

The Causes of Seam Effects in Panel Surveys

Annette Jäckle

Institute for Social and Economic Research
University of Essex

No. 2008-14
April 2008



INSTITUTE FOR SOCIAL
& ECONOMIC RESEARCH

Non-technical summary

Measures of change from panel surveys are typically biased by seam effects. The term “seam effects” refers to the finding that the number of transitions between different states of interest, from one time period to the next, is far higher when the reports for each period come from two different interviews, than when the reports come from the same interview. The causes of this excess of transitions at the “seam” between reference periods are not well understood. As a result, data collection methods designed to reduce seam effects appear to work for some types of items, but not for others. Understanding why seam effects are larger for some types of items and why data collection methods designed to reduce seam effects work only for some items requires an understanding of the causes of seam effects, including whether and how the causes differ for different types of items and question formats. The present paper contributes to this aim by deriving a theoretical framework of the causes of seam effects that unifies existing theories and evidence. The framework is based on the idea that the characteristics of events of interest, the respondent’s situation and the question format and wording determine errors that can occur during the survey and response process. These errors in turn interact with the relative length of time spent in a particular state of interest, to produce biases in estimates of seam and off-seam change leading to the concentration of transitions at the seam. The framework can be used to predict the likely causes of seam effects for different types of items, and the relative magnitude of seam effects. The predictions from the framework are tested using data from the British Household Panel Survey and find support.

The Causes of Seam Effects in Panel Surveys

Annette Jäckle

April 2008

ABSTRACT

For some domains, panel surveys collect information about the period between interviews. Such data are typically affected by “seam effects”: transition rates from one month (or week) to the next are typically far higher if the months were covered in two different interviews, than if they were covered in the same interview. The causes of seam effects are not well understood. As a result, data collection methods designed to reduce the problem appear to work for some types of items, but not for others. This paper presents a theoretical framework of the causes of seam effects that unifies existing theories and evidence. The predictions from the framework are tested using data from the British Household Panel Survey and find support.

Keywords: response process, measurement error, processing error, estimates of change, transition rates, history data, event data, temporal questions.

JEL Codes: C42

Acknowledgement: I would like to thank Frederick Conrad for his time and comments whilst I was visiting the Survey Research Centre at the University of Michigan in spring 2007. The conversations with him about seam effects and whether and how they can occur for different types of items and question formats prompted me to write this paper and his enthusiasm kept me going. I am grateful to the ESRC-SSRC for a visiting fellowship which gave me this opportunity. I would also like to thank Peter Lynn, whose clear views helped me focus my ideas, to Annelies Blom and the participants, including Robert Belli, at a seminar on the “Optimal Timing of Data Collection Waves and Quality of Retrospective Recall Data” at the University of Essex in July 2007 for comments on an early version of this paper.

Contact: Annette Jäckle, Institute for Social and Economic Research, University of Essex, Colchester CO4 3SQ, UK. Email: aejack@essex.ac.uk.

1 Introduction

Estimates of change based on data from panel surveys are notoriously biased. The origin of such biases is however not well understood. As a result, data collection methods designed to reduce biases are somewhat ad hoc and produce mixed results, which are so far unexplained. This paper proposes a theoretical framework of the causes of biases in measures of change that integrates existing theories and empirical findings. The framework can be used to predict the likely causes and relative magnitude of biases for different types of events. The predictions of the framework were tested using data from the British Household Panel Survey (BHPS). The framework has implications for the design of questions to reduce biases in measures of change from panel surveys.

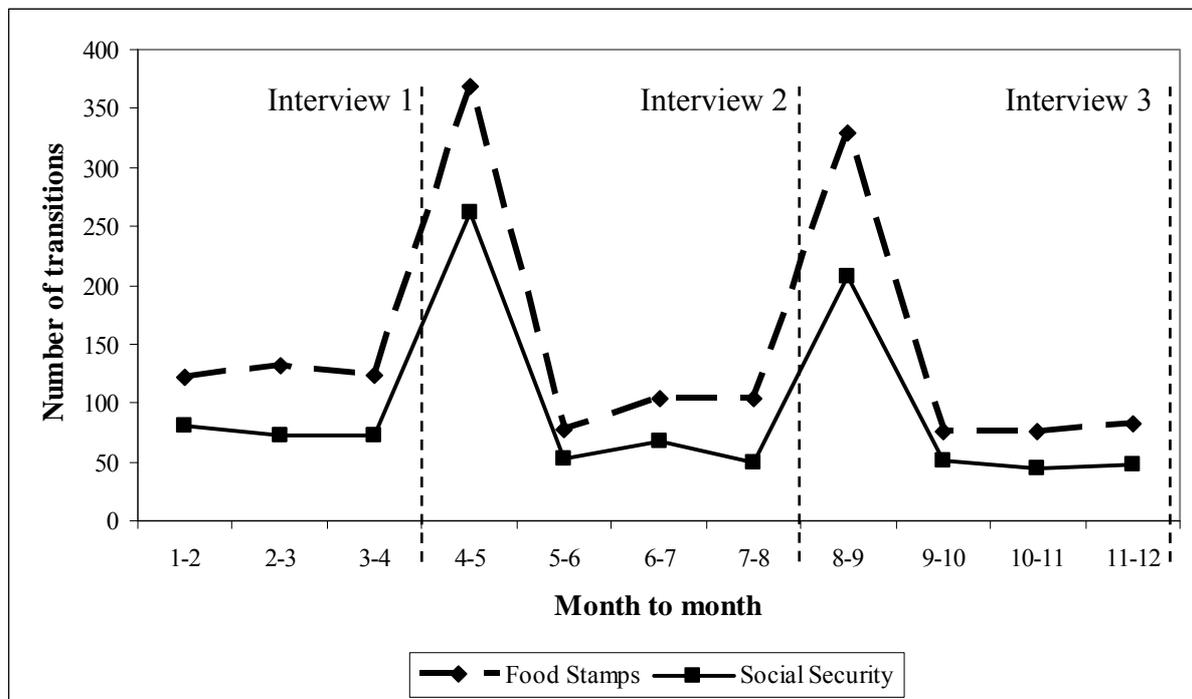
Panel surveys collect repeated measures by surveying sample members at fixed intervals (waves) over a period of time. The bulk of repeated measures consist of information about the respondent's situation at the time of each interview. This information can be used to estimate wave-on-wave change, for example, annual, quarterly or monthly change, depending on the length of the interval between waves. For a small number of domains, surveys in addition ask retrospective questions about the occurrence and dates of all events during the period between interviews. Retrospective histories can be used to estimate change for time periods that are shorter than the interval between interviews. For example, reports on a one-year reference period between interviews may be used to estimate monthly or weekly change. Retrospective histories therefore make it possible to estimate change for sub-periods of the reference period.

Estimates of change based on repeated measures from panel interviews are often biased, due to inconsistencies in individual reports from one wave to the next. These inconsistencies lead to spurious change in wave-on-wave estimates and "seam effects" in estimates from retrospective histories. This paper focuses on seam effects, but I return to wave-on-wave measures in the concluding section, to show that the framework of the causes of seam effects also applies to wave-on-wave measures.

The problem of seam effects is best illustrated with Burkhead and Coder's (1985) classic example from the Survey of Income and Program Participation (SIPP), shown in Figure 1. The interval between waves in the SIPP was four months. At each interview respondents were asked retrospective questions about receipt of income sources during the reference period since the previous interview: in the first interview respondents reported income receipt for months 4, 3, 2 and 1; in the second interview they reported receipt for months 8, 7, 6 and 5, and so on. When the retrospective histories were used to calculate

month-to-month changes in receipt status, the following pattern was observed: the numbers of changes for months 4 to 5 and months 8 to 9 far exceeded the numbers of changes for all other month pairs. This is referred to as the seam effect, since months 4/5 and 8/9 constituted the seams between reference periods: the report for month 4 was from the first and the report for month 5 from the second interview; similarly, the report for month 8 was from the second and the report for month 9 from the third interview. For all other month pairs, the reports for both months were from the same interview. In short, the term “seam effect” refers to the observed excess of transitions at the seams *between* reference periods, compared to transitions for sub-periods *within* the reference periods.

Figure 1: An illustration of the seam effect: numbers of month-to-month transitions onto and off income receipt



Source: Burkhead and Coder (1985), Table 2 and Table 5. Based on waves 1 to 3 of the 1983/4 Survey of Income and Program Participation.

The problem of seam effects has been documented for many panel surveys and seems to exist regardless of the survey design or type of variable (see Callegaro 2007 for a recent review). Seam effects have been documented for discrete states (e.g. labour market activity, income receipt status) and continuous variables (e.g. income amounts). The magnitude of the seam effect seems to vary for different types of events. The extent to which transitions are concentrated at the seam can be measured in different ways: for example, as the proportion of total transitions observed at a seam, or as the ratio of seam to off-seam transition rates.

Martini (1989 Table 2), for example, examined labour market transitions in eight waves of the SIPP, covering a 32 month period. In the absence of seam effects the expected proportion of total transitions at a seam would have been $7/32=.22$. That is, one would expect 22% of all month-to-month transitions to be observed in seam months. Seam effects were largest for transitions between unemployment and inactivity (over 60% of total transitions at seams) and smallest for transitions between unemployment and employment (between 36% and 44%). Using the same data, Young (1989 Table 1) documented that between 40% and 53% of total transitions in marital status, employment status, personal earnings, family income, Aid to Families with Dependent Children and Food Stamp receipt were at a seam, compared to 69% of transitions in receipt of Social Security. No attempts have been made to understand why seam effects for some types of events seem to be larger than for others.

While there are a number of theories about the causes and empirical findings about the nature of seam effects, the problem is generally not well understood. Lemaitre (1992 p15), for example, concluded his review stating that “studies that have attempted to identify the origin of seam effects have not managed to identify a clear underlying cause of the observed results”. This seems mainly to be because existing studies have either focused narrowly on few potential causes of seam effects, or because they have aggregated over different types of items, for which, as will be shown, the sources of problems can be quite different, and have therefore not found any conclusive results about the causes of seam effects.

