

RESEARCH NOTE

Assessing recall of early life circumstances: evidence from the National Child Development Study

Matt Brown

Centre for Longitudinal Studies, Institute of Education, London, UK

M.Brown@ioe.ac.uk

(Received January 2013 Revised September 2013)

DOI: <http://dx.doi.org/10.14301/llcs.v5i1.232>

Abstract

Cross-sectional studies and longitudinal studies alike make regular use of retrospective questions about childhood circumstances. However, little is known about the accuracy with which adults can recall this kind of information. This paper seeks to address this topic by comparing retrospective reports of the number of people and the number of rooms in one's household at age 11 provided by 50 year old members of a birth cohort study, with responses provided contemporaneously by their parents. The paper demonstrates encouraging levels of consistency between retrospective and contemporaneous reports. By examining reports of number of rooms provided by parents living at the same address in two earlier sweeps of the study (at ages 7 and 11), the paper shows that responses to contemporaneous questions may also be inconsistent, suggesting that retrospective questions of this nature may not be hugely less reliable. A retrospective measure of overcrowding at age 11 is derived using the two variables, and compared with a contemporaneous measure. The two measures lead to the same estimate of the extent of overcrowding, but when used in a model examining the odds of experiencing lung problems as an adult, the two measures behave differently. The paper also demonstrates that there are particular groups who are more likely to provide inconsistent responses than others. Around one in five participants were identified as having particularly poor recall, and the likelihood of being in this group was considerably higher amongst those whose childhood circumstances were more complex. The paper also finds that performance in a delayed memory assessment at age 50 was associated with better recall of childhood circumstances.

1 Introduction

Childhood circumstances are of vital importance to understanding adult outcomes. Birth cohort studies which collect details of childhood experiences contemporaneously, play a crucial role in understanding the processes by which childhood experiences affect later life. Cross-sectional studies, or longitudinal studies beginning in adulthood, can only gather this information by using retrospective questions. For example, the English Longitudinal Study of Ageing (ELSA) (Banks, Nazroo & Steptoe, 2012) and the Survey of Health,

Ageing and Retirement in Europe (SHARE) (Börsch-Supan, Hank, Jürges & Schröder, 2008), two longitudinal studies following adults aged 50 and over, have both recently conducted retrospective life history questionnaires which gathered information about life experiences prior to joining the studies (Ward, Medina, Mo & Cox, 2009, Börsch-Süpan & Schröder, 2011).

It is recognised that use of retrospective questions will inevitably lead to a degree of recall bias, caused by differences in the accuracy with

which study members can recall experiences or events from their past. Within the field of survey methodology there is a long tradition of attempting to understand the factors affecting a respondent's ability to accurately recall past events and circumstances (e.g. Gray, 1955; Cannell, 1977; Sudman & Bradburn, 1973).

Past research has generally concluded that longer reference periods, defined as the period over which one is asked to recall events or circumstances, will typically result in reduced accuracy (Sudman and Bradburn, 1973; Cohen, Erickson & Powell, 1983; Bound, Brown & Mathiowetz, 2000). Research has also shown that respondents are typically able to recall events of greater salience or psychological importance with greater accuracy (Linton, 1982; Loftus, 1979; Mathiowetz & Duncan, 1988). Sudman and Bradburn hypothesised that events of greater salience leave a stronger memory trace which reduces the effort required to recall information about such events from memory. Events of great salience to one respondent may be of less importance to another, meaning salience can be difficult to evaluate, but Mathiowetz and Duncan (1988) offer a useful definition, defining salient events as those which 'evoke emotion at the time of occurrence, mark a transition point, have economic or social costs or benefits, or have continuing consequences after the event'. 'Telescoping' is a further source of bias, whereby events that happened in the past are recalled as having occurred more recently (forward telescoping) or earlier (backward telescoping) than they actually did (Bradburn, Huttenlocher & Hedges, 1994). Related to this is the 'accessibility principle' (Brown, Rips & Shevell, 1985), which refers to the phenomenon in which events most easily recalled are assumed to have occurred more recently (and perhaps more frequently), and conversely that events recalled with difficulty are assumed to have happened longer ago. A final source of bias is the difficulty of the reporting task; classical interference theory (Crowder, 1976) contends that individuals who have experienced multiple similar events will be less likely to recall any particular one. In a study seeking to evaluate ability to recall periods of unemployment over a two year period, Mathiowetz and Duncan (1988) found that the difficulty of the reporting task, measured by the number of spells of

unemployment experienced, had a bigger impact on the accuracy with which dates associated with a particular spell of employment could be recalled, than the time which had elapsed since that period.

To date, the number of studies attempting to assess the accuracy with which adults can recall information about their childhood have been very limited. The few studies that have sought to evaluate ability to recall childhood circumstances typically rely on one of three strategies. The first is to employ some form of test-retest, whereby retrospective reports are collected on two occasions and compared for consistency. For example, Haas (2007) compared retrospective reports of childhood health (rated from poor to excellent) provided on two occasions, two years apart, by panel members of two longitudinal studies in the United States (U.S.), and concluded that retrospective measures of childhood health were reasonably reliable.

A second strategy is to assess whether retrospective reports on several measures are logically consistent. For example, Elo (1998) examined the extent to which retrospective reports of overall childhood health provided by members of the Health and Retirement Study in the U.S., a longitudinal study of ageing, were correlated with responses to other questions about childhood health limitations, and found that that the self-reported health measure was highly consistent with these other measures.

The third strategy involves the comparison of retrospective reports with some form of administrative data, often viewed as the "gold" standard for validation, although it must of course be acknowledged that this approach is only feasible when appropriate records are available. Auriat (1993) compared retrospective reports of all changes of address since the age of 14, which were collected in a Belgian survey of couples aged between 41 and 57, with the Belgian National Population Register. When focusing on the first three moves after marriage (if applicable), Auriat found that around 30 per cent of couples misreported the dates of moves by more than three months.

