LIVES
Swiss National Centre of Competence in Research

FNSNF
Swiss National Science Foundation

Brzinsky-Fay, C. (2016)

Surveys, Memories and Sequences: The Role of Recall Bias and Survey Mode

in G. Ritschard & M. Studer (eds), Proceedings of the International Conference on Sequence Analysis and Related Methods, Lausanne, June 8-10, 2016, pp 335-337.

UNIVERSITÉ DE GENÈVE

FONDATION POUR L'UNIVERSITÉ DE LAUSANNE

UNIVERSITÉ DE GENÈVE
GENEVA SCHOOL OF SOCIAL SCIENCES

UNIL | Université de Lausanne
Institut des sciences sociales

Surveys, Memories and Sequences: The Role of Recall Bias and Survey Mode

Dr. Christian Brzinsky-Fay, Wissenschaftszentrum Berlin für Sozialforschung (WZB)

Extended Abstract

Sequence analysis has relatively high demands regarding data properties: the information to be analysed needs to be longitudinal, is not allowed to have any gaps, and should contain categorical information for the statuses, which is exhaustive and mutually exclusive. Such kind of information is provided by either retrospective life course surveys, or administrative register data, or panel datasets. These kinds of datasets have different advantages and disadvantages regarding sequence data. The validity of sequence information in a particular dataset depends to a large extent on the recall bias and how the survey mode moderates it.

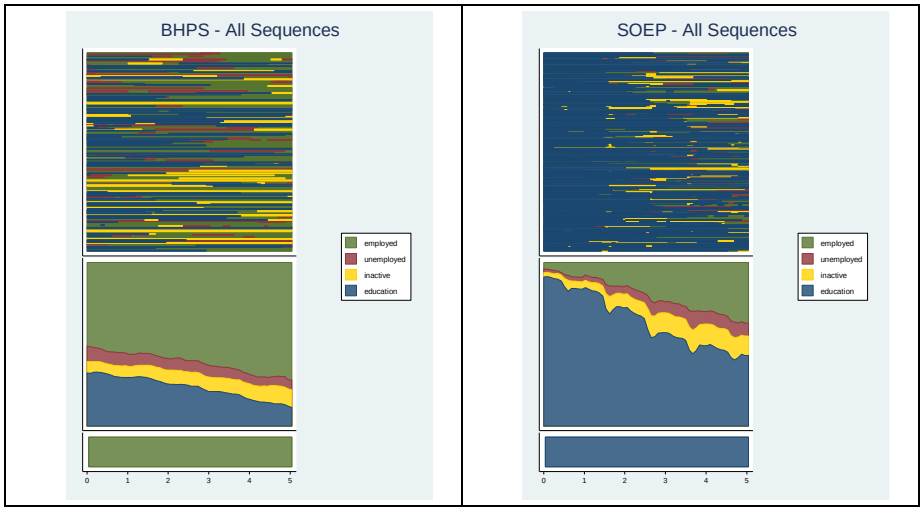
In retrospective data, people were surveyed on one time point and have to remember events of their life course, which might be decades ago. Since this circumstance, there is a large literature on recall error regarding employment careers (e.g. Dex & McCulloch, 1998; Horvath, 1982; Jacobs, 2002). However, whether there is an effect of recall bias on the properties of sequences generated from retrospective data is not researched so far. Administrative data (e.g. on employment histories) are continuously collected, even if the intervals are varying. In general, they are collected by institutions or third persons, so that recall bias doesn't play a role. Of course, administrative data have other disadvantages – such as limited number of variables available or distortion by the data generation process. However, a comparsion of administrative and retrospective data of the same individuals could help to find out, how recall bias influence sequence data.

Furthermore, panel data, where people are surveyed repeatedly, collect the longitudinal information on fixed time points, generally each year. Typical representatives are household panels, such as the German Socio-Economic Panel (SOEP) or the British Household Panel Study (BHPS). The recall bias in these datasets is comparatively low, because the distance between the event and the survey of the event remains small, i.e. people have to take information from their memory after 1.5 years on average. Although the expected recall bias should be the same between different yearly surveyed household panels, we find remarkable differences between the properties of sequences between the SOEP and the BHPS (see figure).

2

In order to assess this problem, I have created comparable samples of 16- to 18-year old individuals from both datasets and look at their employment statuses for five years using only very basic statuses, i.e. employment, unemployment, inactivity and education. First of all, there are structural differences as one would expect: In the UK, young people enter the labour market earlier than in Germany, where young people stay longer in education. This explains the higher share of employment (green) in the UK and the higher share of education (blue) in Germany. The second thing that is eye-catching, are the regular 'waves' in the SOEP data. These regularity doesn't have its origin in the real phenomenon of employment statuses, but in the survey mode applied: in the SOEP, employment careers are surveyed as calendar data, while in the BHPS there are surveyed in an episode format, where in each year there are questions regarding the beginning and the end date of an episode. In the former instance it seems that respondents simply make their crosses for every twelve months, whereas this 'wave' structure seems to be much flatter in the BHPS data. However, the survey mode seems to influence crucially the properties of the data.

The paper proposed aims at clarifying two issues: First, I want to estimate the recall bias of retrospective life course data compared to register data. Here, the data from the National Education Panel Study (NEPS, starting cohort 6) provides the unique possibility to combine the same individuals from the retrospective survey data with administrative data from the German employment service provided by the Institute for Employment Research (IAB). This allows assessing exactly the influence of recall bias on sequence characteristics. Second, with the comparison of the two datasets mentioned above (BHPS and SOEP), I would like to find out more about the effects of different survey modes on sequence characteristics, such as turbulence, episode number etc.



References

Dex, S., & McCulloch, A. (1998). The Reliability of Retrospective Unemployment History Data. Work, Employment and Society, 12(3), 497-509.

Horvath, F. W. (1982). Forgotten Unemployment: Recall Bias in Retrospective Data. Monthly Labor Review(March), 40-43.

Jacobs, S. C. (2002). Reliability and Recall of Unemployment Events Using Retrospective Data. Work, Employment and Society, 16(3), 537-548. doi: 10.1177/095001702762217489