, 2014). The aim of this article is to explain to potential users how the objectives, structure and content of the study were shaped by the history of its forerunners in previous British cohorts and by the context of its funding. Its primary goal was to create a rich and easily accessible scientific resource, shaped by expert input. It aimed to provide continuity with the past and to reflect the present. The purpose of this paper is to provide information on its background for users of the resource. This should also be useful for anyone contemplating founding a cohort study in the future. In reviewing how objectives have so far been met, we hope to indicate the scope for further analysis.

Cohort studies have come to be regarded as an important way for research and policy to take account of the life course – the sequence of events and experiences in individual lives through a number or domains, in the context of historical developments at the level of society (Elder, 1985). A framework recognising the family and social

contexts that surround individual development (Bronfenbrenner, 1979) also contributed intellectual underpinning for the multi-dimensional approach taken by MCS. It was always intended to serve multiple purposes, which themselves grew as the sources of funding increased. The study's objectives, listed in box 1, emerged over its first few years. The first six were incorporated in the project as initially funded by the Economic and Social Research Council (ESRC). Co-funding by Government departments enhanced the resources available to achieve these six aims. It also extended the scope of the study to three further objectives, seven to nine in box 1.

Box 1: MCS Objectives (from MCS report to funders, 2001)

1. To chart the initial conditions facing new children in the new century in terms of social, economic and health advantages and disadvantages, building evidence for future research on individual development.
2. To provide a basis for comparing processes of development with the preceding British cohorts.
3. To collect information on previously neglected topics, such as the role of fathers, non-parental childcare and ethnicity.
4. To focus on the experience and aspiration of the children's parents as the immediate 'background', of the child's early years
5. To emphasise intergenerational links including those back to the parents' own childhood.
6. To investigate the wider social ecology of the family: social networks, civic engagement, community facilities and services, splicing in geo-coded data as available.
7. To cover the whole of the United Kingdom, providing big enough samples for analysis within Wales, Scotland and Northern Ireland.
8. To provide evidence for use in the National Evaluations of Sure Start and of the Children's Fund.
9. To enhance the content of the survey by collecting data from sources beyond survey interviews, drawing on supplementary sources of funding if necessary.

Our first section gives an account of the early history of the study around 2000-2001. The second describes the building up of the data resource since then, summarised in table 1. The third section reviews some results that have emerged in relation to the first seven objectives, and also addresses the eighth objective – the study's novel use in impact evaluations.

We leave a detailed treatment of objective nine, enhancements, beyond the scope of this paper. There have been a number of ways in which supplementary information has been drawn into the database, indicated in table 1. Its funding has tended to be 'added-on', mainly from government funders. These include an important and growing set of consented linkages to administrative records, which will continue to be an important part of the study. In due course these linkages of administrative records to the study should yield enough material for a separate account, including the challenges that have been faced in their implementation. Likewise, the collection of biomedical samples and measurements (such as height, weight and physical activity) are not discussed here. Some of them are described by Connelly and Platt (2014) and others are ongoing. Though their story is not told here, both types of enhancement add greatly to the wealth and complexity of this multi-faceted asset.

Box 2 points towards further information about the MCS datasets and research using them. An appendix provides further detail of the innovative use of MCS in the National Evaluation of Sure Start.

Table 1. UK Millennium Cohort Study (MCS): main sources of information

Born 2000/01 [a]

Fieldwork		2001/2	2003/4	2006	2008	2012	2015
Age (Sweep)	**Collection Mode (see note)**	9 months (MCS1)	3 years (MCS2)	5 years (MCS3)	7 years (MCS4)	11 years (MCS5)	14 years (MCS6)
Source	b	Mother Resident partner	Mother Resident partner	Mother Resident partner	Mother Resident partner	Mother Resident partner	Mother Resident partner
	c		Assessments	Assessments	Assessments	Assessments	Assessments
	d		Height/weight	+waist circumference	+waist circumference/body fat	+body fat	+body fat
	e				Cohort member	Cohort member	Cohort member
	f		Biological samples				Biological samples
	g				Accelerometry		Accelerometry
	h						Time use diary
	i		Older siblings	Older siblings			
	j			Teacher	Teacher	Teacher	
	k	Birth registration Hospital records[n]		School records[l]	School records [m]	School records[m]	
Size	o	18,818	15,808	15,460	14,043	13,469	11,938
Response rate [%]	p	96.4	81.0	79.2	72.0	69.0	61.2

Notes to table 1:

a. Eligible if born between 1/9/2000 and 31/8/2001 (England & Wales), or between 23/11/2000 and 11/01/2002 (Scotland and Northern Ireland), and resident in a stratified sample of electoral wards.
b. Home visit face-to-face and self-completion. Main respondent is almost always the mother.
c. Home visit, assessments of cognitive skills directly administered to cohort members.
d. Home visit, height and weight measured by interviewer at each survey + items indicated at MCS3 to MCS6.
e. Home visit, cohort member self-completion.
f. Biological samples collected at home visit: oral fluid (for immunities) at MCS2; saliva of cohort member and co-resident biological parents (for DNA) at MCS6.
g. Record of physical activity collected via accelerometer devices. In-home placement via post after the interview.
h. In-home placement; diary self-completion outside home visit.
i. Home visit, sibling self-completion.
j. Postal survey, outside England only MCS3; UK MCS4; England and Wales only MCS5.
k. Linkage to administrative records. Consent obtained for linkage to health records to age 14, education records to age 16, and parents' DWP and HMRC records. Data linkages are ongoing and further consents are planned.
l. Routine records of the Foundation Stage Profile, state schools in England.
m. England, Wales, and Scotland, state schools only.
n. Hospital episode of delivery.
o. Unweighted achieved sample of children, including 702 in new families added at sweep 2
p. 'Response rates' expressed as percentage of families responding out of the 19,244 ever interviewed – no adjustment for death or emigration of cohort.

Other supplementary data collection: Health visitor survey, births after assisted fertility, nursery observation and shed milk teeth.

Origins

By the 1990s, Britain already had three national birth cohort studies, 1946, 1958 and 1970 (Pearson, 2016). Their value as multi-purpose research resources and as documents of social change, mobility and inequality was increasingly appreciated, but their long-term continuation was not assured. It would have been hard for the research community to give priority to starting a new cohort if that diverted resources from the existing studies. Apart from the Avon Longitudinal Study of Parents and Children (ALSPAC) that had recruited children born around 1991 in the Bristol area of the south west of England, there was little up-to-date evidence on inequalities in early child development – a topic in which the New Labour government, elected in 1997, was particularly interested. There was thus delight and surprise in the research community when, as part of plans to mark the millennium, the government allocated additional funding, outside the regular research budget, for a new cohort. Two crucial factors in that decision shaped the study. One was the condition that at least some (preferably at least half) of the cohort members should be born in the year 2000, which meant that decisions about design and implementation had to be made to a very short timescale for such a large undertaking. The second factor was that the study was to be commissioned by the Economic and Social Research Council (ESRC). This implied that the study would be a multi-disciplinary research resource, in content and access, on the model of other ESRC investments – in contrast to what was then the 'medical model', applying to the Medical Research Council's 1946 cohort, with health-oriented hypotheses investigated primarily by in-house researchers.

A tender was published as late as February 23rd 2000 calling for a Principal Investigator (PI) on the basis of a scoping study commissioned in 1999 (Pearson, 2016). The ESRC called for a study that would enable comparison and continuity with the previous cohorts, but also build on them, producing a rich long-term research resource of use to social science and social policy, documenting children's early years in the first instance. A crucial

requirement was to spread the births over a year rather than, as in the three previous national birth cohorts, over a week. This reflected both scientific considerations - allowing for variations by season of birth – and practical ones. It was no longer feasible to deploy National Health Service (NHS) staff in the data collection. Cost considerations aside, they would not have the expertise to administer computer-assisted instruments (CAI), and there were not enough trained CAI interviewers in the UK to mobilise for a one-week swoop.

Another implication of taking births from a whole year was that they would have to be sampled. The original specification aimed for a cohort of 15,000 children from a population of births expected to be around 700,000. Sampling provided the opportunity for spatial clustering to contain fieldwork costs and create scope for multi-level modelling. Stratification of the sample would permit disproportionate representation of particular groups. The sample design was left to the bidder.

Several academic teams submitted bids over the five weeks to March 31st 2000. Six weeks later the group asked to proceed was the one based in the Centre for Longitudinal Studies (CLS), home of the 1958 and 1970 cohorts at the Institute of Education. The bid was headed by John Bynner, the director of CLS, in an interdisciplinary and inter-institution partnership with the International Centre for Health and Society, University College London (UCL) (Michael Marmot), the Institute of Child Heath, UCL (Catherine Peckham), and the Department of Psychology, City University (Dermot Bowler). Heather Joshi become the scientific director of the study, supported by a number of colleagues at CLS with experience in running and analysing cohort studies, notably John Bynner, Director of Methodology Ian Plewis, and Neville Butler, with his expertise on paediatrics and his experience of the 1958 and 1970 studies. In the short time available the team had drawn on experts from a number of disciplines in health and social sciences to propose a broadly based survey of health and development in its social context. These partnerships continued to develop, as noted below.

The award of the scientific contract for sweep one was finalised August 1st 2000, well into the Millennium year. There was then a several-month process of tendering for the fieldwork contract, alongside the following developments, which proceeded in parallel:

- negotiations with government departments for substantial co-funding
- design of the data collection instruments
- finalisation of sample design
- development of a sampling frame
- setting a fieldwork timetable.

Co- funding: The Office of National Statistics (ONS) consortium

During 2000 various government departments made a financial commitment to the first two sweeps. These permitted the enlargement of the cohort, in terms of size and content, beyond the original budget. These plans were co-ordinated by the ONS, which played an important role in maintaining the scientific integrity of an extended questionnaire, and balancing the interests of the departments, which were in turn balanced against scientific considerations through the study's governance structure.

Funding from the three devolved administrations increased the target sample size from 1,500 families in each country to 3,000 in Wales, 2,500 in Scotland and 2,000 in Northern Ireland. After these additions, it was Wales, rather than the more populous Scotland, which was the 'smaller' country with the largest presence in MCS.[i] In England, the sample was boosted by approximately 2,600 families drawn from extra wards in disadvantaged areas, as potential controls in the evaluation of Sure Start, funded by the Department for Education and Skills (DfES). Altogether the target sample size rose to over 20,000.

The Department of Health funded add-ons to the first survey included the piloting of parent-held primary care records as a source of data (not taken forward), postal surveys of mothers who had used assisted fertility treatment, and of Health Visitors[ii] and linkage of survey births to birth registration and hospital episode statistics which were all successfully undertaken (respectively Hawkes, 2006; Redshaw, Hockley & Davidson, 2007; Brasset-Grundy et al. ,2007; Hockley et al., 2008).

The consortium, which also included the Department of Work and Pensions (DWP), funded an extension of interview and, crucially, the cost of producing basic reports, not covered by ESRC funding for resource creation. In later sweeps other government departments also contributed to the study – as noted in our acknowledgements.

Design of the data collection instruments

As noted above, the main data collection from parents at the first survey was determined at the outset, by ESRC, to be by computer assisted interview and self-completion (CAPI and CASI). Telephone interviews were not considered suitable for initiating a new cohort, and in any case the quantity of information to be collected would have been too much for a single telephone contact. The internet was not, at that time, an option, when only a minority of families were online. To create a new multi-purpose longitudinal dataset, with the objectives and theoretical framework described in the introduction – namely to capture the diversity of backgrounds from which the Children of the New Century were setting out on life – the content of the interviews was developed, at some speed, in discussions with an extended group of collaborators. Their expertise spanned demography, developmental psychology, economics, epidemiology, geography, midwifery, paediatrics, public health, social psychology, sociology, statistics and survey methodology. The 17 external advisers or collaborators involved for the first sweep (from nine institutions) are listed by Dex and Joshi, (2004, p 6-7) along with 23 members of the CLS internal team of academics and professionals. Although longitudinal funding was not yet fully confirmed, the scientific content was designed with a view to a long-term future. After deliberation among the internal and external team, the content of the first survey was debated by 55 potential users of the dataset, from academe and government departments, in a one-day consultative conference organised by the CLS on October 11th 2000. This led into the CAPI development and piloting by the fieldwork contractor the National Centre for Social Research (Natcen). The resulting instrument covered a rich set of information on household demographics, pregnancy and delivery, physical health of child and parents, their mental health, the family's income, education, employment and housing, neighbourhood, parenting, childcare, various aspects of lifestyle, attitudes, and relationships. The full set of questions, typically lasting 70-75 minutes went to a main informant, almost always the mother, and a shorter set of questions, taking around 30 minutes, went to the main informant's partner – almost always the child's father. This combination is reflected in the entry for the first survey in table 1.

Sample design

The population from which every child should have a known, non-zero chance of selection was defined as all those born within eligible dates (specified in table 1 note a), alive and living in the UK at age nine months, and eligible to receive Child Benefit at that age (Plewis, 2007a). The sampling strategy was to make a selection of areas of residence, and within them to recruit 100% of the children born in the eligible period. The statistical geography available for such clusters in 2000-1 was the boundaries of electoral wards as they stood before updating at the 2001 census.

Details of how these wards were sampled, and disproportionately stratified by ethnicity, area disadvantage, and UK country, are reported in the Technical Report on Sampling (Plewis, 2007a). Wards with a relatively high proportion of minority ethnic population (in England), were identified from the 1991 census. The allocation of the rest according to economic disadvantage was based on the most recent Child Poverty Index, recording the proportion of children in a ward whose families received means-tested benefits. The cut-off value to define a disadvantaged ward was 38.4% receiving such benefits, the bottom quartile of wards in England and Wales in 1998[iii]. This gave nine strata overall, 'disadvantaged' and 'advantaged' in each of England, Scotland, Wales and Northern Ireland, plus the minority ethnic wards in England. The target samples by country and the relative size of the disadvantaged stratum within them emerged from the negotiations about co-funding. Before sampling with the aim of reaching a 20,000 target, it was necessary to forecast the number of births in each of the nine strata and estimate what proportion of their parents would respond.