This paper makes use of a rare opportunity to compare retrospective survey responses with survey data collected contemporaneously. We examine responses to two questions about childhood experiences that were answered by 50

year old members of the National Child Development Study (Power & Elliott, 2006), a British birth cohort study following the lives of all those born in one week in 1958. The questions re-collected information about the number of people lived with and the number of rooms in the household at age 11, which was first provided contemporaneously by study members' parents in 1969. The accuracy with which individuals can recall these details some 40 years later was assessed by comparing the data recorded at age 50 with that recorded at age 11. Such evidence can be used to inform decisions about the validity of including these kinds of questions in other surveys. The paper makes a significant contribution to the literature on recall bias by exploiting the rich life-history information collected from and about study members, to investigate the factors that might cause variation in recall of childhood circumstances. The impact of performance in a delayed memory test, conducted at age 50, was also explored.

2 Data and methods

The National Child Development Study (NCDS) began in 1958 and has followed the lives of around 17,000 infants born in one week in March ever since. Over the years there has inevitably been some attrition from lost contact, refusals, emigrations and deaths, but response rates remain high. Just under 9,800 respondents participated in the age 50 follow-up, which took place in 2008/9.

Two retrospective questions asking about the number of people and rooms in the household at age 11 were asked of a random subset of approximately 25% of respondents ($n=2,498$) in the self-completion section at the end of interview. The questions were specifically added to facilitate the analysis of recall bias and were selected on the basis that they were widely used basic socio-demographic variables. When taken together they can provide a measure of over-crowding, a commonly used measure of living standards. Furthermore, the two questions were both included in the ELSA and SHARE life history questionnaires, so an assessment of the accuracy with which individuals can recall these precise measures was considered particularly useful. Finally, both questions elicited straightforward numerical responses which could be directly compared with responses provided by their parents (typically mothers) to the same questions in 1969, when study members were aged 11.

The retrospective number of people questions was: *"For the next two questions we would like you to think back to your childhood, specifically to the time when you were 11-years-old. When you were 11, including yourself, how many people normally lived in your household? Please exclude any children or others who only lived at home for short periods such as school holidays"*. In 1969 the contemporaneous information was collected from the parent by an interviewer who completed a quasi-household grid. The contemporaneous question asked *"Who normally lives in the Study Child's household?"* and the same instruction about excluding those who only lived at home for short periods was included. An additional definition of the household was used in the pre-ambule to this question that was not included in the retrospective question which stated that *'a household comprises the group of persons living together partaking of meals prepared together and benefitting from a common housekeeping'*. The interviewer collected the details of each household member (name, age and relationship to study member) and the total number of individuals was recorded.

The retrospective number of rooms *"How many rooms did your accommodation have? Please exclude bathrooms, kitchens or sculleries unless they were used as a living room."* The contemporaneous question to the parents was exactly the same, but was framed in the present tense.

Both questions include fairly detailed instructions about people or rooms which should be included or excluded. The number of people question mentions excluding those who only lived at home for 'short periods' but as no precise definition of a 'short period' was provided, it was potentially open to multiple interpretations. The retrospective question included this instruction within the question text which, given that it was administered via self-completion, should have been read by all respondents. At age 11 this instruction was provided on the form the interviewer used to complete the household grid and interviewers may therefore have varied in the extent they made this instruction clear to respondents. Similarly the retrospective question included the instruction about rooms to exclude within the text of the question, whereas the contemporaneous measure included the instruction as supplementary information for interviewers, and so it may not always have been read out.

Recall error may therefore not be the only source of any inconsistency between the retrospective and contemporaneous questions. The differential way in which supplementary information was provided to respondents may also have had an effect.

Survey questions asking respondents to report the number of rooms in their household are always difficult as they inevitably involve the use of detailed instructions which are open to differential interpretation. There may have been rooms in the house that either the parent or the study member were unsure about whether to include. There is, perhaps, therefore a question as to whether the contemporaneous measure of number of rooms should necessarily be treated as the most reliable measure or 'gold-standard'. Using data from earlier sweeps, the paper will examine this question more closely.

Given that the contemporaneous measure of number of people was collected using a quasi-household grid, it seems reasonable to assume that, despite the potential problem discussed above, the contemporaneous measure will in the vast majority of cases be accurate.

3 Hypotheses

Specific hypotheses are as follows:

It is anticipated that both the number of people lived with and the number of rooms in one's household will be of reasonably high salience to the respondent, and therefore, that consistency between the retrospective and contemporaneous measures will be high.

H1: Consistency between retrospective and contemporaneous measures will be high.

However, it is envisaged that the people one was living with in childhood will have greater salience to the respondent than the rooms in one's household, so we expect to see greater consistency between retrospective and contemporaneous measures for number of people. Additionally, the fact that the contemporaneous measure of number of people is assumed to be more reliable than the measure of number of rooms is also thought likely to contribute to a higher level of consistency.

H2: Consistency between retrospective and contemporaneous measures will be greater for number of people than number of rooms.

By identifying a subset of cases living at the same address when surveyed at ages 7 and 11 it is possible to assess the reliability of the age 11

measure of number of rooms, by comparing the consistency of responses to the number of rooms question at these two time points. It is anticipated that the level of consistency between these responses will be greater than the level of consistency between the study members' retrospective responses and their parents' contemporaneous responses.

H3: For those living in the same address at age 7 and 11 years, it is anticipated that consistency of parental responses on the number of rooms at these two time points will be higher than between the study members' retrospective response and their parent's contemporaneous response.

Given that consistency between retrospective measures and contemporaneous measures is expected to be high, it is anticipated that analyses will not be affected by substituting an overcrowding measure based on retrospective responses, with one based on contemporaneous responses.

H4: Analyses will not be affected by the substitution of a retrospective measure of overcrowding for a contemporaneous measure.

It seems likely that those living in smaller households in terms of people and rooms would have an easier task than those living in larger households or houses where one could overlook a particular individual or room.

H5: Consistency between retrospective and contemporaneous measures will be lower for those in larger households.

Similarly, it seems likely that if one's childhood circumstances had been stable in terms of family and housing circumstances, then recalling details about a particular period (i.e. age 11) would be more straightforward than recalling details from a childhood involving multiple changes of household composition or address.

