Panel Conditioning in the Consumer Expenditure Quarterly Interview Survey

Ting Yan[1], Kennon Copeland[1]
[1]NORC at the University of Chicago, 55 E. Monroe Street, Chicago, IL, 6063

Abstract

Panel conditioning is one of important sources of measurement error unique to panel surveys. It refers to changes to either respondents' true value or their report of the true value caused by being interviewed multiple times. Earlier studies on expenditures found that people interviewed for a second time reported significantly less expenditures than those interviewed for the first time. This paper examines panel conditioning effects in the Consumer Expenditure Interview Survey and found no evidence of panel conditioning in the total amount of expenditure reported and the different number of expenditure types reported. In addition, reluctant respondents and respondents with higher reporting burden did not seem to be more prone to panel conditioning effects than cooperative respondents or those with lower reporting burden.

Key Words: panel conditioning, longitudinal surveys, measurement error, Consumer Expenditure Quarterly Interview Survey, reluctant respondents, reporting burden

1. Introduction

Panel conditioning is a source of measurement error unique to longitudinal surveys or panel surveys where sample members are subject to multiple rounds of interviews. It refers to the phenomenon where panel participation in repeated interviews changes respondents' behavior and attitudes or their report of their behavior and attitudes. Even though investigation of panel conditioning effects has been conducted as early as 1950s, literature on panel conditioning is quite thin and lacks definitive conclusions on the presence, the direction, the magnitude, and the mechanisms of panel conditioning effects (Bailar, 1989; Sturgis, Allum, & Brunton-Smith, 2009).

Typically, examinations of panel conditioning effects on behavioral/factual questions other than voting behaviors have failed to yield strong evidence of the presence of conditioning effects; in cases where conditioning effects were found, they tended to be small (sometimes, even negligible) and biased downward against later waves. For instance, Bailar (1975) found that sample persons in housing units interviewed for the first time produced an estimate of the number of unemployed 20% higher than the average of estimates from all rotation groups and that the estimate from the last wave is 9% lower than the average of all rotation groups. On examining home alterations and repairs expenditure, Neter and Waksberg (1964) showed some evidence of panel conditioning effects that led to fewer jobs and expenditures reported in the 3rd wave as compared to the reports in the 2nd wave. However, only 6 comparisons (out of 36) between responses at the 2nd and 3rd waves achieved statistical significance (see Table 1 in Neter & Waksberg, 1964). Similarly, Wang, Cantor, and Safir (2000) found 7 significant differences among 32 comparisons between responses from respondents who participated in an earlier round of survey and those from a fresh sample; but only 4 remained significant after controlling demographic differences. Corder and Horvitz

(1989) did not observe any obvious conditioning effect in the National Medical Care Utilization Survey. Three studies looked into the panel conditioning effect in the Survey of Income and Program Participation (SIPP); none of them demonstrated the presence of a panel conditioning effect (Bassi, 1998; McCormick, Butler, & Singh, 1992; Penell & Lepkowski, 1992).

It is a somewhat different picture when attitudinal questions and questions about voting behavior are involved. Two studies examined change of responses to attitudinal questions over time (Sturgis, Allum, & Brunton-Smith, 2009; Waterton & Lievesly, 1989). Both studies revealed a panel conditioning effect in responses to attitudinal questions that resulted in somewhat better data quality in later waves. For instance, Waterton & Lievesly (1989) demonstrated that respondents reported more honestly and became less likely to use "don't know" option and that their answers were more consistent. Sturgis and his colleagues (2009) showed that attitude items became more reliable and stable over time and that respondents were more likely to express an attitude in the later waves. Traugott and Katosh (1979) found significantly higher levels of registration and voting turnout among sample respondents who went through a pre-election interview than those who did not. This trend of higher reporting of voting turnout with post-election respondents was observed in Clausen (1969), Kraut & Katosh (1979), and Yalch (1976), even though they attributed this trend to different causes.

Three mechanisms have been proposed to account for panel conditioning effect. One mechanism attributes panel conditioning effect to the change in respondents' real behaviors and attitudes as a result of repeated measurements. In other words, repeated measurements are hypothesized to affect the true value of the variable of interest (Y_i). Studies on attitudinal and voting items seemed to provide some evidence to support this real-change mechanism. Interview experience causes respondents to deliberate on the subject matter of the questions; as a result, respondents either produce strengthened and crystallized attitudes or change their voting behavior (Sturgis, Allum, & Brunton-Smith, 2009; Traugott & Katosh, 1979; Waterton & Lievesly, 1989).