Sampling Frame and Recruitment

DWP's support of the study was crucial in providing access to the Child Benefit register as a sampling frame. Officials operated the initial contact with respondents, offering an opt-out, if a child of the relevant age was identified at an address of interest. The claiming of Child Benefit was then near-universal among the resident population[iv]. It was feared that going through the written opt-in approach required via birth registration would have led to considerable bias[v]. However, the use of the opt-out was questioned by the NHS Multi-Centre Research Ethics Committee (MREC). Upon incorporating a supplementary

verbal opt-in to the interviewer, the survey received ethical approval to proceed in May 2001.

Fieldwork period

As the basic shape of the survey emerged, the fieldwork was put out to tender (August 4th 2000). The fieldwork contractor, NatCen (appointed Sept 28th 2000), then joined the deliberations with advisers and funders, about the timing of fieldwork in the first sweep and embarked on programming and piloting the questionnaire.

By this point it was clear that the aspiration to survey children born in 2000 could only be met by starting cohort birth dates in September 2000 and fixing the interview age at nine months. This put the start of fieldwork back to June 2001, which would be clear of the census and, as it turned out, a general election. Although medically focused studies put priority on collecting information at (if not before) birth, from the point of view of social research, data on conditions during the child's first year are also very valuable. Some information about the birth was collected retrospectively from the mothers, and some from routine hospital statistics. On the 'millennium' timetable there was no question of collecting biological specimens at delivery, ruling out funding from the Medical Research Council. The September to August birthdays of the cohort in England and Wales coincide with a school year. However in Scotland and Northern Ireland, fieldwork was deferred to start with births in the last week of November 2000, to avoid double sampling with a survey on infant feeding. In these two countries, recruitment was extended for an extra six weeks of birth dates in the light of fewer births than expected in the months after the start date (Plewis 2007a, Appendix 1). It was decided not to extend birth dates in England and Wales, although there were also fewer births over the relevant period, which meant that the overall target sample would not have reached 20,000.

Building up the data resource

The prospects for continuation became more secure as time went on. Funding for a second sweep (MCS2) at age three was announced by ESRC in 2001 and awarded to CLS on a single tender although on competitive bidding for fieldwork a new agency was brought in. In 2004, ESCRC confirmed funding through CLS for two more surveys, MCS3 at age five and MCS4 at age seven. In 2010 the ESRC confirmed funding for the survey at age 11 (MCS5), at age 14 (MCS6) in 2015 and indicated support for age 17 (MCS7), which was confirmed in 2015[vi]. Government funders generally offered co-funding once ESRC support was established.

The structure and progress of the study over 15 years are summarised in table 1. It shows the major sources and types of information collected, by various modes, in each of the first six surveys, spanning nine months through 14 years. They come from a variety of informants, starting with both parents, and adding in data collection from the cohort children themselves, teachers, and external sources. Some of the scientific content is discussed in the next section.

Response

By the end of sweep 1 fieldwork, 18,552 families had been interviewed, and the cohort included 18,818[vii] children, allowing for 246 sets of twins and 10 set of triplets. This represents a response rate of 72% of all the families with eligible children living at nine months in the sampled wards (Plewis, 2007a; table 7.4) and 81% of (presumed) eligible cases released by the DWP for issue to fieldwork. Though there were some differences in response at MCS1 according to characteristics known from the Child Benefit sample, they were not judged to be substantial[viii].

The uncertainty about exactly how many children should have been eligible arose from families moving in and out of sampled wards in the normal course of events. This could not be completely anticipated in the tight timetable between the four-weekly scanning of the Child Benefit register for children aged seven months and the issuing of batched assignments to interviewers. The age three survey in 2003-4 (MCS2) provided an opportunity to catch up with families who should have been in MCS1 but had been missed because they had only recently moved to an eligible address. This group of 'new families' was only recruited in England [ix]. There is no sample refreshment by immigrants. All the children in the Millennium Cohort were resident in the UK at nine months, and would cease to be eligible if they leave the UK[x].

The boost to the sample from the 'new families' brought the total of all families ever interviewed to 19,244 (and the number of children ever taking part up to 19,519). As the response from the newly issued addresses was somewhat lower, the overall

response rate, ever interviewed out of ever eligible, went down to 71%. Such response rates have become familiar in cross-sectional surveys, but are nowhere near as high as the recruitment of mothers to the 1958 and 1970 cohorts at maternity hospitals (Pearson, 2016).

Attrition

A key question is how many members are retained over time. The last two rows of table 1 take the story forward from the maximum of 19,244 families ever interviewed. Latest results show that MCS6 covered 11,938 14 year-old children in 11, 779 families (61% of the ever-interviewed). However the other 39% are not necessarily permanent losses to follow-up. The pattern of MCS response is not a one-way drain of drop-outs, particularly between the second and third surveys when 1,444 families returned (Mostafa, 2013). By the fifth survey, 54% of the families who had ever responded had done so on all five occasions, but another 20% had participated intermittently. Attrition bias, as Plewis (2007b) showed at MCS2, is more likely than initial response bias. One solution for analysts is to supplement the survey design weights with attrition weights (Mostafa, 2013).

The extensive efforts and procedures used to keep track of the cohort families are described by Calderwood (2013). There are also considerable efforts to maintain good relations with informants by regular feedback, and by limiting respondent burden. There have been small gifts for children as gestures of appreciation, but no cash incentive for participation.

Content and Coverage

Building on the interviews with mothers and fathers, established at the first survey, the structure and coverage of the surveys evolved as the child grew older. Table 1 shows the sources of information expanded to include direct measurements and questioning of the cohort child (neither of which could have been done by telephone). Box 2 indicates where to find more detail on the surveys' content. The tradition of consultation with the research community in the design of each survey has continued, with consultative conferences before each survey. The involvement of scientific peers in the governance of the study, as a collective resource, also continues.

BOX 2. How to find further information

For more information on the dataset readers are referred to the extensive documentation on the study website (www.cls.ioe.ac.uk/mcs) including technical reports on fieldwork, and all questionnaires. The data can be accessed at the UK Data Archive (https://www.ukdataservice.ac.uk).

Hansen (2014) provides a user guide to the structure and content of each sweep.

MCS Guides to Initial Findings accompany each sweep, and can be found on the study website, for extensive descriptive information on a number of topics.

Two sourcebooks have been published by the Policy Press: *Children of the 21st century: from birth to nine months* and *Children of the 21st century: the first 5 years* (edited respectively by Dex and Joshi, 2005, and Hansen, Joshi, and Dex, 2010). Each has chapters by many of the experts who collaborated in designing the study.

Our review in this profile of research uses and findings is far from comprehensive. We have not revisited the epidemiological material reviewed by Connelly and Platt (2014). Neither have we elaborated on the teacher surveys, mentioned in table 1. We draw mainly on material collected up to age seven, as it takes time for research to emerge in peer-reviewed form. These represent a small and unsystematic selection of the 700 items recorded for MCS in the CLS online bibliography (www.cls.ioe.ac.uk/Bibliography) as of May 2016, half of which are dated since 2012. Kneale et al. (2016) provide a more detailed and systematic review of published research on selected themes collected by the survey up to age seven (child behaviour, diet, BMI, immunisation, screen time, hobbies, and child self-reports).

The CLS website (www.cls.ioe.ac.uk/mcs) also gives news of current developments, training and dissemination events and the latest research findings

HOW HAVE THE FIRST EIGHT OBJECTIVES BEEN MET?

Objective 1 – Social Advantage and Disadvantage

A leading objective of the study was to chart the social inequalities at the baseline from which the cohort set out, and to track these inequalities as the cohort members progressed through childhood. They were doing so at a time of unprecedented political interest in equalising life chances in the early years. New Labour policies to support family incomes and early years' services included setting a target to cut child poverty, Working Family Tax Credits, Sure Start and the National Childcare Strategy. Yet from many angles the achievement gaps between children with more and less advantaged backgrounds remained. In respect of cognitive scores, Brown and Sullivan (2014) report that parental education and family income were the most important predictors across the board at age 11, as they had been at all surveys since age three. The gap between rich and poor children on vocabulary at ages three and five was roughly equivalent to one year's progress at both ages three and five. It is not clear if the income-related gaps opened further, as cognitive abilities have been measured in various ways (Dearden, Sibieta & Sylva, 2011; Sullivan, Ketende & Joshi, 2013; Waldfogel & Washbrook, 2010 for example). Children whose families have been in persistent poverty show the worst outcomes (Dickerson & Popli, 2016; Schoon, Hope, Ross & Duckworth, 2010)

Child and parental health outcomes, also show gradients by socioeconomic risk (see Connelly & Platt, 2014). Among these, the most marked are in the indicators of child mental health derived from Goodman's Strengths and Difficulties Questionnaire (SDQ). A large body of research on these scores up to age seven, systematically reviewed by Kneale et al. (2016), establishes a significant relationship between increased risk of children's emotional and behavioural problems and disadvantageous life circumstances. Various factors, such as cognitive ability and self-regulation, have been shown to moderate socioeconomic risks for SDQ scores, indicating possibilities of resilience, but again, complexity rather than any simple panacea. At age 11 there were still wide gaps by socioeconomic status in SDQ scores (Gutman, Joshi, Parsonage & Schoon, 2015).

Adverse associations of child outcomes with such factors as having a teenage mother, being born after an unintended pregnancy, living with two parents who were cohabiting rather than married, or moving house, have all been statistically explained by differences in socioeconomic background (respectively, Hawkes & Joshi, 2012; De la Rochebrochard & Joshi, 2013; Crawford, Goodman, Greaves & Joyce, 2012; Gambaro & Joshi, 2016).

These findings suggest that the study is meeting its main objective as a research resource documenting the dynamics of social advantage and disadvantage, an objective which is all the better served by the enlargement in the size and content of the data base permitted by government co-funding.

Objective 2 – Comparison with other cohorts

One of the principles behind the design of the MCS was that it should enable comparison with the previous British cohorts on prevalence, and particularly, processes. MCS has been used to update the historical picture on such phenomena as the shortening of mothers' employment breaks (Hansen, Hawkes & Joshi, 2009), the marginalisation of social tenants (Feinstein et al., 2008) and increased child adiposity (Johnson, Li, Kuh & Hardy, 2015) across all four national cohorts. Comparing MCS with the second generation studies of the 1958 and 1970 cohorts, whose offspring were surveyed in 1991 and 2004, showed that the social gradient in child development (vocabulary and total difficulties scores) had not widened since the early 1990s (Blanden & Machin, 2008). This was an important interim piece of evidence for the debate about trends in the process of social mobility, which will not be resolved until the Millennium cohort reaches adulthood. The scope for inter-cohort comparisons can only grow.

What was not envisaged, but has also proved fruitful, is comparison of MCS with other datasets beyond the national cohort studies. ALSPAC, the Longitudinal Study of Young People in England (LSYPE, now Next Steps) and the National Pupil Database were used alongside MCS and BCS70 to chart the social gap in educational attainment in a quasi-cohort (Goodman, Gregg & Washbrook, 2011). Another example finds a falling prevalence of psychological difficulties in seven year olds comparing MCS with the British Child and

Adolescent Mental Health Surveys (Sellers, Maughan, Pickles, Thapar & Collishaw, 2015).

International comparison has been an unanticipated opportunity. Using roughly contemporary cohorts in the USA, Canada and Australia, along with MCS, Bradbury, Corak, Waldfogel and Washbrook (2015) found a wider gap in achievement between 11 year olds with least and most educated parents in the UK than in Canada or Australia, but not as wide as in the US. Among other factors, this reflects the income gaps between the low and highly educated parents – widest in US and narrowest in Australia. MCS data has been used in international comparisons of maternal employment around the time of a birth (Crosby & Hawkes, 2007; Huerta et al., 2011). MCS has been compared with the Fragile Families and Child Wellbeing Study in the US on native/ migrant health differentials (Jackson, Kiernan, & McLanahan, 2012) and residential mobility (Lennon, Clark and Joshi, 2016). Another international study finds maternal education more protective to the risk of children's overweight in MCS and Sweden than in China (Lakshman et al., 2012).

MCS was at the forefront of a new generation of child cohorts in Europe and further afield (Pirus & Leridon, 2010). The newcomers, such as Etude Longitudinal Française dès l'Enfance (ELFE) in France, the National Education Panel (NEPS) in Germany, Growing Up in Ireland, Growing Up in Scotland, and Growing up New Zealand took note of the design, content and practices of MCS. They will in due course provide more material for international comparison, making due allowance for the different context of each survey in time and space. Communication between the studies was facilitated by the European Child Cohort Network (EUCCONET), 2008-2013, and the Society for Longitudinal and Life Course Studies, founded in 2010.

Objective 3 - Previously neglected topics

Season of birth

MCS (along with ALSPAC) advanced our understanding of the differences in educational attainment by season of birth (Crawford, Dearden & Greaves, 2014). It was already well known that children born earlier in the academic year – in September or soon after – tend to perform better academically than those born later – i.e. in the following August or other summer months, but there was little evidence on whether this is driven by differences in age at which tests are taken or in the age of starting school. Crawford et al. (2014) exploit a key feature of the MCS that the children born at the start and end of the academic year are very similar in age at the survey assessments, whereas they are a year apart in sitting the national tests. They conclude that age at test is the most important factor behind the difference between the oldest and youngest children in an academic cohort, which suggests allowing for age at test is important when interpreting children's test scores.

Ethnic group

The earlier birth cohorts had few members from ethnic minorities. By 2000 there were substantial numbers of children being born in the UK to parents with an immigrant heritage. The sample of wards was boosted in areas in England with concentrations of Black or Asian population, but this did not boost the chances of families in minority groups living outside such areas being chosen, for example, very few Chinese families were selected. Neither would the cohort cover immigration after 2000-1, from, for example, Eastern Europe. There was, as anticipated, disproportionate dropout by minority groups once recruited. That said, the data gathered has been an important source of evidence on the extent of differentials in health, health behaviour, and social circumstances of families in the larger ethnic minority groups, and their experience of racism (Bécares, Nazroo & Kelly, 2015). Although ethnic minority children tended to have poorer scores on cognitive assessments, all else equal, at age three and five, by age seven there was no apparent penalty to ethnic background over and above that of any other social disadvantages (Sullivan, Ketende & Joshi, 2013; Taylor, Rees & Davies, 2013). Responses from the children themselves provide insight into how different ethnic groups feel about their lives, organise their friendship groups, and use their free time (Collingwood & Simmonds, 2010).

