Modelling socioeconomic trajectories: An optimal matching approach

This item was submitted to Loughborough University’s Institutional Repository by the/an author.


Additional Information:

- This paper was accepted for publication in the journal International Journal of Sociology and Social Policy and the definitive published version is available at https://doi.org/10.1108/01443330810881268

Metadata Record: https://dspace.lboro.ac.uk/2134/36478

Version: Accepted for publication

Publisher: © Emerald

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: https://creativecommons.org/licenses/by-nc-nd/4.0/

Please cite the published version.
MODELLING SOCIOECONOMIC TRAJECTORIES: AN OPTIMAL MATCHING APPROACH

JULIET STONE, GOPALAKRISHNAN NETUVELI AND DAVID BLANE

SOCIAL SCIENCES RESEARCH UNIT, DEPARTMENT OF PRIMARY CARE AND SOCIAL MEDICINE, IMPERIAL COLLEGE LONDON, LONDON, UK

ABSTRACT

PURPOSE – The aim of this paper is to describe the use of sequence analysis to model trajectories of life-course economic activity status, within a broader research agenda aimed at improving understanding of the relationship between socioeconomic position and health.

DESIGN/METHODOLOGY/APPROACH – The analysis used data on 288 participants of the Boyd Orr Stratified Sub-Sample, comprising a combination of prospective and retrospective information on economic activity status, as well as health in early old age. Economic activity was coded as a time-based sequence of states for each participant based on six-month periods throughout their lives. Economic activity was classified as: pre-labour market; full-time employment; part-time employment; housewife; made redundant; stopped work due to illness; retired; other unemployed; or not applicable. Optimal matching analysis was carried out to produce a matrix of distances between each sequence, which was then used as the basis for cluster analysis.

FINDINGS – The optimal matching analysis resulted in the classification of individuals into five economic activity status trajectories: full-time workers (transitional exit), part-time housewives, career breakers, full-time workers (late entry, early exit), and full-time housewives.

ORIGINALITY/VALUE – The paper presents the case for using sequence analysis as a methodological tool to facilitate a more interdisciplinary approach to the measurement of the life-course socioeconomic position, in particular attempting to integrate the empirical emphasis of epidemiological research with the more theoretical contributions of sociology. This may in turn help generate a framework within which to examine the relationships between life-course socioeconomic position and outcomes such as health in later life.

BACKGROUND

SOCIOECONOMIC POSITION AND HEALTH

The relationship between socioeconomic position and health is of ongoing interest on many levels, particularly in terms of policy development. Research indicates that health inequalities are continuing to widen (Acheson et al., 1998) and, moreover, it appears that these inequalities persist across the life course and into older age (Arber and Ginn, 1993).

With regard to explaining such health inequalities, there is increasing evidence that adult health may be affected by past exposures (Barker et al., 2001; Poulton et al., 2002) – including indicators of socioeconomic position such


as occupational class, material circumstances and status within the community. This has been an important
driving force behind the development of a life course approach within epidemiology (Ben-Shlomo and Kuh,
2002).

LIFE COURSE EPIDEMIOLOGY
Life course epidemiology generally centres around three major models of life-course disease development. The
critical period model, informed by David Barker’s “foetal programming hypothesis” (Barker et al., 1990),
contends that the health consequences of exposure to hazards depend on the age at which such exposures
occur. A pathway model is based on the idea that exposures in early life are relevant to later health due to the
way in which they influence susceptibility to subsequent exposures. Finally, an accumulation model assumes
that there is a dose-response relationship between exposure to “risks” and the development of disease, with
gradual, long-term effects (Power et al., 1996).