The other two mechanisms both argue that prior interview experiences affect the way that respondents report their behavior or attitudes; that is, rather than changing the true behavior or attitude (Y_i), prior interview experience changes the observed y_i through the measurement error associated with the reported value (Bailor, 1989; Waterton & Lievesly, 1989). However, they predict biases of opposite directions. According to the "better respondents" speculation, exposure to and experience with a prior interview allows sample respondents to understand more about the interview task and the interview requirement; thus, they become "better" respondents at the next wave. For instance, sample respondents might have a better understanding of the meaning of the survey questions, producing better responses at a later wave (Biderman & Cantor, 1984; Cantor, 1995). According to the "worse respondents" speculation, participation in an earlier interview could tip off sample respondents on interview procedure and on the possible burden of answering survey items (e.g., respondents might learn that a "yes" answer might lead to a series of extra questions) or decrease respondents' motivation to participate in future interviews; therefore, they become "worse" respondents at the next wave (van der Zouwen & van Tilburg, 2001). There is no conclusive evidence supporting either mechanism (Cantor, 1989; Willmans & Mallows, 1970).

To summarize the existing work, it seems that there is no (or limited) panel conditioning effect for behavioral items other than voting, and a somewhat stronger effect for voting

items and attitudinal items. It is generally believed that panel conditioning has a negative rather than positive effect on data quality. However, little is known about how best to control negative conditioning effects and to encourage positive conditioning effects.

One weakness of current research on panel conditioning is that it generally ignores possible subgroup differences. Most of the work so far only attempted to demonstrate an across-the-board conditioning effect, implicitly assuming that prior survey experiences have the same impact on everybody in terms of their response strategies to survey questions in later interviews. However, the direction and the size of panel conditioning could be dependent on respondent characteristics, survey design features, and the question items of interest. It is very likely that some respondent groups, for instance, exhibit a positive conditioning effect whereas other groups a negative conditioning effect. It is also possible that different respondent groups show different conditioning effects to different survey questions. Therefore, it is misleading at least to only talk about an across-the-board conditioning effect. The conflicting empirical evidence on the size and the direction of panel conditioning could be due to potential subgroup differences in terms of their tendency to exhibit panel conditioning.

This paper has to two objectives. The first is to examine whether the data from the Consumer Expenditure Quarterly Interview Survey (CEQ) are affected by panel conditioning. A general concern with the CEQ data is underreporting of expenditures and panel conditioning is believed to be one main factor causing underreporting in later waves. Therefore, the first objective of the paper is to examine whether or not respondents in later waves tend to underreport on expenditure amount and expenditure types.

The second goal is to extend the current research on panel conditioning by introducing a subgroup analysis. Specifically, the paper focuses on respondents requiring significantly more effort for contacting and recruiting and those with higher reporting burden and compares them to those requiring less effort for contacting and recruiting and those with lower reporting burden. We chose to study respondents requiring more effort and respondents with higher reporting burdens because we speculate that they are more likely to exhibit panel conditioning effects, supporting the "worse respondents" mechanism. In other words, if the "worse respondents" mechanism were at work, respondents requiring more effort or with higher reporting burden would be more likely than those requiring less effort and with lower burden to become "worse respondents" and to underreport in later waves.

2. Data and Analysis Method

2.1 Consumer Expenditure Quarterly Interview Survey

The U.S. Consumer Expenditure Quarterly Interview Survey is a longitudinal survey conducted for the Bureau of Labor Statistics by the Census Bureau. The CEQ, together with the diary survey, provides information on the buying habits of American consumers, including data on their expenditures, income, and consumer characteristics. In 2008, about interviews were attempted on 37,303 households and 27,545 households completed CEQ with a response rate of 73.8% (AAPOR RR1) (Bureau of Labor Statistics, 2009).

One important feature of the CEQ is the rotation panel design. Each interview quarter, a national representative sample is selected and become a new panel. The panel is interviewed every three months for a total of five times. The panel retires after the 5th interview and a new panel starts. The first interview is used for bounding purpose and is not included in statistical estimates. Because of this rotation panel design, in any given interview quarter, there are altogether five panels in the field with each at different stages of their panel life. In other words, in any given interview quarter, one panel will start the first interview and one panel has the second interview. One panel has the third interview, one panel the fourth, and one panel the fifth and the last interview.