Fathers

Collecting data directly from resident partners as well as a main respondent was another innovation for a national cohort (although this had been done in ALSPAC). Especially at the start, the main respondent was almost invariably the child's natural mother and any resident partner was almost always the natural father. This extended information on the family's circumstances (e.g. employment, the

child's grandparents) and fathers' health and activities with the child. The 'partner' data was however limited, in that only around four in five of the mothers (in the unweighted sample) had resident partners, and about one tenth of these did not respond. It was also limited by the length of the interview and self-completion time that could be asked of the household or the fieldwork budget. The relatively little analysis of the responses from resident fathers includes: the impact of father involvement with the child on children's emotional adjustment; couples' employment patterns, fathers' mental health, and the child's paternal grandparents (respectively: McMunn, Martin, Kelly & Sacker (2015) and references therein; Kanji (2013); Carson, Redshaw, Gray & Quigley, (2014); and Moulton, Flouri, Joshi & Sullivan (2015). There is also evidence, albeit more limited, about fathers' involvement with the cohort child when they live apart (Kiernan, 2006). An attempt to survey non-resident parents was piloted at MCS3, unsuccessfully.

Childcare

There was very little about non-maternal childcare in the 1958 and 1970 studies: they had no surveys in the pre-school years and mothers' employment was then far less common. By 2000, childcare had become a government priority. There was support for maintaining employment around maternity leave, and for various ways of meeting the costs of formal care. MCS1 asked about the childcare arrangements at nine months. The age three survey had a very detailed module about the sequence and type of arrangements, including any overlaps, since the first survey. A subset of group settings identified in MCS2 was followed up to assess the quality of care they provided (Mathers, Sylva & Joshi, 2007). Unfortunately the childcare module in the MCS2 questionnaire was over-complicated and under-piloted. The use of the term 'childcare' seems to have deterred some parents from reporting attendance at the growing number of nursery- or pre-schools. The questions at age five aimed to repair the history of attending early education and day care settings, which almost all the cohort experienced, given new policy to provide free, part-day nursery education. MCS has been used to relate the arrangements at nine months to child outcomes (Coté, Doyle, Petitclerc & Timmins, 2013; Hansen & Hawkes, 2006). There is evidence of early education raising the academic achievement of seven year olds, but only among children from poor families (George, Stokes & Wilkinson, 2012). It would have been hard for an observational study of a near-universal policy change to produce a reliable estimate of the policy's impact. The data does however provide evidence of children's experience of early years provision at a historic juncture.

Objective 4 – Focus on Parents

Given repeated contacts with both parents throughout the early years, MCS provides much more information about parents' attitudes, activities and practices in bringing up their children than the previous national cohorts. All of these factors, along with behaviours like alcohol consumption, smoking and breast-feeding, are thought to contribute to child health and development, and to contribute to the social gradient in child outcomes. Although some studies are interested in parenting per se (e.g. Haux & Platt, 2015), many link parental beliefs and behaviour to child development, also factoring in differences in the socioeconomic circumstances of the family (a few examples: Connelly & Platt, 2014; Ermisch, 2008; Goodman, Gregg & Washbrook, 2011; Hartas, 2011; Kiernan & Mensah, 2011). These all find that aspects of parenting (especially the home learning environment and a warm rather than a harsh parenting style) have some positive association with children's cognitive or behavioural outcomes. The power of parenting practices to enhance development can however be exaggerated if their correlates, such as material resources and mental health, are not also considered. There has been, and remains, scope for research on the parent's employment and/or worklessness in relation to child outcomes (eg, Hope, Pearce, Whitehead & Law, 2014; Parsons, Schoon & Vignoles, 2014). The rich dataset enables social scientists to explore the complex relationships involved.

Objective 5 – Intergenerational links within the cohort

There was originally an idea to have a three-generation stratum within the MCS, recruiting millennial births to members of the 1970 cohort, and perhaps the 1958 cohort. Despite the attractions of linking into existing rich data, this met with the objection that a cohort of a cohort's offspring is not representative of others born at the same time to older and younger parents, or of the

cohort's children born in other years. The concern about generalisablity has also applied to the second generation sub-studies of the 1958 and 1970 cohorts – those who happen to have been born before the age 33 survey of NCDS in 1991 or the age 34 survey of BCS70 in 2004, respectively. Following them up has also fallen off the priority list.

Nevertheless there is material on intergenerational experience that can be taken from retrospective material on the respondents' own childhood, and about their own parents' partnerships and occupation (Hawkes & Joshi, 2012; Moulton et al., 2015), Grandmothers also feature prominently in the research on informal childcare. Two studies found similar results, linking grandparental care with higher risks of children being overweight or obese (Pearce et al., 2010; Tanskanen, 2013).

Objective 6 – Wider social ecology

The idea of contexts beyond the immediate family affecting individual development (Bronfenbrenner, 1979) is partly served by survey questions touching on social capital and perceptions of the neighbourhood. It is also addressed by linking in geo-coded data on statistically, rather than subjectively, defined local area. The sampling frame on which the survey is based became out of date almost immediately. The irregularly sized ward was replaced for statistical purposes with more uniformly sized zones (Lower Super Output Areas in England and Wales). Indices of Multiple Deprivation became available from 2004, in country-specific versions. On the whole, indices of deprivation did not show much additional predictive power in multivariate analyses of child outcomes once the material circumstance of the family were taken into account (see for example Tzavidis, Savati, Schmidt, Flouri & Midouhas, 2016, and a number of other papers by Flouri cited therein). In one study, the 'neighbourhood effects' apparent in the cognitive scores of seven year olds were largely explained as school effects, once the school attended was included in the model (Heilman, Kelly, Stafford & Watt, 2013). Besides indicators of deprivation and other measures of social composition, information has been linked on more 'ecological' variables such as water quality, rurality and urban green space (respectively, Molitor, Best, Jackson & Richardson, 2009; Taylor et al 2013; Flouri, Midouhas & Joshi, 2014).

Although clearly contributing to knowledge and enhancing the data resource, many of these geographically linked variables are available for further research use only under secure conditions. This to prevent disclosure of localities, which could indirectly identify individuals. In any social survey, it is important to protect the anonymity of informants. In a longitudinal study, needing to maintain the continued confidence of participants, it is paramount. Hence the exploitation of its geographical potential faces this additional challenge.

The original structure of the survey, tightly clustered in electoral wards, where all births were supposed to have been recruited, rapidly lost its value for analysis. Not only did the boundaries change, but the families moved. Over half had moved home by the time of the age five survey[xi]. Although most moves were relatively short distance, the cohort became geographically dispersed. This was probably not the only reason that by the time they went to school, or even pre-school, the children were not tightly clustered in classrooms with other cohort members. The original sampled areas did not map neatly into school catchment areas, which themselves overlap. A few primary school teachers in the teacher surveys had as many as ten cohort children in their class, but the majority of children were the only member of the cohort in their class. This is not the ideal sample design for investigating classroom or peer-group effects. It has however proved useful for the study of residential mobility (Lennon et al., 2016).

Objective 7 – Analysis within and across the smaller UK countries

MCS was the first national cohort study to cover all four countries of the United Kingdom[xii]. Sample sizes in the smaller countries were boosted to yield sufficient cases for within-country analysis – of particular interest given the devolution of much domestic policy to their governments.

Analyses of MCS regularly allow for differences across Scotland, Wales and Northern Ireland (if only to address survey design). The devolved governments have commissioned several research reports from CLS describing results within countries. The four for Scotland include a report on the drivers of unhealthy weight in children (Connelly, 2011). The Welsh Assembly Government also commissioned four reports from CLS and used MCS

data in its own publications (Welsh Assembly Government, 2011). Two reports for Northern Ireland drew strength from comparisons with other countries for the relatively small cohort observed in that country (Sullivan, Joshi, Ketende & Obolenskaya, 2010).

Among academic research on 'home international comparisons', and regional differences, is a study of education-related outcomes across different jurisdictions by Taylor et al., (2013). Among regions within England, London showed unexpectedly good literacy in MCS children at primary school age, adding to other evidence of a 'London effect' in educational achievement at secondary schools. Differences in policies were expected to offer examples of 'natural experiments' that the MCS might 'exploit' to detect policy impact. The legislation banning smoking in enclosed public spaces and workplaces, introduced in Scotland in 2006, ahead of England, offered one such opportunity. Hawkins, Cole and Law (2011) found little sign that it reduced the overall level of smoking among the cohort's parents in Scotland, but some evidence that it reduced the social gradient in the practice. MCS was also used to explore health inequalities across Scotland and regions of England (Cruise & O'Reilly, 2015). While there was no evidence of regional differences at birth, these authors found some evidence that geographic health inequalities may develop cumulatively during the life course.

Due to the nature of the testing system in the UK, so far it is only from state schools in England, where Standard Attainment Tests (SATs) are carried out at the ends of Key Stages 1 and 2, that such scores are available to supplement or substitute for the cognitive tests collected in the homes (e.g. Jones, Gutman & Platt, 2013). Test score data from the other countries will be linked as and when national tests are undertaken.

Objective 8 – The use of MCS in policy evaluation

National Evaluation of Sure Start

Sure Start was an area-based intervention targeted at children under four in the most deprived areas in England, and run by local cross-sector initiatives (Sure Start Local Partnerships, SSLPS). Its roll out began with 60 SSLPs in 1999, initially planned to reach 250[xiii]. MCS was used in the impact evaluation. To evaluate effectiveness of SSLPs, a comparison (control) group was required, in other words a group of children not exposed to the programme, but living in other deprived areas. It was decided to draw controls from the MCS. This was a novel application for a general purpose cohort study, and for the evaluation, the use of such a resource was also novel. Further details are given in the Appendix.

Its use in the evaluation did not affect the MCS adversely, and indeed brought with it some benefits. One was the boost of the initial sample in disadvantaged areas in England, increasing the sample size by 35 disadvantaged wards, expected to include 2,600 families. Certain aspects of the early surveys were also strengthened through collaboration with the Sure Start evaluation team, such as measures of cognitive, behavioral and emotional adjustment, and the home learning environment. On the other hand, the joint efforts at designing the MCS2 childcare questions were not, as noted above, a triumph.

However, from the viewpoint of the impact evaluation, using a cohort study as a control group is challenging – for reasons expanded in the Appendix. We maintain that the evaluation of Sure Start was adversely affected by this design. There are many practical considerations in synchronizing two different surveys. A key tension is that the measurements embedded in a longitudinal study are guided by constructs that are important to measure through life, rather than short-term interests. Central to well-designed evaluations of early childhood development (ECD) interventions, on the other hand, is the robust estimation, in the short and long run, of causal impacts of the programme on specific outcomes it targeted, and understanding the behavioural changes contributing to these impacts. This made it challenging to align the two surveys.

Cohort studies can, however, be instrumental in the estimation of causal impacts – when combined with serendipitous events/policies affecting some cohort members and not others (for instance Fitzsimons & Vera-Hernandez, 2014; Kelly, 2011). This 'natural experiment' approach overcomes many of the issues discussed in the Appendix.

National Evaluation of the Children's Fund

Another national evaluation exercise in which the MCS played a part was the impact evaluation of the National Evaluation of the Children's Fund (NECF), (Edwards, Barnes, Plewis, Morris, et al., 2006). The programmes offered by the trans-

departmental Children's Fund, announced in 2000, for children aged five-13, were also localised and just as diffuse as those offered to children aged zero-four by Sure Start. Their emphasis was on 'prevention' – of poor behavioural and academic outcomes – and provisions typically included breakfast clubs and other out-of-school activities. The evaluation had built-in fluidity, as it was intended to help improve the services as it rolled out rather than wait until there was a clear outcome to be evaluated.

The role of MCS in the evaluation was, in the first instance, to identify take-up of Children's Fund (CF) services in selected areas, rather than to provide a set of controls. The older siblings of MCS children in England, at sweeps 2 and 3, formed a 'convenience sample' of the children within the client age group, who came ready provided with a mass of data on family background. In due course it was expected to estimate the impact of service use by comparing older sibs in CF and non-CF areas as well as those who did and did not report using CF services, on the questionnaire augmented in CF areas. But this never happened. The evaluation was terminated in 2006, with two years still to run, when the organisation of the programme and services was restructured. As often happens, the aim to evaluate an intervention was overtaken by events in the policy arena. We shall never know what the data collection might have contributed to evaluation had the programme continued.

The dowry from NECF became legacy for MCS in a considerable enhancement of the data resource: the extra 692 'new families' in England recruited at MCS2, interviewer assessments of neighbourhoods, extra data on older siblings, including, across the UK, the parent-rated SDQ for up to two older siblings at MCS2 and MCS3, and consent to link older siblings to education records in England. Without the ongoing resource from the evaluation budget to check and develop these resources, relatively little use has been made of them, but their potential remains.

Conclusion

Following the evidence already built up from preceding cohorts, it was anticipated in the CLS proposal to ESRC in 2000 that

> 'The children born at the start of the new century will be setting out with different degrees of handicap and advantage in the race ahead. They will be drawn from across the social and geographical spectrum, from diverse ethnic backgrounds. Some will be well placed to build upon advantages; others will have the cards stacked against them at birth. The follow-up sweeps will reveal who overcomes inauspicious beginnings, and who fails to capitalize on inherited advantage'.

Sixteen years on, we can claim, of course not totally impartially, that the MCS is answering these and many other questions. It has delivered a rich multi-purpose resource for cross-disciplinary research of the sort that was hoped for. On the basis of impressive teamwork, MCS is now a major data resource.

Initial recruitment through a universal register on an opt-out basis was key to getting the study off the ground. Regular contact and engagement with cohort members outside of the study sweeps has been critical. It has helped maintain high response rates and remains a key priority of the study. Innovation in modes of data collection, with potential to enhance the collection and quality of data, will become a major consideration in an increasingly digital age, along with linkage to administrative records.