Survey organisations have nonetheless been experimenting with different data collection techniques to improve the longitudinal consistency of individual reports, such as dependent interviewing (see Jäckle in press) and different calendar methods (see Glasner and Vaart 2007). Event history calendars have been shown to reduce seam effects in labour market transitions (Callegaro 2007). Dependent interviewing appears to reduce seam effects for some types of items effectively, especially labour market activity histories (Jäckle and Lynn 2007; Murray et al. 1991), but not for others, such as histories of benefit receipt (Moore et al. in press). Many questions therefore remain unanswered, including: Why are seam effects larger for some types of items than others? Why do the data collection methods designed to improve longitudinal consistency work for some types of items but not for others? In which situations may certain methods actually make things worse? And how could one best design data collection methods for those items for which they do not seem to be working?

What is needed is an understanding of the different types of problems that contribute to longitudinal inconsistencies and an understanding of whether and how these problems differ for different types of items and question formats. This paper contributes to this aim, by proposing a theoretical framework of the causes of seam effects, which readily incorporates existing theories and empirical findings. The framework is based on the idea that characteristics of the event, the respondent's situation and the question format and wording determine the likelihood of different errors occurring during the survey and response process. These errors interact with the relative length of spells, that is, the length of time actually spent in a state of substantive interest. The interaction of errors and spell lengths produces biases in the numbers of observed status changes at the seam and in off-seam periods, that lead to seam effects. The framework explains the stylised fact from previous validation studies that seam effects are caused by a combination of under-estimated off-seam change and over-estimated seam change. The framework also explains the finding that the magnitude of seam effects varies for different types of items and offers some clues why previous studies testing for correlates of seam effects have not found any clear patterns. The predictions of the framework tend to be supported in tests using the BHPS: the causes of the seam effect depend on the mean length of spells relative to the reference period; the magnitude of seam effects is predicted by characteristics of the event and the respondent's situation. The predictions about the effect of question format characteristics cannot be tested using existing data. Instead I suggest experimental questions that could be used to this end.

In Section 2 I develop a typology of panel history questions based on a review of some of the main household panel surveys. This typology forms the basis for illustrations in later sections. Section 3 describes the theoretical framework of the causes of seam effects and how it relates to the existing literature. I derive the implications and testable hypotheses based on the framework in Section 4 and test these in Section 5. Section 6 contains a summary and discussion of implications.

2 A typology of panel history questions

A review of the questionnaires used for some of the main household panel surveys (BHPS, German Socio-Economic Panel (SOEP), Household Income and Labour Dynamics in Australia (HILDA) and SIPP)¹ suggests that panel histories are typically updated at each

¹ Information about these surveys, including questionnaires, can be found at: HILDA <http://melbourneinstitute.com/hilda/>, SOEP <http://www.diw.de/english/soep/29012.html>, BHPS www.iser.essex.ac.uk/ulsc/bhps/, SIPP <http://www.bls.census.gov/sipp/>.

interview in one of two ways: either by first finding out about the respondent's current situation and then asking about changes during the reference period, or by asking about events and changes during the reference period and inferring the current situation from this information. Dates of events are either requested explicitly, by asking for the dates of occurrence, or implicitly by asking for other information from which the dates can be derived, such as the elapsed time since an event, or the respondent's status in all sub-periods of the reference period. The following typology of panel history questions extends the typology of temporal questions proposed by Tourangeau, Rips, and Rasinski (2000) using an adapted version of their schemata. To illustrate the variety of questions used, optional elements are placed in parentheses; alternative versions are listed vertically in square brackets. After presenting the different question types, they are illustrated with some examples currently used in household panel surveys.

Stability or change questions

Questions about change or stability during the reference period are typically of the following form, where the response categories are 'yes' and 'no':

$$A. \left[\begin{array}{l} \textit{Have you been} < \textit{status} > \textit{ since before} \\ \textit{Did you} < \textit{current status} > \textit{ in} \\ \textit{Has your} < \textit{status} > \textit{ changed since} \\ \textit{Has} < \textit{event} > \textit{ happened since} \end{array} \right] \left[\begin{array}{l} < \textit{start of reference period} > \\ < \textit{date of last interview} > \end{array} \right] ?$$

The first two versions ask about stability, that is, whether the current status has been the same throughout the reference period. The third and fourth versions ask about changes and events during the reference period. The start of the reference period is either a fixed date or the date of the respondent's previous interview, which will be different across respondents.

Time of occurrence questions

If respondents report a change, they are then asked about the timing of events. One type of temporal question identified by Tourangeau, Rips, and Rasinski (2000) are 'time-of-occurrence' questions. (This corresponds to the notion of 'spell-date' questions in Jäckle in press) Respondents are explicitly asked to report the date, for example month and year, of an event:

$$B. \text{ (Within reference period,)} \left[\begin{array}{c} \text{when} \\ \text{in what < dated unit >} \end{array} \right] \text{did} \left(\left[\begin{array}{c} \text{first} \\ \text{next} \\ \dots \\ \text{last} \end{array} \right] \right) \text{event} \left[\begin{array}{c} \text{occur} \\ \text{begin} \\ \dots \\ \text{end} \end{array} \right] ?$$

<Dated unit> can refer to an exact date (e.g. day, month, year) or to a less precise date (e.g. only month, year or only year).

Elapsed time questions

Alternatively, respondents may not be asked explicitly for dates, but for related temporal information from which the dates of events can be derived. This can be done using ‘elapsed time’ questions (Tourangeau, Rips, and Rasinski 2000), where respondents are asked to report how long ago an event started or ended:

$$C. \text{ (Within ref period,)} \left(\left[\begin{array}{c} \text{exactly} \\ \text{about} \end{array} \right] \right) \left[\begin{array}{c} \text{how long since} \\ \text{how long ago since} \\ \text{how many < temp units > since} \end{array} \right] \left(\left[\begin{array}{c} \text{first} \\ \text{next} \\ \dots \\ \text{last} \end{array} \right] \right) \text{event} \left[\begin{array}{c} \text{occurred} \\ \text{began} \\ \dots \\ \text{ended} \end{array} \right] ?$$

Elapsed time questions treat time as either continuous or discrete. The continuous version asks how much time has elapsed using phrases such as ‘how long’ or ‘how long ago’. The discrete version asks how many temporal units, for example months or weeks, have gone by.

Period-status questions

The second implicit method of collecting date information is with ‘period-status’ questions (Jäckle in press). In this case, respondents are asked about their status or the occurrence of events for every sub-period (e.g. month or week) of the reference period:

$$D. \text{ (In which < dated units >)} \left[\begin{array}{c} \text{did < event > occur} \\ \text{were you < status >} \\ \text{did you < activity >} \end{array} \right] ? \text{ (In < unit}_1 \text{ > ?)} \text{ (In < unit}_2 \text{ > ?)} \dots$$

In some surveys this is formulated as an open-ended question where respondents have to volunteer the relevant time units, for example: “In which months did you receive X?” In other surveys this is formulated as a ‘yes/no’ question for each dated unit. That is, the first part of the question is followed up by “Did this occur in <dated unit₁>?” [yes/no] “In <dated unit₂>?” [yes/no], etc.

Finally, there are survey questions where respondents are directly asked a time-of-occurrence question about the start of their current situation, without first being asked about

change. The start date of their current spell is then used to derive whether a change has occurred during the reference period, and whether or not follow-up questions need to be asked to obtain a complete history of events.

Examples

The following are some examples of how different combinations of question types A. to D. are used to collect histories in different household panel studies. Histories of residential moves in the BHPS (and similarly in SOEP) are collected by asking about stability during the reference period, followed by a time-of-occurrence question:

A-Q1. "Have you yourself lived in this (house/flat) for more than a year, that is before September 1st <previous year>?" [yes/no]

IF NO:

B-Q2. "In what month did you move here?" [month, year]

Questions about legal marital histories in the BHPS first ask about the respondent's current status, then whether this has changed during the reference period, and if so about the time of occurrence:

Q3. "What is your current legal marital status, are you married, separated, divorced, widowed or have never been married?"

IF MARRIED, SEPARATED, DIVORCED OR WIDOWED, ASK:

A-Q4. "Has your marital status changed in the last year, that is since September 1st <previous calendar year>?" [yes/no]

IF YES:

B-Q5. "So you have recently been <READ Q0 MARITAL STATUS>. When did that happen?" [month, year]

The BHPS labour market activity history is an example where respondents are not asked about stability or change during the reference period. Instead, they are asked a time-of-occurrence question about the start of their current activity. This date is then used to derive whether or not a change has taken place during the reference period and whether follow-up questions should be asked to obtain complete activity histories:

IF CURRENTLY WORKING:

B-Q6. "What was the date you started working in your present position?" [day, month, year]

IF CURRENTLY NOT WORKING:

B-Q7. "On what date did your present spell of being <current non-employment status> begin?" [day, month, year]

Period-status questions are used extensively in the SIPP, but can also be found in other surveys. The BHPS questions about unearned income sources are an example of open-ended period-status questions:

A-Q8. "Please look at this card and tell me if, since September 1st <previous year>, you have received any of the types of income or payments shown, either just yourself or jointly?"

IF YES: ASK "WHICH ONES?"

D-Q9. "And for which months since September 1st <previous year> have you received <income source> ?"

(RING CODES FOR MONTHS WHEN PAID, IF ALL UP TO THE CURRENT MONTH RING "ALL")

Similarly, the SIPP asks a series of questions about people's labour market activities in open-ended format:

D-Q10. "Please look at the calendar. In which weeks [were you not working? / did you work at a job or business or do any work at all for pay or profit? / were you looking for work? / ...]"

Most period-status questions in the SIPP however ask a closed 'yes/no' question for each month of the reference period, for example:

A-Q11. "At any time between <Reference Month 1> 1st and today were you covered by Medicare?" [yes/no]

IF YES:

D-Q12. "In which months were you covered by Medicare?"

