Temporal Patterns of Survey Response Rates and Reporting Rates in the U.S. Consumer Expenditure Interview and Other Panel Surveys

Moon J. Cho, John L. Eltinge, Barry P. Steinberg
U.S. Bureau of Labor Statistics
Moon J. Cho, 2 Massachusetts Avenue NE, Washington, DC 20212
(Cho_M@bls.gov)

Key Words: Graphical methods, Interviewer effect, Quarterly interview effect, Calendar effect, Incomplete data, Missing data.

Abstract:

In panel surveys, it is often important to evaluate response rates at several levels, including the aggregate (unit) level, panel-specific and wave-specific levels and item-specific levels. This paper considers these issues in two ways. First, we provide a detailed review of the relevant literature, with special attention devoted to the abovementioned distinctions, and to survey characteristics and predictor variables associated with variability in response rates. Second, we present analyses of patterns of reporting rates in the U.S. Consumer Expenditure (CE) Interview Survey. In the CE Interview Survey , consumer units (roughly equivalent to households) are asked to provide month-by-month reports of the amounts of money spent on each of a large number of items. Reported expenditures are recorded at a relatively fine level of detail defined by the six-digit Universal Classification Code (UCC). The analyses place special emphasis on the relationships of reporting rates with interviewer characteristics, the wave, the recall month and the distance between the recall month and the interview month.

1. Introduction

1.1 The CE Survey

One major use of the CE survey data is to provide the basis for revising weights and associated pricing samples for the Consumer Price Index (CPI). In addition, the BLS uses the data to produce estimates of mean expenditures and to produce public data sets of expenditures and income. The CE Survey uses a stratified multistage probability sample of households which represent the total U.S. civilian noninstitutional population. The primary sampling units (PSU) are groups of counties, or independent cities. The set of sample PSU's used for the survey consists of 101 areas; from which 87 urban areas were selected by BLS for the CPI program (BLS Handbook, 1997, p.163). Within each selected primary sample unit, a given sample consumer unit (CU, roughly equivalent to a household) is randomly assigned to one of two modes of data collection: interview or diary. The remainder of this paper will consider only data from the quarterly CE Interview Survey. The purpose of the CE Interview Survey is to obtain detailed data on relatively large expenditure items such as property, automobiles, or major appliances, or on expenses which occur on a fairly regular basis, such as rent, utility bills, and insurance premiums. The CE Interview Survey includes rotating panels: each consumer unit (CU) in the sample is interviewed every 3 months over five calendar quarters and then is dropped from the survey. Approximately 20 percent of the addresses are new to the Survey each quarter. The interviewer uses a structured questionnaire to collect both demographic and expenditure data in the Interview survey. In the CE Interview Survey, approximately 8910 addresses are contacted in each calendar quarter, and the number of completed interviews per quarter is targeted at 6160.

1.2 The CE Interview Survey Data

The CE Interview Survey collects information on demographic and family characteristics and on the inventory of major durable goods of each consumer unit at the initial interview. Expenditure information is also collected in this interview, using a 1-month recall. Numerous adjustments are made to the data after collection, and they affect the amount of mean expenditures and the number of reports. For example, one expenditure of $120 listed for DINE_MOX, the usual monthly expense for din-

The authors thank Jay Ryan and David Swanson at BLS, and Vicki McIntire at the Census Bureau for helpful comments on the U.S. Consumer Expenditure Interview Survey. 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.

ing out, could be split out ("allocated") into various subcategories at a later stage of data processing. Some of the adjustments which are discussed in Silberstein and Jacobs (1989) are the current-month adjustment, imputation, and aggregation. When Silberstein worked on the 1984 CE data (apparel and home furnishing), she found that the analyses based on reports and expenses were consistent. The data used for this study is a processed data, generated from the CE Phase 3 databases. Note that the data analyzed in Eltinge and Cho (2003) covered the similar quarters, but was an earlier data processing stage. However, both datasets should have the same total amount of expenditure on the respective quarter. Our example data comes from the 190 expenditure items of the 2000 U.S. CE Interview Survey.