Research outputs took some time to gain momentum, but a complex set of findings has been built up, and continues to grow. The prospects for the asset to accumulate in scientific and policy value are excellent. Users are exploiting the rich data from many perspectives using a wide range of methods. The scope for further analysis, including international comparisons, increases enormously as the evidence accumulates.

We note some disappointments. The perhaps underestimated dispersal of the sample from selected wards made it less useful than hoped to study effects within communities, schools or classrooms. The near universal roll-out of early education, also unanticipated, made it difficult to detect the impact of this policy on child outcomes, a problem which also faced the (near) parallel cohort constructed for the National Evaluation of Sure Start. The uses of the study as a resource for evaluating policy impacts did not meet expectations, but do provide lessons for the future. Any attempt to combine evaluation with a general observational study should beware of the pitfalls that beset the MCS links with NESS and NECF. The evaluation of Sure Start got going too late to make

the best use of MCS. Designing an impact evaluation around the use of a cohort study as a control group brought with it many challenges and may have missed the opportunity for a clean and robust impact evaluation. On the plus side, the additional data collection due to the national evaluations still enhances the long-term value of the research resource, as care was taken that they added to the scientific content rather than distorting it.

As we write, the cohort is moving through adolescence, and age 14 data will be made available at the end of 2016. The next sweep is scheduled for 2018, when the cohort members will be 17 years of age, entering an important new phase, on the threshold of adulthood, taking divergent paths that will greatly influence their future wellbeing. The seventh sweep will be the first opportunity to observe such diverging decisions and the factors influencing these. On the one hand there will be much interest in these, and other, outcomes as a function of earlier life experiences; on the other hand age 17 will provide an important baseline to collect constructs to be followed through adulthood. The contribution of the study to understanding the evolution of the life course, and life, in 21st Century Britain, will only grow as the cohort marches on.

Acknowledgements

We would like to thank our colleagues Morag Henderson and Dylan Kneale for comments. We are also grateful for comments received in the reviewing process. We thank the users of the MCS data for their work in analysing the resource and suggesting material for inclusion in the surveys, especially those on scientific advisory groups. We apologise that it has not been possible to refer to all the research or include all their suggestions. The ESRC has provided the main funding for the study throughout, and is gratefully acknowledged for this. The co-funding from government departments has also been very important. The original consortium supporting the study since MCS1, was led by the Office of National Statistics, with the Departments of Education and Skills, of Health, and of Work and Pensions, the Northern Ireland Executive, the Scottish Government and the Welsh Assembly Government (and their departmental successors). The Ministry of Transport, the Home Office, and Ministry of Justice joined the government funders at later stages up to and including MCS6. Other funding has been contributed by the Wellcome Trust, the Sir Siegmund Warburg Voluntary Settlement, the United States Department of Health and Human Services, National Institutes of Health, National Institute on Alcohol Abuse and Alcoholism and the International Centre for Child Studies. We are grateful to the cohort maintenance, communications, data management and survey management teams at CLS, and the staff of three fieldwork agencies. Most of all we appreciate the participation in the study by the cohort families.

References

Belsky, J., Melhuish, E., Barnes, J., Leyland, A., Romaniuk, H. & the NESS Research Team. (2006). Effects of Surestart Local Programmes on Children and Families: Early Findings. *British Medical Journal*, *332*, 1476-1478. http://dx.doi.org/10.1136/bmj.38853.451748.2F

Belsky, J., Leyland, A.H., Barnes, J., & Melhuish, E. (2009). Sure Start in England. *Lancet, 373*- 381. http://dx.doi.org/10.1016/s0140-6736(09)60132-x

Bécares, L., Nazroo, J. & Kelly, Y. (2015). A longitudinal examination of maternal, family, and area-level experiences of racism on children's socioemotional development: Patterns and possible explanations. *Social Science & Medicine*, *14,* 128–135. http://dx.doi.org/10.1016/j.socscimed.2015.08.025

Blanden, J. & Machin, S. (2008). Up and down the generational income ladder in Britain: past changes and future prospects. *National Institute Economic Review*, *205(1*), 101-116. http://dx.doi.org/10.1177/0027950108096594

Bradbury, B., Corak, M., Waldfogel, J. & Washbrook, E. (2015) *Too Many Children Left Behind. The U.S. Achievement Gap in Comparative Perspective.* New York: Russell Sage Foundation.

Brassett-Grundy, A., Joshi, H., Butler, N. & Bryar, R. (2006). Health Visitors' Awareness of Local Service Accessibility. *Community Practitioner*, 79(3), 84-88.

Bronfenbrenner, U. (1979). *The Ecology of Human Development: Experiments by Nature and Design*. Cambridge, MA: Harvard University Press.

Brown, M. & Sullivan, A. (2014). Cognitive Development. In Platt, L .(ed), *Millennium Cohort Study: Initial findings from the Age 11 survey.* London: Centre for Longitudinal Studies.

Calderwood, L. (2013). The role of respondent characteristics in tracking on longitudinal surveys: evidence from the UK Millennium Cohort Study. *Longitudinal and Life Course Studies, 4(2*), 105-118.

Carson, C., Redshaw, M., Gray, R. & Quigley, M.A. (2014). Psychological distress in the first year of fatherhood: the influence of preterm birth on father's later wellbeing. *Journal of Epidemiology and Community Health*, 68(Suppl.1), A2-A22. http://dx.doi.org/10.1136/jech-2014-204726.42

Collingwood A. & Simmonds N. (2010). Child Self-Report. In: Hansen K., Jones E., Joshi H., Budge D., (Eds.) *Millennium Cohort Study Fourth Survey: A User's Guide to Initial Findings.* Centre for Longitudinal Studies**,** Institute of Education.

Connelly, R. (2011). *Drivers of Unhealthy Weight in Childhood: Analysis of the Millennium Cohort Study.* Scottish Government Social Research Report. Edinburgh: Scottish Government.

Connelly, R. & Platt, L. (2014). Cohort Profile: UK Millennium Cohort Study (MCS). *International Journal of Epidemiology*, *43(6),* 1719-1725. http://dx.doi.org/10.1093/ije/dyu001

Coté, S.M., Doyle, O., Petitclerc, A. & Timmins, L. (2013). Child Care in Infancy and Cognitive Performance until Middle Childhood in the Millennium Cohort Study. *Child Development*, *84(4)*, 1191–1208. http://dx.doi.org/10.1111/cdev.12049

Crawford, C., Goodman, A., Greaves, E. & Joyce, R. (2012). Cohabitation, marriage and child outcomes: an empirical analysis of the relationship between marital status and child outcomes in the UK using the Millennium Cohort Study. *Child and Family Law Quarterly*, *24(2*), 176-198.

Crawford, C., L. Dearden & Greaves E. (2014). The drivers of month-of-birth differences in children's cognitive and non-cognitive skills, *Journal of the Royal Statistical Society: Series A*, *177*, 829-860. http://dx.doi.org/10.1111/rssa.12071

Crosby, D.A. & Hawkes, D.D. (2007). Cross-National Research Using Contemporary Birth Cohort Studies: A Look at Early Maternal Employment in the United Kingdom and United States. *International Journal of Social Research Methodology, Special Issue - Child Cohort Studies*, *10(5),* 379-404. http://dx.doi.org/10.1080/13645570701677151

Cruise, S.M. & O'Reilly, D. (2015). Are the differences in adulthood ill-health across the north-south divide and between Scotland and England also evident in early childhood health indicators? *Social Science & Medicine*, *130(April*), 277–283. http://dx.doi.org/10.1016/j.socscimed.2015.02.021

De La Rochebrochard, E. & Joshi, H. (2013). Children Born After Unplanned Pregnancies and Cognitive Development at 3 Years: Social Differentials in the United Kingdom Millennium Cohort. *American Journal of Epidemiology*, *178(6*), 910-920. http://dx.doi.org/10.1093/aje/kwt063

Dearden, L., Sibieta, L. & Sylva, K. (2011). The socio-economic gradient in early child outcomes: evidence from the Millennium Cohort Study. *Longitudinal and Life Course Studies*, 2(1), 19-40.

Dex, S. & Joshi, H. (2004). *Millennium Cohort Study First Survey: A Users' Guide to Initial Findings.* London: Centre for Longitudinal Studies.

Dex, S. & Joshi. H. (Eds), (2005). *Children of the 21st century: From birth to nine months.* Bristol: Policy Press

Dickerson, A. & Popli, G. (2016). Persistent poverty and children's cognitive development: Evidence from the UK Millennium Cohort Study. *Journal of the Royal Statistical Society: Series A*, 179(2), 535-558. http://dx.doi.org/10.1111/rssa.12128

Edwards, A., Barnes, M., Plewis, I., Morris, K. et al. (2006). *Working to prevent the social exclusion of children and young people: Final lessons from the National Evaluation of the Children's Fund.* Birmingham: University of Birmingham.

Eisenstadt, N. (2012). *Providing a Sure Start: How government discovered early childhood*. Bristol: Policy Press

Elder, G.H. (Ed.), (1985) *Life Course dynamics: Trajectories and transitions, 1968-1980*. Ithaca. N.Y.: Cornell University Press

Ermisch, J. (2008). Origins of social immobility and inequality: parenting and early childhood development. *National Institute Economic Review 205(1*), 62-71. http://dx.doi.org/10.1177/0027950108096589

Feinstein, L., Lupton, R., Hammond, C., Mujtaba, T., Salter, E., Sorhaindo, A., Tunstall, R., Richards, M., Kuh, D. & Johnson, J. (2008). *The Public Value of Social Housing: A longitudinal analysis of the relationship between housing and life chances.* Smith Institute Report. London: The Smith Institute.

Fitzsimons, E. & Vera-Hernandez, M. (2014). *Food for Thought? Breastfeeding and Child Development.* DoQSS Working Paper 14-04. London: Institute of Education.

Flouri, E., Midouhas, E. & Joshi, H. (2014). The Role of Urban Neighbourhood Green Space in Children's Emotional and Behavioural Resilience. *Journal of Environmental Psychology*, 40, 179–186. http://dx.doi.org/10.1016/j.jenvp.2014.06.007

Gambaro, L. & Joshi, H. (2016). Moving home in the early years: What happens to children in the UK? *Longitudinal and Life Course Studies,*7(3*)* 265-287. http://dx.doi.org/10.14301/llcs.v7i3.375

George, A., Stokes, L. & Wilkinson, D. (2012). Does Early Education Influence Key Stage 1 Attainment? Evidence for England from the Millennium Cohort Study. *National Institute Economic Review*, 222(1), R67-R80. http://dx.doi.org/10.1177/002795011222200106

Goodman, A., Gregg, P. & Washbrook, E. (2011). Children's educational attainment and the aspirations, attitudes and behaviours of parents and children through childhood in the UK. *Longitudinal and Life Course Studies*, *2(1),* 1-18.

Gutman, L.M, Joshi, H., Parsonage, M. & Schoon, I. (2015). *Children of the New Century: Mental health findings from the Millennium Cohort Study.* Centre for Mental Health Report. London: Centre for Mental Health.

Hansen, K. (ed) (2014). *MCS Guide to the datasets for the first, second third, fourth and fifth surveys.* 8th edition. London: Centre for Longitudinal Studies.

Hansen, K., Joshi, H. & Dex, S. (Eds), (2010). *Children of the 21st century: the first 5 years.* Bristol: Policy Press.

Hansen, K. & Hawkes, D. (2009) Early childcare and child development. *Journal of Social Policy*. *38*, 211-239. http://dx.doi.org/10.1017/S004727940800281X

Hansen, K., Hawkes, D. & Joshi, H. (2009). The Timing of Motherhood, Mothers' Employment and Child Outcomes. In Stillwell, J., Coast, E. & Kneale, D. (Eds), *Fertility, Living Arrangements, Care and Mobility.* Dordrecht: Springer. http://dx.doi.org/10.1007/978-1-4020-9682-2_4

Hartas, D. (2011). Families' social backgrounds matter: socio-economic factors, home learning and young children's language, literacy and social outcomes. *British Educational Research Journal*, *37(6)*, 893-914. http://dx.doi.org/10.1080/01411926.2010.506945

Haux, T. & Platt, L. (2015). *Parenting and contact before and after separation.* Nuffield Foundation Report, June 2015. London: Nuffield Foundation

Hawkes, D. (2006). *Women's Experiences of Successful Infertility Treatment.* MCS Technical Report on Response. London: Centre for Longitudinal Studies

Hawkes, D. & Joshi, H. (2012). Age at motherhood and child development: evidence from the Millennium Cohort Study. *National Institute Economic Review*, October, R52-R63. http://dx.doi.org/10.1177/002795011222200105

Heilmann, A., Kelly, Y., Stafford, M. & Watt, R. (2013). The Contribution of Neighbourhoods and Schools to Cognitive Test Performance at Age Seven - Findings from the UK Millennium Cohort Study. *Journal of Epidemiology and Community Health*, *67(1),* A28-A28. http://dx.doi.org/10.1136/jech-2013-203126.57

Hawkins, S.S, Cole, T.J & Law, C. (2011). Examining smoking behaviours among parents from the UK Millennium Cohort Study after the smoke-free legislation in Scotland. *Tobacco Control*, *20(2*), 112-118. http://dx.doi.org/10.1136/tc.2010.037028

Hockley, C., Quigley, M. A., Hughes, G., Calderwood, L., Joshi, H. & Davidson, L.L. (2008). Linking Millennium Cohort data to birth registration and hospital episode records. *Paediatric and Perinatal Epidemiology*, 22(1), 99-109.

