8 Data Quality: Three Examples of Consistency Across SHARE and SHARELIFE Data

Christelle Garrouste and Omar Paccagnella

8.1 The importance of data accuracy

As shown in the second chapter, memory bias can constitute a serious problem in the analysis of retrospective data. Autobiographical memory research has shown that recalling information is an active reconstruction process that is likely to distort past experiences (Gorin & Stone, 2001; Stone et al., 2004). Recall bias has important implications for the measurement of change over time of individuals as they lead to over- or under-estimation of change. It is therefore important to ascertain the presence and extent of this bias in order to determine the quality of a dataset (Goode, 2007).

The presence of such a distortion depends on the ability and the willingness of respondents to remember past events accurately (ibid.). For instance, certain individual characteristics, such as age and gender, may play a role in the accuracy of responses (Auriat, 1991). But certain busy lifestyles or life circumstances may also play a role. The more events there are to remember, the harder it may be to remember all of them accurately. For individuals who have irregular employment patterns, it may therefore be harder to remember what state they were in at every point in time, as opposed to an individual who has been in steady employment throughout the period in question. Another important factor in this context is the length of time elapsed between the interviews and the events that need to be recalled. Research suggests that there is significant negative effect on recall accuracy, as time elapsed lengthens between interviews (Jürges, 2007; Paull, 2002). Moreover, the importance that respondents attach to the event they are being asked to recall and/or its social desirability also plays an important role. For instance, using the British Household Panel Survey data, Paull (2002) found that shorter spells of unemployment are less likely to be recalled than other types of spells. Moreover, in the health economics literature, self-reported health is often said to suffer from "justification bias" in that persons who are unemployed tend to overstate how bad their health is in order to "justify" their unemployment (Crossley & Kennedy, 2000).

There are two main methodological streams in the literature to measure recall bias. The first one consists in assessing basic memory processes in the laboratory, revealing factors influencing memory in general, like the affective valence effect (i.e. information with positive affect is more easily remembered than that associated with negative affect) and the mood congruent memory effect (i.e. current mood state facilitates the processing of material with a similar emotional valence and impairs the processing of material with the opposite valence) (Kennedy et al., 2004). This research stream has for instance revealed differences in recall accuracy between sick and healthy respondents (Ebner-Priemer et al., 2006). On the other hand, the second approach is referred to as "ecological

momentary assessment" or "experience sampling method" and defines recall bias as the difference between the multiple momentary ratings assessed at specific moments in time and the retrospective rating of the same period of time.

Usually, the main difficulty with this second approach is the availability of an objective source of data or information against which to compare survey responses. Moreover, another difficulty lies in that there is no guarantee that respondents will respond in the same way for the same event in subsequent interviews despite the fact that they are being asked to recall the exact same previous event(s) in every interview (Goode, 2007). For example, a respondent may give one answer in *t* and another answer for the same event in *t+1* (Horvath, 1982). It is therefore important to be aware of the extent to which people make 'mistakes' and whether these 'mistakes' are made randomly or not. For instance, if the same or same types of people are consistently recalling the same thing 'wrongly' in the retrospective survey and in the longitudinal panel data over time, this will introduce systematic recall bias in the data, which will be harder to deal with than if the error is committed randomly.

The structure of the SHARE survey allows us to compare responses from its retrospective data (i.e. SHARELIFE) to responses from SHARE wave 1 and SHARE wave 2 for the same individuals on several common indicators. This exceptional setting allows us to assess the quality of the sample by both ascertaining whether recall bias is present in the data and whether this bias is random or systematic. For the purpose of this chapter, we have chosen to exemplify the strategy by using three main variables referring to events contemporary to SHARE wave 1 or SHARE wave 2, namely the employment status, the presence of a cohabitating spouse and the number of children alive. These variables were retained because they allow investigating recall bias in different fields (economic, social networking, demographics), and represent common types of events tested in the literature on memory effect (e.g., Poulain et al., 1992; Goode, 2007).

