RECOMMENDATION REGARDING THE USE OF A CE BOUNDING INTERVIEW

VERSION DATE: 08/01/2012

EDITED FOR PUBLIC RELEASE: 05/01/2013

Ian Elkin

U.S. Bureau of Labor Statistics
Division of Consumer Expenditure Survey
Branch of Research and Program Development

Executive Summary

The Consumer Expenditure Interview Survey (CEQ) is a five interview rotating panel design. The first interview is a bounding interview designed to minimize telescoping errors in the reporting of expenditures, although its effectiveness as a bounding interview has been questioned over the years. In addition, the first interview collects household roster data and inventory information. However, expenditures collected in the first interview are not used in the production of official estimates, nor are they released as part of the microdata files.

Given evidence that the bounding interview may not be effective in minimizing telescoping errors, and since its expenditure data are not used for publication or microdata purposes, eliminating it from the series of panel interviews is one method of reducing data collection costs without adversely affecting the quality of the data. The purpose of this report is to present 1) findings from research into eliminating the bounding interview, 2) an estimate of the net impact on length and cost from eliminating the bounding interview, 3) additional research options for determining bounding interview impact, 4) methods proposed to address the potential for higher burden associated with adding the roster and inventory items to the second interview, and 5) a consensus recommendation regarding the elimination of the bounding interview from the CEQ Interview Survey.

1. Research Findings

Based on a review of the relevant bounding literature, CE recommends eliminating the CEQ bounding interview. This decision is based on research findings that include higher wave one expenses being attributed more to the one month recall period rather than telescoping effects (Silberstein 1990); that, after accounting for time-in-sample, evidence of a bounding impact existed for only some expense reporting and when this did occur, the bounded group reported higher spending when compared to the unbounded group (indicating that the bounding interview was ineffective in curtailing telescoping) (Tseng and Tan 2011); a shortened first interview did not provide sufficient cost savings to warrant further pursuit (Elkin 2011); the bounding interview does not have a significant impact on overall reporting levels (Hackett 2011); and, finally, after adjusting for the elimination of the bounding interview, the National Crime and Victimization Survey (NCVS) has not reported any degradation in data quality (Killion 2011).

2. Impact on Length and Cost

An analysis was then conducted to determine the net impact on length and cost from eliminating the bounding interview, adding roster and inventory questions to what is the current second interview, and eliminating the second interview "inventory" questions. The analysis determined that these changes

would lead to the addition of approximately eight to eleven and a half minutes of interview time at a .9 percent to 1.2 percent per successful interview cost increase.

3. Additional Research Options for Determining Bounding Interview Impact

Additional research options for determining the extent of the impact on telescoping error and respondent burden from the elimination of the CEQ bounding interview were considered as well: (i) a traditional field test with sufficient sample for statistical power was determined to be too expensive to be a feasible option for CE; (ii) an analysis of bounded and unbounded wave 2 respondents was established as a worthwhile analysis that did not fall within the proposed project schedule; (3) and an examination of the data from the current CE Research Section regarding respondent burden is not feasible because sufficient data are not currently available. Otherwise, it was determined that the research findings summarized in (1) comprised a comprehensive investigation of bounding effects in CE.

4. Methods Proposed to Address the Potential for Higher Burden in Wave 2

Methods to address the potential for higher burden associated with an increase in the number of questions in the second interview are discussed, and three types of questions that might be removed from the second interview are proposed and summarized. They are questions (i) for which the published estimates are sourced from the Diary Survey; (ii) pertaining to income and credit liability; and, (iii) on recurrent expenditures. Methods for imputing data for these questions are also discussed, with development and testing to be conducted using historical data.

5. Recommendation to Eliminate the Use of a Bounding Interview

Representing 11.8 percent of total CEQ costs for calendar year 2011, the bounding interview represents a significant proportion of the CEQ costs. Consequently, it is recommended CE eliminate the use of a bounding interview, after completing simulations using historical CE data to determine which of the proposed methods best mitigates respondent burdens without sacrificing data quality.

Table of Contents

I. Project Background

The Consumer Expenditure Interview Survey (CEQ) is a five interview rotating panel design. The first interview is a bounding interview designed to minimize telescoping errors in the reporting of expenditures, although its effectiveness as a bounding interview has been questioned over the years. In addition, the first interview collects household roster data and inventory information. Importantly, expenditures collected in the first interview are not used in the production of official estimates, nor are they released as part of the microdata files.

Given evidence that the bounding interview may not be effective in minimizing telescoping errors, and since its expenditure data are not used for publication or microdata purposes, eliminating it from the series of panel interviews is one method of reducing data collection costs.

II. Summary Justification

The following is a summary justification for the elimination of the CEQ bounding interview based on previous CE research and other examples, including Neter and Waksberg (1965), Silberstein (1990), Tseng and Tan (2009), Yan and Copeland (2010), Tseng and Tan (2011), Hackett (2011), Elkin (2011), and Killion (2011).

II.I Summary Justification Recommendation

Based on the research described below, CE recommends eliminating the bounding interview.

- Silberstein (1990) finds that the first wave of the panel surveys studied exhibited higher means than the overall means for subsequent waves, even after estimated telescoping effects were deducted. However, *the one month recall period was viewed as the major reason for the higher estimates, other factors not excluded*.

- Yan and Copeland (2010) *suspect that the lack of a significant main effect of time in sample is partially because the data examined was from waves 2 through 5.* The first interview was used for bounding purpose to reduce recalling error and was not used for statistical estimates. Yan and Copeland noted that panel attrition is the worst after the first interview; a larger percentage of panel respondents dropped off after the first interview than at the later waves. *Therefore, those respondents who stayed beyond the first interview are probably the better respondents, and, thus, are less likely to exhibit panel conditioning than those who dropped off after the first interview.*

- Hackett (2011) determined that the mean with no bounding is less than the mean with bounding for about 40% of Universal Classification Codes (UCCs) in 2008 and about 42% of UCCs in 2009. Thus, Hackett finds that for CUs that completed all four quarters of the CE302, *the bounding interview does not have a significant impact on overall reporting levels*.