2.2 Analysis Methods

To study panel conditioning effects in the CEQ, we compare the means in the reports by different panel groups on the variables of interest to see if there exist differences in reporting by panel respondents who are at different stages of their panel life.

Mathematically speaking, the answer to a question reported by respondent *i* in panel *j* at the t^{th} interview (y_{ijt}) can be decomposed into four parts:

$$y_{ijt} = Y_{ij} + \alpha_{ijt} + \beta_{ijt} + \epsilon_{ijt} \qquad (1)$$

where Y_{ij} is the true value for respondent *i* in panel *j* at the beginning of the panel, α_{ijt} is the change to the true value at the t^{th} interview, β_{ijt} the measurement error associated with being in sample for the t^{th} time; in other words, β_{ijt} is the non-random error in the reported values for respondent *i* in panel *j* at the t^{th} interview. ϵ_{ijt} is a normally distributed error term with a mean of zero and a variance of one.

The mean for panel *j* at the t^{th} interview is:

$$\overline{y_{jt}} = \overline{Y_j} + \overline{\alpha_{jt}} + \overline{\beta_{jt}} \qquad (2)$$

and the difference between any two panels in a given interview quarter is the sum of three difference terms:

$$\overline{y_{jt}} - \overline{y_{j't}} = \left(\overline{Y_j} - \overline{Y_{j'}}\right) + \left(\overline{\alpha_{jt}} - \overline{\alpha_{j't}}\right) + \left(\overline{\beta_{jt}} - \overline{\beta_{j't}}\right) \qquad (3)$$

Since each panel is a nationally representative sample, the first difference term is zero in expectation. Assuming the second difference term is 0 (that is, the change in true values is the same for different panels interviewed in the same interview quarter), the difference in means between two panels reflects the differences in the measurement error in the reported values. In other words, the mean difference between panels for a given interview quarter reflects the extent of panel conditioning for being in sample for different number of times. Therefore, if the means in the reported values are not significantly different across different panels, then there is no panel conditioning effect. However, if the means do differ across different panels, then panel conditioning exists. In particular, if the mean values reported by panels who have had more interviews are less than panels who have had just one interview before, then it is an evidence that underreporting due to panel conditioning plagues the CEQ data.

To control for the possible confounding of panel attrition, we restricted our analyses to panel respondents who completed all five interviews. We further restricted analyses to households with the same person reporting for the households for the 2nd to the 5th interview.

For this paper, we examined data collected in the second quarter of 2008. We chose this quarter simply because we were able to construct the four full panels that were in field in the second quarter of 2008, using 2008 and 2007 public use data available to us. A total of 3,907 households were included in the analyses that met both criteria – completing all five interviews and having the same person reporting for the household for the 2nd to the 5th interview.

3. Results

3.1 Total Amount of Expenditure and Number of Expenditure Types Reported

We focused on two variables – the total amount of expenditures and the number of different expenditure types with a non-zero expenditure reported by respondents. As shown in Table 1, even though the panel at a later stage of their panel life (i.e., those who had their fourth and fifth interview) reported lower expenditure totals and fewer number of expenditure types than those who had their second or third interview, the differences are not statistically significant at the traditional .05 level. It seems that respondents who were interviewed more times are *not* more likely to underreport on expenditures than those who went through only one interview before.

Table 1. Mean of Total Expenditure Amount and Mean Number of Expenditure Types Reported by Panel

Panel	Mean of Total Expenditure Amount	*Mean Number of Expenditure Types*
Panel starting 2nd interview (n=961)	$12,683	10.2
Panel having 3rd interview (n=963)	$12,179	10.1
Panel having 4th interview (n=996)	$11,877	10.1
Panel having 5th interview (n=987)	$11,673	10.1
F-Test	F(3,3903)=2.4, p=0.07	F(3, 3903)=0.4, p=0.72

3.1 Total Amount of Expenditure by Respondent Subgroups

As stated earlier, a lack of across-the-board panel conditioning effects does not necessarily mean that all respondent subgroups behave in the same way or adopt the same response strategies. We continued the research by comparing respondents requiring more efforts for contacting and recruiting to those requiring less efforts, and respondents with higher reporting burden to those with lower reporting burden.