H6: Consistency between retrospective and contemporaneous measures will be lower for those who moved home most and whose family circumstances changed more frequently.

Retrieval of information from memory is a key stage of the cognitive process required to respond to any form of survey question (Tourangeau, 1984). It is reasonable therefore to assume that one factor associated with being unable to provide accurate reports of childhood circumstances would be poor memory. The age 50 survey included an assessment of memory (Brown, Elliott, Hancock, Shepherd & Dodgeon, 2010) which took the form of

a word-list recall task, where study members were read ten simple words and then asked to recall as many as they could, both immediately and after about five minutes. The nature of this task is very different to recalling information from some 40 years previously, but it is nevertheless hypothesised that those with better ‘delayed memory’ as measured at 50 will be better at storing and retrieving information, and will therefore recall childhood circumstances with greater accuracy.

H7: Consistency between retrospective and contemporaneous measures will be higher for those with better ‘delayed’ memory as measured at 50.

4 Results

Section 4.1 examines the level of consistency between the retrospective measures and the contemporaneous measures (referred to as R-measures and C-measures) in terms of both the proportion providing consistent responses and the magnitude of difference in the case of any discrepancy. The reliability of the contemporaneous measure of number of rooms is also

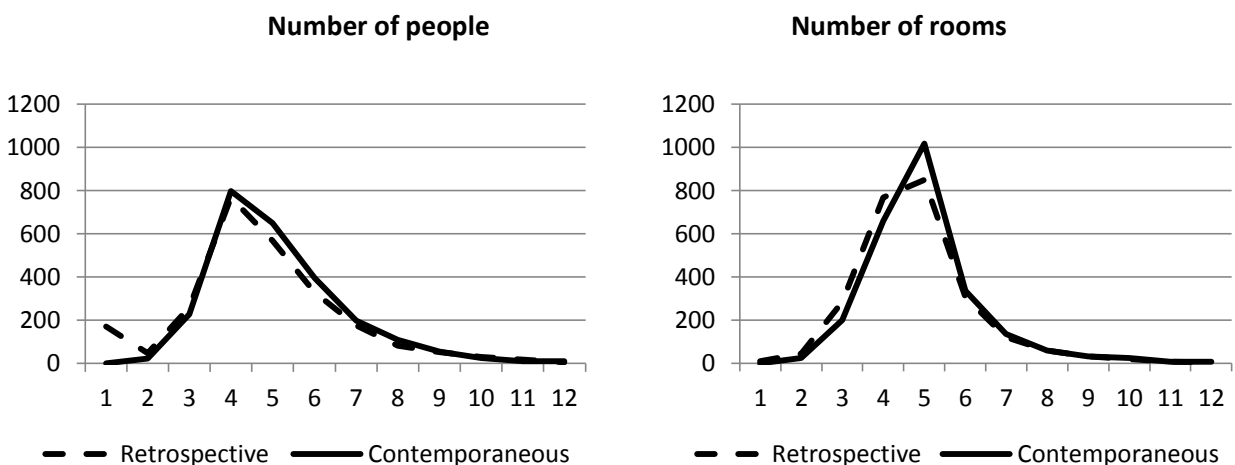
explored. As noted above, the two measures are often used in combination to create a single measure of overcrowding; an overcrowding variable based on the R responses is derived and compared with a measure based on the C responses, in order to identify the extent to which individuals could potentially be misclassified if the R measure were to be used. The extent to which any mis-classification could impact on substantive findings is also examined by running logistic regression models to examine the associations between the R and C measures of overcrowding and adult respiratory problems.

Section 4.2 explores the impact of the complexity of one's childhood circumstances and section 4.3 examines impact of performance in the delayed memory task.

4.1. Consistency between retrospective and contemporaneous measures

Figure 1 compares the distribution of R and C responses of number of people and number of rooms at age 11. Reassuringly it is immediately evident that the distributions are similar.

Figure 1. Number of people and number of rooms – comparison of distributions of R-measures and C-measures



*Base: All in random sub-set (n=2,498)

** A small number of respondents reported more than 12 people / rooms on both the R and C measures. These responses are not shown above.

R-measures and C-measures were highly correlated for both number of people ($r=0.69$, $p<0.001$) and number of rooms ($r=0.68$, $p<0.001$), which provides initial support for Hypothesis 1, that consistency between the R and C responses will be high.

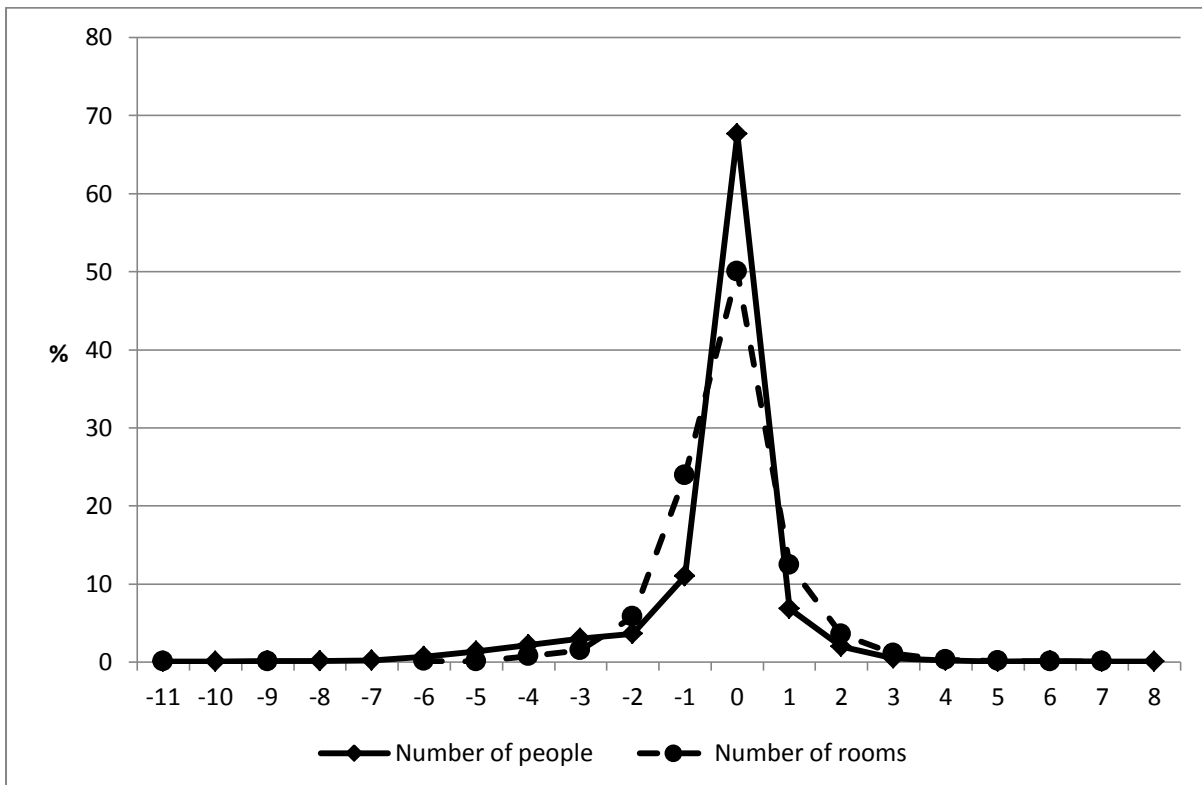
Despite the similarity of the distributions, the mean R responses were lower on both measures than the mean C responses, and two sample paired t-tests show the differences are both highly significant ($p<0.001$). The mean C response to number of people was 5.12 ($SD = 1.65$) compared with a mean R response of 4.74 ($SD = 1.95$). The mean C and R responses to number of rooms were 5.01 ($SD= 1.45$) and 4.81 ($SD= 1.46$) respectively.