Hope, S., Pearce, A., Whitehead, M. & Law, C. (2014). Family employment and child socioemotional behaviour: longitudinal findings from the UK Millennium Cohort Study. *Journal of Epidemiology & Community Health*, *68(10),* 950-957. http://dx.doi.org/10.1136/jech-2013-203673

Huerta, M., Adema, W., Baxter, J., Corak, M., Deding, M., Gray, M.C., Han, W-J. & Waldfogel, J. (2011). *Early Maternal Employment and Child Development in Five OECD Countries.* OECD Social, Employment and Migration Working Papers, No. 118. Paris: OECD. http://dx.doi.org/10.1787/5kg5dlmtxhvh-en

Jackson, M.I., Kiernan, K. & McLanahan, S. (2012). Immigrant–Native Differences in Child Health: Does Maternal Education Narrow or Widen the Gap? *Child Development*, *83(5*), 1501-1509. http://dx.doi.org/10.1111/j.1467-8624.2012.01811.x

Johnson, W., Li, L., Kuh, D. & Hardy, R. (2015). How Has the Age-Related Process of Overweight or Obesity Development Changed over Time? Co-ordinated Analyses of Individual Participant Data from Five United Kingdom Birth Cohorts. *PLoS Medicine*, *12(5):* e1001828. http://dx.doi.org/10.1371/journal.pmed.1001828

Jones, E., Gutman, L. Platt, L. (2013). *Family stressors and children's outcomes*. Department for Education Research Report DFE RR254, London.

Kanji, S. (2013). Do fathers work fewer paid hours when their female partner is the main or an equal earner? *Work, Employment & Society*, 27(2), 326-342. http://dx.doi.org/10.1177/0950017012460321

Kelly, E. (2011). The Scourge of Asian Flu: In utero Exposure to Pandemic Influenza and the Development of a Cohort of British Children. *Journal of Human Resources*, *46(4),* 669-694. http://dx.doi.org/10.3368/jhr.46.4.669

Kiernan, K. (2006). Non-residential fatherhood and child involvement: evidence from the Millennium Cohort Study. *Journal of Social Policy*, *35(4)*, 651-669. http://dx.doi.org/10.1017/S0047279406000122

Kiernan, K.E. & Mensah, F.K. (2011). Poverty, Family Resources and Children's Early Educational Attainment: The Mediating Role of Parenting. *British Educational Research Journal*, 37(2), 317-336. http://dx.doi.org/10.1080/01411921003596911

Kneale, D., Patalay, P., Khatwa, M., Stansfield, C., Fitzsimmons, E., Thomas, J. (2016). Piloting and producing a map of Millennium Cohort Study Data usage: Where are data underutilised and where is granularity lost? EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London. http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3502

Lakshman, R., Zhang, J., Zhang, J., Koch, F. S., Marcus, C. Ludvigsson J., Ong, K. K., Sobko, T. (2013). Higher maternal education is associated with favourable growth of young children in different countries. *Journal of epidemiology and community health*, 67(7), 595-602. http://dx.doi.org/10.1136/jech-2012-202021

Lennon, M.C., Clark, W.A.V. & Joshi, H. (2016). Residential mobility and well-being: Exploring children's living situations and their implications for housing policy. Symposium in *Longitudinal and Life Course Studies.* 7(3) 197-200. http://dx.doi.org/10.14301/llcs.v7i3.393

Mathers, S., Sylva, K. & Joshi, H. (2007) *Quality of childcare settings in the Millennium Cohort Study.* Report for Surestart. Nottingham: HMSO.

McMunn, A., Martin, P., Kelly, Y. and Sacker, A. (2015). Father's involvement: Correlates and Consequences for Child Socioemotional Behaviour in the United Kingdom. *Journal of Family Issues.* http://dx.doi.org/10.1177/0192513X15622415

Melhuish, E., Belsky, J., Leyland, A., Barnes, J. & the Ness Research Team. (2008). Effects of fully-established Sure Start Local Programmes on 3-year-old children and their families living in England: a quasi-experimental observational study. *Lancet*, *372(9650*), 1641-1647. http://dx.doi.org/10.1016/S0140-6736(08)61687-6

Molitor, N.T., Best, N., Jackson, C. and S. Richardson. (2009). Using Bayesian graphical models to model biases in observational studies and to combine multiple sources of data: application to low birth

weight and water disinfection by-products. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(3), 615-637. http://dx.doi.org/10.1111/j.1467-985X.2008.00582.x

Mostafa, T. (2013). *Technical Report on Response in the Millennium Cohort Study Wave 5.* London: Centre for Longitudinal Studies.

Moulton, V., Flouri, E., Joshi, H. & Sullivan, A. (2015). The influence of grandparents' social class on children's aspirations. *British Journal of Sociology of Education*, advance online access.

NESS (The National Evaluation of Sure Start Team). (2012). *The impact of Sure Start Local Programmes on seven year olds and their families.* Research Report DFE-RR220. London: Department for Education. http://dx.doi.org/10.1080/01425692.2015.1093407

Parsons, S., Schoon, I. & Vignoles, A. (2014). Parental worklessness and children's early school achievement and progress. *Longitudinal and Life Course Studies*, *5(1*), 19-41. http://dx.doi.org/10.14301/llcs.v5i1.230

Pearce, A., Li, L., Abbas, J., Ferguson, B., Graham, H., Law, C. & the Millennium Cohort Study Child Health Group. (2010). Is childcare associated with the risk of overweight and obesity in the early years? Findings from the UK Millennium Cohort Study. *International Journal of Obesity,* 34, 1160-1168. http://dx.doi.org/10.1038/ijo.2010.15

Pearson, H. (2016). *The Life Project*. London: Penguin.

Pirus, C. & Leridon, H. (2010). Large Child Cohort Studies across the World. *Population*, 65(4), 671-730. http://dx.doi.org/10.3917/popu.1004.0671

Plewis, I. (Ed). (2007a). *Millennium Cohort Study First Survey: Technical Report on Sampling (4th Edition).* London: Centre for Longitudinal Studies, Institute of Education.

Plewis, I. (2007b). Non-Response in a Birth Cohort Study: The Case of the Millennium Cohort Study. *International Journal of Social Research Methodology,* 10(5), 325-334. http://dx.doi.org/10.1080/13645570701676955

Redshaw, M., Hockley, C. & Davidson, L.L. (2007). A qualitative survey of the experience of treatment for infertility among women who successfully became pregnant. *Human Reproduction*, 22(1), 295-304. http://dx.doi.org/10.1093/humrep/del344

Sellers, R., Maughan, B., Pickles, A., Thapar, A. & Collishaw, S. (2015). Trends in parent- and teacher-rated emotional, conduct and ADHD problems and their impact in prepubertal children in Great Britain: 1999-2008. *Journal of Child Psychology and Psychiatry*, 56(1), 49-57. http://dx.doi.org/10.1111/jcpp.12273

Schoon, I., Hope, S., Ross, A. & Duckworth, K. (2010). Family hardship and children's development: the early years. *Longitudinal and Life Course Studies*, *1(3*), 209-222. http://dx.doi.org/10.14301/llcs.v1i3.109

Sullivan, A., Joshi, H., Ketende, S. & Obolenskaya, P. (2010). *The consequences at age 7 of early childhood disadvantage in Northern Ireland and Great Britain.* A report to the Northern Ireland Office of the First Minister and Deputy First Minister, December 2010. London: Institute of Education.

Sullivan, A., Ketende, S. & Joshi, H. (2013). Social Class and Inequalities in Early Cognitive Scores. *Sociology*, 47(6), 1187-1206. http://dx.doi.org/10.1177/0038038512461861

Tanskanen, A.O. (2013). The Association between Grandmaternal Investment and Early Years Overweight in the UK. *Evolutionary Psychology*, *11(2*), 417-425. http://dx.doi.org/10.1177/147470491301100212

Taylor, C., Rees, G. & Davies, R. (2013). Devolution and geographies of education: the use of the Millennium Cohort Study for 'home international' comparisons across the UK. *Comparative Education*, *49(3*), 290-316. http://dx.doi.org/10.1080/03050068.2013.802927

Tzavidis, N, Salvati, N, Schmidt, T, Flouri, E and Midouhas, E. (2016). Longitudinal analysis of the Strengths and Difficulties Questionnaire scores of the Millennium Cohort Study children in England using Quantile and M-quantile multilevel models *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(2), 427–452. http://dx.doi.org/10.1111/rssa.12126

Waldfogel, J. & Washbrook, E. (2010). *Low Income and Early Cognitive Development in the UK.* London: Sutton Trust.

Welsh Assembly Government (2011). *2011 Children and Young People's Wellbeing Monitor for Wales*. Report from the Welsh Assembly, March 2011. Cardiff: Welsh Assembly Government.

Appendix

The role of MCS in the national evaluation of Sure Start

Sure Start was a set of area-based programmes of services for children under four in the most deprived local areas in England, rolled out between 1999 and 2003. It started in 60 Sure Start localities in 1999, initially intended to rise to 250 in 2003 (Eisenstadt, 2012). Areas serving about 600 children, were conceptualised in terms of 'pram-pushing' distance from their centre. They were chosen for intervention on the basis of proposals from local cross-sector partnerships, (SSLPs). The impact evaluation, National Evaluation of Sure Start (NESS), was led by Edward Melhuish (see Belsky et al. 2006; Melhuish, Belsky, Leyland, Barnes & The NESS Research Team, 2008; Belsky, Leyland, Barnes & Melhuish, 2009; and the reports contained in http://www.ness.bbk.ac.uk/). The MCS provided a control group for the impact evaluation.

The evaluation finished in 2012, with a report on the impact on seven year olds. This concluded that the SSLPs had made some beneficial impact on four out of fifteen target outcomes, mostly on mothers' parenting and well being (NESS, 2012). However they also report considerable methodological difficulties, on which we comment below.

MCS was incorporated in the evaluation strategy because a randomised controlled trial (RCT) was ruled out, mainly for political reasons. The timing of the MCS, in following a cohort of children born in 2000/01, as Sure Start was getting under way, appeared to match well with the requirements of the evaluation sample. Furthermore, the Department for Education and Skills was funding the evaluation of Sure Start, alongside contributing funds to MCS. It was a condition on their contract that the evaluators work together with the MCS team. It was also a condition for MCS to balance the needs of NESS against those of other stakeholders.

In relying on a longitudinal study rather than a RCT to provide a control group (to measure what might have happened in the absence of the intervention), a number of methodological complexities arose. Despite the best efforts on the part of the evaluation team to overcome them, these challenges were often insurmountable and ultimately detrimental to the impact evaluation. As a result, the estimates of the programmes' impacts came with a series of caveats, undermining their interpretation as causal effects and, thereby, their utility for policymaking.

Choice of control groups

The critical bedrock of an impact evaluation is the choice of a suitable control group. The rapid expansion of the SSLPs limited the availability of possible control areas. In 2002, before any results were available, the target number of SSLPs for 2004 was doubled. One challenge was that Sure Start areas were more disadvantaged than potential control areas. The NESS team selected areas in England in which children from the MCS lived but which did not have Sure Start, using well-executed propensity score matching. However, matching did not eliminate demographic differences between the NESS sample and those selected from MCS. This exacerbates concerns about unmeasured confounding factors, and means that it cannot be ruled out that the 'impact' of the programme reflects the influence of other differences between the areas (Melhuish et al., 2008). Showing comparability across treatment and control areas in pre-programme trends in outcomes would have been reassuring, though has not, to our knowledge, been done. Furthermore, it was not possible to find suitable control areas for the 57 most deprived of the Sure Start areas (of a total of 150 available for evaluation from rounds one to four of Sure Start), so over one third of them had to be excluded from the main evaluation. The final samples consisted of 5,883 three-year-old children in 93 SSLP areas (out of the 150) and 1,879 children in 72 non-SSLP areas.

Data collection

For two separate studies with distinct objectives to track each other is extremely challenging, even in the short-term. Coordination is required on the timing and content of surveys, consistent measurement of constructs across them, preservation of confidentiality, tracing of respondents and attrition. Moreover, coordination is difficult to sustain and likely to fade out in medium to long term, rendering the estimation of long run effects impossible. The NESS and MCS research teams coordinated closely from the outset, though some problems, some unforeseen, could not be overcome.

In both studies, families were visited when the child was nine months old, three, five and seven years of

age. However, given the challenges of implementing the innovation on an unprecedented scale, it was a necessary for a SSLP to have 'bedded down' for at least three years, before surveying its areas. This meant that the NESS sample was born two years after the MCS, and the timing of data collections was not aligned. NESS surveys started in 2003. This two-year gap had major repercussions. It can never be ruled out that any NESS-MCS differences are due to time effects (e.g. a national policy may apply to just one of them), making it difficult to defend the estimated impacts as *causal* - the central aim of the evaluation. Furthermore, NESS was constrained to follow the timing of MCS rounds in terms of the children's ages whereas other ages might have been more suited to its own purposes. The surveys had common elements, but they were not the same. For example MCS interviewed partners while NESS did not, and NESS asked more questions about social support in the neighbourhood. The fact that interviewers conducting the two surveys came from different agencies and had different training also raised concerns about the comparability of data they collected, for example on cognitive assessments. Using linked administrative data helped to some extent, but this raises another possible source of bias in differential consent to data linkage across two studies.

The long-term impacts of an early years programme are of central policy importance. However, coordination on data collection across two study teams becomes increasingly difficult in the medium and long term. Whilst administrative education records were used up to age seven in NESS, we know of no plans to follow these into the future.

Other considerations

A well-designed impact evaluation should not only measure the impact of the programme on targeted outcomes, but also understand behavioural changes that contribute to measured impacts. For instance, in the case of Sure Start, how did the intervention change mothers' behaviour with children in their homes? Why were the poorest families initially least likely to benefit? Data collection and measurement must be designed accordingly, which can be challenging on two different surveys. A forced marriage has its drawbacks.

Sure Start in England took on a new form in 2004 when it was substantially increased in scale, extended to the under five age group and changed the emphasis of content. Its organisation transferred to Children's Centres run by Local Authorities, more integrated with existing services. It was a popular initiative that Labour, and subsequently the Coalition Government, sought to bring to as many places as possible. The target of 3,500 Children's Centres was achieved in 2010. Their survival as the legacy of the Sure Start continues, subject to local decisions, and less fiscal protection.

Endnotes

[i] The Welsh boost was drawn from disadvantaged wards.

[ii] Health visitors are community nurses specialising in support to young families. The charity founded by Neville Butler, the International Centre for Child Studies, also contributed to the Health Visitor survey.

[iii] The indices for Scotland and Northern Ireland were not available in time to draw a line based on a UK distribution. The proportions of wards in these countries covered by this cut-off were higher, especially in Northern Ireland.