However, it is often difficult to disentangle the different processes of disease development in practice (Hallqvist
et al., 2004) and the best explanations may be found by exploring the ways in which these processes interact
(Bartley, 2004). In order to better understand the development of disease in a life-course perspective, it may
therefore be necessary to take a more holistic and theory-driven view of the life-course and its interdependent
pathways. In this respect, epidemiology has been somewhat lacking, in that work in this field has not been
caracterised by a theoretical approach to the life course as a concept. Instead, the life course in epidemiology
has been used more as a framework within which to situate hypotheses regarding disease development, rather
than as a problematic concept that itself needs discussion. With this in mind, it may be useful to take an
interdisciplinary approach and draw upon the more theoretical ideas on the life course from sociology in order
to enhance the more empirical focus that is often observed in epidemiology.

TRANSITIONS AND TRAJECTORIES
In sociology, the concept of the life course has been strongly developed in qualitative research and is closely
associated with the life history technique, whereby people’s lives are “reconstructed” through interviews,
letters and other sources to provide insights into the social world (Thomas and Znaniecki, 1996). The resulting
“life stories” are then used as the basis for examining the relationship between individuals and society.

Developing these concepts further, Elder describes the life course as referring to “social patterns in the timing,
duration, spacing and order of events” (p. 21), encompassing multiple, independent pathways (Elder, 1978).
More recently, Giele and Elder have argued that a “life-course paradigm” has emerged in the social sciences,
with central concepts of transition and trajectory (Giele and Elder, 1998).

Transitions are described as changes in state and are particularly related to “the adaptive problems of entry and
exit from a single role” (p. 24) (Elder, 1978). Such transitions may include school examinations, entry into the
labour market, leaving home, marriage, parenthood, departure of the last child from home, employment
changes, onset of chronic illness and retirement (Bartley et al., 1997; Elder, 1978).

A trajectory is described as linking states and transitions over time and also considers duration in terms of the
waiting times or spells between a change in state (Elder, 1985). The latter point is especially important in terms
of the relationship between socioeconomic position and health, as there is evidence that in this context the
duration of exposure to disadvantage may be even more important than the timing (Power et al., 1999). With
this in mind, this paper therefore attempts to incorporate the theoretical concepts of trajectory and transition
into an empirical study of life-course socioeconomic position.
SOECONOMIC TRAJECTORIES

In epidemiology, models of life-course socioeconomic position have often been driven by the availability of data and have not necessarily used the concept of a trajectory. For example, where information is available from three points during participants’ lives, this has been used to construct a simple, cumulative “disadvantage score” (Davey Smith et al., 1997) or to look at social mobility between these three points (Naess et al., 2006). However, as much of such research is based on secondary analysis of data, the timing of the points of measurement has necessarily been informed by practical rather than theoretical considerations. Therefore, although existing models have identified significant associations between life-course socioeconomic position and health outcomes, which is important in that it supports further investigation of such associations, there is an ongoing need to develop more comprehensive models.

Using the concept of a socioeconomic trajectory, rather than considering life-course socioeconomic position as a series of isolated events, may help to uncover processes of chronic disease development that would fail to emerge using more conventional strategies. The first step towards this is to try to identify whether distinct patterns of life-course socioeconomic position emerge empirically. This information can then potentially be used to develop measurement models that capture a more valid picture of what is occurring during the life course. In order to achieve this, a method of analysis that considers life-course trajectories as holistic units is required. Optimal matching, a method of data classification originating in microbiology, is one such method.

OPTIMAL MATCHING

Optimal matching is a type of sequence analysis that was introduced into sociology towards the end of the twentieth century by Abbott (Abbott and Forrest, 1986). It has subsequently been used to analyse a range of social and historical phenomena, from the structure of English folk dances (Abbott and Forrest, 1986) to patterns of work-life social mobility (Halpin and Chan, 1998). Optimal matching differs from more commonly used methods such as event history analysis because it provides a means to examine entire sequences of data simultaneously. In theoretical terms, it therefore allows investigation of entire trajectories as the basic unit of analysis, rather than series of events.

The technique works by measuring the “distance” between pairs of sequences that are coded as a series of consecutive states, the distance being based on the minimum number of changes or “operations” needed to turn one sequence into the other. In sociology, the three operations that are usually used are insertion, deletion and substitution. For a detailed explanation of this process, see (Macindoe and Abbott, 2004).