1.3 Previous findings

Bailar (1989) note that "Bias is observed when a significantly higher or lower level for a characteristic occurs in the first interview than in subsequent interviews, when one would expect the same level. This is the case for the number of employed and unemployed, the number of persons victimized by crime, the number of people with illnesses who visited a doctor, the number of people who worked more than 35 hours a week, the number of alterations and repairs on a housing unit, and the amount of money spent on consumer goods." For daily or weekly expenditure diaries, several authors have reported systematic declines in expenditure estimates. They observed that the estimates of the first week are higher than those of subsequent week(s), that the average estimate of the first day of the first week is greater than the overall average estimate, and that the average estimate of the first day of the week is greater than the average estimate of the week. See, e.g., Silberstein and Scott (1991), Dippo et al. (1977), Pearl (1977) for the CE Diary Survey; and Kemsley and Nicholson (1960), Kemsley (1961), and Turner (1961) for the United Kingdom Family Expenditure Survey. Bailar (1989) mentioned that Mooney's finding (1962) followed the above mentioned pattern; there was a higher number of reports on medically attended illnesses in the initial interview compared to the subsequent interviews. Mooney's evaluation was also based on the diaries kept by respondents, and the number of reports in the subsequent interviews were biased severely downward in his case. In addition, Bailar (1989) reported that the unemployment rate was significantly higher in the first interview compared to later interviews for the Current Population Survey for the time period January 1983 to October 1984. Silberstein (1990) reported that the first wave of the panel survey in unbounded interviews in the CE Interview Survey exhibited higher expenditure means than the overall means for the subsequent four waves, even after estimated telescoping effects were deducted. In the same study, she showed that there was a significantly greater number of reported expenses of $100 or more in the first wave than the subsequent four waves, but reporting rates for smaller expenses were not significantly different across the waves. Several researchers have studied the impact of interviewers on response rates. Groves and Couper (1998) found that survey interviewers play a crucial role in reducing the unit nonresponse. They speculated on the possible mechanisms through which greater interviewer experience and a heightened sense of confidence may translate into success in the field. Groves and Couper assured that it is less the stable characteristics of interviewers that determine the likelihood of success than the interaction of these characteristics and interviewer behavior with householders that important in gaining cooperation. Gray (1956) examined interviewer variability using examples of items from two surveys. In his examples, he observed that when adequately defined, the factual items showed no significant interviewer variability. He also observed that opinion questions showed appreciable variability that related to the interviewers' own opinions. Kemsley (1965) studied whether expenditure questions are susceptible to interviewer variability, and whether variability varies for different methods and for different types of questions in expenditure surveys. He concluded that many of the conclusions he drew were tentative because of the compounding with the respondents' factors.

1.4 Summary of results

The remainder of this paper presents methodology and empirical results for reporting rates observed in the Consumer Expenditure Interview Survey. For an individual consumer unit, we will define a reporting rate to equal the proportion of CE questionnaire items that receive a nonzero expenditure report from that consumer unit. Section 2.1 defines a finite-population average of these CU-level reporting rates and presents a probability-weighted estimator of this average. Section 2.2 defines related interviewer, interview number and period-specific averages and presents related point estimators. Reporting rates defined at various levels of cross-sectional and temporal aggregation are of practical interest because they may provide relatively simple and accessible indicators of data quality. For example, if a given interviewer has a reporting rate that is unusually low (relative to other interviewers covering consumer units with similar demographic and economic characteristics), this may indicate that this interviewer is not probing for expenditures as much as other interviewers are probing. Section 3 through 6 explore relationships between observed reporting rates and interviewer characteristics, interview number, reference month within reference period, and calendar month, respectively. In particular, Section 3.1 compares reporting rates across the four principal CE interviews (labeled interviews 2 through 5), and notes that reporting rates for interview 5 are slightly higher than reports for the previous three interviews. Section 3.2 presents related results on the distribution of interviewer-level reporting rates. Section 4.1 reports that the average reporting rate for supervisory field representatives tend to be substantially lower than those for standard field representatives. For standard field representatives, Section 4.2 explores the relationships between interviewer-level reporting rates and interviewer workload. Section 5 presents related results of overall reporting rates for the most recent, middle, and most distant reference months in a three-month reference period. The higher reporting rates are observed in the more recent months. Consequently, the reporting rate was the highest in the most recent month. Section 6 presents reporting rates across the twelve calendar months, of special interest are a decidedly elevated rate for December and a slightly elevated rate for August, possibly related to true increases in purchase rates associated with holiday shopping and back-to-school shopping, respectively. Section 7 provides some additional discussion of these empirical results and possible follow-up studies.