[iv] It may be worth noting that the few families not claiming Child Benefit for children born in the UK at the time could well have been classified as 'non-resident' if they were diplomats, members of foreign armed forces or newly arrived asylum seekers.

[v] There were some restrictions put on the use of the Child Benefit register, in particular a minority (under 3%) of families were excluded if deemed 'sensitive cases' by the DWP, or if they had been approached for another survey (Plewis, 2007a). In the nature of things little is known about the various circumstances of the 'sensitive cases'. Their exclusion means that children who had already been taken into social care for example would not have been recruited, but they were not numerous. As far as can be told the social bias towards more advantaged families opting out offset the less advantaged cases being withheld as 'sensitive cases' (Plewis, 2007b).

[vi] Lucinda Platt took over from Heather Joshi as PI in 2011, during the early stages of preparations for MCS5, Emla Fitzsimons succeeded her in 2013 as MCS6 was being developed to go into the field in 2015.
[vii] This total differs from the 18,827 quoted by Connelly and Platt (2014) that appears to include twins and triplets in the 'new families' not recruited till sweep 2. A total of 18,819 children is also sometimes quoted. This reflects the N that was thought to have been collected at MCS1 before elimination of one invalid case.
[viii] Plewis (2007b) estimated non-response weights for MCS1 that made little change to the sampling weights.
[ix] The funding came from the English Department for Education and Skills.
[x] However, at age 11, there was some contact with families known to have emigrated.
[xi] 58% unweighted and including non-respondents at sweep 3.
[xii] The 1946 and 1958 cohorts had been confined to Great Britain, and the small number recruited in Northern Ireland in 1970 were excluded from follow-up because of the political 'troubles'.
[xiii] The roll out actually accelerated after the evaluation started, to reach 550 areas by 2004 when the programme's scope was further extended and transformed (see Appendix).

A dialogue between research, politics, administration and the general public: an interview with Paul Bradshaw, Rainer Bromme, John Bynner, Manfred Egner, Harvey Goldstein and Alexander Renner

Jutta von Maurice[i] Leibniz Institute for Educational Trajectories, Germany
jutta.von-maurice@lifbi.de

(Received May 2016 Revised August 2016) http://dx.doi.org/10.14301/llcs.v7i4.416

Introduction

The interrelations between research, politics, administration and the general public are highly complex in all content areas. In longitudinal research it might be especially necessary to focus on this relationship in more detail, as scientifically profound analyses regarding the development, processes and transitions in life courses often require large, representative, carefully drawn, diligently tracked and surveyed samples of participants. Therefore, longitudinal research needs persistent engagement of researchers, substantial and reliable funding, and a long-term commitment of participants.

The Policy Forum of the Society for Longitudinal and Life Course Studies (SLLS) is a highly interesting initiative in discussing and further developing such aspects. It is based on four fundamental principles, which are sketched by Bynner and Schuller:[ii]

- policy is developed within a political framework in which complex research evidence is only one element
- the timetable for action on policy rarely harmonises with that for producing findings at fixed intervals from longitudinal research
- past evidence from long term longitudinal research may be seen as of 'fading relevance' to the key policy questions of the 'here and now'
- translating findings into policy-relevant messages and conclusions for practice requires communication and brokerage skills which are often lacking in research teams

Even though aspects concerning the dialogue between researchers, politicians, administrators and the general public are generally missing at many international research conferences, much interest in this topic became clear at the 2014 SLLS conference in Lausanne, Switzerland, with six symposia related to the research-policy interplay. This interest might have been triggered by the carefully chosen conference title 'Lives in Translation – Life Course Research and Social Policies'. Fortunately, we then noticed at the 2015 SLLS conference in Dublin, Ireland, that the policy-research interface was an ongoing topic for life course researchers – again, five symposia focused on this topic – although this time around the conference title 'Life Courses in Cross-National Comparison: Similarities and Differences' no longer hinted at the policy-research-link quite so openly.

This paper aims to contribute to the process of further fostering this discussion through a compilation of different aspects of the interplay between research, politics, administration and the general public. In some arguments we took the German situation and the German National Educational Panel Study (NEPS) as a focal point for the discussion. NEPS collects data about competence development and educational processes over the life span by following the life trajectories of more than 60.000 participants in six cohorts (covering the whole life span from newborns to adults; for more details see Blossfeld, Roßbach, & von Maurice, 2011). NEPS is seen as a rather typical example of the aspects discussed in this paper because an ongoing exchange with policy makers and administrators as well as the continued support from participants and the general public is absolutely vital for keeping this huge panel study

successful. The arguments made here are not limited to Germany or NEPS, and they are not limited to educational research only. The presented arguments are especially valuable and generalisable for large-scale longitudinal research – irrespective of the country, study or research topic in question.

As points of view are not always identical and are not 'official statements', this paper uses an interview format and brings together different people – most of whom have been involved in the SLLS policy group or in some of the policy symposia at the 2014 or 2015 SLLS conferences. All of the interview partners contribute unique, equally valuable perspectives to this discussion.

Interviews

Alexander Renner, representing the Federal Ministry of Education and Research, Berlin, Germany, at the 2014 SLLS conference, you described the special German situation from your personal point of view. What is special about it? What might be different to other countries? Would you like to share some aspects?

In Germany researchers and policy makers (and the administration) cooperate closely in the field of education policy. It is not unusual that policy makers contact researchers and ask for a solution to a certain problem. But the close contact is limited to the field of empirical education research.

The close cooperation started in the late '90s. From my point of view there are three main reasons:

First, education policy in Germany was much ideologized. Politicians from a conservative or social democratic background stuck to their ideological positions and there was no movement at all. Research was used to cut the Gordian knot. New thoughts could be introduced without losing face.

Second, in the German federal system nearly all decisions concerning schools are taken by the 16 Federal States – the so-called 'Laender'. The formal influence of the federal level is very small. Therefore the federal government tries to influence the direction of education policy by fostering research in fields considered important.[iii]

Third, the federal ministry is a ministry for education and research. That makes the short cut between research and education policy easy. Educational research contributes to both tasks of the ministry. People dealing with education policy and research policy know each other. For example, I was engaged in the institutionalisation of the NEPS as part of the newly founded Leibniz Institute for Educational Trajectories. It was very helpful that the person responsible for the Leibniz Association works in the same ministry.

And Alexander Renner you mentioned some formal requirements such as timelines for politics and research. What are your experiences regarding this aspect?

Politicians often need quick responses. But I think there is a general understanding that research needs time. Sometimes it is a bit difficult to find the right level of detail. Politicians are very often interested in a broad direction. Researchers are used to caring about the second and third digit behind the decimal point.

Even if there is enough time for a research project, which could be financed by the ministry, we in the administration have the problem of finding out how big the project needs to be. We cannot always rely on the project plans of the researchers. Their ideal project is an ambitious long-lasting research project, while we often just need some information that gives us a hint in the right direction. It is important to find a compromise.

Us researchers usually think that we have results and we take a lot of effort to prepare those results 'in bite-sized' pieces for political 'consumption'. Manfred Egner from Bamberg, Germany, former school principal and working in school inspection, you support the NEPS team in communicating results politicians, administrators, and the general public. What are the measures that the team takes and are they effective?

There is an increasing demand for the scientific community to provide practical and politically relevant knowledge. Simultaneously, there are complaints about a decreasing trust in the sciences, and calls for problem- and user-oriented research. Between these conflicting demands, numerous questions emerge about the foundations and possibilities of science as well as the relevance of scientific expertise for decision-making in politics, administration, and in society in general.

Public relation activities of the Leibniz Institute for Educational Trajectories addressing the political system, the administration, educational facilities, and science ensure that communication is intensified with teacher, parent, and student

unions, and also with cooperating facilities, as well as elected officials on the federal and state level in order to maintain or even, if possible, increase participation in the NEPS.

It is the contact with associations of German teachers, principals, and parents in particular that has expanded and basic information on the NEPS is conveyed to members of the Bavarian state and federal parliaments as well as to local politicians.

The ultimate goal of the above-mentioned initiatives is to optimize response rates and maintain panel participation. This task is highly important and effective because it successfully generates a link between practical application and politics, fostering supporting relationships with various institutions such as ministries and educational facilities.

And Alexander Renner, you are one of those people for whom the NEPS team writes a special 'policy/funder newsletter'. During our discussion we recognised that we really don't know whether this is an effective tool. Out of 10, how many do you really read?

To be honest: I don't read newsletters like yours regularly. If I read all the newsletters, information brochures, etc. sent to me, I would do nothing else but read all day long. In addition there are the results and reports from the projects commissioned by us. I work in the field of large-scale assessments. The latest results on the competencies of 15 year olds in schools within the OECD Programme for International Student Assessment (PISA) are published in several volumes. It has thousands of pages. But there is also a national PISA Report and there are several other large-scale studies, a huge number of smaller studies, and last but not least the National Education Report.

I believe that inside an organisation such as the ministry, research findings are noticed best/most when they are published in a way that the publication can be used as a kind of reference book. It is important that information can be found fast.

Manfred Egner – coming from a very different field of experience – what is your view regarding the relationship between politics and administration?

A lot has certainly been done. No political speech can do without praising science and without mentioning the key significance of education.

The world of politics creates its attitudes, opinions, objectives, etc. by drawing on all kinds of sources, but only in the rarest of cases does it refer to science and its results! But especially in terms of empirical findings it is highly remarkable – almost absurd – that it is still not 'standard practice' for politicians to automatically refer back to science and research when searching for solutions to their problems.

However, the need for dialogue should also be motivated by science's own interests. It requires money and personnel, to name but two factors. But only talking to policy makers and politicians in order to receive funding is too short sighted.

I am advocating a world of science that promotes itself by allowing politics and administration to participate in it through continuous dialogue, turning politics and administration thus into co-actors, which then leads to a triangular system including the general public. Informed citizens and taxpayers calling for more resources in science and research, paired with the world of politics that increasingly sees science and research as a permanent source of its own knowledge-creating process. The world of science needs a new strategy of communication and participation – aimed not only at politicians but also at the general public. And, for starters, the willingness to do so!

This willingness comes from the insight that I would like to summarise in the following four points:

- In a democracy, citizens have a right to participate in the use of public money.
- Democracy demands transparency and participation in relation to science and society.
- In such a society, science has a commitment to transparency.
- Through the dialogue with politics and administration, science will increase the attention paid to its findings as well as the likelihood of creating impact.

We heard a lot about efforts to bring research and its results to policy. And we had to hear that this is far from perfect. Harvey Goldstein, Centre for Multilevel Modelling at the University of Bristol, United Kingdom, what do you think about the challenges of bringing research results to policy makers?

Most governments are keen to emphasise that their decisions are 'evidence informed' and based

upon sound research. Yet, by the beginning of the 21st Century, many researchers could be found complaining that, on the contrary, governments only cared about evidence if it suited their existing views and fitted in with their plans. In education, for example, this was often heard with respect to government responses to the OECD PISA league table rankings of countries; if a ranking position appeared to be too low, such as happened in countries like Wales and Germany in 2008 and 2012, this was used as a justification for implementing whatever reform was being prepared – despite the caveats about interpreting such rankings that were made by many researchers.

My own view is that such complaints against governments have considerable justification, yet the issue is actually not clear-cut. To illustrate what I mean let me refer to a debate about the importance of 'homework' in promoting student achievement in secondary (high) schools.

Research in the UK by Durham University academics (Farrow, Tymms, & Henderson, 1999) claimed that among 11 year olds there was a negative association between the amount of homework done and educational attainment. Such a conclusion appeared to be counter-intuitive and the then Secretary of State for Education, David Blunkett, was recorded as attempting to rubbish the research and those who carried it out – a clearly unethical approach. The researchers' findings were published shortly after a previous report partly authored by Prime Minister Tony Blair's then principal adviser on education, Michael Barber, which concluded that homework was associated with improved performance and lent support to the Labour Party's current policy (January 1997) in favor of mandatory periods of homework. While that research had quite serious flaws, it did resonate with received opinion and accorded with Labour Party policy.

In fact, the Farrow et al. report itself was also seriously flawed in that it did not properly adjust for prior achievement so that one could make no real judgment about the direction of causation – for example, it may have been the case that poorer performing students were given more homework.

The irony here is that, had the politicians acted responsibly, they may well have been able to substantiate their policies through critical peer review of the Farrow research, and this could have provided a more secure basis for their own homework policies – they would not have needed to rely upon the original flawed research by Barber and colleagues. As we all know, research is prone to error and researchers make mistakes and, sadly, sometimes claim far too much for their findings. The research process has built-in mechanisms that attempt to cope with this through peer review and replication. While researchers may need to show somewhat more humility in public, what policy makers need to realise is that they too need to adopt a responsible and transparent approach to the evaluation of evidence. If they can succeed in doing this it would open up many opportunities for a creative collaboration with the research community and help to avoid the cynicism that is expressed all too often about policy makers' motives.[iv]

Paul Bradshaw, group head at NatCen Social Research, is active in the management and development of the Growing Up in Scotland study. At the 2015 SLLS conference you talked about a 'two-way conversation' between policy and research. Can you describe the way from policy to research?

In my experience, which is predominantly in undertaking research commissioned by government departments, the dialogue from policy to research usually works in one of two ways: representatives of policy teams engage directly with researchers, or they engage with an internal intermediary who then liaises with researchers. There are positives and negatives to both approaches.

When policy teams engage directly with researchers, those researchers benefit from the opportunity to explore the policy question in great detail. This can provide researchers with a deeper understanding of the evidence need which allows them to provide a research design which best fits the requirement. However, it may not be straightforward to set up this direct dialogue unless there is a forum or context in which both parties can meet and engage or some reliable lines of communication are in place. There may also be an issue with having to 'translate' the discussion for both parties; policy will not necessarily be familiar with the technical language of research (or the limitations of research evidence) and research may not appreciate how the evidence is to be used for policy making. It is therefore necessary to use a common language and not all researchers or policy

makers will be capable of doing this (nor will they necessarily want to either). Furthermore, direct engagement can be demanding for research teams on studies that have wide-ranging substantive, and thus policy, remits – such as many birth cohort studies. On such studies, direct engagement can mean researchers have to work with – and balance the needs of – a large number of policy teams. For a typical birth cohort study this may include liaising with health (which could have separate teams for children and parents), education, communities, justice, environment, and employment policy teams, amongst others.