One clear difference on the number of people measure is that seven per cent retrospectively reported that at age 11 they lived in one-person households (i.e. lived alone) whereas there were no C reports of living alone. As noted, the retrospective questions were included in the self-completion section at the end of the questionnaire. It is possible that respondents could have entered an answer of 'one' as a way of skipping the question, but it seems likely that if they had done this to the number of

people question they would also have done so on the number of rooms question, but that was not the case. It seems clear therefore that these individuals had misunderstood the question in some way. A series of four cognitive assessments were completed as part of the age 50 survey (Brown & Dodgeon, 2010) and those who retrospectively reported living in one person households achieved significantly lower scores on all assessments than those who did notⁱ, which could be indicative of a greater propensity to misunderstand although no evidence of any systematic misunderstanding was foundⁱⁱ.

The level of agreement between the C and R measures are shown in Figure 2. For number of people, the measures were consistent for just over two thirds of individuals (68 per cent), whilst for number of rooms the figure was 50 per cent, thus providing support for Hypothesis 2, that consistency will be greater for number of rooms. On both measures, retrospective under-reporting was around twice as common as over-reporting (23 per cent compared with 10 per cent for people, and 32 per cent compared with 18 per cent for rooms).

Figure 2. Magnitude of inconsistencies - difference between C and R responses



*Base: All in random sub-set (n=2,498)

** The differences here are calculated by subtracting responses to the C-measures from responses to the R-measures.

On both measures it is encouraging that where R and C measures did differ, the magnitude of the difference was typically small; it was most common for R and C responses to differ by one. Cases where R responses differed from C responses by more than 3 people or rooms were rare (9% for number of people and 4% for number of rooms).

Table 1 shows that just over a third of cohort members (36 per cent) provided consistent R and C responses on both measures, around half (47%) provided consistent responses on one of the two, leaving less than one in five (18 per cent) providing inconsistent responses on both. This final group

could potentially represent the group with the poorest recall ability.

More than half (55%) of those providing inconsistent responses to the number of people also provided inconsistent responses to the number of rooms. However, amongst those providing inconsistent responses to the number of rooms, only a third (36%) also provided inconsistent responses to the number of people. This perhaps suggests that an inability to 'correctly' recall the number of people lived with at 11, is a better indicator of overall poor recall than an inability to recall the number of rooms in one's house.

Table 1. Consistency of R and C measures of number of rooms vs number of people

		Number of rooms		
		R <or> C	R=C	TOTAL
Number of people	R <or> C	446 (17.9%)	362 (14.5%)	808 (32.3%)
	R=C	802 (32.1%)	888 (35.5%)	1690 (67.7%)
	Total	1248 (50%)	1250 (50%)	2498 (100.0%)

As noted above recall bias may not be the only source of error in measurement. Both questions included detailed instructions about which people and rooms should be included or excluded, and in the C measures in particular, these were not necessarily provided to respondents in a uniform way. It is therefore possible that a number of C responses may have included people or rooms which should have been excluded, which could explain at least some of the apparent under-reporting in the R responses.

Collecting information about number of rooms using survey questions is always difficult as will inevitably involve the use of detailed instructions which are open to differential interpretation. Both parents and study members may have had difficulty deciding whether to include particular rooms in their responses, which raises the question as to whether the C number of rooms response is necessarily more reliable than the R response. Using data from earlier sweeps, it is possible to examine this matter more closely. The 'number of rooms' question put to parents in the age 11 survey was also asked four years earlier (with exactly the same

wording) in the age 7 survey. There were 619 cases (25 per cent) where it was established that at age 11 they were living at the same address as they had been at the time of the age 7 survey. Amongst these cases, seven in ten (70 per cent) provided the same response on both occasions, and the Pearson's correlation coefficient between number of rooms reported at the two time points was 0.81. Although this finding confirms Hypothesis 3, that consistency between responses at 7 and 11 will be higher than the level of consistency between retrospective responses at 50 and parental responses at 11, it is illuminating to note that the difference is not huge, which suggests that the responses collected by a retrospective question may not be significantly less reliable than those from a contemporaneous question.

Additionally, C responses were collected from parents on one particular date during the year when study members were aged 11. In contrast, the two R questions asked respondents to think back to the "time you were 11-years-old". A proportion of the R responses which seemed inconsistent might therefore have been provided in

respect of a different period within that same year, when a change in housing situation or household composition had occurred. A greater degree of specificity within the wording of the question may have reduced the amount of inconsistency.