Indeed, variables on labour force participation are commonly identified as a major source of recall bias when collected retrospectively. For instance, Goode (2007) investigated the employment recall bias in reported events across the five first waves of the survey on Household, Income and Labour Dynamics in Australia (HILDA). The results from the descriptive analysis and multivariate analysis show that there is systematic employment recall bias present in the HILDA data. Out of all respondents, some 30% make more than one mistake. The probability that a respondent makes any mistake is statistically significantly associated with being in full time education, the number of children, the number of jobs in the last financial year, possibly the time elapsed between interviews and the number of jobs reported in the employment calendar. Overall, the most important factor associated with the recall bias is the employment state that individuals are in at interview *t* and interview *t+1*. Further, factors associated with the probability of an individual making a mistake will change, both in magnitude and direction, depending on their exact employment state at interview *t* and interview *t+1*.

Using the German Socio-Economic Panel (GSOEP) study, Jürges (2007) compared current and 1-year retrospective data on unemployment. He found that 13% of all unemployment spells are not reported one year later, and another 7% are misreported. He also showed that the ratio of retrospective to current unemployment increased in recent years and is related to salience of unemployment measures such as the loss of life satisfaction that is associated with unemployment. This result is consistent with evidence on retrospective bias found by cognitive psychologists and survey methodologists (affective valence effect). Individuals with weak labour force attachment, such as women with children or individuals who are close to retirement, have for instance the greatest propensity to un-report unemployment retrospectively.

Mathiowetz and Duncan (1988) drew on Panel Study of Income Dynamics (PSID) validation data and compared individual respondent reports with company records. They found that two-third of spells remain unreported and that a strong negative relationship exists between the length of spell and the degree of under-reporting (i.e. the shorter the period of the unemployment spell the higher the probability of under-reporting). This result is confirmed by Manzoni et al. (2009) using the GSOEP data and comparing it to the German Life History Study. They also found that the lower transition rates reported in the life-course study can be explained by short spells recall bias.

Moreover, Elias (1997) compared unemployment rates calculated from 9-year employment biographies from the British Household Panel Survey with corresponding unemployment rates from the British Labour Force Survey. The results showed that under-reporting becomes serious if a spell dates back more than 3 years, which contradicts the result by Mathiowetz & Duncan (1988) that the length of the recall period is not significant. Elias (1997) also found that men under-report less than women. This last result about cluster differences in employment recall bias confirms earlier findings by Akerlof & Yellen (1985), comparing the US Current Population Survey and its annual supplement on work experience, the Work Experience Survey, who found differences both by gender and by age group (with men and older respondents under-reporting less than women and younger respondents).

Furthermore, the literature provides also evidence of the existence of recall bias in social network variables related to the respondent's children and their marital status. The degree of bias varies, however, a lot across countries. Comparing responses from a retrospective life history survey, the 3B-B Survey by INED-UCL, and the Belgian Population Register, Poulain et al. (1992) show that errors of reporting were found even for life vital events (such as birth, marriage, divorce, death, as well as the birth of children living or having lived in the household). In all cases, these recall biases varied significantly by gender with, for instance, only 2% of error for women vs. 7.2% for men on the report of the date of birth of children and only 1% of error for women vs. 7% for men on the report of the date of marriage. In both cases, some of these mistakes could be corrected during the joint interview. Looking at the annual U.S. National Longitudinal Survey of Work Experience (NLS) and its retrospective modules of 1978 and 1983, Peters (1988) checked for the accuracy of the changes in marital status.

Taking the “at-the-time” recordings of the panel data as the true data, Peters found a concordance between the panel and the retrospective data of only 76%. The main determinants of the recall error were found to be the time distance between interview and event (positive correlation) and the level of education (negative correlation).

Hence, given the evidence-based increased risk of recall bias in the retrospective surveys and given the length of the recall period covered by the SHARELIFE survey (i.e. at least 50 years), we aim at assessing the quality of the SHARELIFE data exemplified by investigating the extent to which SHARELIFE respondents remember three of their past events (in)correctly.

8.2 Data and descriptive results

This analysis compares SHARELIFE data with the information collected in the respondent’s first completed SHARE interview. Thus, in case of respondents having completed both wave 1 and wave 2 surveys, we compare wave 1 and SHARELIFE data. This leads to a potential sample composed by 16870 wave 1 respondents and 9875 wave 2 respondents (both based on the first internal release of the SHARLIFE data). In this section we aim at testing for recall bias by comparing contemporaneous information collected in one of the SHARE waves with retrospective information collected in SHARELIFE.