2. Reporting Rates for CE Survey Items

2.1 Aggregate Reporting Rate

Define J_{ic} as the number of non-zero reports obtained from an interview c, by an interviewer i, and w_{ic} is the associated probability weight. Then a probability-weighted estimator of the overall proportion of nonzero reports is:

$$\hat{\pi} = \frac{\sum_{i=1}^{I}\sum_{c=1}^{n_i} w_{ic}J_{ic}}{\sum_{i=1}^{I}\sum_{c=1}^{n_i} w_{ic}J}$$

where I is the number of interviewers, n_i is the total number of interviews conducted by an interviewer i, and J is the total number of item categories in our data. We also define $\hat{\pi}_{uw}$ as an unweighted estimator of the overall proportion of nonzero reports:

$$\hat{\pi}_{uw} = \frac{\sum_i\sum_c J_{ic}}{\sum_i\sum_c J} \quad .$$

Also define $\hat{\pi}_Q$ as a vector whose element is the weighted reporting rate for each interview number such that $\hat{\pi}_Q = [\hat{\pi}_{Q2}, \hat{\pi}_{Q3}, \hat{\pi}_{Q4}, \hat{\pi}_{Q5}]'$, $\hat{\pi}_{Q2} = \{\sum_i\sum_c w_{ic}J_{ic} \times I_{Q2}\}/\{\sum_i\sum_c w_{ic}J_{ic} \times I_{Q2}\}$ where I_{Q2} is the indicator for the second interview. The aggregate-level reporting rates for recall months, calendar months, and workload groups are defined in the similar way. The variance estimates are obtained by the balanced repeated replication method with 44 replicate weights.

2.2 Interviewer-level Reporting Rate

Define $\hat{\pi}_{i,uw}$ as an unweighted estimator of the proportion of nonzero reports for an interviewer i:

$$\hat{\pi}_{i,uw} = \frac{\sum_{c=1}^{n_i} J_{ic}}{n_i J} \quad .$$

Analysis on the coefficient of variation estimates of associated final weights for interviewers demonstrated that weights are generally homogeneous within interviewers. This implies that design effects are relatively small, and analysis on the unweighted reporting rate at the interviewer level is approximately equivalent to the weighted reporting rate.

3. Interview Number Effect

In the CE Interview Survey, each selected consumer unit is asked to participate in a total of five interviews. The first interview collects data for only bounding purposes (BLS Handbook 1997, pp. 161-162), and will not be considered here. The second through fifth interviews are conducted at three-month intervals. In each of these interviews, the consumer unit is asked to report expenditures for the past three months. Only the second and the fifth interviews collect data on a financial profile of the consumer unit; thus, those financial data are excluded from further analysis here.

3.1 Overall Interview Number Effect

For panel surveys of this type such as CPS (see section 1.3); response rates have been observed to vary across the interview number (wave). Consequently, for our current work, we compared the distributions of weighted item reporting rates across interview numbers (2 through 5). Figure 1 displays point estimates and pointwise 95% confidence intervals for the weighted reporting rates. Note especially that the fifth interview has somewhat higher reporting rates. To explore further the association between reporting rates and interview number, consider the null hypothesis $H_0 : \pi_{Q2} = \pi_{Q3} = \pi_{Q4} = \pi_{Q5}$, and define the quadratic-form test statistic,

$$T_Q = (A\hat{\pi}_Q)' \left\{ A\hat{V}_{BRR-Q}A' \right\}^{-1} (A\hat{\pi}_Q)$$

where A is the 3×4 contrast matrix, $\hat{V}_{BRR-Q}$ is the design-based covariance matrix estimator for $\hat{\pi}_Q$, computed using the balanced repeated replication method. Under the null hypothesis, T_Q is distributed approximately as a scalar multiple of $F_{(3,38)}$ In our example, T_Q is 45.43 which exceeded the cutoff point of 9.01 at $\alpha = 0.05$. Hence, we conclude that the association between the interview number and the reporting rate is statistically significant. The total numbers of completed interviews are, respectively, 7632, 7717, 7625 and 7689.