It must also be acknowledged that some of the more radical disparities between R and C responses may have resulted from data entry errors at one or other time point. It seems unlikely that someone whose parent reported living in a one-room house would genuinely recall living in an eight-room house as a child.

On the basis of these initial comparisons it is concluded that, at least on these particular measures, an individual in adulthood is typically able to recall childhood circumstances with a level of accuracy that will be useful to most researchers. Researchers requiring very precise measures would need to approach these retrospective measures with a degree of caution. However researchers who are content with classifying individuals into groups (e.g. large, medium and small households) could fairly confidently identify these groups. There remains a small percentage of respondents whose R and C responses are dramatically different. Such individuals would likely be mis-classified by a researcher relying solely on retrospective measures.

As mentioned previously, a commonly used practical application of measures of number of people and rooms is to use them in combination to define a measure of overcrowding. Many researchers have looked at the impact of overcrowding in childhood on subsequent adult outcomes. In one example, Marsh, Gordon, Pantazis & Heslop (1999) used NCDS data to examine the impact of poor housing during childhood (including overcrowding) on adult health, and demonstrated a link between overcrowding in childhood and respiratory problems in adulthood.

A way of examining the potential impact of the typically minor inconsistencies between R and C responses is to derive a childhood overcrowding variable using the R responses, and then to derive a second measure based on C responses as a validity check. A commonly used definition of overcrowding, and that used by Marsh and colleagues, is there being more than one person per room in a household. Applying this definition to the R responses gives an estimate that 37% lived in overcrowded conditions at 11, and the C estimate is almost exactly the same (38%). However, Table 2 shows that the R and C measures do not necessarily result in the same cases being classified as living in overcrowded conditions.

Table 2. Overcrowding at age 11 as measured by C and R measures of number of people and number of rooms in household

		Contemporaneous			
		Cut off = 1+ per room		Cut off = 1.5+ per room	
		Not overcrowded	Overcrowded	Not overcrowded	Overcrowded
Retrospective	Not overcrowded	1322 (52.2%)	250 (10.0%)	2073 (83.0%)	140 (5.6%)
	Overcrowded	224 (9.0%)	702 (28.1%)	115 (4.6%)	310 (12.4%)

Of the utmost importance is establishing whether any misclassification that might occur as result of using the R measure of overcrowding has an impact on analyses that might make use of the variable. Table 3 compares the results of logistic regression models which use the R and C overcrowding measures and a similar range of controlsⁱⁱⁱ to those used by Marsh and colleagues, to assess the impact

of childhood overcrowding on the odds of reporting asthma or wheezy bronchitis at age 50. Model 1 uses the one or more persons per room definition of overcrowding. It is concerning to observe that using the C measure of overcrowding would lead us to conclude that the odds of suffering from asthma in adulthood are significantly increased by experiencing childhood overcrowding, whereas

using the R measure would suggest there is no significant association. However, if we use a 1.5 person per room threshold to define overcrowding, then the C measure would show no significant association with adult respiratory problems whereas the R measure suggests that overcrowding in childhood is associated with lower odds of adult

lung problems. We therefore find no support for Hypothesis 4, that substituting a retrospective measure of overcrowding for a contemporaneous measure will not affect analyses. This somewhat puzzling finding must in fact lead us to conclude both the R and C measures of overcrowding must be treated with caution.

Table 3. Odds of reporting asthma or wheezy bronchitis at age 50 using R and C measure of overcrowding

		R measure of overcrowding		C measure of overcrowding	
		Exp(B) (Odds)	Sig.	Exp(B) (Odds)	Sig.
Model 1	Overcrowding at age 11(1+ people per room)	0.87	0.40	1.58	0.01***
Model 2	Overcrowding at age 11 (1.5+ people per room)	0.56	0.04**	1.01	0.96

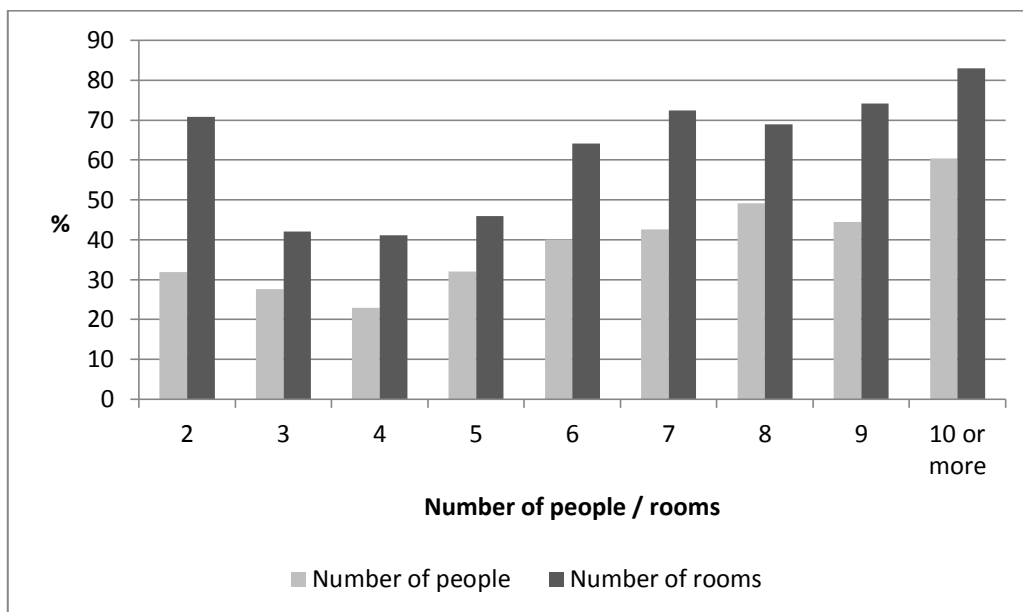
*Included in analysis (n=2,047)

4.2. Is an inability to provide accurate reports of childhood circumstances the result of the complexity of the task?

Figure 5 shows how inconsistency varied by the number of people and rooms reported at age 11. For number of people, the lowest level of inconsistency was found among those where the

parent's C response was 4 (mainly 'nuclear' families), where inconsistent responses were provided by just under a quarter (23 per cent). Inconsistent responses increased as C responses increased, so that in the largest households (10 or more), six in ten (60 per cent) R responses were inconsistent. For number of rooms the pattern was very similar.