We focus our analysis on three different types of collected information, covering economic, demographic as well as social networking areas: being married and living together with a spouse; being a worker (employee or self-employed); and the number of (any) children being alive of singles (to focus on singles is necessary, because in waves 1 and 2, only one member of the couple is asked about her/his children).

These variables highlight the living condition of respondents at the time of the first completed SHARE interview and are collected through a direct question (“What is your marital status?”, “In general, which of the following best describes your current employment situation?”, “How many children do you have that are still alive?”, respectively). In SHARELIFE this information is collected through a life history approach that allows reconstructing, for instance, relationship and job spells over the whole life of the respondents.

The analysis is restricted to the subsample of SHARE respondents that report the characteristics under investigation. Therefore, recall bias in SHARELIFE is tested by comparing whether the individual history spells on marriage, employment status and children includes the year of the SHARE interview or not. An *error* (recall bias) is defined as a dummy variable equal to 1 when the event is not recorded correctly in SHARELIFE.

These restrictions on the definition of the variables of interest imply different sample sizes for each of them out of the 26745 potential observations: being married and cohabitating with the spouse (N=19165), having a job (N=8281) and the number of living children for singles (N=4857).

Table 1 reports the sample composition and the percentage of errors for each of the three variables, by wave and gender. Overall, the table reveals good results

and, as expected, the percentage of recall bias is larger when comparing SHARELIFE with wave 1 than with wave 2 (the longer the time elapsed between interviews, the greater the percentage of recall bias).

Table 8.1: Percentage of error across gender

	Wave 1					
	Sample composition			Percentage of error		
Variable	Total	Men	Women	Total	Men	Women
Being married (living together with spouse)	11941	5810	6131	1.58	1.48	1.68
Having a job	5054	2616	2438	7.7	5.31	10.25
Number of living children for singles	3177	718	2459	10.26	12.26	9.68

	Wave 2					
	Sample composition			Percentage of error		
Variable	Total	Men	Women	Total	Men	Women
Being married (living together with spouse)	7224	3516	3708	1.05	0.97	1.13
Having a job	3227	1656	1571	6.38	4.89	7.96
Number of living children for singles	1680	368	1312	8.15	10.87	7.39

While being married and living together with the spouse shows very low rates of recall biases for both waves (1.58% in wave 1 and 1.05% in wave 2), men appear to remember relationship spells slightly better than women (1.48 vs. 1.68% of errors in wave 1 and 0.97 vs. 1.13% in wave 2). However, this result does not take into account the different sample compositions by age. As highlighted in Figure 8.1, we find that errors are more equally distributed across age classes for women than for men (where the less than 50 year-olds inflate the error propensity by almost 0.2 percentage points in each wave). Surprisingly, for both men and women, recall accuracy increases with age. Moreover, the Netherlands appears to be the main contributor to the magnitude of this recall bias (without this country the overall percentage is lower than 1% in each wave).