- Tseng and Tan (2011) report that, after accounting for time-in-sample, evidence of group differences existed for reporting on child care and cash contributions and weakly for footwear and electricity. However, the group differences did not occur in Wave 2. In addition, *when group differences did occur, the bounded group reported higher spending when compared to the unbounded group*.

- Elkin (2011) determined that a shortening of Interview 1 could approximately save .2 percent of total CEQ costs per year. Subsequent meetings with *CE management determined that these saving were not sufficient to warrant further testing of a shortened Interview 1 survey as a cost savings measure*.

- Starting in 2006, the Department of Justice dropped the bounding interview from the NCVS. *Adjusting for the elimination of the bounding interview from the outset, NCVS has not reported any degradation in data quality* (Planty, 01/31/2012).

III. Net Impact on Survey Length and Cost

The following is an overview of the net impact on length and cost from eliminating the CEQ bounding interview, adding roster and inventory questions to the second interview, and eliminating the second interview "inventory" questions.

III.I First Impact Analysis (Interview Time)

In order to determine the net impact on interview length from eliminating the CEQ bounding interview, six years of CEQ data were analyzed, January 2006 through December 2011. Analyses were restricted to complete interviews completed during the second wave. Completed second interviews were chosen because they are most representative of interviews that will be affected by the modifications to the interview structure due to the dropping of the CEQ bounding interview. Initially, these interviews were divided into two groups: 1. Households whose first interview occurred in the second wave (H2) and 2. Households whose first interview occurred in the first wave (H1). An H2 household is defined as a household whose original consumer unit has been replaced by a new and completely separate consumer unit (CU) during the five quarter interview period. For example, when one CU leaves a home and a new CU moves into that home during the interview period. Second interview replacement households were chosen for analysis because their structure most closely matches the proposed unbounded first CEQ interview. In addition, an H2 household also includes CUs that were non-interviews in the first wave and good interviews during the second wave. That is, they include the entirety of the second interview as well as all rostering and inventory questions that would normally appear in the CEQ bounding interview.

Initial analyses focused on comparing mean and median total interview time concerning H2 and H1 households.

Table 1. Mean and Median Analyses: Total Interview Time (TOT_TIME)

	H2 Households (a) n=3,671	H1 Households (b) n=34,400	Interview Time Difference (a-b)
Mean	4,215.01 sec. (70.25 min.)	3,812.44 sec. (63.54 min.)	402.57 sec.
Median	3,791 sec. (63.18 min.)	3,434 sec. (57.23 min.)	357 sec.

The mean analysis resulted in the finding that on average H2 household second interviews were 402.57 seconds (≈6.50 minutes) longer than H1 household second interviews. These results were confirmed, while controlling for outliers, through median analysis where H2 household second interviews were 357 seconds (≈6.0 minutes) longer than H1 household second interviews. These results are as expected because the H2 households receive more questions, in the form of the rostering and inventory questions.

In order to better ascertain differences between H2 and H1 households, demographic analyses were performed on the data, concentrating on household size and household tenure.

Table 2. H2/H1 Household Demographic Analyses

	H2 Households	**H1 Households**
Median Household Size	2	2
Frequency of Household Tenure	Owner – 51.81% Owner: Condo/Co-op – 3.32% Renter – 41.65% Occ. w/ no payment – 2.21% College Housing – 1.01%	Owner – 64.78% Owner: Condo/Co-op – 3.90% Renter – 29.72% Occ. w/ no payment – 1.05% College Housing – 0.55%

Due to similar household size, analysis of household size does not sufficiently emphasize the demographic differences in H2 and H1 households. However, analysis of household tenure does show significant variance between H2 and H1 household tenure. This variance is most apparent when comparing the percentage of owners and renters; since home ownership information is one of the sections that requires the greatest initial outlay of respondent time, it is plausible that the average total time for H2 households would be higher if the percentage of H2 households who were characterized as owners was similar to that of H1 households.

In order to accurately examine interview times, H2 and H1 households were divided into separate cohorts, with each cohort representing household type (H2 versus H1) and household tenure. Mean section times for each section were examined, with Tables 3 and 4 representing sections with large numbers of

inventory and rostering questions for owners (excluding condo/co-op owners) and renters as these household tenure groups make up over 90% of both household types.

Table 3. Mean Analysis: Inventoried and Rostering Section Interview Times (Owners)

	H2 Households (c) n=1,902	**H1 Households** (d) n=22,285	**Interview Time Difference** (c-d)
Section 03	410.16 sec.	220.57 sec.	189.60 sec.
Section 10	40.48 sec.	29.24 sec.	11.24 sec.
Section 11	248.41 sec.	43.25 sec.	205.16 sec.
Section 13	142.29 sec.	124.36 sec.	17.93 sec.
Section 14	144.64 sec.	105.52 sec.	39.11 sec.
Coverage Section	43.12 sec.	13.83 sec.	29.29sec.
Control Section	280.15 sec.	86.63 sec.	193.53 sec.
Total (all Sections)	4,652.64 sec.	4,087.34 sec.	565.53 sec.

Table 3 and Table 4 confirm that interview time for the second interview is greater for H2 households regardless of household tenure status. In addition, both tables illustrate that the outlay of time is significantly greater for many of the inventory and rostering sections for both owner and renter H2 households. However, the total interview time (aggregate of interview time for all sections) difference for owner H2 households is 1.24 times higher than that for renter H2 households demonstrating the influence that the home ownership questions have on interview time. This difference combined with the differences in household tenure makeup for H2 and H1 households, as well as the group size difference between household types, makes unadjusted group comparison as a whole unadvisable.

Table 4. Mean Analysis: Inventoried and Rostering Section Interview Times (Renters)

	H2 Households (e) n=1,529	H1 Households (f) n=10,226	Interview Time Difference (e-f)
Section 03	42.73 sec.	19.85 sec.	22.88 sec.
Section 10	35.68 sec.	27.17 sec.	8.52 sec.
Section 11	153.35 sec.	33.12 sec.	120.23 sec.
Section 13	80.42 sec.	71.30 sec.	9.12 sec.
Section 14	108.76 sec.	78.59 sec.	30.17 sec.
Coverage Section	64.16 sec.	17.74 sec.	46.42 sec.
Control Section	265.93 sec.	89.55 sec.	176.37 sec.
Total (all Sections)	3,700.11 sec.	3,242.48 sec.	457.63 sec.