In contrast, where there is an internal intermediary engaging with policy teams, researchers engage with only that intermediary. This is a more straightforward approach for researchers but introduces a stage between policy and research that can influence what researchers are ultimately asked to do. The intermediary takes different forms in different places. Within the Scottish Government they are called 'analysis' teams and consist of social researchers – (or 'analysts') who collate, commission and manage external research – and statisticians – who prepare and analyse internal data sets. Apart from removing the need for researchers to engage with multiple policy divisions on certain research projects, these intermediaries can have other positive influences on the dialogue from policy to research. For example, because the intermediary divisions are embedded within government and have legitimate and direct links to policy, it is easier for them to engage with policy than it would be for researchers outside of government. Furthermore, having a research background within a policy environment, they can translate policy needs into research terms avoiding some of the 'language' issues discussed above. However, these benefits are only realised if the intermediary is effective at liaising with internal policy teams and external researchers. In addition, they must have a good understanding of research and be able to clearly specify what is required.

Paul Bradshaw, you sketch positive and negative aspects of both approaches. What are the critical factors of success in the dialogue between policy and research in your point of view?

Irrespective of the approach, the best discussions between policy and research are those that are conducted as soon as possible after the policy question arises and involve researchers at the earliest possible stage. All too often research to inform policy is commissioned reactively, at short notice with a flawed specification, an unrealistic timescale and an inappropriate budget. The research produced as a result will often be flawed having been conducted by those who have the time and using methods which meet the specification rather than those who fully understand the policy question and using the methods most likely to answer it robustly. Researchers are happy to engage in early general dialogue around the best approaches to meet the evidence needs, timescale and budget. If policy allows research such opportunities, this ensures that both parties will have a more positive engagement process and be more content with the end products – better research and better informed policy.

A two-way conversation sounds meaningful. What does this look like in daily collaboration? Alexander Renner, as a person working in the ministerial setting, you mentioned that you once tried to understand DIF analyses (differential item functioning) – admittedly, a rather complex statistical characteristic in test development, which indicates whether a test item works differently in different groups. What was the reaction of the researcher? Did he praise this initiative and did he provide support?

I don't think he would be happy if I knew more about DIF analyses. The researcher used DIF in a discussion as hegemonic knowledge, trying to muzzle me. This is a quite common reaction when we question projects and research plans or when we ask why projects were not as successful as promised. Researchers hide behind scientific terminology hoping that it will intimidate us and that we will leave them alone. I hope that we who are working in the administration are experienced enough to identify this.

Most researchers have an agenda of their own. In some cases it is, as mentioned, a research agenda, in other cases it is a policy agenda. Sometimes it is both. This is no secret. There are enough articles about this published even in journals for educational science (most recently de Moll, Riefling, & Zenkel, 2014; Tenorth, 2015). The difficulty is not that there are different agendas and interests, but how to deal with this. I think during

the last years we were well on track. But yes, there were also disappointments for both sides.

Rainer Bromme, University of Muenster, Germany, you recently published a book together with Manfred Prenzel talking about the way from research to evidence-based decision-making in the area of educational research (Bromme & Prenzel, 2014). What opportunities and constraints of evidence-based policy do you see?

In this book (unfortunately it is available in German only) we argue that evidence-based policy first of all requires that there can be evidence at all (Bromme, Prenzel, & Jäger, 2014). This is a truism, but it has also a practical implication. In order to base policy on evidence it is necessary to establish research activities (research agendas, institutions, funding lines) focused on the critical synthesis of the research available. This includes the need for meta-analytical techniques of all kinds and it requires what we called *evaluative research syntheses*. Just because it is not possible to run experimental designs (for example, double-blind randomised trials) in many educational fields and with regard to many educational questions, evidence-based policy often must be based on further, less controlled evidence. This evidence is often heterogeneous and sometimes even contradictory, and the very idea of 'evidence-based policy' does not mean that research could be unequivocally 'transferred' to policy. Therefore this evidence has to be collected, compared, and fed into the public discourse on educational issues. To put it into one claim: evidence-based policy requires not only the research providing this evidence, but also critical research syntheses.

Second, we have argued that any implementation (sometimes also called transfer) of research-based evidence (or of the results of such evaluative research syntheses, respectively) is a process of science communication. The notions of *Science Communication, Public Understanding of Science (PUS),* or *Public Engagement with Science (PES)* often refer to campaigns for the improvement of understanding science. Sometimes these campaigns primarily aim for *trust* in science, not mainly for an improvement of knowledge and understanding in the sense of knowledge-based reasoning. In the following I do not refer to such campaigns. Instead, I refer to the communication of science and to the understanding of science held by different strata of the public (i.e., by laypeople). When it comes to educational issues, many actors are relevant (politics, administration, school principals, teachers, students and the general public). Nearly all 'applications' of research findings inherently require the communication of these findings. For example, research findings about the effects of homework – to take up the example by Harvey Goldstein – will have an impact on policy decisions as well as on the daily life of teachers and students *only* if these research findings are communicated and understood by the relevant stakeholders. Therefore, implementation requires communication, understanding and trust.

Manfred Egner, let us turn to the communication of research results to the general public? What is your evaluation from a non-researcher's perspective if you are really honest?

The world of science and research works just like any other area – it follows its own internal 'laws' and rules, its own formal and informal forces and powers.

But: It is not (always) aware of this fact. Without explicitly intending to do so, the world of science actually appears to be rather excluding, discriminating and isolating against the outside world. There can be hardly any talk of targeting or including the general public – that is, citizens and taxpayers. Viewed from the outside, the world of science and research appears to be lacking in transparency. Interested outsiders may question whether the work of science actually makes any sense, calling for justifications and questioning the adequacy of resources, but science usually does not hear them.

The old formula or metaphor of 'science as an ivory tower' still largely dominates public perception.

Well then, one may argue, is not the club of polo players also mostly exclusive and nontransparent to the rest of society in general? The difference is not just that one affects only the private sphere and the other (mostly) belongs to the public sphere.

No, the key difference should be that the activities of the polo players are irrelevant to the rest of the world. The work of science and research however is relevant to the public. And if science and research were interested in securing their own future, especially in regard to public funding, then it must be in their own interest to increase – through

public participation – their significance compared to other areas of society, economy, and the state.

Providing public money for science and research is ultimately linked to the public's positive opinion about scientific activities. And politicians are also fully informed by public opinion research institutes about what the population 'thinks' and 'knows' about this particular area. This prevailing positive mood and, hence, approval by the public is accompanied by the degree of transparency and, thus, comprehensibility of scientific research and its findings. This applies, by the way, particularly to the area of empirical research findings, which we are 'producing' at the Leibniz Institute for Educational Trajectories in Bamberg.

It is not only the value of research results measured in terms of academic criteria of scientific best practice but also their attention and recognition among the general public that should therefore be established as a benchmark – a constant research objective as it were. The excellence of the findings is not only a result of scientific criteria but also of the impact force and attention in the nonacademic world.

If anyone should think that this was a plea in favor of the popularisation of science or a demand for subjugating the freedom of science and research according to popular taste they are mistaken.

Rainer Bromme, let's go back to the researchers perspective – should we invest more in actively fostering the communication between research and the 'outside world' – politics as well as the general public?

Whenever science-based knowledge is distributed these are actual cases of science communication. Whenever nonscientists process such knowledge these are actual cases of the public understanding of science. In this sense, science communication and the public understanding of science could occur as deliberate and planned processes. When, for example, researchers report on their findings regarding the effects of violent computer games on a website for parents they are actively doing science communication. When parents read about scientific evidence on such a website this would be a case of public understanding of science. When they are then willing to believe what they have read, this is a case of *trusting* science-based knowledge.

Science communication and the public understanding of science could also occur as implicit and accidental processes. Then, providers as well as users and recipients of such knowledge are oblivious to the fact that they actually communicate and process *scientific or science-based knowledge* and are even more oblivious to the *teaching* and *learning* processes that are inherently embedded in the provision and the processing of such knowledge (Bromme & Goldmann, 2014).

But what are the implications or – to put it more critically – what are the advantages of conceiving the implementation of evidence as science communication, Rainer Bromme?

Well, this understanding of implementation as science communication opens up the pathway of using the conceptual and empirical tools of research on science communication and on the public understanding of science for a better understanding and also a better handling of implementation problems. The most important implication here is that establishing evidence-based policy should be subject to social science research. And this research could be done fruitfully if it conceived of the issue of implementing social science results as a case of science communication.

And Rainer Bromme, could you exemplify the advantage of such a perspective with regard to NEPS?

First of all, NEPS is not only an endeavor to provide data for evidence-based policy, but also a project to provide data for fundamental research. But it is a good example of the advantages of a science-communication perspective. From this perspective, it would be necessary to start with an analysis of the goals for the communication about NEPS. There are different kinds of goals as well as different kinds of audiences, and each goal as well as each audience must be addressed differently. In the following, I refer to the conceptual differentiation between knowledge *from* science (its theories and results) and *about* science (e.g., about its institutions, procedures, and methods for assuring the quality of its results).

I think there are at least four goals for science communication about and from NEPS.

- Providing a wide range of stakeholders (educational policy makers, administrations,

the general public) with information about NEPS, in order to maintain public support and to justify the public funds spent on NEPS.
- Providing the above-mentioned audiences with information from NEPS, that is, with data and results that could be either 'directly applied', for example, by informing decision makers in the context of science-based policy – or that could be subject to further research. The latter is the core objective of NEPS.
- Providing a range of scientific communities (different disciplines) with information about as well as from NEPS in order to foster the use of NEPS data and also as part of the regular scientific communication (in peer-reviewed journals, at conferences, etc). Typically, this kind of communication is not conceived in the context of public engagement with science and public understanding of science, but it is seen as the everyday business of research. But due to the multidisciplinary relevance of the NEPS, the limits between communicating to the scientific community and communicating to more general audiences are blurred. Furthermore, even science communication targeted at more general audiences matters for the visibility within the scientific community.
- Providing information about NEPS and from NEPS for the panel under study. I would suggest conceiving the involvement of research participants (panel members) in the research of NEPS as a kind of Public Engagement with Science, a concept that is nowadays becoming more and more important in research about science communication.

I would like to argue that these goals and these activities of science communication should be based on research about science communication in the context of NEPS. Neither the practical activities nor the planning and reflecting of these activities could be done without treating the issue of science communication as a generic research topic.

John Bynner, UCL Institute of Education, United Kingdom, you organised – together with Tom Schuller – the 2014 and 2015 policy meetings and followed all policy sessions at the 2014 and 2015 conference. Where did we make progress and what are the next steps to be taken?

Longitudinal study began as a small-scale scientific enterprise directed at particular topics in human development, most typically early childhood and adolescence, then more recently, the aging process and old age. The major change came with the recognition that in a rapidly changing technological-driven and increasingly globalised society there was a need to track the progress of individuals across the life course. By this means the routes to success and failure in all the domains of life could be identified and what shaped them understood. Such an evidence basis supplies the means of exposing developmental problems and offers the pointers to the means of solving them on which effective policy rests.

Growing government interest was accompanied by new ideas of national investment needed to ensure the country had what were now to be called longitudinal research resources that were needed for the policy development that was seen as contingent upon them. The language of investment and resources is followed by that of effective policy returns and realisable impact – for which the researchers in receipt of government money were to be held accountable. Such discourse brought to the fore the undoubted tension between scientific inquiry within the framework of a long-term life course perspective and the policy drivers founded in political priorities dominated most typically by the 'here and now'.

The issue has become a growing priority in the work of the Society for Longitudinal and Life Course Studies. In the Lausanne conference (2014) as well as the Dublin conference (2015) the SLLS Policy Group's meetings and several symposia devoted to the longitudinal research-policy interface identified the need for much better understanding of the concept of 'knowledge transfer' and the complex communication processes involved.

The symposium by Jutta von Maurice and Hans-Günther Roßbach on the link between policy, administration and research was particularly valuable in addressing the issues head on by bringing policy users into the frame as speakers alongside academic longitudinal researchers. Discussion revealed what amounts to a degree of impatience for usable findings to feed the policy process. Thus, one of the drivers of the NEPS project had been Germany's poor showing on educational attainment in one of the PISA surveys. These results had been perceived in almost crisis

terms in government circles with recognition of the need to discover their origins in terms of differential effects on different cohorts passing through the educational system at different ages. The NEPS national survey design was constructed primarily to meet such policy needs, offering opportunities to illuminate the educational processes leading to success and failure from preschool through elementary school, secondary school and ultimately the tertiary education and beyond.

In theory, translating the wealth of data into effective action to put right what was going wrong might be seen as straight forward. But here the disjunctions between the language and strategies of science and that of pedagogy, educational management, and organisation became clearly apparent. Usable results were taking longer to emerge than policy makers were prepared to wait and when they did come, though interesting, were often not supplying the precise answers in action terms that were expected.

Participating in and weighing up the discussions that took place in the symposium provided invaluable insights into both the prospects and the difficulties in building bridges across the policy-research divide. The SLLS Policy Group is taking matters forward through identification of two major outcomes for the work of the group:

- to document across the world examples supplied by members of effective translational research;
- to use the database constructed to identify particular projects exemplifying good translational practice and disentangle the processes, outcomes and political contexts of knowledge transfer and impact in ethnographic depth

These case studies could be expected to begin over an extended period and the society's annual conference provides a means of communicating them as they emerge to the wider body of longitudinal researchers. They could feed into what is seen as another priority for the society – the promotion within the longitudinal research community of the training and capability building in translational research that the next generation of longitudinal researchers is increasingly going to need.

Conclusion

Taking all the aspects discussed above together it seems necessary to intensify further the dialogue between researchers, politicians, administrators and the general public in order to reach common goals. Appropriate ways to achieve this intensified dialogue can only be defined by active involvement of all these groups. The argument that underlines the benefits of an intensified research-policy link holds true for any area of research. But it might be especially valuable in the area of longitudinal research because such large-scale designs can only succeed with the aid of excellent researchers as well as the permanent commitment and support of the many stakeholders outside research.