An important aspect of optimal matching is the assigning of “costs” to insertion/deletion (generally referred to as “indel”) and substitution operations. These costs are the basic unit used to calculate the distance between sequences and so are and varying them will, therefore, potentially influence the results of the analysis.

A particularly significant issue with regard to the use of indel operations is the notion of “temporal warping” (Lesnard, 2006). This is essentially based on the idea that inserting or deleting states from a sequence will change the timing of the remaining states. This means, therefore, that the meaning of the sequence as a whole may be altered – for example, in an analysis of employment status, the deletion of a state could “shift” the timing of entry to the labour market significantly, perhaps moving it to before statutory school leaving age.

While such issues need to be considered, it should nevertheless be remembered that the idea of making insertions, deletions and substitutions is not to produce a new set of sequences for analysis, but instead to estimate how closely they resemble other sequences. Using indel can be useful if the aim of the analysis is to...
identify similarity in subsequences, regardless of where they occur in the sequence as a whole. In such cases, “shifting” of the time-line in order to align subsequences is desirable rather than being a problem.

Substitution costs can also be specified in optimal matching. It may be the case that for theoretical reasons, changing from state A to state B is assigned a lower cost than changing from state A to state C. For example, it might be considered that being in part-time work is “closer” in cost terms to full-time work than to unemployment. Therefore, a substitution cost matrix can be constructed. Care needs to be taken when specifying such a matrix, however, unless there are good theoretical reasons for changing the costs.

In order to interpret the results of optimal matching, an additional stage of data classification is generally carried out, often involving a technique such as cluster analysis. In the case of cluster analysis, the distance matrix produced by the optimal matching analysis is used to determine how the sequences join together. The ultimate aim of such analysis is to maximise similarity within clusters and difference between clusters, thereby identify distinct and homogenous patterns within the sequence data.

This paper describes the use of optimal matching and cluster analysis to identify patterns of life-course economic activity status within a population of adults currently in early old age. This process is part of a broader research agenda, the ultimate aim of which is to use these empirically emerging patterns to inform the development of theoretical models of the relationship between life-course socioeconomic position and disease development, which can in turn be used to predict and potentially modify health in later life.

**METHODS**

**DATASET**

Data from the Boyd Orr Stratified Sub-sample were used for the analysis. Full details of the study design and progress can be found in (Blane, 2005). Figure 1 shows the key stages of development of the study.

Briefly, the study originated from Sir John Boyd Orr’s Carnegie Survey of Family Diet and Health in the 1930s, for which data were collected on a range of social, dietary, health and anthropometric factors from 1,352 British families.
In the mid-1990s, participants of the survey were traced for an NHS record linkage study. Survivors were then re-contacted in 1997, when they were aged 63-78 years. A random sample, stratified by socioeconomic circumstances in childhood, was drawn with the aim of recruiting 100 participants from each of three strata. The 294 interviewees eventually included in the study were found to be broadly representative of the general population at those ages in the UK (Blane et al., 1999). During 1997-1998, the 294 participants were interviewed in their homes, providing detailed retrospective information on their residential, occupational and family circumstances throughout their lives as well as their current health. During 2000-2001 and again in 2004-2005, the surviving members of the cohort were re-contacted to collect follow-up information. The resulting dataset includes detailed prospective information on socioeconomic circumstances in childhood, from original Carnegie Survey, as well as prospective information from early old age, with retrospective data on the intervening years. It therefore provides a valuable and unique resource for investigating life course processes of disease development.

This paper concerns participants’ economic activity status over the life-course. Overall, there were 288 cases with complete data for this trajectory. Of the six excluded cases, 50 per cent were female and their mean age was 67.2 years. This was comparable to the included cases, among whom 53 per cent were female and the mean age was 67.9 years. There was a difference between the two groups in terms of class of last main occupation, which was defined as manual in all of those with incomplete data, compared with 43 per cent manual and 57 per cent non-manual in those with complete data. However, given the small number of excluded cases, it was considered that this difference was unlikely to have biased the results.