Figure 5. Inconsistent retrospective responses by number of people / rooms in household at age 11



*Base: All in random sub-set (n=2,498)

Figure 6 shows how inconsistent R responses on the 'number of rooms' varied by the number of household moves experienced by age 11. Those

who moved more, were most likely to provide an inconsistent R response.

Figure 6. Inconsistent retrospective 'number of rooms' responses by number of moves by age 11



*Base: All in random sub-set who reported number of moves (n=2,466)

Table 4 shows that the one in five respondents who did not provide a consistent R response to either measure, lived in larger homes (in terms of both rooms and people) than those who provided either one or two consistent responses. Comparing the means with t-tests showed that the magnitudes of all the differences were highly significant ($p < 0.001$). The differences between the three groups in terms of number of moves was in the

anticipated direction, and the difference between those who provided inconsistent R and C responses to both measures, and those who provided two consistent R and C responses, was marginally significant ($p = 0.1$).

These analyses provide support for Hypotheses 5 and 6, that consistency between R and C responses will be greater for those who lived in smaller households and had more stable childhoods.

Table 4. Number of people, number of rooms and number of moves by inconsistent R and C responses.

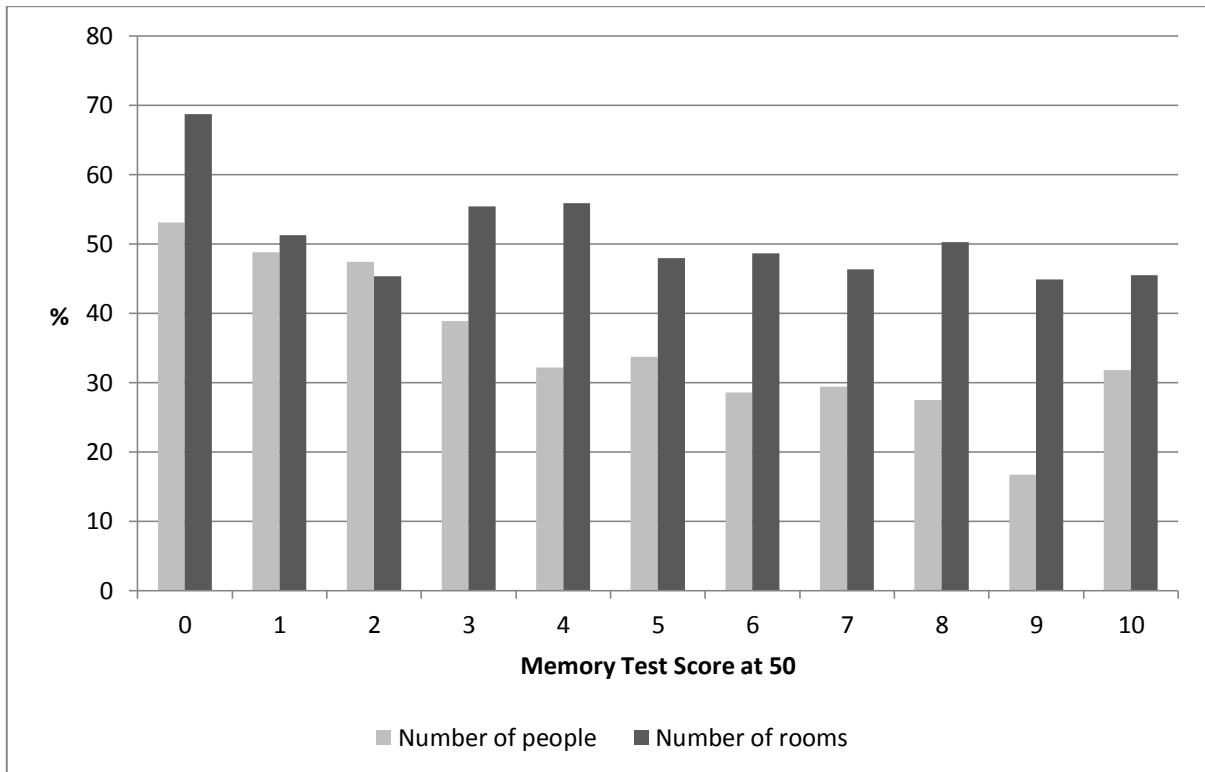
	Number of people in household (C response)		Number of rooms in household (C response)		Number of household moves by age 11	
	Mean	SD	Mean	SD	Mean	SD
R and C responses inconsistent on both measures	5.58	2.00	5.43	1.95	1.67	1.95
R and C responses inconsistent on one measure	5.12	1.6	5.09	1.43	1.64	1.71
R and C responses consistent on both measures	4.89	1.45	4.69	1.07	1.51	1.58
Base	2,498		2,498		2,466	

4.3 Is an inability to provide accurate reports of childhood circumstances the result of poor memory?

Figure 7 shows how levels of inconsistency varied on the two measures by performance in the delayed word-list recall task at age 50. Higher

scores indicate better memory and we see that, at least on the ‘number of people’ measure, there is evidence of a bivariate association between performance in the test and recall ability. For the ‘number of rooms’ measure, the association is not so apparent.

Figure 7. Inconsistent retrospective responses by delayed memory test score at age 50



*Base: All in random sub-set who completed memory test (n=2,474)

One might also hypothesise that the magnitude of any inconsistencies would be higher amongst those with the poorest scores. This is confirmed, for number of people, by Table 5 which shows the mean difference and absolute mean difference (i.e. taking no account of the direction of any discrepancy) between the R and C responses broken down by performance in the delayed memory test.

The mean absolute difference reduces from 1.5 among those scoring zero, to 0.3 among those scoring nine. The mean difference reduces from -1.2 to -0.1. As shown above, for 'number of rooms' the level of inconsistency was not greatly associated with performance in the test and so there is similarly no clear association between test scores and the magnitude of any inconsistencies.