"In this month?" [yes/no]

"In <Reference Month 4>?" [yes/no]

"In <Reference Month 3>?" [yes/no]

“In <Reference Month 2>?” [yes/no]

“In <Reference Month 1>?” [yes/no]

Elapsed time questions are rarely used. One example is in the HILDA self-completion questionnaire, which asks questions about life events during the reference period, covering amongst other things partnership and family, health, financial and labour market events:

C-Q13. “We now would like you to think about major events that have happened in your life over the past 12 months. For each statement cross either the YES box or the NO box to indicate whether each event happened during the past 12 months. If you answer “YES”, then also cross one box to indicate how long ago the event happened or started.” [0 – 3, 4 – 6, 7 – 9, 10 – 12 months ago]

These examples illustrate how different question types are used to collect panel histories for different domains. The one thing they have in common is that they all suffer from seam effects. As the framework derived in the next section suggests, however, the causes are likely to be different for different types of items and question formats.

3 Theoretical framework of the causes of seam effects

This section presents a theoretical model of the causes of seam effects, which integrates existing theories and empirical findings. The model explains the stylized facts documented by previous validation studies of the causes of seam effects, namely that the excess of seam transitions is caused by a combination of under-reported change *within* a reference period and over-estimated change *between* reference periods. The model also explains the finding that the magnitude of seam effects varies for different types of events. The framework, illustrated in Figure 2, is based on the idea that characteristics of the event of interest, the respondent’s circumstances and the question wording and format determine errors that can occur during the survey and response process. The errors in the reporting and dating of events interact with the length of spells, relative to the length of a reference period, to create biases in estimates of within and between wave change, which contribute to the seam effect.

3.1 Errors, spell lengths and within and between wave change

Marquis, Moore and Huggins (1990) compared SIPP survey reports for two waves (covering four months each) with individual administrative records for the receipt of eight State benefit programmes. The record check showed that month-to-month change in receipt status within

the reference period was under-estimated by between 6% and 64% depending on the item, while change at the seam was over-estimated by between 20% and more than 200%. These biases were thought to be due to a combination of omitted events, misdated events and misclassified events (Lemaitre 1992; Martini 1989). Note that the terms *event* and *spell* are used somewhat interchangeably in the discussion here, although strictly speaking the event refers to the date of a change and the spell to the resulting state until the occurrence of the subsequent event. *Events* can refer both to discrete states (e.g. labour market activity status, benefit receipt status) and to continuous variables (e.g. amount of earnings). *Misclassification* can refer both to an inappropriate response category for a discrete state and to a value for a continuous variable that differs from the true value by more than some predefined threshold.

Figure 3: Interaction of error type, length of spell and position of error

Case	Error Type	Interview 1	Interview 2	Interview 3	Estimated Change
1	Omission	✕.....✕	✕.....✕		Under-reported within-wave start and end
2a		✕.....◇	—		Under-reported within-wave start, misplaced to seam
2b		—	◇.....✕		Under-reported within-wave end, misplaced to seam
3		—	○.....○	—	Spurious end at I1 seam and spurious start at I2 seam
4	Misclassification	≡≡≡			No effect
5		≡≡≡	○	—	Spurious end and start at I1 seam
6		—	○	≡≡≡	Spurious ends and starts at I1 and I2 seams
7	Misdating		◇	✕	Under-reported within-wave start, misplaced to seam
Legend		— Correct report	✕ Under-reported within-wave change	◇ Within-wave change misplaced to seam	
	 Omission		○ Spurious seam change	
		≡≡≡ Misclassification			
		— — Misdating			

Martini (1989), Martini and Ryscavage (1991) and Young (1989) pointed out that the effect of omissions, misclassifications and misdating on estimates of change depends on *how many* reference periods the actual spell spans, and at *which point* the measurement error occurred. The following discussion also illustrates that the over-estimation of seam changes can be due to both *spurious* seam changes and true changes within the reference period that are *misplaced* to the seam. The following scenarios are illustrated in Figure 3:

- Omission of an entire spell leads to under-estimation of within-wave change for both start and end dates (case 1 in Figure 3). This type of problem is most likely for short spells that started and ended during one reference period, but can also happen for spells spanning multiple reference periods. Omissions of entire spells may also lead to under-estimation of seam changes. Since the proportion of seam months, and hence of spells starting or ending in a seam month, is usually small, this occurs less frequently. The net effect of omissions of entire spells is therefore an under-estimation of within-wave change.
- Omission of a spell that covers two periods and was correctly reported in one of the periods, leads to both under-estimation of within-wave change and a misplacement of the within-wave change to the seam. In case 2a the spell was not reported in the first interview, but reported in the second. As a result, the start date was misplaced forward in time to the seam. In case 2b the spell was reported in the first but not the second interview, and the end date misplaced backward in time to the seam.
- Omission of a spell that spans three or more reference periods and is correctly reported for previous and subsequent periods, leads to a spurious end at one seam and a spurious start at the next seam. In case 3 the spell was correctly reported in the first and third interview, but omitted in the second. As a result, the wave 1 spell ends at a seam and the wave 3 spell starts at the following seam, when in reality the spell continued through three reference periods.
- Misclassification of a spell has no effect on within or between wave changes in a particular domain, if the spell is entirely contained within the reference period (case 4). Reporting receipt of unemployment benefit instead of income support, for example, will not affect aggregate estimates of transitions in benefit receipt. If the individual benefit types are of interest, however, within-wave transitions will be under-estimated for one and over-estimated for the other.
- If the spell spans two reference periods and is misclassified for only one of the periods, this will lead to a spurious end and a spurious start change at one seam. In case 5 the spell is misclassified in the first interview and correctly reported in the second interview. As a result, the misclassified spell ends and the correctly reported spell starts at the seam between the reference periods for the first and second interview.
- If the spell spans three or more reference periods and is misclassified only in one of the middle waves, this will lead to spurious start and end changes at two seams. In

case 6 the true spell type ends at the first seam. At the same time, there is a spurious start date for the misclassified spell type. At the second seam there is a spurious start for the true spell type, and a spurious end for the misclassified type.

- Finally, misdating the start as having occurred before the start of the reference period leads to under-reporting of within-wave change and misdating of change to the seam. In case 7 the spell is correctly reported, but incorrectly reported for the start of the reference period. As a result the start date is misdated backward in time to the seam.

3.2 Causes of errors in the reporting and dating of events

Having illustrated different ways in which omissions and misclassifications can contribute to the excess of seam transitions, the next question is what causes these errors? The causes suggested in previous studies can be grouped as respondent errors at different stages in the response process, interviewer errors and data processing errors. The process of constructing a response involves a number of stages (Cannell, Miller, and Oksenberg 1981; Tourangeau, Rips, and Rasinski 2000). Respondents must first understand the survey question and identify what information is requested. The second stage involves searching for and retrieving information from memory. Respondents then make judgements about the completeness and relevance of the retrieved information and complete and integrate this information to compute a response. Finally, respondents map the retrieved and computed information onto one of the response options and may edit their response. As Willis (2001 p28) summarized, “answers to survey questions are often not so much reported from storage as they are synthesized on the spot from a variety of information sources”. Errors leading to omission, misdating and misclassification can occur at each of these stages.

Understanding the survey question

Respondents may understand a question differently or misinterpret a question in some interviews, which would lead to volatility across waves in what respondents perceive and report their status to be. The resulting response variation is thought to be one of the causes of seam effects (Clarke and Tate 1999; Young 1989), since it would lead to different classifications across waves (Lemaitre 1992; Martini and Ryscavage 1991) and to the omission of events in some interviews. Question understanding is influenced, amongst other things, by the clarity of concepts, the complexity of sentence structures, question wording and reporting instructions. A cognitive study of the causes of seam effects by Marquis, Moore and Huggins (1990), however, suggested that comprehension was not a general problem. In

contrast Burkhead and Coder (1985 p352) thought questionnaire wording and design were important because they are directly related to recall errors, “since the procedures used to ask questions must have some effect on the way the answers are given”.

Retrieving information from memory

Errors and biases in respondent recall are thought to be a second cause of seam effects (Burkhead and Coder 1985). Retrieval of information from memory may fail because the information was not encoded in the first place, because of memory decay or because of lack of effort by the respondent. Proxy reporting has been suggested as one of the causes of seam effects (Martini and Ryscavage 1991), but empirical studies have found that changes between self and proxy reporting do not increase seam problems (Clarke and Tate 1999; Marquis, Moore, and Huggins 1990). Forgetting seemed to be the main cause of inconsistencies in reported labour market activities in a study by Jäckle and Lynn (2007). Tests of whether forgetting is caused by the decay of memory over time have produced mixed results. Forgetting theory would predict that errors are pre-dominantly omissions, that the extent of omissions increases with elapsed time since the event and that recent events are reported accurately. As a result, more changes would be reported for the most recent period than for the start of the reference period. None of these hypotheses were supported in Marquis, Moore and Huggins’ (1990) validation study: omissions were the most common type of error, but there was also considerable over-reporting of events, the number of omissions did not increase with time and recent events were not reported more accurately than past events. As a result there was no difference in the number of changes reported for the most recent and most distant periods. Martini (1989) and Martini and Ryscavage (1991) however found that the level of change did fall with time. Kalton and Miller (1991) also showed that a change in the amount of benefit income was reported less frequently as time since this change increased. How quickly memory decays with time depends on the item in question. Validation studies have shown that shorter spells are more likely to be omitted, and that even for similar durations, some types of spells are more likely to be omitted than others (e.g. Paull 2002).

Constructing a response

Judgement errors in the inference and estimation strategies used to complete the retrieved information and transform it into a response, are thought to be a third source of seam effects. As described by Tourangeau, Rips, and Rasinski (2000), attempts to reconstruct memories by inferring missing details may be based on information about the retrieval process, for

example the fact that an event is hard to recall might suggest that it is an infrequent type of event or happened a long time ago, or may be based on other ‘typical’ events of the same kind. Estimation strategies, used to compute a response while adjusting for missing pieces in the retrieved information, may include recalling events in the recent past from memory and extrapolating from these to the past; averaging; rounding to prototypical values; or simply guessing. The inference and estimation strategies are based on information recalled from memory – although the judgement stage may sometimes replace the retrieval stage completely. These strategies will inevitably produce errors of omission, misdating and misclassification. In addition, these strategies are likely to produce responses that are consistent for the sub-periods within a wave, but not consistent across waves. And as Martin said, “any factor or process that increases the consistency of reporting across weeks within a wave, and/or that reduces the consistency of reporting between waves, could produce a seam effect” (Martin 2001 p32).