3.2 Interview Number and Interviewer Effect

We computed the unweighted reporting rates $\hat{\pi}_{i,uw}$ separately for each interviewer for each interview number. For interviewer and interview number combinations that had fewer than five sample CUs, the estimates $\hat{\pi}_{i,uw}$ tended to be very unstable and thus are excluded from the analysis here.

Figure 2 presents side-by-side boxplots of the unweighted reporting rate estimates separately for each interview number. For each interview number group, the middle line in the box corresponds to the median, the dot corresponds to the mean, the upper and lower bounds of the box correspond to the sample upper and lower quartiles, and the upper and lower whiskers correspond to the largest and smallest sample values respectively. Note that the median values of the groups, 0.144, 0.144, 0.142, and 0.147 respectively are approximately equal contrast with increase in mean for 5th interview. When we include all interviewers regardless of their workloads, the median values of the four groups are 0.145, 0.143, 0.141, and 0.147 respectively.

4. Interviewer Effect

The CE Interview Survey has approximately 500 field representatives (FR) and 200 supervisory field representatives (SFR) working in the program at any given month. FR are interviewers who carry regular assignments of interviews. SFR do not carry a regular monthly assignment, and are usually considered to be the best FR. In addition to doing supervisory work, SFR conduct re-interviews for quality assurance purposes, and conduct interviews for some of the more difficult cases. They sometimes fill in for field representatives on vacation or who have resigned. Due to their supervisory responsibilities, SFR often interview households across more than one PSU in the region especially if the SFR is in a large metropolitan area that has been divided into more than one self-representing PSU. Although not as frequently as SFR, FR also may interview in more than one PSU in the region. Each FR has a unique code inside the same regional office boundary. However, distinct FR may have the same code if they belong to different regions. For example, Mary's code is B06 at Region A, and Tom's code can also be B06 at Region B. New FR are given new codes; old codes are not reused for approximately 3 years. With the CAPI instrument for data collection in April 2003, the CE Survey has different variable fields for FR and SFR which provide more detailed information on interviewers. Prior to that period, office clerks, supervisors, and SFR were classified as the same code.

4.1 Interviewer's Status

For the current work we compared the distributions of observed item reporting rates across interview numbers (2 through 5) for each interviewer. Because SFR generally have fewer interviews, we include all interviewers regardless of workloads. In Figure 4.1, we compare the median unweighted reporting rates of SFR and FR. It can be shown that the mean unweighted reporting rates have the similar pattern as the median values. Median reporting rates for FR for the interview number 2 to 5 are: 0.148, 0.146, 0.146, and 0.152, respectively. The corresponding median reporting rates for SFR are: 0.138, 0.132, 0.132, and 0.131 respectively.

We observed that the median reporting rates for FR are consistently higher than those of SFR. There are several possible explanations. Firstly. although SFR are considered to be as the best interviewers, the cases they dealt with were harder ones. Also, because interviewer ID codes are not reused for approximately three years, FR who were promoted to SFR still carry their FR ID codes. Note that median reporting rates for the 5th interview are somewhat higher for FR.

4.2 Interviewer Workload

We also explored the relationship between unweighted reporting rates and interviewer-workloads for FR where workload is defined to equal n_i, the number of interviews performed by interviewer i.

The numbers of interviews, n_i, covered by an FR range from 1 to 188, with a mean of 44, a first quartile of 5, a median of 35, and a third quartile of 70. Consequently, we ordered the 596 field representatives according to their workloads from smallest to largest. Second, we computed the 20th percentile (3), 40th percentile (21), 60th percentile (47), and 80th percentile (79) of workloads. Third, we divided the data into five quintile groups. Figure 4.2 presents side-by-side boxplots of the unweighted reporting rates within each of the five quintile workload groups. Note that there is considerable variability in the first workload group. This is because the number of FR are fewer in this group. The median values of the reporting rates for the five groups are, respectively, 0.139, 0.147, 0.146, 0.151, and 0.148.