References

Aljets, E. (2015). *Der Aufstieg der Empirischen Bildungsforschung – Ein Beitrag zur institutionalistischen Wissenschaftssoziologie*. Wiesbaden: Springer VS.

Blossfeld, H.-P., Roßbach, H.-G., & von Maurice, J. (Eds.). (2011). *Education as a Lifelong Process: The German National Educational Panel Study (NEPS)*. (Zeitschrift für Erziehungswissenschaft; Special Issue 14). Wiesbaden: VS Verlag für Sozialwissenschaften.

Bromme, R., & Goldman, S. (2014). The public's bounded understanding of science. *Educational Psychologist, 49*(2), 59–69. http://dx.doi.org/10.1080/00461520.2014.921572

Bromme, R., & Prenzel, M. (Eds.) (2014). Von der Forschung zur evidenzbasierten Entscheidung: Die Darstellung und das öffentliche Verständnis der empirischen Bildungsforschung [Special issue]. *Zeitschrift für Erziehungswissenschaft, 27*.

Bromme, R., Prenzel, M., & Jäger, M. (2014). Empirische Bildungsforschung und evidenzbasierte Bildungspolitik. Eine Analyse von Anforderungen an die Darstellung, Interpretation und Rezeption empirischer Befunde. *Zeitschrift für Erziehungswissenschaft, 27*, 3–54. http://dx.doi.org/10.1007/s11618-014-0514-5

De Moll, F., Riefling, M., & Zenkel, S. (2014). „Bin wohl etwas naiv gewesen“ – Zur Rezeption empirischer Bildungsforschung in der Öffentlichkeit – Das Beispiel ELEMENT. *Zeitschrift für Pädagogik*, *60*, 368–389.

Farrow, S., Tymms, P., & Henderson, B. (1999). Homework and attainment in primary schools. *British Educational Research Journal, 25*, 323–341. http://dx.doi.org/10.1080/0141192990250304

Goldstein, H. (2008). Evidence and education policy – Some reflections and allegations. *Cambridge Journal of Education*, *38*, 393–400. http://dx.doi.org/10.1080/03057640802299643

Tenorth, H.-E. (2015). Bildungsforschung und -politik im Dialog?: Lernprozesse und Irritationen. *schulmanagement, 1*, 33–36.

Endnotes

[i] I would like to express my gratitude to all the people who have supported the development of this article. Hans-Günther Roßbach has encouraged me to work on this paper and carefully reviewed my first draft. All of my interview partners were willing to share their very personal thoughts in a rather unusual format and without gaining a direct benefit from it, for example in the form of an increased impact factor. My special thanks are extended to those interview partners without a research background. Moreover, to the editors and reviewers of the international journal *Longitudinal and Life Course Studies* who were open to this new format, which certainly cannot be reviewed in a conventional way. The feedback I received has been constructive and has helped me to improve my arguments.

[ii] E-mail by Cat Westlake from September 26, 2014.

[iii] The development of empirical education research and its growing influence on education policy is very well described in Aljets (2015).

[iv] I have drawn upon a more detailed discussion that appeared in Goldstein (2008), where detailed references can also be found.

BOOK REVIEW

Handbook of the Life Course, Volume II

Michael J. Shanahan, Jeylan T. Mortimer, Monica Kirckpatrick Johnson(Eds.) 2016
ISBN 978-3-319-20880-0

Karl Ulrich Mayer Leibniz Association
kum@leibniz-gemeinschaft.de

There are three, partially overlapping sets of expectations with which I set out to examine the second volume of the Handbook of the Life Course. The first set of expectations derives from my own specialties and interests in the study of the life course. I entered the study of the life course primarily because I wanted to unravel the mechanisms generating inequalities between the family of origin and later life and thereby the structure of opportunities in a given society. Since inequalities and inequalities of opportunities play out in the educational system and the labour market, this brought me to view the life course as a set of interrelated institutionalised sequences as educational trajectories, occupational careers and family roles as well as a sequence of roles in relation to the (welfare-)state rather than age roles or life phases. The interest of life courses as the playing ground for inequalities also leads to the comparative interest in historical or cross-national life course regimes. Do we learn more about the effect of politics and policies on life courses in contrast to specific conditions and transitions?

The second set I formulated as the standards of measuring progress in my review article on new trends in the study of the life course (ARS 2009: 413-433): Is there progress in testable and tested theories rather than heuristics? How well-specified are the intersections between organismic and psychological development and socially embedded life courses? Are there new methods overcoming the fragmentary view of successive events? Are we learning more about conditions in early life as distal rather than proximate and immediate causes? What have we learned about the alleged general trends in life courses such as de-standardisation and pluralisation?

The third set of expectations follows some of my criticisms of the first volume of the Handbook (Social Forces 2003). Then I noted a number of conspicuous omissions like aging, longitudinal data collections and inputs from economics and too little on cross-national institutional variation and policy impacts. I also complained about the lack of separate and more complete name register.

In sum, what I was eager to find were theories rather than concepts and heuristics, corroborated results rather than hypotheses or illustrative findings, mechanisms rather than associations, interdisciplinary rather than disciplinary breakthroughs, historical and national specificity rather than universals, institutional structures rather than age norms. I also was curious how US-centric the handbook would turn out to be. And of course I wanted to learn truly new stuff.

In their introduction Shanahan, Mortimer and Johnson not only define the differing philosophies guiding the second and the first handbook (forward-looking and future oriented vs summary review of state of the art), but also chart the outburst of publications in the area and very usefully collect the recommendations for the future of life course studies made in the 31 contributed chapters. The five overall headings (I: Foundations of Life Course Studies and Future Studies, II: Changing Social Contexts and Life Course Patterns, III: Health and Development through the Life Course, IV: Life Course Methodologies , and, V:The Life Course and Policy: Building the Nexus) are only partially informative and, in some cases, clearly do not work well. For instance, the systematic and excellent review by Hagestrand and Dykstra on social policies impacting on the life course would have fitted much better in part V and

the piece by Wadsworth and Kuh on 'Epidemiological Perspectives on the Life Course' much better in part III.

In the introductory chapter the editors also reiterate the concept of the life course as "age-graded roles that structure the (or create patterns) in the biography". This still carries the functionalist baggage of (normative) age roles and life phases and clearly misdirects many of the chapter authors. This is the case, for instance, when Phyllis Moen in her chapter on work starts from the tripartite life course and looks at work as a life phase ("master life course status") rather than work life mobility as job and occupational careers. Above all, the editors stick to the idea of the life course 'paradigm' and obviously authors were instructed to follows Glen Elder's five heuristic principles. I think it is Glen Elder's great merit to have triggered the take off of life course studies by formulating them. But in many chapters we find an 'o mani padme hum' of these principles as mere thematic rubrics – very often much below the level of sophistication of its inventor. It is then the question whether the outburst of publications was associated with a corresponding outburst of theories, findings and methods..

Both the quality and the genre of the 34 chapters are very uneven. They range from (too many) mostly suggestive essays (e.g. Dannefer et.al., Hitlin and Kwon, London and Wilmoth, Mortimer and Moen) to detailed findings concentrating on just single or a few studies – mostly the ones of the authors themselves (Elder/George, Thornberry, Hermanowicz, Macmillan and Furstenberg) to textbook like treatises (Moore/Brand) .

But let me start in a more positive and constructive spirit with a number of very impressive highlights. John Bynner's excellent chapter on the huge importance of the growing and accumulating number of longitudinal studies argues rightly how such research enterprises were and will be crucial in institutionalising life course research as an emergent field. The chapter gives a very good overview of older and current studies and, not least, identifies the incorporation of new kinds of data (like biomarkers and geocoded data) and the challenges in regard to funding, attrition and response burden. He might have, though, added a more sober reflection about the astonishingly modest yield of, for instance, the longest running British cohort studies. The chapter by Kalil, Duncan and Ziol-Guest on the short and long-term consequences of early childhood poverty – especially in regard to health – is not only the shortest, but also the most brilliant contribution. It is theoretically highly developed – e.g. with the distinction between transient and persistent poverty or the importance age sensitivity – and provides an excellent review of empirical and methodological progress in the area. Rather than just suggesting life course effects it gives precise information about effect sizes as well and goes a long way to specifying the causal mechanisms (like family and environmental stress, allostatic load). And it does not – like many chapters in the handbook – just suggest the strengths of life course linkages, but – by reporting on results from large scale interventions and experiments, among other things – also discusses modest or no effects.

Similarly impressive is the chapter by father and daughters Blossfeld on changes in educational inequality in cross-national perspective which uses the OECD-PIACC data on 22 countries to empirically test competing theories (modernisation, cultural reproduction) of trends in inequality of educational opportunity. In a further breakthrough in the area of cross-national life course studies Hagestad and Dykstra not only highlight the difference between US and European traditions on gendered life courses, but also provide a systematic and very fruitful overview of age related social policies. This is well complemented by the both analytical and informative chapter by Pamela Herd On 'Influences of Social Welfare Policies on Health Disparities Across the Life Course'. In regard to theory development Martin Diewald brings together two strands from the previous handbook: on the one hand the explication of risk by O´Rand and the idea of policies as life course risk management by Leisering. Diewald very fruitfully takes up the distinction between risk and adversity and expands on DiPrete's country specifications as systems which either prevent risks (Germany) or deal with adversity, but allow risk (Sweden) or fail in both respects (US). Not least he reconnects these concepts with Tilly's ideas on relational stratification.

Probably the greatest progress of the handbook is documenting and where it also most informative is in the area of health. In addition to the already

mentioned chapter on the effects on early poverty on later life Kalil et al. and the impressive piece by Wadsworth and Kuh on epidemiological perspectives on the life course life, Hayward and Sheehan in their chapter 'Does the body forget? Adult Health, Life Course Dynamics and Social Change' stress the importance of taking a comprehensive view of health change encompassing biological risk, morbidity, functioning, disability and mortality and of using a biologically informed framework for early gains or deficits in capacity as well as differential functional decline in adulthood and old age. One very important contribution of this chapter is the discussion and detailed documentation of the specificity of the historical contexts and thus of social change for health development across the life course. Johnson et.al. and Ferraro in their respective chapters trace two important connections between health and life course: education and aging. Johnson et.al. unravel the multiple ways in which education affects health such as differences in risky behaviour, life styles, differential health knowledge and the indirect linkages between education and health via social support, employment and work and income. Ferraro takes on the Fries/Manton controversy on compression of morbidity by showing that the extension of healthier life is most likely quite different for different kinds of diseases and that the available evidence points more to an extension of survival after the onset of disability rather than compression of morbidity.

Duan Alwin and his co-authors contribute an empirically well-informed as well as theoretically very reflective and methodologically sophisticated chapter on 'Cognitive Development and the Life Course: Growth, Stability and Decline'. They rightly note that developmentalists like Baltes and Bronfenbrenner have treated the environment and especially the institutional characteristics of life courses in a residual fashion. In their own discussion they want to remedy this deficit by focusing on the effects on cognitive ability by educational transitions, transition to adulthood, adulthood retirement an old age. One noteworthy point is that they are actually refuting the celebrated finding by Kohn and Schooler on the impact of occupational flexibility and corresponding trajectories on cognitive function as an artefact. It is unfortunate that in their literature review they miss what I take to be the most impressive evidence on the relationship between institutional macrostructures and cognitive outcomes, namely the study by Rohwedder and Willis on 'Mental Retirement' (Journal of Economic Perspectives 24 (1), 2010 119-38) where these authors show how 'national' reforms in extending the age of retirement delay aggregate cognitive decline.

I was looking forward to read the chapters on the 'new' topics of criminal behaviour, disaster, agency, mental health and longitudinal qualitative studies but found them all relatively disappointing. Both the chapters by Wakefield and Apel and by Laub on criminal behaviour in the life course do not add much if anything beyond the chapters by Sampson and Laub and Uggen and Massaglia in the first handbook. The chapter on latent growth model by Macmillan and Furstenberg is very helpful even to one like myself who used the method, but trajectories of the body mass index (BMI) are probably not the most salient application – income or status trajectories would have been more central. A very curious omission – now in both handbooks – is a chapter on sequence analysis, a method which has made great progress and found many applications in recent years.

One criterion against which to measure some of the contributions to the handbook is whether they live up to the state of the art in the respective disciplines and go beyond it by bringing their fields into the life course arena. This clearly succeeded well in the chapters close to family demography by Hofferth and Goldscheider, the one close to social policy by Hagestad and Dykstra and the one close to developmental psychology by Alwin and co-authors. But I found the chapters on education by Crosnoe and Benner and – as already mentioned – the one on work by Moen not up to what has already been accomplished in their 'home' disciplines. An innovative and important contribution from which I learned a lot about developments in the US proved to be the chapter by Zapata-Gietl and co-authors on the new life phase of 'college for all'. It could have benefitted, however, from comparisons with other countries where – despite huge institutional differences – this life phase after high school seems similarly 'floating'. Focusing on the interdependencies between physical development, cognitive skills and psychosocial development the

chapter by Paul Dornan on the 'Young Lives Study' following cohorts from age one to age 15 and from age eight to age 22 in Ethiopia, India, Peru and Vietnam is a real breakthrough in life course research going beyond advanced societies.

One last note relates to the issue of trends covered in various ways in the chapters by Dannefer et al., Elder and George, Mortimer and Moen, and O`Rand and Bostic. There is a tendency to stress the negative effects of globalisation, increasing 'precariousness' and vulnerability. Although this accords well with the 'social problem' tradition of sociology I found both the explication of these processes and the alleged evidence lacking.

All in all, like the first edition, the Handbook of the Life Course, Volume II is an impressive accomplishment for which the editors can only be congratulated. It shows, however, a puzzling combination of accumulation, innovation and stagnation in this field. The explosive burst of publications is clearly not matched by corresponding progress in theory development and the accumulation of corroborated findings. And – again – both a name and a subject register would have been useful.

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 crandall@slls.org.uk **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 crandall@slls.org.uk. 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

RESEARCH POLICY DIALOGUE

BOOK REVIEW

LLCS Journal can be accessed online at www.llcsjournal.org

Published by the Society for Longitudinal and Life Course Studies
info@slls.org.uk www.slls.org.uk