Marquis, Moore and Huggins (1990) concluded from cognitive interviews that seam effects were caused by respondents using simple heuristics to compute responses, instead of detailed direct recall from memory. One such simple heuristic was suggested by Young (1989) who thought seam effects were in part the result of ‘constant wave responses’, a strategy whereby “respondents may give an answer for earlier months in an interview period, identical with the answer they give for the most recent month or their current state” (p395). Respondents may use this strategy to avoid the effort of recalling information about more distant periods from memory, or because their memory fails and they extrapolate from their current situation to the past, making too little adjustment for changes (Tourangeau, Rips, and Rasinski 2000). As Moore and Kasprzyk (1984) put it, imperfect recall may lead “respondents to report receipt for the entire 3-month period of a single wave as having been more stable than it really was” (p728). Evidence of constant wave responses comes from record check studies that showed that under-reporting of income receipt is more likely if receipt ended during the early part of the reference period. In these cases respondents tend to report their current non-receipt status for the entire reference period. These studies have also shown that respondents tend to report current amounts of income for all sub-periods of the reference period (Goudreau, Oberheu, and Vaughan 1984; Kalton and Miller 1991; Lynn et al. 2004). Constant wave responses are represented by two of the cases in Figure 3. The reports from Interview 2 for cases 2b and 7 correspond to constant wave responses, where the current situation of receipt (case 7) or non-receipt (case 2b) is correctly reported for the date of the interview, but incorrectly reported for all previous months in the reference period.

The choice of strategies used at the judgment stage depends on the task difficulty and required effort, the reporting instructions, the respondent's motivation and most crucially the information accessible from memory (Conrad, Brown, and Cashman 1998; Tourangeau, Rips, and Rasinski 2000). These ideas are combined in the cognitive model of the causes of seam effects proposed by Rips et al. (2003), according to which forgetting leads respondents to either guess or provide a constant wave response based on their current situation. Their laboratory experiment and theoretical model are based on period-status questions, where respondents are asked about their status in each sub-period (e.g. each month) of the reference period. The strategy used by the respondent to report on the status of an item in each sub-period, depends on the information the respondent has for that time period. Since memory is thought to decay with time, this implies that events for recent sub-periods are more likely to be recalled from memory, while reports about the distant sub-periods are more likely to be based on guesses or constant wave responses. Transitions at the seam reflect the difference between events recalled (accurately) from memory for the sub-period before the seam and constant wave responses for the sub-period after the seam. The seam effect is therefore seen as the difference in reports produced by different response strategies, based either on memory or estimation. The findings from their laboratory study suggest that there is "a direct relationship between difficulty of the respondents' retrieval task and the size of the seam effect: The harder it is for respondents to recall the queried information, the larger the effect". In addition, they conclude that "conditions that favour constant wave responses also produce larger seam effects" (p545).

Formulating a response

At the final stage respondents map the computed response onto the response categories or formulate an open-ended response. Errors may occur at this stage because respondents deliberately choose an inappropriate category, because of different forms of satisficing or because the response options do not describe the respondent's circumstances well. Respondents may edit their response if they perceive a risk of disclosing sensitive, embarrassing or threatening information, or information diverging from their self-image (Lemaitre 1992). Alternatively, errors may occur because respondents satisfice and shortcut the response process. They may choose the first acceptable response category instead of attempting to retrieve information from memory (Krosnick 1991). Similarly respondents may learn from past interviews that certain responses lead to follow-up questions (e.g. 'yes' in response to questions about income receipt triggers follow-up questions about timing and

amounts) and choose the response options that do not lead to follow-up questions (Burkhead and Coder 1985; Martini 1989). Finally, errors can occur even if the respondent has carefully retrieved information from memory and is making the best effort to select an appropriate response category. This could happen if the categories are not mutually exclusive and more than one category could describe the respondent's state. If there is no unambiguous definition of how to select the "main" category, the respondent may choose different response categories at different interviews, to describe a situation that has in fact not changed. For example, if the respondent has multiple jobs, or is inactive (e.g. retired) but also in employment, or unemployed but also looking after family, they might report one activity in one interview and the other in the next. Alternatively, the respondent's circumstances may be unusually complex and not map onto the available response categories at all.

Question instructions

Burkhead and Coder (1985) illustrated how the question instructions might interact with retrieval, judgement and response selection to create seam effects. SIPP respondents are first reminded of the list of unearned income sources reported at the previous interview, then asked which of these sources they have received at any time during the current reference period and only when asked about the amounts asked to report the precise months of receipt. The authors believe that this "procedure that makes a determination of receipt at any time during the current 4-month reference period before probing for statuses in individual months is probably a major contributor toward the large gross monthly changes from recipient to nonrecipient status" (Burkhead and Coder 1985 p352). Firstly, the question about any receipt during the reference period is not precise enough and does not probe the respondent to think about when changes took place. Secondly, respondents may learn from past interviews that responding "no" to the question about any receipt during the reference period will considerably shorten the interview, as follow-up questions about each individual income source will be skipped. Thirdly, when respondents are asked about any additional income sources, the months and amounts of receipt are recorded in backward chronological order, starting with the most recent month. This, the authors think, "may lead respondents to 'extend' the months of reciprocity beyond the time during which the income was actually received" (p352).

Interviewer and data processing errors

In addition to errors made by the respondent during the process of answering a survey question, interviewer and data processing errors are also thought to contribute to the seam effect. Interviewers may misunderstand or mis-record the response, which could lead to omissions, misdating or misclassifications (Burkhead and Coder 1985; Clarke and Tate 1999; Lemaitre 1992). Imputation for unit or item non-response, which will typically be done independently for each wave, would contribute to misclassifications and omissions (Burkhead and Coder 1985; Moore and Kasprzyk 1984). So would incorrect matching of respondents across waves (Martini and Ryscavage 1991; Moore and Kasprzyk 1984). Finally, errors in coding and variability in items coded to complex frames lead to misclassifications (Kalton, McMillen, and Kasprzyk 1986; Martini and Ryscavage 1991).

4 Hypotheses

The framework implies that 1) the composition of the seam effect (that is, whether the excess of seam transitions is predominantly caused by under-reporting of off-seam change, by misdating of changes to the seam or by spurious seam changes) is likely to be different for different types of items, depending on which types of errors are most frequent and how many reference periods an average spell spans; 2) event types which are more likely to be omitted, misdated or misclassified have larger seam effects; 3) the likelihood of errors is determined by characteristics of the event, the respondent's situation and the question format and wording; and therefore 4) the determinants of errors can be used to make predictions about the expected size of seam effects. These implications lead to the following hypotheses:

Hypothesis 1: The composition of the seam effect is determined by prevalent error type and average spell length.

The seam effect is likely to be predominantly caused by

- under-reporting of within-wave change, if there is a high likelihood that the entire spell is omitted. This is more likely if the spell is contained within one reference period, but also possible if the spell spans multiple reference periods (case 1 in Figure 3);
- misplacement of within-wave change to the seam, if the spell spans two or more reference periods and there is a high likelihood of omission in the reports for the first or last period (cases 2a and 2b), or if the spell is shorter than one reference period and there is a high likelihood of reporting the current receipt for the entire period (case 7);

- spurious seam transitions, if the average spell length spans two or more reference periods and there is a high likelihood of misclassification (cases 5 and 6), or if the spells span three or more periods and there is a high likelihood of omission for periods other than the first or last (case 3).

Hypothesis 2: Event characteristics predict seam effects.

Seam effects are larger for events which

- are less memorable: for example, shorter spells, spells that have ended, events which do not impact on life at present, mildly unpleasant or neutral events, regular/less distinct events;
- are more sensitive: for example, events which are embarrassing or threatening or do not match the respondent's self-image, but where the social desirability bias is not strong enough for the events never to be reported;
- have concepts which are less clear: for example, where different interpretations of which information is required are possible, or where respondents' self-image may change and they therefore report different states at different waves;
- have different definitions over time: for example, State benefits that have changed name or the way they are paid out;
- have fuzzy onsets: for example, spells for which different start dates could be reported, depending on the criteria used to define the start of the spell.

Hypothesis 3: Respondent situation predicts errors contributing to seam effects.

Seam effects are larger for reports by respondents with

- more complex circumstances: for example, respondents experiencing more events of the same type, or more intervening events of different types;
- lower cognitive ability: for example, respondents with worse memory, or less able to compute responses;
- less motivation: for example, respondents who are less interested in the survey topic, or less convinced that the survey is useful.

Hypothesis 4: Question format and wording predict seam effects.

Seam effects are larger for questions for which

- the definition of concepts is unclear: for example, if the question does not define which types of events to include or exclude, if the question does not define how to

identify the date of an event, if the question wording is too difficult for respondents to understand what is required;

- the date request is implicit: for example, asking for elapsed time since an event, duration of a spell or period-status questions, instead of exact date of occurrence;
- response categories are not mutually exclusive: for example, if different categories could adequately describe respondent's situation and the criteria for selecting the 'main' category are ambiguous;
- the implementation encourages errors: for example, features of the instrument that produce interviewer errors, implementation that encourages shortcutting by the interviewer or respondent;
- responses are open ended and involve coding: for example, different coding of the same verbatim description, or different description of the same status.

Testing hypotheses 2, 3 and 4 requires data where the characteristics of one determinant vary in the dimensions identified here, but the other two determinants are held constant. For hypotheses 2 and 3 this condition is fulfilled in any existing survey data: information about *different types of events* is collected from the same sample of respondents using the same question format and wording; information is collected from *different types of respondents* within the sample, about the same types of events and using the same question wording and format. Testing hypothesis 4 however requires experimental data, since surveys do not usually collect information about the same events from the same respondents using different question wording and formats. In the following sections, I test hypotheses 1, 2 and 3 using some examples from the BHPS and suggest an experimental setup that could be used to test hypothesis 4.

5 Illustration

5.1 Composition of the seam effect is determined by prevalent error type and average spell length

The hypothesis tested here is that the composition of the seam effect, that is whether the excess of seam transitions is predominantly caused by under-reporting of off-seam change, by misdated of events or by spurious seam changes, will depend on the mean length of spells relative to the length of the reference period between interviews:

Hypothesis 1a: Seam effects for short spells (relative to the length of the reference period) are more likely to be due to under-reporting of within-wave change and misdating of transitions to the seam. Seam effects for long spells are more likely to be due to spurious seam transitions.