In order to accurately consider the impact of eliminating the bounding interview and moving inventory and rostering questions to what is currently the second interview, an adjustment factor has been applied to the interview time difference between H2 and H1 households. Table 5 shows adjusted interview time difference which is calculated as interview time difference for each cohort adjusted by the household tenure frequency of the larger H1 household group.

Table 5. Mean Analysis by Housing Tenure for Total Interview Times (QTENURE) and Adjustment

	H2 Households (g) (Weighted Time)	H1 Households (h) (Weighted Time)	Interview Time Difference (i = g-h)	H1 Frequency of QTENURE (k)	Adjusted Interview Time Difference (l = i*k)
Owner	4,652.64 sec. (2,410.53 sec.)	4,087.34 sec. (2,647.78 sec.)	565.53 sec.	64.78%	366.20 sec.
Owner: Condo/Co-op	4,816.16 sec. (159.90 sec.)	3,945.01 sec. (153.86 sec.)	871.15 sec.	3.90%	33.98 sec.
Renter	3,700.11 sec. (1,541.10 sec.)	3,242.48 sec. (963.67 sec.)	457.63 sec.	29.72%	136.01 sec.
Occ. w/ No Payment	3,385.79 sec. (74.83 sec.)	3,197.26 sec. (33.57 sec.)	188.53 sec.	1.05%	1.98 sec.
Student Housing	2,212.73sec. (22.35 sec.)	2,428.04 sec. (13.35 sec.)	-215.31 sec.	0.55%	-1.18 sec.
Total (Weighted Time)	4,208.71 sec. (70.15 min.)	3,876.76 sec. (64.61 min.)	331.95 sec. (5.53 min.)	100%	
Total Adj. Difference					536.99 sec. (8.99 min.)

Adjusted interview time difference aggregated for all five cohorts illustrates that an H2 household interview, when the group is adjusted to resemble H1 households, takes approximately 536.99 seconds (8.99 minutes) longer. *(Note: median analysis was also performed without substantively different results)*

Since first interviews have a much higher incidence of personal visit (PV) collection than later waves, and PV interviews take longer than telephone interviews, it is known that, regardless of any additional questions, the change in interview mode from telephone to PV will result in a longer second interview. In order to determine the impact of interview mode, additional analyses of interview time were performed taking into account interview mode by sub-setting for PV collection. The adjusted interview time difference for all five cohorts (PV only) is 393.56 seconds (6.56 minutes) longer for H2 households when compared to their H1 household counterparts. As a result, of the total increase of 536.99 seconds, it can

be determined that 393.56 seconds of interview time is directly associated with the greater number of questions asked to respondents in the H2 group and an increase of 143.43 seconds (2.39 minutes) is associated with the proportional shift in interview mode from telephone to PV.

III.II Second Impact Analysis (Interview Time)

A second analysis to determine the net impact on interview length from eliminating the CEQ bounding interview was performed. This analysis also analyzed six years of CEQ data (January 2006 through December 2011) and were restricted to consumer units with complete interviews in both the first and second waves of the interview survey. For this analysis, section times for the "part B" inventory sections from the first interview were compared to the section times in their "part A" section counterparts in the second wave. For example, in the first interview, households report all insurance policies associated with the consumer unit (e.g. life, auto, home, etc.) in Section 13B of the interview instrument; then, in Section 13A of the second interview, households report if they still own the policies detailed during the first interview.

For this analysis, the marginal increase of moving the detailed "part B" questions to the second interview and dropping the "part A" questions was acquired by taking the difference between the total time for "part A" inventory sections from the total time for "part B" sections. Since the CAPI instrument provides timing information by section and not by part, for this analysis second wave data were additionally subset to only include households that reported no inventory section alterations to what they reported in the first interview. This subsetting of the data allowed for a more accurate section analysis of households who answered the detail questions in first interview and only the inventory questions during the second interview. The analysis was performed on all inventoried sections (including the rostering sections) and the time per section, for these sections, was summed by household and the mean additional time per household was calculated.

Initial analyses focused on comparing mean and median time difference for the inventoried and rostering sections as well as the total time difference by household for all inventoried and rostered sections.

Table 6. Mean and Median Analyses: Inventoried and Rostering Section Interview Times Differences

	Mean Time Difference	**Median Time Difference**
Section 03 n=28,219	257.08 sec.	218 sec.
Section 10 n=1,358	138.91 sec.	114 sec.
Section 11 n=60,519	308.19 sec.	254 sec.
Section 13 n=70,839	91.66 sec.	73 sec.
Section 14 n=31,095	84.11 sec.	69 sec.
Coverage Section n=34,403	44.20 sec.	22 sec.
Control Section n=34,402	185.79 sec.	149 sec.
Total Time Difference (Inventory/Rostering Sections)	692.28 sec. (11.55 min.)	601 sec. (10.02 min.)

The mean analysis resulted in the finding that, on average, completing the detailed questions from the inventoried/rostering sections increased interview time by 692.28 seconds (≈11.50 minutes) when compared to answering only the inventory questions. These results were confirmed, while controlling for outliers, through median analysis where completing the detailed questions from the inventoried/rostering sections increased interview time by 601 seconds (≈10.0 minutes) when compared to answering only the inventory questions. These results are similar to the total adjusted time difference from the first impact analysis on interview time.

Table 7. Mean and Median Analyses: Total Interview Times Differences (Housing Tenure)

	Mean Time Difference (Owners)	**Median Time Difference** (Renters)
Owners Total Time Difference n=22,285	805.99 sec.	727 sec.
Renters Total Time Difference n=10,226	466.26 sec.	390 sec.