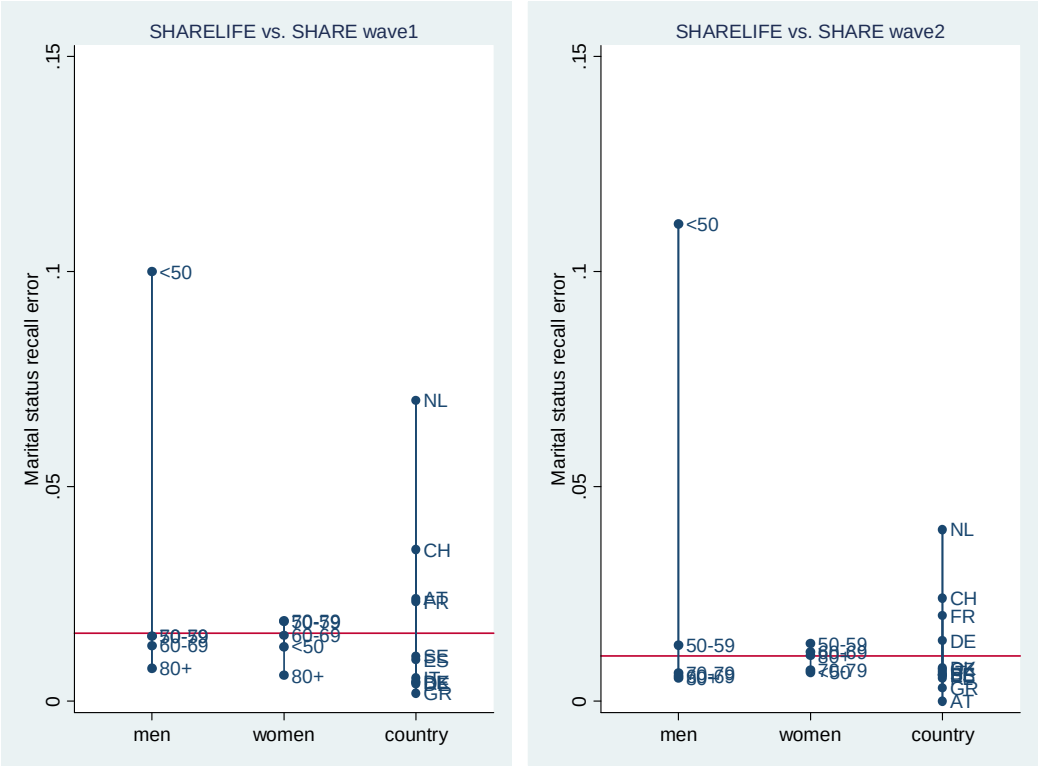

Figure 8.1: Recall bias on marital status

Recall biases for jobs are larger than the ones related to marriage (7.7% and 6.4% comparing wave 1 and wave 2 data respectively). However, these results may overestimate the true percentage of bias because of some slight differences in the definition of the employment status used by SHARE and SHARELIFE surveys. On the one hand, SHARE asks for a self-evaluation of the current regular employment status in wave 1 and 2 (question EP005_). If the respondent does not declare her/him-self employed or self-employed (including working for family business) in question EP005_ then question EP002_ is asked ("Did you do nevertheless any paid work during the last four weeks, either as an employee or self-employed, even if this was only for a few hours?). Hence, in SHARE, the focus is on capturing any paid work, regardless of the length or type of work. On the other hand, SHARELIFE collects information on every paid job that lasted *at least 6 months*. In order to reduce the potential discrepancies between surveys, in SHARE we consider a respondent as a worker only when he/she defines him/herself as a regular employed or self-employed (question EP005_). This definition however implies losing respondents who defined themselves differently from employed or self-employed (e.g. retired), but who still might have had a paid job that lasted for at least 6 months.