Table 5. Mean differences between retrospective and contemporaneous responses by performance in delayed memory test score

Delayed memory test score	Number of people		Number of rooms		N	%
	Mean difference (absolute)	Mean difference (R-C)	Mean difference (absolute)	Mean difference (R-C)		
	R-C	(R-C)	R-C	(R-C)		
0	1.5	-1.2	0.9	-0.5	32	1.3%
1	0.9	-0.5	0.7	-0.3	41	1.7%
2	0.9	-0.5	0.7	-0.2	97	3.9%
3	0.8	-0.4	0.7	-0.3	193	7.8%
4	0.7	-0.4	0.8	-0.3	367	14.8%
5	0.8	-0.5	0.7	-0.2	534	21.6%
6	0.6	-0.3	0.7	-0.2	539	21.8%
7	0.6	-0.3	0.6	-0.2	378	15.3%
8	0.5	-0.3	0.6	-0.2	193	7.8%
9	0.3	-0.1	0.6	0.0	78	3.2%
10	0.6	-0.3	0.6	-0.1	22	0.9%

*Base: All in random sub-set who completed memory test (n=2,474)

One in five respondents provided inconsistent R responses on both measures and it was suggested earlier that this group could be those with the poorest recall. If so, we would also expect this group to achieve lower scores in the delayed memory test (see Table 6). The mean score of those providing two inconsistent R responses was

compared with the mean score of those providing one consistent response with a t-test, and found to be significantly lower ($p < 0.001$). The mean score of those providing one consistent response was in turn significantly lower than those providing two consistent responses ($p < 0.001$).

Table 6. Delayed memory test score at age 50 by inconsistent R and C responses

	Mean memory test score
R and C responses inconsistent on both measures	4.97 (<i>SD</i> =1.98)
R and C responses inconsistent on one measure	5.37 (<i>SD</i> =1.87)
R and C responses consistent on both measures	5.58 (<i>SD</i> =1.80)

*Base: All in random sub-set who completed memory test (n=2,474)

On the basis of these bi-variate analyses, we do find support for Hypothesis 7, that those with poorer memory at 50 appear to have poorer ability to recall circumstances from their childhood. On the number of people measure at least, there was also evidence that poorer performance on the memory test was associated with a greater magnitude of inconsistency between C and R responses.

5 Summary and discussion

Cross-sectional studies and longitudinal studies alike make regular use of retrospective questions about childhood circumstances. However, to date, little has been known about the accuracy with which adults can recall this kind of information. This paper therefore makes a vital contribution to the literature on recall accuracy by comparing retrospective questions about childhood circumstances answered at age 50, with questions answered contemporaneously by parents, and thereby demonstrating that 50-year-olds were typically able to recall these aspects of their childhood with reasonable accuracy. As hypothesised, the level of consistency between retrospective and contemporaneous responses was greater for number of people than for number of rooms.

Although a significant proportion failed to provide retrospective responses that matched precisely those collected contemporaneously, the difference between responses was typically small. Retrospective responses tended to 'under-report' when compared with contemporaneous responses. One cause of this possibly stems from the fact that, unlike the retrospective questions, the contemporaneous questions in the age 11 survey

did not present clarifying information about which people and rooms should be included in a consistent way, which may well have resulted in a degree of contemporaneous over-reporting. This underlines the importance of ensuring when designing questions, whether contemporaneous or otherwise, that all the relevant information required to answer a question is provided to all respondents.

Using a subset of participants whose address had remained stable between ages 7 and 11 made it possible to assess, to an extent, whether the reliability of the retrospective measure of number of rooms at 11 was any more reliable than the responses collected contemporaneously. Comparing the level of consistency between number of rooms reported at 7 and 11 with the level of consistency between the retrospective report of number of rooms at 11, and the number recorded contemporaneously, led to the encouraging conclusion that the retrospective measure was only marginally less reliable. This finding provides strong support for the use of retrospective questions of this nature.

The two retrospective measures were used in combination to produce a measure of overcrowding, which was compared with a measure of overcrowding using contemporaneous information. Estimates of the proportion living in overcrowded conditions in childhood did not differ between the retrospective and contemporaneous measures. However, substituting the retrospective over-crowding measure for the contemporaneous measure in a logistic regression model, examining the odds of suffering from lung problems at age 50, was found to have a significant impact on how the

variable would be interpreted. Altering the definition of overcrowding was also found to affect how the two measures performed in the models. We must take from this that researchers using retrospective measures of this nature must be cautious, but that this applies equally when using contemporaneous measures.

Previous evidence about the characteristics associated with poor recall has been very limited. This paper has demonstrated that there are particular groups who are more likely to provide inconsistent and therefore potentially inaccurate responses than others. In particular around one in five respondents were identified as having the poorest recall on the basis of providing inconsistent retrospective responses to both questions. The likelihood of being in the group identified as having poorest recall was considerably higher among those with less stable family backgrounds and those living in larger households. Researchers making use of retrospectively collected data on these measures

would therefore be advised to use a degree of caution when interpreting the responses of these particular groups.

Additionally, the paper made use of a unique opportunity to examine the impact of memory as measured by a simple delayed recall task on ability to recall childhood circumstances correctly. Performance in this test at age 50 was found to have a significant impact on ability to 'correctly' recall details from childhood. If feasible, researchers conducting a study with a significant focus on the collection of retrospective information about childhood might be advised to consider the inclusion of a brief objective assessment of memory, as a way of identifying individuals whose responses might potentially be less reliable.

It is highly probable that future follow-ups of NCDS will involve repeated tests of cognition similar to those included at age 50. It will be of interest to see whether poor retrospective recall at age 50 is predictive of later life problems with cognition.

Acknowledgements

The National Child Development Study is funded by the Economic and Social Research Council (ESRC) and run by the Centre for Longitudinal Studies (CLS). This research was carried out as part the Resource Centre funding for CLS from the ESRC (Ref: RES-579-47-001). I would like to thank Professor John Micklewright for helpful comments on drafts of this manuscript.