In order to determine if additional interview time is significantly impacted by housing tenure and/or interview mode, mean and median analyses were performed. Tables 7 and Table 8 confirm that regardless of housing tenure or interview mode, replacing the inventory questions in the current second interview with detailed questions adds interview time. In addition, as expected, Table 7 illustrates that the outlay of time for owner households completing all the detailed questions is nearly twice the outlay of time of renter households. These results are as expected due to the expectation of owner households completing the lengthy homeownership information section. However, Table 8 shows little differentiation, approximately half a minute, between the total time difference added when taking survey mode into account. Thus, confirming previous results that interview time is impacted more by the number of questions a respondent receives rather than by the mode of the interview.

Table 8. Mean and Median Analyses: Total Interview Times Differences (Interview Mode)

	Mean Time Difference (Owners)	**Median Time Difference** (Renters)
Personal Visit/Mostly Personal Visit Total Time Difference n=11,140	649.44 sec.	558.50 sec.
Telephone/ Mostly Telephone Total Time Difference n=5,284	625.90 sec.	545 sec.

III.III Cost Impact Analysis

First Impact Analysis:

To determine the net impact on cost from eliminating the bounding interview, adding roster and inventory questions, and eliminating second interview "inventory" questions, the weighted average wage of Census Field Representatives (FR) and Senior Field Representatives (SFR) is multiplied by the aggregate of interview time difference, expressed in minutes, divided by sixty. The increased interviewing costs associated with eliminating the bounding interview and moving the roster and inventory questions to the current second interview is .9 percent per successful interview.

Second Impact Analysis:

Using the results of the second impact analysis to determine the net impact on cost from eliminating the bounding interview, adding roster and inventory questions, and eliminating second interview "inventory" questions, the weighted average wage of Census Field Representatives (FR) and Senior Field Representatives (SFR) is multiplied by the aggregate of interview time difference, expressed in minutes, divided by sixty. The increased interviewing costs associated with eliminating the bounding interview and moving the roster and inventory questions to the current second interview is 1.2 percent per successful interview.

III.IV Net Impact on Survey Length and Cost Conclusion

In summary, the analysis was conducted to determine the net impact on length and cost from eliminating the bounding interview, adding roster and inventory questions to what is the current second interview, and eliminating second interview "inventory" questions. These steps would lead to the addition of approximately eight minutes to eleven and a half minutes of interview time to what is the current second interview at an increased cost of .9 percent to 1.2 percent per successful interview.

IV. Other Options for Confirming Hypotheses about the Elimination of the CEQ Bounding Interview

This section examines additional research options for determining the extent of the impact on telescoping error and respondent burden from eliminating the bounding interview from the CEQ Survey.

IV.I Field Test

Traditionally field tests have offered a number of advantages, such as being an effective tool for identifying troublesome questions, difficult concepts, respondent reactions to new data collection techniques, identifying telescoping error, and allowing for observation of indicators of respondent fatigue; however, there are disadvantages. Some problems may go by unnoticed; respondents may misunderstand questions, use inappropriate judgment strategies, provide socially desirable answers, etc. without giving off any signals that these error sources are occurring. In addition, relevant concepts such as respondent fatigue as well as respondent attrition add additional complexities and costs to an already burdensome testing option.

Subsequently, the complex nature necessitated to fully examine the impact of eliminating the bounding interview has on respondent burden and telescoping error as well as the considerable costs of fielding such a test, a field test is not currently a feasible option for CE.

IV.II Analysis of Bounded and Unbounded Respondents

A bounded respondent is defined as a person that serves as a survey respondent for his/her CU over two consecutive interviews. An unbounded respondent is defined as a CU where two different members of a CU serve as the respondent over two consecutives waves. An unbounded interview would normally be defined as an interview where respondents are asked for expenditures made since a given date where no control is exercised over the possibility that respondents may erroneously shift some of their expenditure reports into or out of the recall period.

Under the proposed scenario, both sets of CUs would be defined as "bounded interviews" under the classic definition; however, this scenario would compare expenditure levels as well as expenditure reports between bounded and unbounded respondents. The purpose of this method is to determine if CU expenditures/reports of unbounded respondents, as defined, increase as proposed by classic bounding literature or have little or no change as more recent bounding literature suggests.

Unfortunately, project time constraints made the inclusion of all potential analyses unfeasible; however, analysis of bounded and unbounded respondents remains as a potentially informative analysis fir determining the prevalence of telescoping error in bounded and unbounded interviews.

IV.III Burden Index[1]

A method, slated to start in the October 2012 CEQ Survey instrument, for addressing the potential for higher burden associated with a longer second interview consists of additional questions added to the CE Research Section on burden perception. These burden questions are modeled after the Post Survey Assessment Questions used in the analysis of Fricker, et al. (2011). One potential outcome of this section is the assessment of which dimensions of burden are most relevant to a redesigned second interview. Essentially, the goal is to verify that an additional eight to eleven and a half minutes in interview length translates into higher (perceived) burden.

In order to maximize the results from this method, the burden questions need to be asked to a group of first interview nonrespondents who complete the second interview as well as a group of "regular" second interview respondents, as currently scheduled the burden questions will only be asked during the fifth interview. The rationale for using the former group is that these respondents would be administered a survey interview that most closely resembles the proposed redesigned second interview survey. The "regular" second interview respondents are defined as those who are administered the current second interview. Results to the questions would then be compared across the two groups. If no statistically significant differences exist, then evidence may exist that simply adding the Inventory and Roster questions to the second interview might not increase the perception of burden (even though the average lengths of the interviews may differ). This could potentially serve as justification for having the only change to the second interview to be adding the inventory and rostering questions.

If, however, there are differences in the responses across the two groups, then assuming that the questions are framed after the multi-dimensional concept of burden identified above, then the dimensions of burden that are statistically significant would be used to identify classes of questions to consider for elimination from the second interview. The hope is that by eliminating questions from the redesigned second interview on the basis of the significant burden dimensions, the perception of burden for the redesigned second interview would be the same as the current second interview.