We use three different methods to divide respondents based on the amount of effort required to contact and to recruit them. The first method grouped together those respondents who didn't need refusal conversion efforts at the second quarter of 2008 as cooperative respondents and those who required refusal conversion efforts as converted refusals. We then examined whether or not these two groups of respondents differ in the extent that they exhibit panel conditioning effect. As shown in Figure 1, x-axis represents the four different panels being interviewed in the second quarter of 2008 and the y-axis is the mean value in total amount of expenditure reported. The dashed line refers to the

cooperative respondents whereas the solid line the converted refusals. It seems that panel respondents who had more interviews (who had their fourth or fifth interview in the second quarter of 2008) tended to report lower expenditure totals than those who had their second or third interview regardless of whether they required refusal conversion efforts. However, the differences by panels were not statistically significant at the .05 level for either group. In addition, there is neither a significant main effect of time in sample nor significant interaction effect. Therefore, the converted refusals were not more likely to show panel conditioning effects than cooperative respondents even though converted refusals reported significantly less expenditure than cooperative respondents ($F(1, 3899)=8.6$, $p<0.01$).

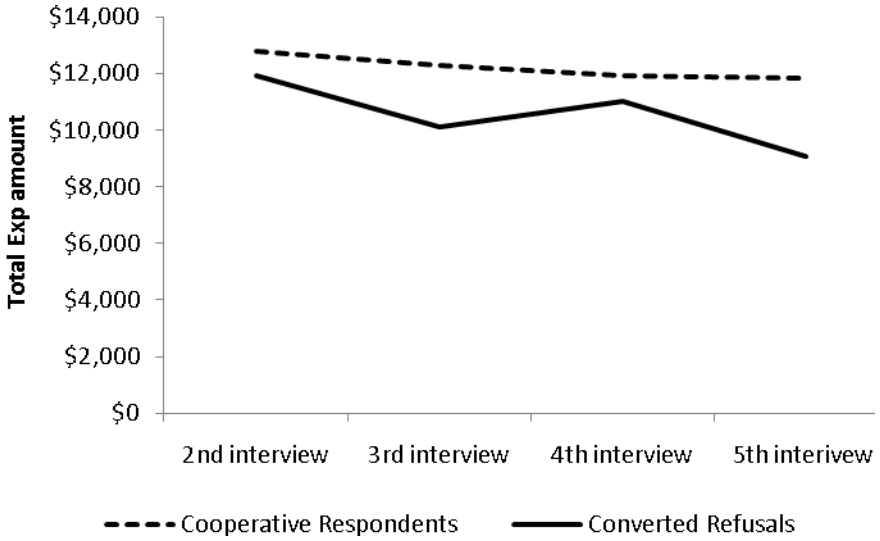


Figure 1: Mean Values of Total Expenditure Amount by Panel and by Whether or Not Refusal Conversion Efforts were Required at the Current Wave

Unlike cross-sectional respondents, panel respondents have more chance to refuse the survey request. They can refuse the survey request at earlier waves of interviews as well. Therefore, we compared respondents who never expressed refusal to those who had indicated refusals either at an earlier wave or at the current wave. Their responses were shown in Figure 2. The dashed line refers to respondents who never indicated refusal and the solid line refers to those who required refusal conversion efforts either at an earlier wave or at the current wave. Similar to Figure 1, those respondents requiring refusal conversion efforts at an earlier or the current wave are not more likely to show panel conditioning effects than those who cooperated at all waves, even though those who needed at least one refusal conversion efforts reported significantly lower expenditure totals than those who never needed refusal conversion efforts.