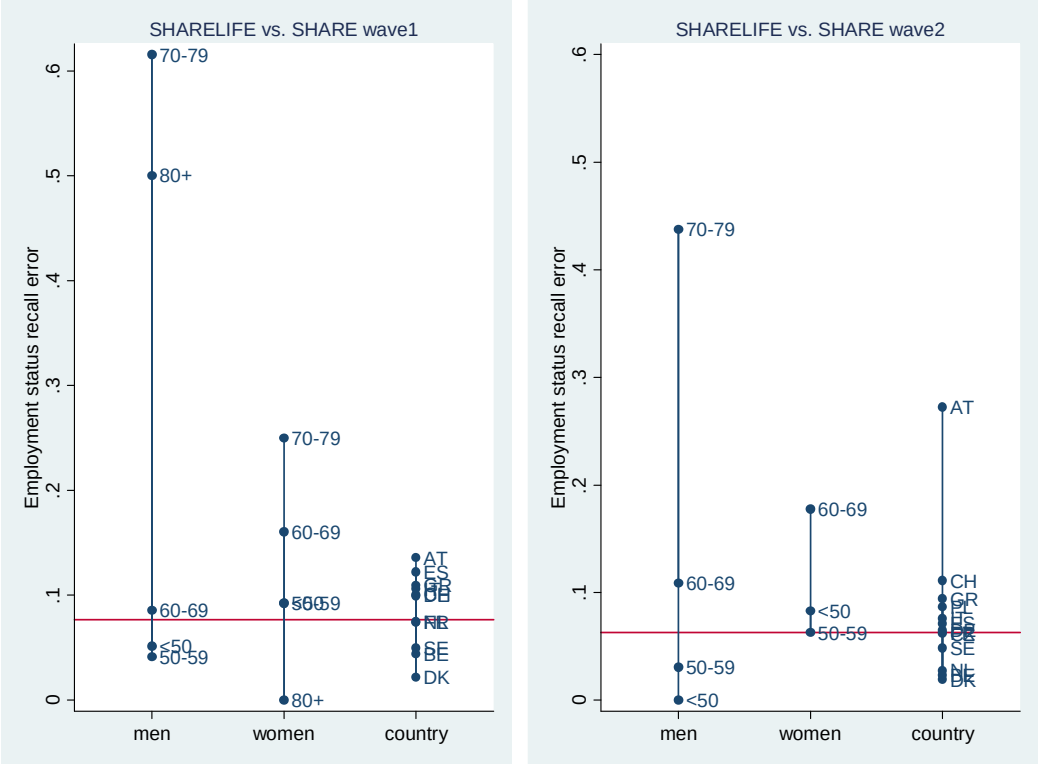

Figure 8.2: Recall bias on employment status

Table 1 confirms that the recall bias in employment is stronger for women than men (10.25% vs. 5.31% in wave 1 and 7.96% vs. 4.89% in wave 2). When looking at the age distribution of this bias by gender (Figure 8.2) we notice that it increases with age. These results support the idea of an overestimation of recall bias, since women and workers older than 65 years old are more likely to have a paid job even though retired or homemaker. When looking at the distribution of frequency of errors across countries, we see that Denmark, Belgium, the Netherlands and Sweden recall on average better this event than Mediterranean countries, Switzerland and Austria. It is also interesting to note the stability over time of these percentages for Nordic Countries.

Finally, the findings on recall bias with regard to the number of living children for singles show the worst performances. However, similarly to the job history, collection of the number of living children might suffer from some discrepancies between surveys. While in SHARE the number of children counts all natural children – fostered, adopted and step-children – and also includes any children of the spouse or partner, in SHARELIFE only information on natural and adopted children are collected, addressing these questions to each respondent. The choice of considering only those respondents who lived as a single at the time of the SHARE interview is expected to limit the effects of these potential discrepancies, but the difference in asking the question is likely to cause an upward bias (i.e. more children being mentioned in wave 1 and wave 2 vs. SHARELIFE). Indeed, when looking at the size of the error, about 75% of the differences are positive, and about one third of these are off by one child.

The largest percentage of errors is due to women, but these results do not take into account the fact that men constitute only one fifth of this sample. Figure 8.3 disentangles the sources of these errors, showing that recall biases are actually mainly due to men. An other interesting result is the lack of any age-gradient (especially when comparing SHARELIFE with wave 2): younger respondents have recall bias percentages close to the ones reported by older respondents. Moreover, this indicator shows a large cross-country variability. Spain and Austria (not reported for wave 2 because of a small sample size) are again the main contributors to the magnitude of this recall bias. Italy performs best when comparing SHARELIFE with wave 1 and worst when comparing with wave 2, while the opposite relationship applies for Germany and Greece.

In view of these results, it would have been interesting to further investigate the role played by individual characteristics in the probability of these recall errors through the estimation of some multivariate models. However, the low percentage of errors, particularly for being married, prevents us from doing such analyses. Instead, some logistic regressions on the accuracy of the number of living children for singles and of the employment status have been estimated, but their results provide very little information since almost all variability is captured by country dummies. Multivariate regression analysis was applied to analyze the probability of occurrence of an error, estimating a logistic regression model on two of our focus variables (i.e. employment status and number of living children) and defining a dummy variable equal to 1 when the year of the event is recorded in SHARELIFE differently than reported at the time of occurrence, i.e. in SHARE wave 1 or SHARE wave 2.