DATA PREPARATION
For the present study, the data first needed to be prepared as sequences for use in optimal matching analysis. Two coders carried out the data preparation, with dual coding of every tenth case for quality control.

In order to keep all of the sequences the same length, the time-line began at the year of birth of the oldest participant (1919). For younger participants, the pre-birth period was assigned the code “not applicable”. For those not included in later waves of the survey, the post-inclusion period was also assigned the code “not applicable”. Therefore, every sequence in the original dataset ran from 1919 until 2004. These sequences were then converted to personal time, based on participants’ chronological age, in order to be able to assess the effects of timing in terms of age as well as cohort. The preliminary results presented below pertain to the personal time-line only, with comparison with the historical time-line planned as part of future work.

MEASURES
The potential states for the economic status trajectory were:

(1) Pre-labour market.
(2) Fulltime employment. (3) Part-time employment. (4) Housewife.
(5) Made redundant.
(6) Stopped work due to illness.
(7) Retired.
(8) Other unemployed.
(9) Not applicable.
In the raw data, the dates of employment were continuous with the year a job ended tending to be recorded as the same year that the next job began. The time-line was therefore coded in six-month periods to take account of this overlap.

DATA ANALYSIS
The optimal matching analysis was carried out using SPSS 14.0 and TDA 6.41.

As the preliminary analysis was exploratory and there were no clear theoretical reasons to alter the costs, substitution costs were initially set at a default of 1.0 for any substitution and 0.5 for any insertion or deletion. Using the distance matrix produced by the optimal matching analysis, cluster analysis was subsequently carried out in SPSS 14.0 using Ward's method, which produces clusters based on a weighted average of distances between variables.

RESULTS

CLUSTER ANALYSIS
Figure 2 shows the agglomeration schedule of the cluster analysis plotted as the difference between cluster coefficients at each stage against the number of clusters.

The convention is to look for the “elbow” of the curve in order to identify the point at which sequences that are relatively distant from each other must be joined in order to create a new cluster. This point gives an indication of the optimal number of clusters.

In this case, the elbow initially appears to emerge at around the seven-cluster point, with a steeper change at the four-cluster point. This suggests that the optimum number of clusters is between four and seven. Solutions
based on between three and eight clusters were examined and it was concluded that a five-cluster solution provided the best balance between internal homogeneity and external heterogeneity of the clusters, as well as providing substantively meaningful classifications.

**CLUSTERS**

The five clusters were given the following labels based on their characteristics (see Appendix):

1. Full-time workers (transitional exit) (Figure A1).
2. Part-time housewives (Figure A2).
3. Career breakers (Figure A3).
4. Full-time workers (late entry, early exit) (Figure A4).
5. Full-time housewives (Figure A5).

**Table 1: Basic Demographic Characteristics of the Five Clusters**

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Men n (%)</th>
<th>Women n (%)</th>
<th>Median year of birth</th>
<th>Median age at entry to labour market</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>117 (85%)</td>
<td>21 (15%)</td>
<td>1929</td>
<td>14 years</td>
</tr>
<tr>
<td>Full-time workers (transitional exit) n=138</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cluster 2</td>
<td>0 (0%)</td>
<td>58 (100%)</td>
<td>1929</td>
<td>15 years</td>
</tr>
<tr>
<td>Part-time housewives n=58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cluster 3</td>
<td>0 (0%)</td>
<td>38 (82%)</td>
<td>1929</td>
<td>15 years</td>
</tr>
<tr>
<td>Career breakers n=38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cluster 4</td>
<td>19 (49%)</td>
<td>20 (51%)</td>
<td>1929</td>
<td>17 years</td>
</tr>
<tr>
<td>Full-time workers (late entry, early exit) n=39</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cluster 5</td>
<td>0 (0%)</td>
<td>15 (100%)</td>
<td>1929</td>
<td>16 years</td>
</tr>
<tr>
<td>Full-time housewives n=15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 shows the main demographic characteristics of the five clusters.