To test this hypothesis I used survey data about the receipt of State benefits and tax credits, linked to individual level administrative records from the Department for Work and Pensions (DWP). The survey data were from the BHPS 1999, 2000, 2001 interviews and the ‘Improving Survey Measurement of Income and Employment’ (ISMIE) follow-up survey in spring 2003. (See Jäckle et al. 2004 for details of the ISMIE survey and record linkage.) The sample used here included only respondents who were successfully linked and who were interviewed in all four panel waves. Histories of income receipt were collected in all four interviews using questions A-Q8 and D-Q9 outlined in Section 2. The administrative records included dates of receipt for 13 types of income sources between January 1999 and the spring 2003 interview date, covering a window of observation of on average 49.5 months. This window covered three seams: for each respondent these were the interview months of the 1999, 2000 and 2001 surveys. I classified Retirement Pensions, different disability related benefits and Child Benefit as long-term benefits (the mean duration during the window of observation according to the administrative records was 41.2 months). Short-term income sources included Income Support, Working Families’ Tax Credit and Job Seeker’s Allowance (mean duration 19.2 months). (Housing Benefit was dropped from the analysis, because of the poor quality of the administrative data.)

Comparing the start dates reported in the survey with the matched administrative data, I classified the survey reports as 1) *correct*, if both the survey and the administrative spell started in an off-seam month, or if both started in a seam month, 2) as *under-reported* off-seam or seam start, if the administrative spell was not reported in the survey, 3) as *misdated*, if the administrative spell started in an off-seam month and the survey spell started in a seam month, or vice versa, and 4) as *spurious* seam or off-seam start, if a survey spell did not correspond to any administrative spell. This classification was repeated for end dates.

The number of spell starts during the window of observation (N=693) exceeded the number of spell endings (N=589), meaning that there were more right than left censored spells in the sample (Table 1). Overall, over half of all transitions were in one of the three seam months: 52.3% of start dates and 55.3% of end dates. In comparison in the administrative data only around 7% of all transitions occurred in a seam month. This is close

to the expected proportion. If transitions were uniformly distributed and there were no errors leading to seam effects, we would expect $3/49.5=6.1\%$ of transitions to be at a seam.

The magnitude of the seam effect was larger for long than short-term benefits: around 62-64% of start and end dates of long-term spells were at a seam, compared to 43-46% of short-term benefits. This is reflected in the finding that a smaller proportion of the long-term benefits were reported correctly.

Among the mis-reported spells, short-term spells were more likely to be under-reported or misdated; long-term spells were more likely to have spurious transitions. Summing up the different types of errors in start dates, 73.3% of short-term spells were either under-reported or misdated, the remainder 26.7% were spurious. For long-term spells 37.3% were under-reported or misdated and 62.7% spurious (Pearson $\chi^2(1)=43.3$, $P=0.000$). The composition of errors in end dates was similar: 59.7% of short-term spells were under-reported or misdated, compared to 19.5% of long-term spells (Pearson $\chi^2(1)=64.6$, $P=0.000$).

Table 1: Composition of errors contributing to the seam effect

	Start dates			End dates		
	Short-term	Long-term	Total	Short-term	Long-term	Total
N spells	354	339	693	369	220	589
% Total changes at seam	42.9	61.5	52.3	45.8	64.2	55.3
% Spells correctly reported	37.5	25.6	32.0	25.5	5.2	16.1
Composition of errors:						
% Under-reported off-seam change	51.5	24.7	38.1	33.2	10.0	21.0
% Under-reported seam change	2.4	1.8	2.1	2.2	0.5	1.3
% Misplaced to seam	18.2	10.2	14.2	21.6	8.0	14.4
% Misplaced to off-seam	1.2	0.6	0.9	2.8	1.0	1.8
% Spurious seam change	21.8	53.0	37.5	21.0	53.0	37.8
% Spurious off-seam change	4.9	9.6	7.3	19.3	27.5	23.6

Notes: The base included all spells in the survey, plus spells in the administrative records under-reported in the survey.

Source: BHPS 1999, 2000, 2001, ISMIE 2003 and linked administrative records.

These findings provide some support for the hypothesis that the composition of seam effects depends on the mean length of spells relative to the length of the reference period.

5.2 Event characteristics predict seam effects

The hypothesis tested next, is that the likelihood of errors of omission, misdating and misclassification, and hence the magnitude of seam effects, is determined by characteristics

of the event. The characteristics examined are 1) factors affecting the memorability of an event, in this case, whether events are of a positive or negative nature, and 2) the ambiguity of event dates, that is, whether events have clearly defined or fuzzy onsets. Cognitive theories of autobiographic recall suggest that positive events are recalled more precisely than negative events (Skowronski et al. 1991). Similarly, we would expect the task of dating events with clearly defined onsets to be easier than the dating of fuzzy events. This results in the following hypotheses:

Hypothesis 2a: Positive events are reported more accurately and hence have smaller seam effects than negative events.

Hypothesis 2b: Events with clearly defined onsets are dated more accurately and hence have smaller seam effects than events with fuzzy onsets.

To test these hypotheses, I used the labour market activity histories from the BHPS. At each wave respondents were asked a showcard question about their current activity. The possible response options were: self-employed, employed, unemployed, retired, on maternity leave, looking after family, full-time student, long-term sick or on a government training programme. Respondents were then asked a time-of-occurrence question (B-Q6 and B-Q7 in Section 2), to determine whether the current activity had started during the reference period. If yes, respondents were asked what their previous main activity had been and a time-of-occurrence question about the start of that activity. These retrospective questions were repeated, until an activity spell was reported which had started before the reference period. The information from successive interviews, about current activities and retrospective histories between interviews, can be combined to construct histories of labour market activities that cover the duration of the panel.

With this type of question, an excess of seam transitions appears if some activity spells are omitted or misclassified, or if dates of changes are misreported. Respondents may, for example, report different current activities at successive interviews, without reporting any change for the period between interviews. This could be because they define the same situation differently at different points in time, for example, saying in one wave that they are long-term sick and in the next that they are looking after family. Misclassification can also occur if interviewers misunderstand or mis-key the activity type. Alternatively, respondents may forget to report an activity or not report it because they misdate it as having occurred outside the reference period. This is particularly likely for transitions between activities with fuzzy onsets. For example, transitions between unemployment and looking after family are

likely to be fuzzy, since they are defined by the date at which respondents start or stop looking for work, which may be a gradual process rather than a precise date.

Table 2: Percentage of transitions between labour market activities observed at seams

Status of origin	Event								
	Self-empl.	Empl.	Unemp.	Retired	Mat. leave	Family care	Ft student	Lt sick	Govt. training
Self-employed	–	63.1 (1,415)	30.4 (309)	54.8 (31)	75.0 (8)	61.2 (227)	35.7 (28)	49.2 (61)	100.0 (1)
Employed	53.1 (1,518)	–	17.2 (3,592)	18.8 (165)	74.7 (669)	35.5 (1,729)	42.7 (529)	54.6 (533)	47.6 (42)
Unemployed	31.6 (402)	22.1 (3,863)	–	85.0 (40)	45.5 (22)	91.6 (536)	45.5 (200)	76.0 (358)	22.4 (143)
Maternity leave	58.8 (17)	63.6 (547)	18.2 (22)	–	–	60.1 (148)	25.0 (4)	80.0 (5)	0.0 (0)
Family care	45.0 (240)	32.0 (2,005)	90.1 (503)	96.5 (85)	97.5 (40)	–	36.8 (125)	91.0 (289)	30.8 (13)
Full-time student	20.0 (40)	26.5 (992)	15.8 (379)	0.0 (0)	75.0 (4)	52.0 (102)	–	57.1 (21)	66.7 (9)
Long-term sick	42.1 (57)	32.6 (322)	67.4 (273)	87.1 (93)	66.7 (3)	91.3 (242)	31.8 (22)	–	78.6 (14)
Govt. training	27.3 (11)	29.3 (82)	41.0 (117)	100.0 (1)	0.0 (0)	30.8 (13)	92.9 (14)	75.0 (12)	–

Source: Pooled transitions from BHPS waves 1991 to 2005, including extension samples.

Sample size: 95,421 wave-pairs for 23,283 respondents.

Notes: Total number of transitions in parentheses.

Table 2 shows the possible transitions between the nine different labour market activities. The rows correspond to the status of origin and the columns to the events. (Transitions from retirement into other activities are not shown, although they are observed in the data.) The table shows the percentage of all observed transitions of a particular type, which were observed at a seam. The numbers were derived from histories of labour market activities based on the 1991 to 2005 BHPS data. The sample included data from 95,421 pairs of waves for 23,283 respondents. For each wave, the sample was restricted to respondents of working age, in this case aged between 20 and 55 at the date of interview. For each transition between two activity types, an indicator was derived that showed whether the date was in a seam or off-seam month. If the date reported at $t+1$ was after the wave t interview, the transition was classified as off-seam change; if the date was misdated to before the wave t interview, or if no change had been reported at all, it was classified as a seam change. (Less than 1% of transitions had to be dropped from the analysis, because the start month was missing and the year was the same as the interview year, making it impossible to tell whether the change would have been on or off the seam.) The transition and seam change indicators

were used to compute the proportion of all transitions of a particular type observed at a seam. If transitions were uniformly distributed and there were no errors leading to seam effects, we would expect to observe 1 in 12 (8.3%) of all transitions at a seam.

The possible transitions between labour market activities from the BHPS questions can be classified according to whether the onset of a state is clearly defined or likely to be fuzzy, and whether an event is likely to be perceived as positive or negative. Transitions from unemployment to employment are likely to be positive events, while the reverse transition is probably a negative event. For both events the dates are clearly defined. Table 3 shows the results of testing hypothesis 2a and 2b. Hypothesis 2a was tested using a χ^2 test of whether the distribution of seam and off-seam transitions was similar for negative and positive events with clear transition dates. The hypothesis was not supported: seam effects were smaller for transitions into unemployment (17.2% of total transitions at the seam) than for transitions into employment (22.1%, $P=0.000$). This is nonetheless consistent with the hypothesis that the nature of spells matter: respondents were more likely to report a previous activity if the activity was positive (i.e. employment) than if it was negative (i.e. unemployment). Unfortunately this was the only pair of transitions that could be compared in terms of saliency. There were likely differences in saliency for some of the other types of transitions. In all cases, these were however confounded with differences in whether the dates were clearly defined in time.