However, as noted previously, the existing burden questions, as scheduled, will only be asked during the fifth interview. In order to determine if there are differences in the responses across the two groups, burden questions would need to be added the second interview research section.

---

[1] The majority of this subsection was contributed by Jeffrey M. Gonzalez, BLS Office of Survey Methods Research

V. Addressing the Potential for Higher Burden Associated with a Longer Second Interview[2]

The purpose of this section is to provide a summary and proposal of methods to address the potential for higher burden associated with a second interview that collects roster and inventory information.

V.I Methods to Address Burden

There are three types of questions that might be removed from the second interview to address the potential higher burden associated with a lengthier interview. They are questions (1) for which the published estimates are sourced from the Diary Survey; (2) pertaining to income and credit liability; and, (3) on recurrent expenditures.

V.I.I Sourcing Published Estimates from the Diary Survey

The goal of sourcing published estimates from the Diary Survey is to explore development of statistical models that would enable the CE to delete questions from the Interview Survey and impute for the missing data by using data from the Diary surveys for similar CUs.

Sourcing published estimates from the Diary Survey could improve data quality in the Interview by using the items in Diary that have higher expenditures. Regarding the elimination of the bounding interview, there are additional benefits, namely the Interview Survey will be shortened, thus reducing field collection costs and also reducing respondent interview time. Subsequently, respondent burden associated with respondent interview time may be reduced.

The Interview and Diary surveys have a number of expense categories that overlap, for example: food at home, food away from home, clothing, etc. The CE and CPI programs have developed a statistical method to select which survey to use for publications and CPI weights. The source selection process is run every two years. The Diary Survey is the source for many items that are also collected in the Interview.

V.I.II Removing Income and Credit Liability Questions from the Second Interview

Historically, the mean time for a successful second interview is 3,886.80[3] seconds (64.78 min.). By eliminating the income and credit liability questions (Sections 21 and 22) from the second interview, a mean time of 548.27[4] seconds (9.14 min.) will be eliminated from the total survey time. The elimination

---

[2] The majority of this section was contributed by Jeffrey M. Gonzalez, BLS Office of Survey Methods Research

[3] Mean of total time for successful second interviews: CEQ Data M01Y2006-M12Y2011

[4] Mean of the sum of income and credit liability sections for successful second interviews: CEQ Data M01Y2006-M12Y2011

of the income and credit liability questions from the second interview represents, on average, a 14 percent reduction in interview length.

On average, successful third interviews take 3,298.26[5] seconds (54.97 min.) a difference of 588.54 seconds (9.81 min.) when compared to the second interview. Subsequently, the income and credit liability questions might be eliminated from the second interview and, instead, be asked during the third interview. The relative increase in time attributed to the inclusion of the income and credit liability questions would be roughly equivalent to the current interview time difference between the second and third interviews. However, a thorough examination of the impact on respondent burden, due to survey length, would be necessary prior to addition of the income and credit liability questions to the third interview. In addition, it is necessary to note that this proposal may not align with the American Community Survey (ACS) income plans as well as interest in one-year change estimates.

A second approach would be to wholly eliminate the income and credit liability questions from the second interview and only collect income and credit liability information during the fifth interview. This approach would, on average, reduce the second interview length by 548.27 seconds without increasing respondent burden in another wave. However, a methodical assessment of the impact on data quality, possibly due to respondent attrition, of simply asking the income and credit liability would be paramount to making an informed decision.

V.I.III Removing Questions on Recurrent Expenditures from the Second Interview

A recurrent expenditure is an expense that a sample unit is likely to have either monthly, quarterly, or over some other regularly spaced time interval. The rationale for proposing to eliminate questions on recurrent expenditures is that they may be viewed as low variability, perhaps even, constant across waves. By capturing the information about the expense at a later wave, the data collectors capture most of the variability that it will ever capture about that expenditure during the sample unit's tenure in the panel.

Of course, a natural first step in proposing to eliminate these types of questions would be to identify those expenditures that can be regarded as recurrent. This should be the extensive study of an officially chartered team. One of the primary goals of the team would be to propose quantitative criteria to use when identifying recurrent expenditures. It is recommended to consider quantitative criteria related (but not limited) to variance of the cross-sectional and longitudinal mean expenditure, prevalence/incidence of incurring the expense during the reference period(s), and frequency of purchase.

---

[5] Mean of total time for successful third interviews: CEQ Data M01Y2006-M12Y2011

V.II Imputation of Item Missing Data

In all three proposals, questions would be eliminated from the second interview; subsequently, there would be item missing data. Depending on the uses of the collected data, it may or may not be necessary to recover the missing data, e.g., through imputation. If CPI can (and does) use the Diary data for their purposes and if quarterly mean expenditure estimates can be calculated using the Diary data, then it is not essential to recover the missing data other than for microdata users. However, if there are other concerns, e.g., having a second interview dataset that mimics the current second interview dataset, then imputation methods are a viable candidate for addressing those concerns.

Various methods can be proposed to impute the information from the questions eliminated from the second interview. Forthcoming deliverables from the Diary-to-Interview Imputation Team can inform the imputation procedures, particularly for those questions that were eliminated based on being sourced from the Diary Survey. If income questions are eliminated and to be imputed, then the current income imputation procedures can be modified to recover the information typically collected via those questions. Also note that income questions can be moved to the third interview so it may not be essential to impute those. In the case of recurrent expenditures, a variety of techniques can be used.

Most imputation procedures can be framed in a regression context. The differences among the procedures arise due to making different assumptions on some combination of the following (and perhaps other features): (1) the underlying distribution of the variable that is being imputed, to include the mean and variance structure, and (2) the functional form of the model. This paper does not delve into the specific details of each of those as significant resources (e.g., teams) should be devoted to developing the exact imputation procedures; however, the paper identifies three sources of information that could potentially be leveraged when developing imputation methods. They are (1) historical data (i.e., external to the data for which the data are being imputed); (2) same survey panel (i.e., using their own data from future interviews); and (3) same quarter (i.e., using other sample units having either their third, fourth, or fifth interviews in that calendar quarter).