We then tested whether an interviewer's workload is associated with the reporting rate, based on the null hypothesis $H_0 : \pi_{L1} = \pi_{L2} = \pi_{L3} = \pi_{L4} = \pi_{L5}$, and the test statistic

$$T_L = (A\hat{\pi}_L)' \left\{ C\hat{V}_{BRR-L}C' \right\}^{-1} (C\hat{\pi}_L)$$

where C is the 4×5 contrast matrix, $\hat{\pi}_L = [\hat{\pi}_{L1}, \hat{\pi}_{L2}, \hat{\pi}_{L3}, \hat{\pi}_{L4}, \hat{\pi}_{L5}]'$, $\hat{\pi}_{L2} = \{\sum_i \sum_c w_{ic} J_{ic} \times I_{L2}\}/\{\sum_i \sum_c w_{ic} J_{ic} \times I_{L2}\}$, and I_{L2} is the indicator for the second workload group, and $\hat{V}_{BRR-L}$ is the design-based covariance matrix estimator for $\hat{\pi}_L$, computed using the balanced repeated replication method. Under the null hypothesis, T_L is distributed approximately as a scalar multiple of $F_{(4,37)}$. In our example, T_L is 2.50 which does not exceed the cutoff point of 11.36 at $\alpha = 0.05$. Hence, we conclude that the association between workload effect and reporting rate is not statistically significant.

5. Recall Month within Reference Period and Related Recall Effects

It has been noted by several studies (see Brennon et al., Silberstein, Silberstein and Jacobs) that respondents' retrospective reports are often inaccurate, consequently resulting in the over- or under-reporting of the past events. There are two major types of questions in the CE Interview Survey: the first asks for the purchase month for each reported expenditure, the other asks for a quarterly amount of expenditure. We did preliminary analysis on the data obtained from both types questions for the current paper. We also analyzed the data which was recorded for a specific month. Figure 3 presents the side-by-side confidence intervals of weighted reporting rate for selected UCCs for each recall month. Weighted mean reporting rates for the selected UCCs were 0.142, 0.123, and 0.118 and unweighted mean reporting rates were 0.145, 0.125, and 0.120 for 1-month, 2-month, and 3-month respectively.

6. Effects of Calendar Month

Weighted mean reporting rates, $\hat{\pi}$, for the calendar months January to December are: 0.140, 0.139, 0.141, 0.141, 0.142, 0.143, 0.143, 0.147, 0.141, 0.143, 0.145, and 0.163, respectively. The corresponding unweighted mean reporting rates, $\hat{\pi}_{uw}$, are: 0.142, 0.142, 0.144, 0.144, 0.145, 0.146, 0.146, 0.149, 0.143, 0.145, 0.146 and 0.166, respectively. We carried a standard quadratic form test for equality of reporting rates across the twelve months, and rejected the null hypothesis of equality at the $\alpha = 0.05$ level of significance. However, for the current explanatory analysis, it would be more informative to show the side-by-side confidence interval plots as in Figure 4.

7. Discussion

Contrary to observations from previous studies, the first interview has no outstanding reporting rates, while the fifth interview has relatively high reporting rates. In Section 4, we observed that the median reporting rates for a FR are consistently higher than those of a SFR. The median reporting rates for the 5th interview are considerably higher for a FR. The reference period for most questions in the second through fifth interviews is "since the first of the month, three months ago" This includes the month of the interview. The current month data is excluded from the CE processes that go into the final Phase 3 database. However they are loaded into the database three months later. More specifically, the data reported in the interview month of the first interview is brought forward and loaded into the database of the second interview. The second, third and fourth interviews follow the same procedure. Only for the fifth interview the CE Interview Survey discards the current month expenditure report. Hence the most distant month data is compounded with the current month data. In future study, analyzing the proportion of current month data to the most distant month data will be interesting. Since CE data is adjusted and aggregated at several different phases, which affects the mean number of reports, it would be interesting to check whether our study on reporting rates is consistent with the study on the CE expenditure. Caution should be given when one attributes what he or she observes in Section 5 to either recall or telescoping. In an experimental setting, studies such as Brennan et al. (1996) have been done to minimize telescoping. Brennan et al. (1996) pointed out the confusion on over-reporting with forwarding telescoping. They suggested referring to the effect (over-reporting) rather than the assumed cause (forward telescoping) when the cause cannot be confirmed.