With regard to the accuracy of the number of living children for singles, we find that being a man and having several children increases the propensity for recall errors. With regard to the accuracy of the employment status, we find that the propensity of recall errors in relation to wave 1 is higher for low educated young women with a fair or poor health status at the time of the SHARE interview, having been married several times and with very few employment spells. The same inferences apply to the recall bias in relation to wave 2, except for the fact that the number of marriages is not significant anymore and is replaced by the number of children (positively and significantly correlated to recall bias). In none of the logistic regressions did the socio-economic status of the respondent, the mental health nor the memory capacity of the respondents explain variability in recall errors.

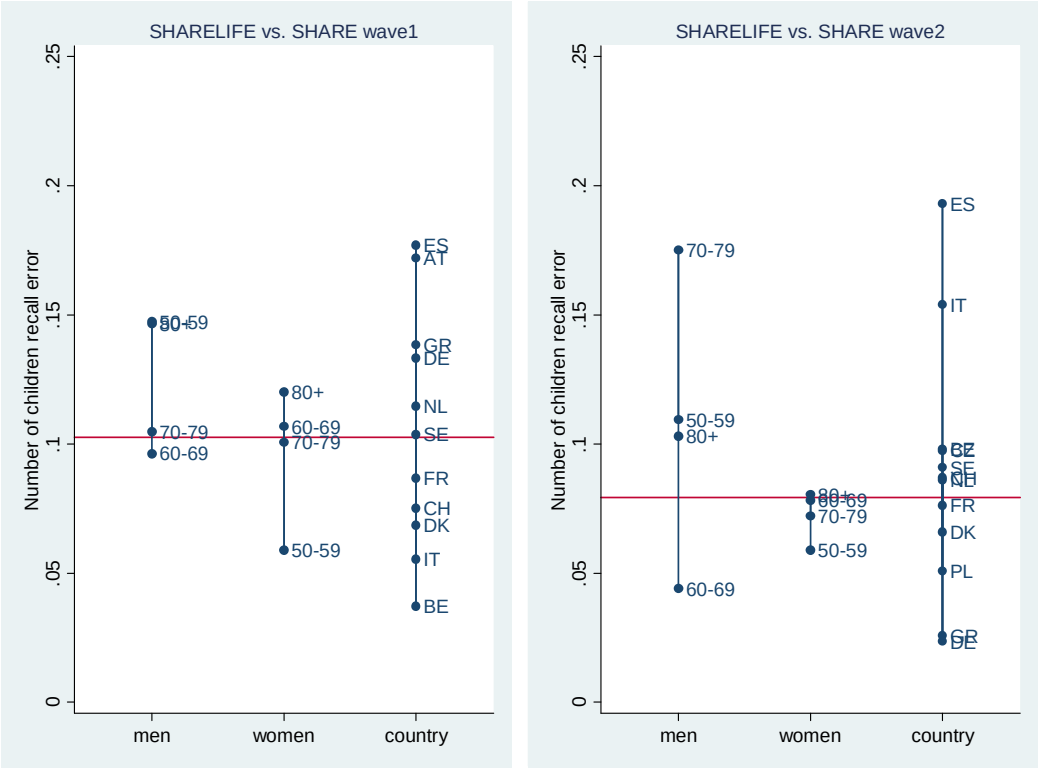

Figure 8.3: Recall bias on number of children of singles

8.3 Final remarks

The reliability of any retrospective survey is based on the accuracy of this collected information. Several studies have tested the presence of recall biases in various contexts of these surveys, such as unemployment spells (Akerlof and Yellen, 1985; Jürges, 2007) or health conditions (Lungenhausen et al., 2007).

The goal of this chapter is to provide a first sight at the quality of SHARELIFE data, investigating the main determinants of recall biases in this retrospective survey. To this aim, information collected independently and in different ways from SHARE and SHARELIFE on some personal events was linked and consistencies between the two surveys were evaluated. In particular, the focus was on an event in each of the following categories: demographics (being married and living together with the spouse), economic status (being employed or self-employed) and social network (the number of living children for singles).

The main result of this analysis is that SHARELIFE data is overall strongly consistent with the information reported at the time of occurrence of the events (with less than 10% recall errors over all events). When investigating further the distribution of the recall bias, we find that gender, age and family status are the main determinants in the recall capacity, which confirm the main findings underlined by the literature. While we find that busy lifestyles or life circumstances, e.g. number of marriages, appear to play a significant role (the more events there are to remember, the harder it may be to remember all of them

accurately), educational attainment is an important determinant only in the recall accuracy of employment status.