Once an imputation procedure is decided upon, the data imputer must evaluate the success of the imputation procedure. In effect, the imputer must specify the goals of the imputation procedure and that, in turn, would assist in indentifying the most appropriate means of evaluating it. One way to evaluate the imputation procedure would be to compare the imputed data to an external data source. This external data source is determined by the goals of the imputation procedure. One might compare the imputed data on the basis of the first and second moments, full empirical distribution, preservation of multivariate

relationships, etc., to the appropriate comparison data source. In addition, sensitivity analyses for alternative specifications of the imputation procedure could be conducted. This amounts to conducting a statistical simulation and computing standard empirical quantities such as the mean, variance, relative bias, etc.

VI. Recommendation and Next Steps

Representing 11.8 percent of total CEQ costs for calendar year 2011, the bounding interview represents a significant proportion of the CEQ Survey costs and time. Emerging survey methodological literature regarding the bounding interview, including literature using CE data, indicates that the bounding interview may not be effective in minimizing telescoping errors. Furthermore, its expenditure data are not used for publication or microdata purposes. Additional research suggests that the elimination of the bounding interview and subsequent transfer of inventory and rostering questions to the current second interview will add approximately eight to eleven and a half minutes to the survey, at an increase in interviewing cost of .9 percent to 1.2 percent per successful interview. In addition, a number of methods have been suggested to mitigate the increase in respondent burden associated with a longer second interview. Consequently, it is recommended CE eliminate the use of a bounding interview, after completing simulations using historical CE data to determine which of the proposed methods best mitigates respondent burden without sacrificing data quality.

Limitations and Potential Risks[6]

A limitation of this analysis is that the bulk of data collection costs for the bounding interview are thought to come from contact attempts and gaining respondent cooperation, rather than the actual time spent conducting the interview. As with all CE cost data, it is difficult to separate out individual interviewing activities. Therefore, the estimated potential savings are not accurately represented as the total current bounding interview expenditures.

Furthermore, by eliminating the bounding interview there are potential risks for degradation in data quality, broadly defined as in Brackstone (1999), among the data used for official estimates and publication, which may outweigh the benefits of eliminating the bounding interview.

In addition, if higher, the new levels of burden experienced by the respondent as a consequence of receiving a longer initial interview (currently Interview 2) might have adverse affects on nonresponse to later waves in the panel. As Yan and Copeland (2010) point out, drop-off is greatest after the first interview of a panel survey. This trend is not unique to the CE interview Survey. A more burdensome first exposure to the survey might result in larger attrition effects.

---

[6] The majority of this subsection was contributed by Jeffrey M. Gonzalez, BLS Office of Survey Methods Research

References

Brackstone, G. (1999). Managing Data Quality in a Statistical Agency. *Survey Methodology*, 25, 139—49.

Elkin, I. (2011). Cost Savings from Shortening Interview 1 Project Report. BLS, Division of Consumer Expenditure Survey internal report.

Fricker, S., Gonzalez, J. M., and Tan, L. (2011). Are You Burdened? Let's Find Out. Paper presented at the 2011 Annual Conference for the American Association for Public Opinion Research, Phoenix, AZ.

Hackett, C. (2011). Consumer Expenditure Quarterly Interview Survey: Effectiveness of the Bounding Interview. BLS, Division of Consumer Expenditure Survey internal report.

Killion, R.A., (2011) National Crime Victimization Survey: Review of Bounding Adjustments for the 2010 Estimates. Bureau of the Census, Demographic Statistical Methods Division internal document.

Neter, J., and Waksberg, J. (1965). Response errors in collection of experimental data by household interviews: An experimental study. Bureau of the Census, Technical Report No.11.

Silberstein, A. (1990). First Wave Effects in the U.S. Consumer Expenditure Interview Survey. BLS, Division of Consumer Expenditure Survey internal report.

Tseng, N., and Tan, L. (2009). An exploration of the effectiveness of the Bounding Interview in the Consumer

Expenditure Interview Survey. BLS, Division of Consumer Expenditure Survey internal report.

Tseng, N., and Tan, L. (2011). Effects of Time-in-sample and Interview 1 Bounding in the Consumer Expenditure Interview Survey. BLS, Division of Consumer Expenditure Survey internal report.

Yan, T. and Copeland, K. (2010). Panel Conditioning in Consumer Expenditure Interview Survey. Presentation made at BLS on December 20, 2010.

Appendix A: CE Literature Review

CE bounding interview research is summarized below, with an emphasis on purpose, methodology, findings, and limitations.

Response errors in collection of expenditure data by household interviews: an experimental study (1965), J. Neter and J. Waksberg (more in-depth summary by Lucilla Tan available[7])

Purpose: To study nonsampling errors in the collection of household expenditures for Survey of Residential Alterations and Repairs (SORAR). Survey designers were concerned about the effect of nonsampling errors on data quality because responses were likely to be based largely on recollection without supporting records. Study outcome would be used improve survey procedures. Nonsampling errors occur when the respondent omits reporting expenditures due to faulty recollection or misstatement of the timing or amount of the expenditures, or the interviewer misclassifies the type of expenditure. The goal of the experimental study was to investigate the existence and magnitude of nonsampling errors.

Methodology: Subject a number of different samples, each a probability sample, to different survey procedures over the same time period, and then compare the estimates of aggregates from each of these procedures. Differences in the estimates reflect the net effect associated with 2 alternative procedures (subject to limitations due to sampling errors). They provide estimates of differences in the response bias (the constant component of response errors). Such comparisons do not provide information on response variability (the variable component of response errors). Major emphasis in the analysis was placed on number of jobs reported, by size of job, because the distribution of expenditures is highly skewed.