We also looked at the number of call attempts needed to complete the interview and compared those who needed less than 4 call attempts to those who required 4 or more attempts. (4 call attempts is the 75^{th} percentile cut-off points.) Shown in Figure 3, the dashed line refers to easy respondents who completed the interview with less than 4 call attempts whereas the solid line those who needed 4 or more call attempts. Even though easy respondents tended to report significantly less expenditure than those who needed more call attempts ($F(1, 3899)=20.7$, $p<0.001$), those who needed more calls were not

more likely to show panel conditioning effects than those who needed fewer calls; there is neither a significant main effect of time in sample nor a significant interaction effect

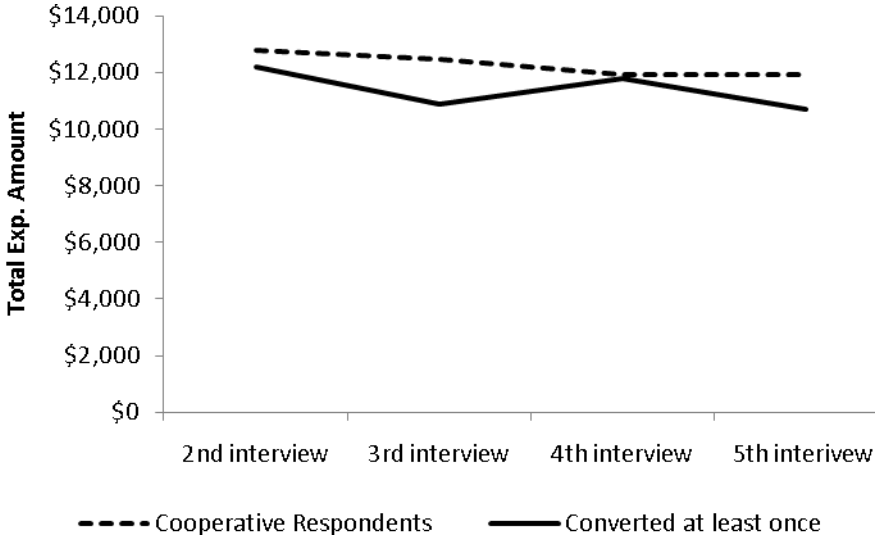


Figure 2: Mean Values of Total Expenditure Amount by Panel and by Whether or Not Refusal Conversion Efforts were Required at either an Earlier or the Current Wave

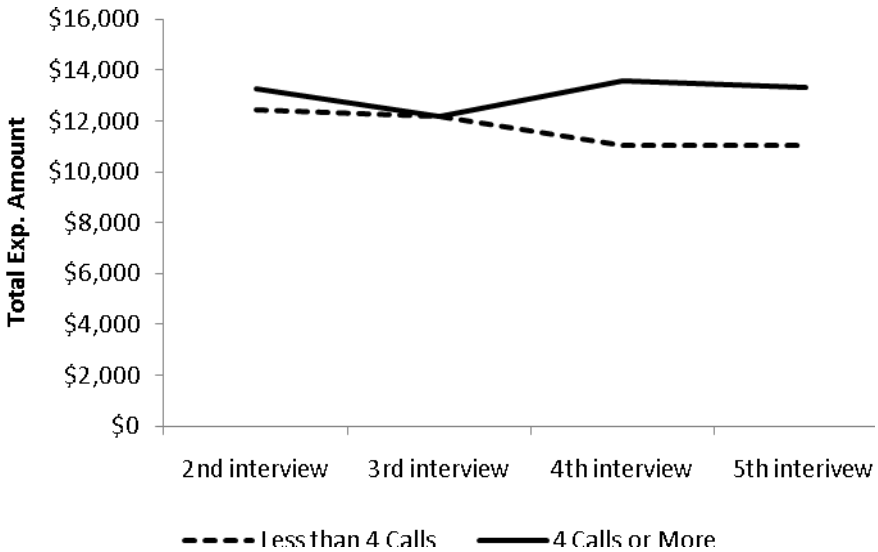
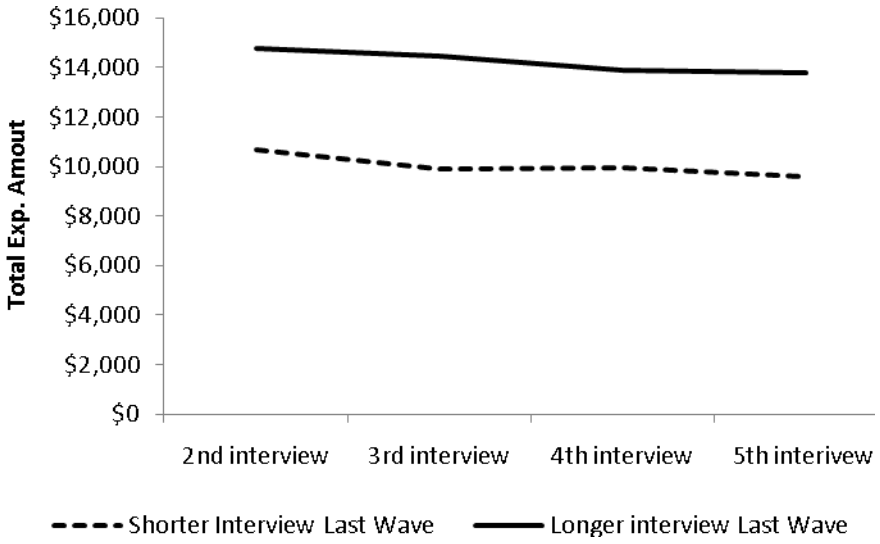