Cluster 1, full-time workers (transitional exit), was characterised by consistent full-time employment during working age, followed by a short transitional period of part-time work, being made redundant or stopping work due to illness, and finally a period of retirement. This cluster was primarily male (85 per cent).

Cluster 2, part-time housewives, showed a short period of full-time employment followed by part-time working or being a housewife for the remainder of working life. Retirement followed without the transitional period observed in cluster 1. This cluster was 100 per cent female.
Cluster 3, career breakers, showed periods of full-time employment at the beginning and end of working life, with an intervening period of part-time work or being a housewife. Most in this cluster had a short period of part-time work before retirement. This cluster was also 100 per cent female.

Cluster 4, full-time workers (late entry, early exit), was similar to cluster 1 in that it consisted of people in consistent full-time employment throughout their working lives. However, the overall period of employment was shorter than for those in cluster 1, with a later entry to and earlier exit from the labour market. Importantly, the transitional period between employment and retirement was not observed in this group, with most people moving directly from full-time work to retirement. This cluster showed a fairly even gender distribution (49 per cent men, 51 per cent women).

Cluster 5, full-time housewives, was characterised by a short period of full-time working followed by being a housewife for the remainder of the time sequence. There was no period of retirement at the end of the time-line in this cluster. This cluster was 100 per cent female. Despite being small (n = 15), this cluster showed extremely good internal homogeneity so was considered to be a valid classification rather than a merely a residual group.

Table 2 shows the distribution of non-manual vs manual class of main occupation in each cluster. In cluster 1, there were slightly more cases in the manual class than in the total population (51 vs 43 per cent). The same was the case in Cluster 2, with 42 percent manual. In clusters 3, 4 and 5, the proportion in the manual class was lower than the average, at 34, 18 and 20 per cent, respectively.

Logistic regression revealed a significant association between cluster membership and class of last main occupation (p < 0.001), although this association appeared to be driven mainly by the high proportion of non-manual workers in cluster 4.

**Table 2: Distribution of class of last main occupation in five clusters of life-course economic activity status.**

<table>
<thead>
<tr>
<th>Main occupational class</th>
<th>Non-manual</th>
<th>Manual</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cluster 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-time workers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Transitional exit)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n (%)</td>
<td>67 (49%)</td>
<td>71 (51%)</td>
<td>138</td>
</tr>
<tr>
<td><strong>Cluster 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part-time housewives</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n (%)</td>
<td>28 (48%)</td>
<td>30 (52%)</td>
<td>55</td>
</tr>
<tr>
<td><strong>Cluster 3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Career breakers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n (%)</td>
<td>25 (66%)</td>
<td>13 (34%)</td>
<td>38</td>
</tr>
<tr>
<td><strong>Cluster 4</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-time workers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(late entry, early exit)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n (%)</td>
<td>32 (82%)</td>
<td>7 (18%)</td>
<td>39</td>
</tr>
<tr>
<td><strong>Cluster 5</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-time housewives</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n (%)</td>
<td>12 (80%)</td>
<td>3 (20%)</td>
<td>15</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>164 (57%)</td>
<td>124 (43%)</td>
<td>288</td>
</tr>
</tbody>
</table>
DISCUSSION

This paper has demonstrated the potential of optimal matching as a tool for modelling life-course socioeconomic position, integrating the sociological concepts of trajectories and transitions within the empirical tradition that is typical in epidemiological research. The analysis identified distinct clusters of socioeconomic sequences that were characterised by particular patterns of transition between different states of economic activity.

These results were not unexpected, showing that men tended to be employed full-time consistently throughout their working lives, while employment patterns among women were more heterogeneous. In particular, men’s working lives appeared to be characterised by just two primary transitions – entry to and exit from the labour market. In women, the results indicated that other transitions such as marriage and family formation may also be important in terms of employment patterns.