Summary: Neter and Waksberg found that unbounded recall of expenditures for preceding month involved substantial net forward telescoping of jobs into the recall period. In addition, the net internal telescoping within the 3-month recall period was forward in direction into the most recent month of the recall period. The magnitude of this effect was larger on larger jobs. The recall loss effect reflects incremental losses in recall incurred by lengthening the recall period from 1 to 3 months. This effect was substantial for smaller jobs, but decreased rapidly as job size increased. Changes in the length of the reference period (with identical recall period) affected reporting of jobs and expenditures. Finally, the reference period effect may be considered to have 2 components: reporting load effect: which reflects the impact of the length of the interview, number of jobs during the reporting period, on the R. and

---

[7] In-depth summary of *Response errors in collection of expenditure data by household interviews: an experimental study* (1965), J. Neter and J. Waksberg by Lucilla Tan is located at \\filer1\dces\dces-brpd\Projects - In Progress\Bounding Memo\Reference\Summary_NeterWaksberg.doc

interviewer and differential telescoping effect: reflects differences in the extent of telescoping with changes in the length of the reference period.

Limitation: This approach only measures differential bias, but does not indicate which alternative procedure is better in terms of accurate reporting.

First Wave Effects in the U.S. Consumer Expenditure Interview Survey (1990), Adriana Silberstein
Purpose: Panel responses to the U.S. Consumer Expenditure Interview Survey are compared, to assess the magnitude of telescoping in the unbounded first wave. Analysis of selected expense categories confirms other studies' findings that telescoping can be considerable in unbounded interviews and tends to vary by type of expense. In addition, estimates from first wave are found to be greater than estimates derived from subsequent waves, even after telescoping effects are deducted, and much of these effects can be attributed to the shorter recall period in the first wave of this survey.

Methodology: Silberstein evaluates telescoping effects at the aggregate level by comparing estimates of unbounded and bounded responses, with certain precautions. The precautions include tracking the experience of several panels to overcome seasonal incompatibilities and account for panel conditioning. Silberstein creates simultaneous confidence intervals for individual comparisons by group were derived using the Boneferroni method. Expenditure means were computed using a log transformation of individual expenses reported in the first recall month. Sample weights included adjustments for nonresponse and subsampling, but excluded final weight factors for population controls, which were not available for the first wave.

Summary: The Silberstein paper provides an investigation of potential telescoping effects in unbounded interviews. Silberstein notes that these effects appear to be considerable. The study demonstrates that external telescoping effects are much greater than internal telescoping effects within a three month recall period of subsequent waves. In addition, the first wave of the panel survey studied was found to exhibit higher means than the overall means for subsequent waves, even after estimated telescoping effects were deducted. Finally, the one month recall period is viewed as the major reason for the higher estimates, other factors not excluded.

An Exploration of the Effectiveness of the Bounding Interview in the Consumer Expenditure Interview Survey (2009), Neil Tseng and Lucilla Tan

Purpose: To determine the effectiveness of the CE301 interview in bounding the second interview by comparing patterns of reporting between bounding and unbounded second interviews using historical data. A successfully bounded second interview is determined when the reporting levels for the unbounded group are higher than the levels for the bounded group.

Methodology: The analysis included only CUs that completed their first and second interviews in adjacent waves and 1 and 2 respectively. The CUs in wave 1 were then classified as unbounded, while CUs in Interview 2 were classified as bounded for the analyses. Tseng and Tan used a multivariate analysis of variance model (MANOVA) to test the significance of group differences in reporting levels of the components. To counter significant group effect between the bounded and unbounded groups, Tseng and Tan used the "Honestly Significant Difference" test (HSD) to control for Type 1 experiment-wise error.

Summary: Group effect was found to not be significant for major household appliances, but it was significant for household appliances, clothing and accessories, and home furnishings. Bounding effects of the first interview seemed to be greater on categories that are more susceptible to external telescoping. High value items were less susceptible to telescoping error when compared to household appliances, clothing and accessories, and home furnishings.

Limitations: Test used historical data with unbalanced group sizes leading to a more restrictive definition of bounded and unbounded interviews. In addition Tseng and Tan used unedited expenditure data, which may have contained data inconsistencies and errors from the survey data collection.

Panel Conditioning in Consumer Expenditure Interview Survey (2010), Ting Yan and Kennon Copeland

Purpose: Yan and Copeland look at panel conditioning which is a source of non-sampling error unique to panel surveys. It refers to a change that occurs because the respondent has had one or more prior interviews. They speculate that there are three mechanisms corresponding to three types of empirical findings: 1) Deliberation on subject matters, 2) Better respondents, and 3) Worse respondents.

Methodology: For a given quarter/year Yan and Copeland compare means of different rotation groups and restricted to households completing all four waves and with the same reporter for all four waves. Yan

and Copeland use this methodology to attempt to identify whether panel condition effects exist for each expenditure category, to determine whether conditioning effects differ by type of expenditure, to identify respondent subgroups which are more prone to panel conditioning, and to test the link between panel condition in wave t to respondents' likelihood to participate in wave t+1.

Summary: Yan and Copeland report that, for the amount of total expenditures and the number of expenditure types with non-zero expenditures, panel respondents who had more interviews reported lower expenditure totals than those who had one or two interviews before. However, the differences were not statistically significant, suggesting that the panels with more interviews were not more likely to be prone to panel conditioning than panels with fewer interviews. There was mixed evidence when testing the link between panel conditioning in wave t to respondents' likelihood to participate in wave t+1. There was not any evidence that respondents at wave t were less likely to participate in subsequent waves; however, there was evidence that panel attriters tended to report less expenditures before they dropped-off than those who continued.

Limitations: Yan and Copeland report that wave 1 data was not available and only accessed public use data. Yan and Copeland suspect that the lack of a significant main effect of time in sample partially is because the data examined was from waves 2 through 5. The first interview was used for bounding purpose to reduce recalling error and was not used for statistical estimates. Therefore, wave 1 data are not included in the public use datasets and were not available to Yan and Copeland. Yan and Copeland noted that panel attrition is the worst after the first interview; a larger percentage of panel respondents dropped off after the first interview than at the later waves. Therefore, those respondents who stayed beyond the first interview are probably the better respondents, and, thus, are less likely to exhibit panel conditioning than those who dropped off after the first interview.