References

- Auriat, N. (1993). "My wife knows best": A comparison of event dating accuracy between the wife, the husband, the couple, and the Belgium Population Register. *The Public Opinion Quarterly*, 57, 165-190. <http://dx.doi.org/10.1086/269364>
- Banks, J., Nazroo, J., & Steptoe, A. (2012). *The Dynamics of Ageing, Evidence from the English Longitudinal Study of Ageing 2002-2010, Wave 5*. London: Institute for Fiscal Studies.
- Börsch-Supan, A., Hank, K., Jürges, H. & Schröder, M. (2008). Longitudinal data collection in continental Europe: Experiences from the Survey of Health, Ageing and Retirement in Europe (SHARE). MEA Discussion Paper (08162). University of Mannheim.
- Börsch-Supan, A., & Schröder, M. (2011). Retrospective data collection in the Survey of Health, Ageing and Retirement in Europe. In M. Schröder (Ed.) *Retrospective data collection in the Survey of Health, Ageing and Retirement in Europe. SHARELIFE Methodology*. Mannheim: MEA.
- Bound, J., Brown, C., & Mathiowetz, N. (2000). Measurement error in survey data. In J.J. Heckman & E. Leamer (Ed.), *Handbook of Econometrics* (pp. 3707-3745). Amsterdam: Elsevier.
- Bradburn, N. M., Huttenlocher, J., & Hedges, L. (1994). Telescoping and temporal memory. In N. Schwarz and S. Sudman (Eds.) *Autobiographical memory and the validity of retrospective reports* (pp. 203-216). New York: Springer Verlag. http://dx.doi.org/10.1007/978-1-4612-2624-6_14
- Brown, M. & Dodgeon, B. (2010). NCDS Cognitive Assessments at Age 50: Initial Results. CLS Working Paper 2010/1. London: Centre for Longitudinal Studies.
- Brown, M., Elliott, J., Hancock, M., Shepherd, P. & Dodgeon, B. (2010). National Child Development Study 2008-2009 follow-up (first deposit): A guide to the dataset. London: Centre for Longitudinal Studies.
- Brown, N. R., Rips, L. J. & Shevell, S. K. (1985). Subjective dates of natural events in very long term memory. *Cognitive Psychology*, 17, 139-177. [http://dx.doi.org/10.1016/0010-0285\(85\)90006-4](http://dx.doi.org/10.1016/0010-0285(85)90006-4)
- Cannell, C. (1977). A summary of studies of interviewing methodology. *Vital and Health Statistics, Series 2- No.69 Data evaluation and methods research*. U.S. Dept of Health, Education and Welfare. U.S. Govt. Printing Office, Washington D.C. DHEW pub. No. (HRA), (pp 77-1343).
- Cohen, B., Erickson, P. & Powell, A. (1983). The impact of length of recall period on the estimation of health events. *Proceedings of the Social Statistics Section of the American Statistical Association*, pp. 497-502.
- Crowder, R. (1976). *Principles of Learning and Memory*. Hillsdale, NJ: Lawrence Erlbaum.

- Elo, I.T. (1998). Childhood conditions and adult health: Evidence from the Health and Retirement Study. Population Aging Research Center, WPS 98-03, University of Pennsylvania. PMCid:PMC19331
- Gray, P. (1955). The memory factor in social surveys. *Journal of the American Statistical Association*, 50, 344-363. <http://dx.doi.org/10.1080/01621459.1955.10501269>
- Haas, S. (2007). The long-term effects of poor childhood health: An assessment and application of retrospective reports. *Demography*, 44, 113-135. <http://dx.doi.org/10.1353/dem.2007.0003>
- Linton, M. (1982). Transformations of Memory in Everyday Life. In U. Neisser (Ed) *Memory Observed* (pp 77-91). San Francisco: W. H. Freeman.
- Loftus, E. (1979) *Eyewitness Testimony*, Cambridge, MA: Harvard University Press.
- Marsh, A., Gordon, D., Pantazis, C. & Heslop, P. (1999). Home Sweet Home? The impact of poor housing on health. Bristol: The Policy Press.
- Mathiowetz, N., & Duncan, G. (1988). Out of Work, Out of Mind: Response Errors in Retrospective Reports of Unemployment. *Journal of Business and Economic Statistics*, 6, 221-229.
- Power, C., & Elliott, J. (2006) Cohort Profile: 1958 British Birth Cohort Study (National Child Development Study). *International Journal of Epidemiology*, 35, 34-41. <http://dx.doi.org/10.1093/ije/dyi183>
- Sudman, S., and Bradburn, N.M. (1973) Effects of time and memory factors on response in surveys. *Journal of the American Statistical Association*, 64, 805-15. <http://dx.doi.org/10.1080/01621459.1973.10481428>
- Tourangeau, R. (1984). Cognitive Sciences and Survey Methods. In Jabine, T., Straf, M., Tanur, J. & Tourangeau, R. (Eds). *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*. (pp 73-100). Washington, D.C. : National Academy Press.
- Ward, K., Medina, J., Mo, M., & Cox, K. (2009) ELSA wave three: Life history interview – A user guide to the data. London: National Centre for Social Research.

Endnotes

ⁱ The four assessments completed were an immediate memory test in which respondents were read a list of 10 words and asked to recall as many as they could. An animal naming task in which respondents were given one minute to name as many animals as they could. A letter cancellation task in which respondents were given a grid of randomised letters and had to cross out as many Ps and Ws as they could in one minute. A delayed memory task where respondents were asked to recall the words from the immediate memory test after a delay of approximately 5 minutes. Scores were compared using an independent samples t-test and all differences were significant ($p < 0.05$).

ⁱⁱ It was not the case that these individuals lived in two person households at 11 and were not including themselves in the total reported. Parents of these individuals provided a full range of responses from two to 12, with the most common responses being four and five. These individuals were also no more likely to live in one person households at age 50 so it does not appear that they were mistakenly reporting their current circumstances.

ⁱⁱⁱ Age 50 characteristics: sex, smoking, social class, highest qualification. Age 11 characteristics: free-school meals, financial hardship, tenure, indoor WC, hot water in household, paternal unemployment