The fact that the patterns that emerged were broadly what would have been expected suggests that the optimal matching approach adopted in the present study is producing a valid representation of patterns of economic activity and could potentially identify equally valid patterns for other socioeconomic trajectories such as occupational class. Nevertheless, the analysis also identified findings that might not have emerged using more conventional methods unable to compare whole sequences of data from the entire life course. The “career breakers” cluster is a good example of this; if those that fell into this cluster were only assessed before and after their “break”, this pattern would not have become apparent.

The analysis also highlighted the transition to retirement as an important driving force behind the way in which the sequences clustered together. Again, the significance of this transition was not entirely unexpected but the strength with which it emerged as a defining characteristic of the life-course economic activity patterns in this population was interesting. Specifically, men were mainly divided between two clusters, one showing a direct transition between full-time work and retirement, and the other showing a period of part-time work, stopping work due to illness and/or being made redundant between full-time work and retirement. The latter group of men were much more likely to have been in manual occupations than the former (51 vs 20 per cent). One could therefore speculate that, as the group showing a direct transition between work and retirement appeared to be more advanced in terms of occupational class, they may have had made more provision for retirement with, for example, private pensions, and would therefore have defined themselves as retired as soon as they stopped work, regardless of why they stopped working. Those in the group showing a less clear-cut transition between work and retirement may have been more likely to rely on the state pension and would therefore have not defined themselves as retired until they reached the age of 65 years.

Confirming such speculation would require further study but nonetheless, the findings emphasise the capacity of sequence analysis to identify significant transitions within the life-course and their relative import. Sequence analysis may therefore prove to be a useful tool for identifying the complexities of socioeconomic trajectories that characterise particular populations, while at the same time allowing these complexities to be distilled into a form amenable to analysis.

LIMITATIONS

There has been debate over whether and how optimal matching can be used in sociological research (Abbott, 2000; Abbott and Tsay, 2000; Levine, 2000; Wu, 2000). Some of the criticisms of the technique concern the way that it “warps” time, while others question the conceptual legitimacy of using a “biological” technique to analyse social phenomena. In particular, it is argued that unlike molecular structures, trajectories such as occupational status tend to develop step by step with distinct phases and are therefore more suited to techniques such as event history analysis than to techniques such as optimal matching that consider sequences as a whole (Levine, 2000; Wu, 2000). However, others argue that it is this very characteristic that marks sequence
analysis out as an ideal method to deal with life-course data, allowing empirical identification of differential trajectories without a priori assumptions of causality between different “steps” (Abbott, 2000).

There are also related criticisms of cluster analysis – the arguments being that it does not necessarily produce a unique solution (Morgan and Ray, 1995) and that it may mask variability. While these criticisms are valid, there are ways to produce more robust solutions using cluster analysis. As with any research, it is important to carry out tests of reliability and validity in order to be confident about the results achieved. In cluster analysis, one needs to consider both internal homogeneity and external heterogeneity in order to evaluate the validity of the clusters. In addition to assessing these criteria qualitatively, as was the case in the present study, it is also possible to apply statistical tests such as analysis of variance within the clusters and discriminant analysis (Peseau and Tudor, 1988) in order to appraise the validity of the solution, including both internal and external validation studies (Speece, 1994). One weakness of the present study is that such analyses were not carried out; this was partly due to time constraints, but it was also felt that the qualitative assessment was adequate at this exploratory stage of analysis, especially given that the clustering appeared to be substantively plausible.

In addition, one can look to other studies for external validation. A study in Switzerland using optimal matching to test hypotheses regarding pluralization of the life course in terms of professional and family activities (Widmer et al., 2005) found employment patterns very similar to those identified in the present study. In their sex-specific analysis, the authors identified one dominant trajectory in men characterised by continuous full-time employment. In contrast, women showed a number of distinct employment trajectories, with four main models emerging: continuous full-time employment; continuous part-time employment; full-time “homemaker”; and returning to work after a period as a homemaker. These four trajectories also map fairly consistently onto the patterns identified for women in the Boyd Orr cohort. This goes some way to supporting the reliability of our findings as they appear to have been broadly replicated in a different population.