Figure 3: Mean Values of Total Expenditure Amount by Panel and by the Number of Call Attempts Needed to Complete the Interview at Second Quarter of 2008

Lastly, we examined whether reporting burden has an effect on respondents' tendency to show panel conditioning. We looked at the interview length at the previous interview and divided respondents into those with a shorter interview (the length of the previous interview is shorter than the median cut-off point) and those with a longer interview (the length of the previous interview is equal to or longer than the median cut-off point). Presumably, a shorter interview reflects lower reporting burden than a longer interview.

As shown in Figure 5, the dashed line refers to respondents with a shorter interview at the last interview, indicating lower reporting burden, and the solid line respondents with a longer interview at the previous round, indicating higher reporting burden. Respondents with a shorter interview and lower reporting burden at the last round reported lower expenditure totals than those with a longer interview at the last wave and higher reporting burden ($F(1, 3899)=232.0$, $p<.001$). However, neither group of respondents showed a significant panel conditioning by panel.



We also examined the number of expenditure types with non-zero expenditure reported by these subgroups (not shown in the paper). The results on the number of expenditure types are in parallel with the results on the reported expenditure totals. We didn't find evidence of underreporting on the number of expenditure types due to panel conditioning. We also didn't find evidence suggesting that difficult respondents or respondents with high report burden are more likely to show panel conditioning effects than those who were more cooperative and have less reporting burden.

4. Conclusions and Discussion

This paper examined panel conditioning in the Consumer Expenditure Quarterly Interview Survey collected in the second quarter of 2008. We looked at two summary variables – the amount of total expenditure and the number of expenditure types with non-zero expenditure. For both variables, panel respondents who had more interviews reported lower expenditure totals than those who had one or two interviews before, but the differences were not statistically significant, suggesting that panels with more interviews were not more likely to be prone to panel conditioning than panels with fewer interviews. This is good news to the CEQ as the general concerns is that CEQ data are subject to underreporting and panel conditioning is believed to be one cause. We didn't find evidence of underreporting across panels in different life cycle to substantiate the concern.

We further examined panel conditioning by respondent subgroups. We used three different methods to divide respondents into cooperative respondents and difficult respondents. Regardless of which method we used, difficult respondents do *not* seem to be more prone to panel conditionings than cooperative respondents, contrary to what the "worse respondents" mechanism would predict. In addition, we looked at whether reporting burden would have an effect on how likely respondents would be subject to panel conditioning effects. We again didn't find any evidence that those respondents with higher reporting burden are more prone to panel conditioning than those with lower reporting burden. Overall, the "worse respondents" mechanism is not supported.

We did find that, regardless of how many interviews panel respondents had, reluctant respondents who had refused to the survey request at least once reported lower expenditure totals and fewer expenditure types than cooperative respondents who had not indicated refusal. In addition, respondents required four or more call attempts and respondents with higher reporting burden at the prior interview reported more expenditure than those required less than 4 call attempts and those with lower reporting burden at the prior interview. However, neither group became "worse" respondents during the course of their panel life.

We suspected that the lack of a significant main effect of time in sample partially is because the data examined here are from waves 2 to 5. As mentioned earlier, the first interview was used for bounding purpose to reduce recalling error and panel conditioning and was not used for statistical estimates. Therefore, we do not have data from wave 1 interview. Ideally, panel conditioning is detected when we compare responses from respondents at their first interview to responses from those who have had at least one interview before. The unavailability of the first interview makes this comparison impossible, reducing the power to detect panel conditioning. In addition, the large variance in expenditure and the relatively small sample size in each panel and respondent subgroups might have reduced our ability to detect differences.