Effects of Time-in-sample and Wave 1 Bounding in the Consumer Expenditure Interview Survey (2011), Neil Tseng and Lucilla Tan

Purpose: Tseng and Tan reported evidence of wave 1 bounding effect on reporting in wave 2 for some expenditure categories they analyzed. However, their analysis did not account for the repeated measurement over time-in-sample. The purpose of this study was to examine the effect of the wave 1 bounding interview on reporting levels in wave 2, after controlling for time-in-sample.

Methodology: The study examined sample units that were eligible for first interviews in 2009, and for whom their final dispositions were known for the entire five wave survey panel. Tseng and Tan defined

the bounded group as those sample units who completed all five interviews in the survey panel, and the unbounded group as those sample units that completed interviews two through five in the survey panel. Using a subset of summary variables, Tseng and Tan analyzed average dollar values and frequency in reporting in eleven categories and used longitudinal analysis methods to account for the repeated measurement nature of data.

Summary: After accounting for time-in-sample, evidence of group differences existed for reporting on child care and cash contributions and weakly for footwear and electricity. However, the group differences did not occur in Wave 2. In addition, when group differences did occur, the bounding group reported higher spending when compared to the unbounded group. There was evidence of a linear time-in-sample effect for childcare, electricity, and health insurance; reporting trended down for childcare, but increased for electricity and health insurance.

Limitations: Limitations include the study's use of retrospective data such that the treatment conditions are not randomly assigned to sample units. In addition, the group sizes were extremely unbalanced with the bounded group ten times larger than the unbounded group. Expenditures in all categories were highly variable leading to large standard errors. Finally, Tseng and Tan lacked information on if and the extent to which interviewers used wave 1 bounding expenditure information as intended.

Consumer Expenditure Quarterly Interview Survey: The Effectiveness of the Bounding Interview (2011), Catherine Hackett

Purpose: The goal of Hackett's study was to analyze the effect of bounding throughout the history of the CUs in the survey. Since data collected in the first interview are not used to produce expenditure estimates and the significant resources are expended for the collection of expenditures in the first interview, is it an effective means of bounding.

Methodology: Hackett uses mean expenditures as a measure of reporting levels and a program which applies the balanced repeated replication (BRR) method to compute variances of CE expenditure estimates. In addition, it was used to generate annual estimates of mean expenditures and standard errors for the 856 UCCS in the Integrated CE Stub. Hackett computes the mean and standard errors for both bounded and unbounded CUs. Then using the mean expenditure and standard error for each UCC, Hackett takes the difference between the mean expenditure of CUs with a bounding interview and the mean expenditure of CUs with no bounding interview, the percent difference, and a Z-score.

Summary: For the selected CUs, the mean expenditure with no bounding is greater than the mean expenditure with bounding for about 20% of UCCs in 2008 and about 17% of UCCs in 2009. The mean with no bounding is less than the mean with bounding for about 40% of UCCs in 2008 and about 42% of UCCs in 2009. The mean with no bounding is equal to the mean with bounding for about 40% of UCCs in both 2008 and 2009. Hackett finds that for CUs which completed all four quarters of the CE302, the bounding interview does not have a significant impact on overall reporting levels.

Cost Savings from Shortening Interview 1 in the Consumer Expenditure Interview Survey (2011), Ian Elkin

Purpose: The objective of the Cost Savings from Shortening Interview 1 project was to determine the cost savings associated with shortening Interview 1 on the basis of removing a pre-specified set of non-inventory questions from the interview. A secondary objective was to develop a stand-alone method that can be used on an as-needed basis to determine cost savings associated with different configurations of questions identified for removal. It was acknowledged that additional factors should be considered when deciding to shorten Interview 1, such as a possible negative impact on response rates due to a false expectation of shortened Interviews 2 through 5.

Methodology: Elkin analyzed fiscal year 2010, October 2009 through September 2010, of historical Interview 1 Post Phase 2 data. Analyses were restricted to complete interviews completed in the 1st interview that were personal visits or predominantly personal visits. In addition, Elkin focused on expense variables from non-inventoried sections. Finally, Elkin provided means analysis using timing data and cost data from workforce representation information of Field Representatives.

Summary: Elkin determined that a shortening of Interview 1 could approximately save .2 percent of total CEQ costs per year. Subsequent meetings with CE management determined that these saving were not sufficient to warrant further testing of a shortened Interview 1 survey. In its place, a project to determine the feasibility of eliminating the bounding interview was commissioned by CE management.

Limitations: Limitations include Elkin's use of interview time to determine costs when bulk of data collection costs are thought to come from contact attempts and gaining respondent cooperation. The estimated potential savings are likely misrepresented as a result. Another limitation is that both mean and median analyses weighted time equally over all questions. In reality, question length varies by section complexity, question type, number of potential respondents, respondent fatigue etc.

National Crime Victimization Survey: Review of the Bounding Adjustments for the 2010 Estimates (2011), Ruth Ann Killion (and NCVS Conference Call with Mike Planty)

Purpose: Due to monetary constraints, in 2006 National Crime Victimization Survey (NCVS) cut time-in-sample (TIS) 1. Prior research has shown that telescoping error in the NCVS is approximately 25%. To combat this error, NCVS adjusts the data by creating TIS 1 in-house. When determining how to adjust for the loss of the bounding interview, NCVS looked at how its removal would impact sample size and how that would impact point estimate variance and change estimates. In the end, NCVS chose to approach the issue without a break n the sequence; thus, the development of the bounded adjustment.

Methodology: TIS bounding adjustments are computed by the following formula:

$$\text{Adj. Factor for TIS } a = \frac{\dfrac{\text{Weighted number of crimes in TIS } a}{\text{Weighted total of cases in TIS } a}}{\dfrac{\text{Weighted number of crimes in TIS 1}}{\text{Weighted total of cases in TIS 1}}}$$

Where $a = 2\text{-}6$

With this formula, NCVS is adjusting the crime level for first interviews in each of TIS 2-6 to the level of crimes that they would expect for that TIS. The number of crimes generally tends to be underreported in later interviews, due to respondent fatigue and possibly other factors. By using an adjustment in each TIS, NCVS is not only adjusting for bounding bias but also maintaining consistency across time for each TIS.