**FUTURE WORK**

This preliminary analysis described in this paper looked at only one trajectory – employment status – but the Boyd Orr dataset contains enough detailed retrospective information to be able to analyse a range of trajectories related to work, family and material circumstances. Ideally, one would have life-course information on both socioeconomic position and health in order to examine the interdependence of these two trajectories. However, there is a lack of availability of such data and current analyses will therefore probably be limited to using health as a cross-sectional outcome.

Nevertheless, the Boyd Orr Subsample remains a rich resource for studying life-course influences on health. Importantly, for married participants – who represent the majority of the population in the study – information is also available on spouse’s employment, which may allow incorporation of an element of the “social relations” dimension of the life course, drawing on the concept of “linked lives” (Elder, 1998). Moreover, as the participants vary in age, being born over a period of 15 years, it may be possible in future work to examine different “birth cohorts” and thereby investigate the possible influence of sociohistorical context.

Human agency is also recognised as an important influence on life-course trajectories. However, operationalising human agency is more problematic. It could be hypothesized individuals whose sequences appear as “outliers” within the clusters are operating outside the normative, structural influences that shape the life-course for the majority – perhaps agency plays a greater role in their lives than in those who follow with the “normative” trajectories. This may be an interesting topic for future research.

To further investigate the idea of normative trajectories, another option is to develop “ideal types” based on theory and/or evidence and to examine how closely individuals follow these archetypal trajectories. This may give an indication of how people’s lives relate to normative or structural influences.
Ultimately, the work described in this paper aims to pave the way for causal analysis to identify associations between life-course socioeconomic position and health outcomes. Used as an exploratory tool, optimal matching may facilitate this aim as part of an inductive process, providing an evidence-based framework to inform development of theoretical models of the life course. On a more practical level, it may be that such sequence analysis can also help to identify the points during the life course at which it is most important or productive to measure socioeconomic position cross-sectionally, in terms of predicting later outcomes. Nevertheless, more innovative techniques will likely be required in order to translate the empirical models of the life-course produced by optimal matching into analyses that can produce causal inferences, and this should be the focus of future research.

CONCLUSIONS
Using a unique dataset, this paper describes an optimal matching approach to describing people’s life-course economic activity status. We consider that this method represents an improvement on more conventionally used techniques such as event history analysis, as sequence analysis allows the researcher to treat the life course as an interlinked process rather than as a dislocated series of events. In theoretical terms, this in turn allows a focus on the interdependence of transitions and trajectories. More specifically to epidemiological research, the integration of these theoretical concepts within empirical work may be an important step towards improving understanding of the relationship between life-course socioeconomic position and health.

REFERENCES


APPENDIX. VISUAL REPRESENTATION OF FIVE-CLUSTER SOLUTION

Cluster 1: Full-time workers (transitional exit)

- Not applicable
- Other unemployed
- Retired
- Stopped work due to illness
- Made redundant
- Housewife
- Part-time
- Full-time
- Pre-labour market

Cluster 2: Part-time housewives

- Not applicable
- Other unemployed
- Retired
- Stopped work due to illness
- Made redundant
- Housewife
- Part-time
- Full-time
- Pre-labour market
Cluster 3: Career breakers

- Not applicable
- Other unemployed
- Retired
- Stopped work due to illness
- Made redundant
- Housewife
- Part-time
- Full-time
- Pre-labour market

Cluster 4: Full-time workers (late entry, early exit)

- Not applicable
- Other unemployed
- Retired
- Stopped work due to illness
- Made redundant
- Housewife
- Part-time
- Full-time
- Pre-labour market
Cluster 5: Full-time housewives

- Not applicable
- Other unemployed
- Retired
- Stopped work due to illness
- Made redundant
- Housewife
- Part-time
- Full-time
- Pre-labour market