Recalling the number of living children shows the largest biases, especially among men. Overall, the data reveals large cross-country variability in recall bias, and, within some countries, even large cross-wave variability. Spain and Austria are the countries with the largest and most systematic biases. Nevertheless, our findings should be used with caution, particularly because in some cases the information collected in SHARE and SHARELIFE is not exactly the same. Further information needs to be taken into account in the analysis to better isolate the main sources of recall errors, particularly some country-specific characteristics like interviewer effects or information on the time length between the two interviews.

References

Akerlof, G. & Yellen, J. (1985). Unemployment through the Filter of Memory. *The Quarterly Journal of Economics*, *100*(3), 747-773.

Auriat, N. (1991). Who forgets? An analysis of memory effects in a retrospective survey on migration history. *European Journal of Population, 7*, 311-342.

Cremin, I., Mushati, P., Hallett, T. et al. (2009). Measuring trends in age at first sex and age at marriage in Manicaland, Zimbabwe. *Sexually Transmitted Infections*, *85*, 34-40.

Crossley, T.F. & Kennedy, S. (2000). The Stability of Self Assessed Health Status. CEPR Discussion Papers 421. Centre for Economic Policy Research, Research School of Economics, Australian National University.

Ebner-Priemer, U.W., Kuo, J., Welch, S.S., Thielgen, T. et al. (2006). A Valence-Dependent Group-Specific Recall Bias of Retrospective Self-Reports. A Study of Borderline Personality Disorder in Everyday Life. The *Journal of Nervous and Mental Disease*, *194*(10), 774-779.

Elias, P. (1997). Who forgot they were unemployed? Working Paper 97-19. Institute for Social and Economic Research, Colchester.

Goode, A. (2007). Recall Bias in Reported Events. HILDA, Waves 1-5. HILDA project Discussion Paper Series 2/07 (June). University of Melbourne.

Gorin, A.A. & Stone, A.A. (2001). Recall biases and cognitive errors in retrospective self-reports: a call for momentary assessment. In: A. Baum, T.A. Revenson and J. Singer, Editors, *Handbook of Health Psychology*. London: Lawrence Erlbaum Associates, pp. 405–413.

Horvath, F.W. (1982). Forgotten unemployment: recall bias in retrospective data. *Monthly Labor Review, 105*, 40-43.

Jürges, H. (2007). Unemployment, Life Satisfaction, and Retrospective Error. *Journal of the Royal Statistical Society A*, *170*(1), 43-61.

Kennedy, Q., Mather, M. & Carstensen, L.L. (2004). The role of Motivation in the Age-Related Positivity Effect in Autobiographical Memory. *Psychological Science, 15*, 208-214.

Manzoni, A., Luijkx, R. & Muffels, R. (2009). Explaining differences in labour market transitions between panel and life-course data in West-Germany. Quality & Quantity. doi:10.1007/s11135-009-9292-1.

Mathiowetz, N.A. & Duncan, G.J. (1988). Out of work, out of mind: response errors in retrospective reports of unemployment. *Journal of Business & Economic Statistics, 6*, 221-229.

Paull, G. (2002). Biases in the reporting of labour market dynamics. Working Paper 02/10. Institute for Fiscal Studies, London.

Peters, E. (1988). Retrospective Versus Panel Data in Analyzing Lifecycle Events. *The Journal of Human Resources, 23*(4), 488-513.

Poulain, M., Riandey, B. & Firdion, J-M. (1992). Data from a Life History Survey and from the Belgian Population Register: A Comparison. *Population: An English Selection, 4*, 77-96.

Schröder, M. & Börsch-Supan, A. (2008).Retrospective Data Collection in Europe. MEA Working Paper 172-2008. Mannheim Research Institute for the Economics of Ageing, Mannheim.

Stone, A.A., Broderick, J.E., Shiffman, S.S. & Schwartz, J.E. (2004). Understanding recall of weekly pain from a momentary assessment perspective: absolute agreement, between- and within-person consistency, and judged change in weekly pain. *Pain, 23*, 354-362.