The labour market histories offered more potential for testing hypothesis 2b. Transitions from employment to unemployment or retirement are clearly dated in time. The same transitions from self-employment are likely to be fuzzier, for example, if the self-employed retire gradually over a period of time. Similarly, transitions into employment from being long-term sick are likely to be dated clearly in time. Transitions out of employment into long-term sickness may be less clearly defined, for example, if employees have periods of sick leave before leaving employment completely. (One could also argue that transitions into employment are more socially desirable and therefore more likely to be reported correctly than transitions into long-term sickness. The expected effect would be the same.) Finally, within the different inactivity states, there are some for which the start and end dates are clearly defined in time: transitions into and out of full-time education and government training programmes, as well as transitions from employment into retirement. For all possible transitions between unemployment, maternity leave, looking after family and long-term sickness, the transitions are likely to be fuzzy.

Table 3: Tests of hypothesis 2a and 2b

H	% of transitions at seams (N)		Test of independence ¹
2a	Negative events: Emp to Unemp	17.2 (3,592)	$\chi^2(1) = 34.1, P = 0.000$
	Positive events: Unemp to Emp	22.1 (3,863)	
2b	Fuzzy dates: Semp to Unemp	30.4 (309)	$\chi^2(1) = 32.8, P = 0.000$
	Clear dates: Emp to Unemp	17.2 (3,592)	
2b	Fuzzy dates: Semp to Retired	54.8 (31)	$\chi^2(1) = 17.7, P = 0.000$
	Clear dates: Emp to Retired	18.8 (165)	
2b	Fuzzy dates: Emp to LtSick	54.6 (533)	$\chi^2(1) = 47.1, P = 0.000$
	Clear dates: LtSick to Emp	32.6 (322)	
2b	Fuzzy dates: transitions between Unemp, Family, Matern, LtSick	83.2 (2,441)	$\chi^2(1) = 599.2, P = 0.000$
	Clear dates: transitions into/off School, Training, Retirement	42.4 (1,411)	

Source: BHPS waves 1993 to 2006.

Notes: ¹Pearson χ^2 test of independence between the proportion of transitions in seam and off-seam months and event characteristics, adjusted to account for clustering of transitions within respondents. Calculated using the `-svy-` commands in Stata.

The results of testing hypothesis 2b, that seam effects are smaller for events with clearly defined dates than for events with fuzzy dates, suggest support for this hypothesis. The seam effects were larger for transitions from self-employment into unemployment (30.4%) or retirement (54.8%), than for transitions from employment into unemployment (17.2%, $P=0.000$) or retirement (18.8%, $P=0.000$). Similarly, seam effects were larger for transitions from employment to long-term sickness (54.6%) than the other way round (32.6%, $P=0.000$). Finally, seam effects for all transition types between unemployment, maternity leave, looking after family and long-term sickness were larger (83.2%) than between these activities and moving into retirement or into or out of full-time education, government training programmes (42.4%, $P=0.000$).

The results provide some support for the hypothesis that event characteristics predict the magnitude of seam effects. Findings for the hypothesis that positive events are reported more accurately than negative events were somewhat mixed. The hypothesis that events with clearly defined transition dates are reported more accurately than events with fuzzy dates was however clearly supported.

5.3 Respondent situation predicts errors contributing to seam effects

The third hypothesis is that the respondent's situation is predictive of how the response process is executed and therefore of the likelihood and extent of errors leading to seam effects. The specific hypotheses tested here are:

Hypothesis 3a: Respondents experiencing more complex situations face a more difficult reporting task, and are therefore more likely to mis-recall some aspects of their experiences during the reference period. As a result, I expect seam effects to be larger for respondents reporting on more complex situations.

Hypothesis 3b: For respondents with lower cognitive ability, recalling and reporting retrospectively on events during the reference period is likely to be a more difficult task and they are therefore more likely to misreport. As a result, I expect seam effects to be larger for respondents with lower cognitive ability.

Hypothesis 3c: Respondents who are less motivated to participate in the survey are less likely to expend the required effort to recall and report retrospective histories about events during the reference period. Instead they are more likely to shortcut the response process and, for example, use simple heuristics instead of direct recall, making it more likely that their reports will contain errors. As a result, I expect seam effects to be larger for respondents with lower motivation.

To test these hypotheses I again used the BHPS survey reports on unearned income sources, linked to administrative records. I used the administrative records to derive the number of benefit and tax credit spells the respondent experienced during the window of observation, as a measure of the complexity of the respondent's situation. That is, I used the number of spells during the entire 49.5 month window of observation to proxy for the difficulty of the reporting task in each of the four interviews. This was based on findings about autobiographic memory that the more events a respondent has experienced, the more difficult it is to recall and report on any one of them correctly (Eisenhower, Mathiowetz, and Morganstein 1991). In the sample of survey respondents matched to administrative records (N=361), 41.0% of respondents had experienced only one spell during the entire window of observation, 26.9% experienced two spells, 20.1% three spells, and 11.4% four or more spell.

Measures of cognitive ability were unfortunately not available from the survey. Instead I used education as a proxy. The survey classified highest educational qualifications as first or higher degree, teaching, nursing or other higher qualification, secondary school, commercial qualifications, apprenticeship or none of these. The survey sample linked to administrative records used here had a large proportion of respondents without any of these

qualifications (45.4%), since the sample over-represented low income groups (See Jäckle et al. 2004 for details). I therefore used a binary indicator set to 1 if the respondent had any qualifications and 0 otherwise, to proxy for the respondent's cognitive ability.

Some indicators of respondent motivation were measured by the interviewer observations recorded at the end of each interview. For example, interviewers were asked to judge the respondent's cooperativeness with the following question: "In general, the respondent's co-operation during the interview was <very good, good, fair, poor, very poor>"? From this I created a binary indicator of cooperativeness for each respondent, coded 1 if cooperativeness was 'very good' in each of the four interviews (70.6%) and 0 otherwise. This indicator was used as a proxy for respondent motivation.

To test hypotheses 3a, 3b and 3c, I used the sample of unearned income receipt spells reported in the survey (882 spells reported by 361 respondents). Each spell starting during the window of observation, that is, excluding left censored spells, was classified as to whether the start date was in a seam or an off-seam month. Similarly, each spell ending during the window of observation, that is, excluding right censored spells, was classified as to whether the end date was in a seam or off-seam month. This classification resulted in 52.3% of start dates at a seam (N=354 non-left censored spells) and 55.3% of end dates (N=369 non-right censored spells). The hypotheses were tested by testing for independence between the proportion of start or end dates at a seam and the indicators of respondent ability, motivation and complexity of situation. The test of independence accounted for the clustering of spells within respondents, using the `-svy-` commands in Stata. The seam effect was judged to be larger in size for one group, if the proportion of transitions at the seam was larger than for the other group.

Hypothesis 3a that the size of the seam effect, measured as the proportion of total transitions observed at a seam, increases with the complexity of the respondent's situation was not supported (Table 4). There did not appear to be any clear association between the size of the seam effect and the number of spells experienced.

Hypothesis 3b that the size of seam effects is smaller for respondents with higher cognitive ability seemed to be supported. For both start and end dates the proportion of transitions at a seam was lower for respondents with educational qualifications. For start dates the difference was 23 percentage points and significant. For end dates the difference was smaller at around 7 percentage points and not significant.

Hypothesis 3c that the size of seam effects is smaller for more motivated respondents was weakly supported. More co-operative respondents had a smaller proportion of both start and end dates at a seam. Neither differences were however significant.

Table 4: Percentage of total transitions at a seam, by respondent situation

	% of Start Dates at Seam	% of End Dates at Seam
1 Spell	56.1	46.5
2 Spells	51.1	61.8
3 Spells	59.2	56.0
4+ Spells	41.6	55.0
Test of independence ¹	F(2.97,615.03)=1.91, P=0.1274	F(2.99,579.63)=1.19, P=0.3133
No qualifications	65.7	59.6
Secondary school or higher	42.9	52.1
Test of independence ¹	F(1,204)=15.05, P=0.0001	F(1,192)=1.98, P=0.1608
Not always 'very co-operative'	56.4	57.3
Always 'very co-operative'	49.6	53.9
Test of independence ¹	F(1,204)=1.16, P=0.2835	F(1,190)=0.35, P=0.5521

Source: BHPS 1999, 2000, 2001, ISMIE 2003.

Notes: ¹ Pearson χ^2 test of independence between the proportion of transitions at the seam and respondent characteristics, adjusted to account for clustering of spells within respondents. Calculated using the `-svy-` commands in Stata.

These findings provide some support for the hypothesis that the magnitude of seam effects varies with the respondents' situations and characteristics.

5.4 Question format and wording predict seam effects

The question format features for which I suggest testable hypotheses and experimental questions, that could be used to test these, are the definitions of events and instructions for the dating of events. The hypotheses are:

Hypothesis 4a: Questions about events with unambiguous definitions have smaller seam effects than questions about events with ambiguous definitions.

Hypothesis 4b: Questions that explicitly ask for the date of occurrence of an event have smaller seam effects than questions that collect dates implicitly by asking elapsed time or period-status questions.

Testing these hypotheses would require experimental data, where different randomly allocated treatment groups are allocated questions about the same events, but using different definitions (H4a) or dating instructions (H4b). The following are suggestions for experimental questions that could be used.