Underreporting is just one manifestation of panel conditioning. A more dramatic manifestation of panel conditioning is panel attrition. Respondents learn how complex the survey is, how burdensome the reporting task is, and how sensitive some questions are after they complete the first interview. Therefore, they could choose to dodge the second and later interviews completely as an alternative to conditioned underreporting. As a next step, we will investigate respondents who attrited and compare their responses to those who stayed. We will also conduct personal level analysis to looking for evidence of panel conditioning.

Acknowledgements

This work is sponsored by a grant from the Bureau of Labor Statistics. We would like to thank BLS for their financial support and technical guidance and advice. The views expressed in this paper are those of the authors and do not necessarily reflect the policies of the U.S. Bureau of Labor Statistics.

References

Bailar, B. A. (1975). The effects of rotation group bias on estimates from panel surveys. *Journal of the American Statistical Association*, 70, 23-30.

Bailar, B.A. (1989). Information needs, surveys and measurement Errors. In D. Kasprzyk, G. J. Duncan, G. Kalton & M. P. Singh (Eds.), *Panel Surveys*, (pp.1-24). New York: John Wiley and Sons.

Bassi, F. (1998). Gross flows estimation from the Survey of Income and Program Participation. *Journal of Economic and Social Measurement*, 25, 97-110.

Biderman, A. D., & Cantor, D. (1984). A longitudinal analysis of bounding, respondent conditioning and mobility as sources of panel bias in the National Crime Survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 7-8-713.

Bureau of Labor Statistics. (2009). 2008 Consumer Expenditure Interview Survey Public Use Microdata: User's Documentation.

Cantor, D. (1989). Substantive implications of selected operational longitudinal design features: The National Crime Survey as a case study. In D. Kaspryck and G. Duncan (Eds.), *Panel surveys* (pp.1-30). New York: John Wiley.

Cantor, D. (1995). [What respondents learn from questionnaires: The survey interview and the logic of conversation]: Discussion. *International Statistical Review*, 63, 173-177.

Clausen, A. R. (1969). Response validity: Vote report. *Public Opinion Quarterly*, 37, 398-406.

Corder, L. S., & Horvitz, D. G. (1989). Panel effects in the National Medical Care Utilization and Expenditure Survey. In D. Kasprzyk, G. J. Duncan, G. Kalton & M. P. Singh (Eds.), *Panel Surveys*, (pp 304-318). New York: John Wiley and sons.

McCormick, M. K., Butler, D. M., & Singh, R. P. (1992). Investigating time in sample effect for Survey of Income and Program Participation. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pp 554-557.

Neter, J., & Waksberg, J. (1964). Conditioning effects from repeated household interviews. *Journal of Marketing*, 28, 51-56.

Pennell, S. & Lepkowski, J. (1992). Panel conditioning effects in the Survey of Income and Program Participation. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pp 566-571.

Sturgis, P., Allum, N., & Brunton-Smith, I. (2009). Attitudes over time: The psychology of panel conditioning. In Lynn, P. (ed.). *Methodology of Longitudinal Surveys*. Wiley.

Traugott, M.W., & Katosh, J. P. (1979). Response validity in surveys of voter behavior. *Public Opinion Quarterly* 43:359-377.

van der Zouwen, J., & van Tilburg, T. (2001). Reactivity in panel studies and its consequence for testing causal hypotheses. *Sociological Methods & Research*, 30, 35-56.

Wang, K.., D. Cantor, and A. Safir. (2000). Panel conditioning in a random digit dial survey. *Proceedings of the Survey Research Methods Section of the American Statistical Association,* 822-827.

Waterton, J. & Lievesley, D. (1989). "Evidence of Conditioning effects in the British Social Attitudes Panel." In D. Kasprzyk, G. J. Duncan, G. Kalton & M. P. Singh (Eds.), *Panel Surveys,* (pp 319-339). New York: John Wiley & Sons.

Williams, W. H., & Mallows, C. L. (1970). Systematic biases in panel surveys due to differential nonresponse. *Journal of the American Statistical Association*, 65, 1338-1349.

Yalch, R. F. (1976). Pre-election interview effects on voter turnout. *Public Opinion Quarterly*, 40, 331-336.