Definitions of events

As an example, in the BHPS marital history data there appear to be two types of definitional problems: the definition of ‘never married’, which seems to be confused with a ‘single’ category that is not offered, and the definition of when to report separation and divorce. As Fowler (1995 p13) said, “one basic part of having people accurately report factual or objective information is ensuring that all respondents have the same understanding of what is to be reported, so that the researcher is sure that the same definitions have been used across all respondents” – and in a panel context, that respondents have used the same definitions across waves! There are two basic approaches of trying to ensure consistent understanding of terms: either 1) the researcher provides complete definitions, so that all or most of the ambiguities about what is called for are resolved, and the respondent is asked to do the classification work, or 2) the respondent is asked to provide all the information needed in order for the researcher to properly classify information. In the first case, the needed definitions are built into the questions. But when definitions are too complex, a simpler alternative is to add some extra questions to cover commonly omitted kinds of events. In this case, it is not necessary to communicate a complex definition consistently to all respondents. Instead respondents are asked a series of simple questions covering all aspects of what is to be reported and the researcher later applies a classification. According to Fowler this is a sound way of solving many definitional problems: “if investigators identify what simple, easy questions people can answer that will provide the basis for classification, on many occasions better measurement will occur.” (Fowler 1995 p20).

Improving the consistency with which legal marital status is classified would either require a more detailed definition of ‘never married’ and of the stage at which a divorce or separation should be reported, or breaking the question up into a series of unambiguous questions. A suggestion for the second approach would be to replace the question about current status (Q3 in Section 2) with a question about whether the respondent has ever been married (Q14 below), which could then be followed by a series of yes/no questions about the current status of respondents who *have* been married (Q14). The yes/no questions would end as soon as the respondent first replies ‘yes’. That is, respondents who for example say ‘yes’ they are widowed, would then not be asked whether they are divorced, separated or married:

Q14. Have you ever been married? [yes/no]

IF YES:

Q15. What is your current legal marital status, are you...

ASK UNTIL RESPONDENT FIRST REPLIES ‘YES’

... *Widowed? [yes/no]*

... *Divorced or in the process of divorcing?*

[yes, divorced / yes, in process of divorcing / no]

... *Separated or in the process of separating?*

[yes, separated / yes, in process of separating / no]

... *Married? [yes/no]*

These questions would be simpler for respondents to answer than if they were given more detailed definitions and asked to pay more attention to the classification. The questions imply a priority of different legal marital statuses, which can be classified unambiguously by the researcher: only respondents who answer Q14 with 'No' would be classified as 'never married'. All other single respondents would be classified according to their response to Q15. As a result I would expect the classification of 'never married' to be more consistent across waves and therefore to see a smaller number of legally impossible transitions with this question than with the original BHPS question. Similarly, the researcher could identify whether a divorce or separation has been completed. If it is still in process, the respondent could be classified as 'married'. Question A-Q4 in Section 2, about whether the respondent's legal status has changed during the reference period, could then be asked unchanged. Presumably, respondents who at Q15 report being in the process of a divorce or separation would not report a change in their legal marital status, while respondents reporting a completed divorce or separation would. In sum, I would expect to see smaller seam effects with this experimental version of the marital history questions than the standard BHPS version.

Dating instructions

Evidence from studies in which respondents are asked to produce dates of personal events, and are then asked how they arrived at that date, suggests that time information is not stored and recalled directly from memory, but instead reconstructed. As Tourangeau, Rips, and Rasinski (2000 p109) said, "Information of various types is drawn from memory [...] to establish initial constraints on the answer to a time question. Respondents add new information as it becomes available to constrain the answer further until they have satisfied the criteria for an acceptable response."

This suggests that the respondent's understanding of what constitutes an acceptable response is crucial. As Burkhead and Coder (1985) already stated, period-status questions (where dates of changes are only requested implicitly and derived from changes in the status

reported for each sub-period) are not precise enough and do not make it clear that the respondent is expected to provide an exact date.

A test of whether time-of-occurrence questions, which explicitly ask for the dates of events, produce smaller seam effects than period-status questions could be the following. For all income sources reported as received during the reference period (A-Q8 in Section 2), respondents could be asked the following, instead of the period-status question D-Q9:

Q16. Were you receiving X at <start of reference period / date of last interview>?

IF NO:

Q17. In which month since <start of reference period / date of last interview> did you start receiving X?

ASK ALL:

Q18. In which month did you stop receiving X?

IF STILL RECEIVING – ASK Q8A FOR NEXT INCOME SOURCE

IF REPORTS END DATE:

Q19. In which month did you start receiving X again after that?

IF NOT RECEIVED AGAIN – ASK Q8A FOR NEXT INCOME SOURCE

IF REPORTS START DATE, ASK Q8C ETC.

The expectation is that asking respondents explicitly to report dates would make them think more carefully about the timing of events. For example, it would be less easy for them to give a constant wave response by reporting receipt for “all months”. As a result, I would expect to see smaller seam effects with these experimental questions than with the original BHPS question (D-Q9).

6 Conclusion

The framework of the causes of seam effects proposed in this paper (and summarized in Figure 2) assumes that characteristics of the event in question, the respondent’s situation and question format and wording determine the likelihood and nature of errors during the survey and response process. The errors, of omission, misdating and misclassification, in turn interact with the mean length of spells relative to the length of the reference period between interviews, to produce under-reported off-seam changes and over-estimated seam changes. These combine to produce the characteristic seam effects, or concentrations of transitions at the seam between reference periods, in event data from panel interviews. All existing theories of the causes and findings of the empirical nature of seam effects are readily incorporated into this framework.

The implications from the framework were mostly supported in empirical tests using data from the BHPS. First, the composition of the seam effect was clearly different for spells with different lengths relative to the reference period: seam effects in long-term spell types were mainly due to spurious seam changes, caused by omission of part of the spell; seam effects in short-term spell types were mainly due to under-reported off-seam changes, caused by omissions of entire spells, and changes that were misdated to the seam. Second, the characteristics of events were associated with the magnitude of seam effects: event types with fuzzy transition dates were dated less accurately and had larger seam effects than event types with clearly defined dates. Third, some aspects of the respondent's situation were predictive of the magnitude of seam effects: respondents with higher cognitive ability tended to have smaller seam effects than respondents with lower ability; respondents that appeared to be more motivated to participate in the survey tended to have smaller seam effects than less motivated respondents (although this difference was not significant). The fourth hypothesis, that questions providing clear definitions and questions asking about the date of occurrence of events explicitly would have smaller seam effects than questions that are less well defined or ask about the timing of events implicitly, could not be tested using existing data. Instead I proposed some experimental questions that could be used to perform such tests.

The framework and empirical tests provide some evidence why previous studies searching for correlates of seam effects have not found any conclusive results (e.g. Goudreau, Oberheu, and Vaughan 1984; Hill 1987; Weidman 1987). These studies typically grouped measures of change for different types of events and tested for associations of the probability of seam transitions with socio-demographic characteristics of respondents. The framework of the causes of seam effects, however, suggests that causes of errors in the survey and response processes leading to seam effects may be quite different for different types of events. For different types of events, the respondent characteristics associated with higher probabilities of misreporting are therefore likely to be different. These differences are however obscured in analyses of seam effects where different types of events are aggregated.

The framework also provides some clues as to why methods such as dependent interviewing appear to successfully reduce seam effects for some types of items but not for others. Dependent interviewing, where respondents are reminded of previous information, or where survey responses are compared with previous responses during in-interview edit checks, has been shown to reduce seam effects in labour market activity histories (Jäckle and Lynn 2007; Murray et al. 1991) collected with questions such as B-Q6 and B-Q7. Applying similar methods to histories of income receipt, collected with questions such as A-Q8 and D-

Q9, has however not been successful in reducing seam effects (Moore et al. in press). In both cases, reminding respondents of previous answers or verifying whether apparent changes are true reduces under-reporting and inconsistent classification. For labour market activities it appears that these are the main causes of seam effects. For income sources, the dating of events appears to be an additional problem. The income source data are collected with period-status questions, which only ask for dates implicitly. The labour market activities are collected with explicit time-of-occurrence questions and therefore likely to produce more accurate dates. Dependent interviewing, as implemented, helps respondents improve the reporting, but not the dating of events. As a result, the problems leading to seam effects in income histories are not resolved. In some ways, dependent interviewing instead makes things worse: it reminds respondents to report short spells they are otherwise likely to under-report. But at the same time it does not reduce the tendency for short-spells to be misdated. The result is that estimates of spell durations based on dependent interviewing data are more biased than estimates of spell durations based on independently collected data (Jäckle 2008).

Although the framework was explicitly derived to explain the causes of seam effects, that is, of biases in transitions from retrospective questions about events between interviews, the framework can also be used to think about biases in wave-on-wave transitions. The bulk of questions in the household panel surveys reviewed for this study ask only about the respondent's situation at the time of each interview, without collecting information about events between interviews. The resulting repeated measures can also be used to construct measures of change, albeit only of wave-on-wave change, where each interview pair could be thought of as a seam between reference periods. Respondents are not asked to date events and the discussion of dating errors is therefore not relevant. Errors of omission or misclassification can however occur in the same way as for retrospective event histories. As a result, measures of change can also be (grossly) over-estimated. Lynn and Sala (2006), for example, showed that change in employment characteristics, such as occupation, industry or the size of the organisation, from one wave to the next was implausibly high. The framework could therefore also be useful in thinking about the design of questions used to collect wave-on-wave measures of change.

Ideally then, devising data collection methods to reduce biases in estimates of change would involve several steps, which are summarized in Figure 2: first, identifying the types of errors (omission, misdating or misclassification) that are most likely in the measurement of a particular type of event. Second, mapping out details of the likely predictors of these errors, including characteristics of the event, the respondent's situation and the question format.

Third, assessing the likely impact, in terms of the relative frequency, of the different potential error sources, to enable decisions about which problems to focus on. Fourth, adapting the question format and wording to include cues that address the most important problems causing errors that lead to biases in estimates of change.

References

- Burkhead, D., and J. Coder. 1985. Gross Changes in Income Reciprocity from the Survey of Income and Program Participation. Paper read at Proceedings of the Social Statistics Section, at Washington DC.
- Callegaro, Mario. 2007. Seam Effects Changes due to Modifications in Question Wording and Data Collection Strategies. A Comparison of Conventional Questionnaire and Event History Calendar Seam Effects in the PSID, Survey Research and Methodology Program, University of Nebraska, Lincoln.
- Cannell, Charles F., Peter V. Miller, and Lois Oksenberg. 1981. Research on Interviewing Techniques. *Sociological Methodology* 12:389-437.
- Clarke, P. S., and P. F. Tate. 1999. Methodological Issues in the Production and Analysis of Longitudinal Data from the Labour Force Survey. In *Methodology Series No. 17*. London: Government Statistical Service.
- Conrad, Frederick G., N. R. Brown, and E. R. Cashman. 1998. Strategies for Estimating Behavioral Frequencies in Survey Interviews. *Memory* 6:339-366.
- Eisenhower, D., N.A. Mathiowetz, and D. Morganstein. 1991. Recall Error: Sources and Bias Reduction Techniques. In *Measurement Errors in Surveys*, edited by P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz and S. Sudman. New York: Wiley.
- Fowler, Floyd J. 1995. *Improving Survey Questions, Applied Social Research Methods Series Volume 38*. Thousand Oaks, CA: Sage.
- Glasner, Tina, and Wander van der Vaart. 2007. Applications of Calendar Instruments in Social Surveys: A Review. *Quality and Quantity*.
- Goudreau, K., H. Oberheu, and D. Vaughan. 1984. An Assessment of the Quality of Survey Reports of Income from the Aid to Families with Dependent Children (AFDC) Program. *Journal of Business and Economic Statistics* 2 (2):179-186.
- Hill, Daniel H. 1987. Response Errors Around the Seam: Analysis of Change in a Panel with Overlapping Reference Periods. Paper read at Proceedings of the Survey Research Methods Section, at Washington DC.

- Jäckle, Annette. 2008. Measurement Error and Data Collection Methods: Effects on Estimates from Event History Data. In *ISER Working Paper, 2008-13*. Colchester: University of Essex.
- Jäckle, Annette. in press. Dependent Interviewing: A Framework and Application to Current Research. In *Methodology of Longitudinal Surveys*, edited by P. Lynn. Chichester: Wiley.
- Jäckle, Annette, and Peter Lynn. 2007. Dependent Interviewing and Seam Effects in Work History Data. *Journal of Official Statistics* 23 (4):529-551.
- Jäckle, Annette, Emanuela Sala, Stephen P Jenkins, and Peter Lynn. 2004. Validation of Survey Data on Income and Employment: the ISMIE Experience. In *ISER Working Paper*. Colchester: University of Essex.
- Kalton, Graham, D. B. McMillen, and D. Kasprzyk. 1986. Nonsampling Error Issues in the SIPP. In *SIPP Working Paper No. 10*. Washington DC: US Bureau of the Census.
- Kalton, Graham, and Michael E. Miller. 1991. The Seam Effect with Social Security Income in the Survey of Income and Program Participation. *Journal of Official Statistics* 7 (2):235-245.
- Krosnick, J. 1991. Response Strategies for Coping With the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology* 5:213-236.
- Lemaitre, Georges. 1992. Dealing with the Seam Problem for the Survey of Labour and Income Dynamics. In *SLID Research Paper Series*. Ottawa: Statistics Canada.
- Lynn, Peter, Annette Jäckle, Stephen P. Jenkins, and Emanuela Sala. 2004. The Impact of Interviewing Method on Measurement Error in Panel Survey Measures of Benefit Receipt: Evidence from a Validation Study. In *ISER Working Paper*. Colchester: University of Essex.
- Lynn, Peter, and Emanuela Sala. 2006. Measuring Change in Employment Characteristics: The Effects of Dependent Interviewing. *International Journal of Public Opinion Research* 18 (4):500-509.
- Marquis, Kent H., Jeffrey C. Moore, and Vicki J. Huggins. 1990. Implications of SIPP Record Check Results for Measurement Principles and Practice. Paper read at Proceedings of the Survey Research Methods Section, at Alexandria, VA.
- Martin, Elizabeth. 2001. Remarks on Seam Effects for Quantitative and Qualitative Facts. Paper read at Seminar on the Funding Opportunity in Survey Research, at Washington DC.

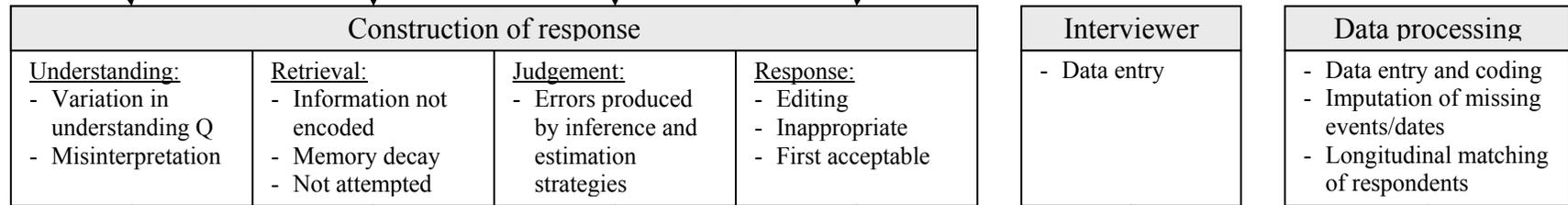
- Martini, A., and P. Ryscavage. 1991. The Impact of Survey and Questionnaire Design on Longitudinal Labor Force Measures. Paper read at Annual Research Conference of the US Bureau of the Census, at Arlington, VA.
- Martini, Alberto. 1989. Seam Effect, Recall Bias, and the Estimation of Labor Force Transition Rates from SIPP. Paper read at Proceedings of the Survey Research Methods Section, at Washington DC.
- Moore, J., and D. Kasprzyk. 1984. Month-to-Month Reciprocity Turnover in the ISDP. Paper read at Proceedings of the Survey Research Methods Section, at Washington DC.
- Moore, Jeffrey C., Nancy Bates, Joanne Pascale, and Aniekan Okon. in press. Tackling seam bias through questionnaire design. In *Methodology of Longitudinal Surveys*, edited by P. Lynn. Chichester: Wiley.
- Murray, T. Scott, Sylvie Michaud, Mike Egan, and Georges Lemaitre. 1991. Invisible Seams? The Experience with the Canadian Labour Market Activity Survey. Paper read at Proceedings of the US Bureau of the Census Annual Research Conference, at Washington DC.
- Paull, Gillian. 2002. Biases in the Reporting of Labour Market Dynamics. In *Working Paper*. London: Institute for Fiscal Studies.
- Rips, Lance J., Frederick G. Conrad, and Scott S. Fricker. 2003. Straightening the Seam Effect in Panel Surveys. *Public Opinion Quarterly* 67 (4):522-554.
- Skowronski, John J., Andrew L. Betz, Charles P. Thompson, and Laura Shannon. 1991. Social Memory in Everyday Life: Recall of Self-Events and Other-Events. *Journal of Personality and Social Psychology* 60 (6):831-843.
- Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. New York: Cambridge University Press.
- Weidman, Lynn. 1987. Examination of Relationships Between Actual and Reported Changes in the SIPP. Paper read at ASA Survey Research Methods Section.
- Willis, Gordon. 2001. Discussion Rips et al.: Seam Effects for Quantitative and Qualitative Facts. Paper read at Seminar on the Funding Opportunity in Survey Research, at Washington DC.
- Young, Nathan. 1989. Wave-Seam Effects in the SIPP. Paper read at ASA Survey Research Methods Section.

Figure 2: A framework of the causes of seam effects

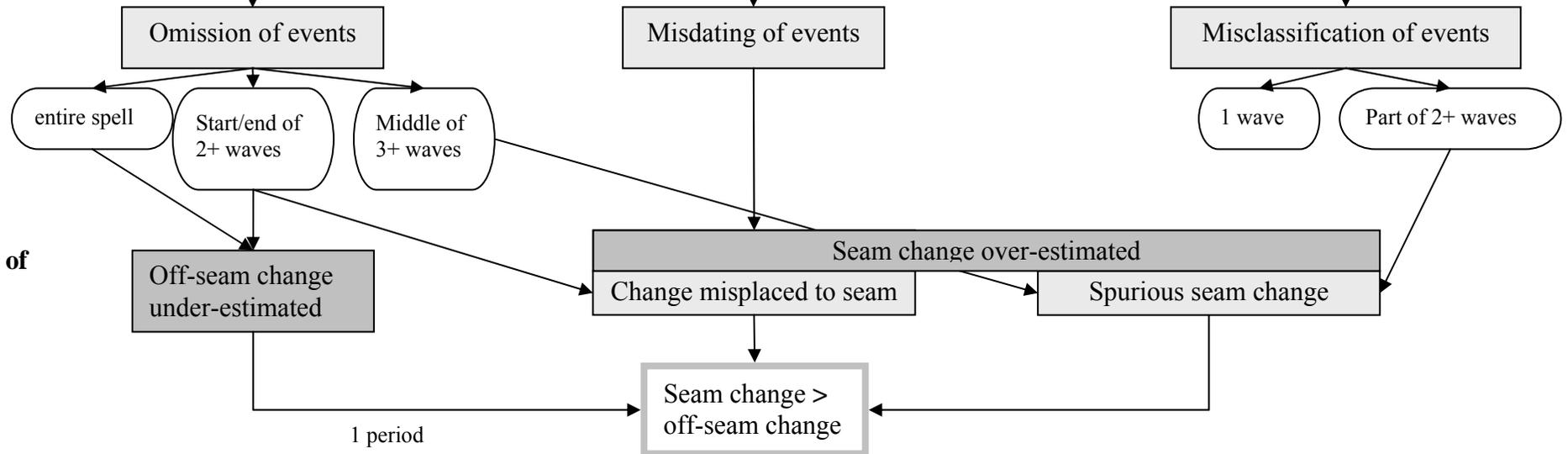
Characteristics of

Event:	- Ambiguous dating of transitions	- Memorability - Elapsed time - Memory encoding	- Task difficulty	- Perceived risk
Question:	- Question wording - Reporting instructions - Clarity of concepts and definitions	- Respondent rules (proxy reporting) - Question terms vs. memory encoding - Cues	- Type of retrieved information - Reporting instructions - Effort/Motivation	- Ambiguity of main category, if not mutually exclusive - Order of categories - Open ended
Respondent:	- Ability and effort	- Ability and effort		- Ability and effort

Survey and response process:



Resulting errors
x
Spell length



Estimates of change