8. References

Bailar, B. (1989). Information Needs, Surveys, and Measurement Errors. In D. Kasprzyk, G. Duncan, G. Kalton and M.P. Singh (eds.), *Panel Surveys*, pp 1-24, New York: John Wiley & Sons, Inc.

Brennan, M., Chan, J., Hini D., and Esslemont, D(1996). Improving the Accuracy of Recall Data: A Test of Two Procedures. *Marketing Bulletin*, 7, pp20-29. Available at http://marketing-bulletin.massey.ac.nz/article7/article3b.asp

Dippo, C., Coleman, J., and Jacobs, C. (1977). Evaluation of the 1972-73 Consumer Expenditure Survey. *Proceedings of the Section on Social Statistics*, American Statistical Association, pp. 486-491.

Eltinge, J., and Cho, M. (2003). Nonidentified Item Nonresponse and Reporting Rate in the U.S. Consumer Expenditure Interview Survey. *Proceedings of the 14th International Workshop on Household Survey Nonresponse*. Belgium. Available at http://www.nonresponse.org

Gabler, S., Haeder, S., and Lahiri, P. (1999). A Model Based Justification of Kish's Formula for Design Effects for Weighting and Clustering, *Survey Methodology*, vol. 25, pp. 105-106, 1999.

Gray, P. (1956). Examples of Interview Variability Taken from Two Sample Surveys. *Applied Statistics*, Vol. 5, No. 2 (June 1956), pp. 73-85.

Groves, R., and Couper, M. (1998). *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons, Inc.

Kemsley, W. (1961). The Households Expenditure Enquiry of the Ministry of Labor: Variability in the 1953-54 Enquiry. *Applied Statistics*, Vol. 10, No. 3 (November 1961), pp. 117-135.

Kemsley, W. (1965). Interviewer Variability in Expenditure Surveys. *Journal of the Royal Statistical Society*, Vol.128, No. 1, pp. 118-139.

Kemsley, W. (1965). Interviewer Variability in Expenditure Surveys. *Journal of the Royal Statistical Society*, Vol.128, No. 1, pp. 118-139.

Kemsley, W., and Nicholson, J. (1960). Some Experiments in Methods of Conducting Family Expenditure Surveys. *Journal of the Royal Statistical Society*, Vol.123, No. 3, pp. 307-328.

Mooney, H. (1962). Methodology in Two California Health Surveys. (*Public Health Monograph* No.70).

Pearl, R. (1977). The 1972-73 U. S. Consumer Expenditure Survey: A Preliminary Evaluation. *Proceedings of the Section on Social Statistics, American Statistical Association*, pp. 492-497.

Silberstein, A., and Jacobs, C. (1989). Symptoms of Repeated Interview Effects in the Consumer Expenditure Interview Survey. In D. Kasprzyk, G. Duncan, G. Kalton and M.P. Singh (eds.),pp 289-303, New York: John Wiley & Sons, Inc.

Silberstein, A. (1990). First Wave Effects in the U.S. Consumer Expenditure Interview Survey. *Survey Methodology*, Vol.16, No. 2, pp. 293-304.

Silberstein, A., and Scott, S. (1991). Expenditure Diary Surveys and Their Associated Errors. In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman (eds.), *Measurement Errors in Surveys*, pp 303-326, New York: John Wiley & Sons, Inc.

Spencer B. (2000). An Approximate Design Effect for Unequal Weighting When Measurement May Correlate With Selection Probabilities. *Survey Methodology*, Vol.26, No. 2 (December, 2000), pp. 137-138.

Turner, R. (1961). Inter-Week Variations in Expenditure Recorded During a Two-Week Survey of Family Expenditure. *Applied Statistics*, Vol.10, No. 3 (November, 1961), pp. 136-146.

Wolter, K. (1979). Composite Estimation in Finite Populations. *Journal of the American Statistical Association*, Vol. 74, No. 367. (Sep., 1979), pp. 604-613.

Wolter, K. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.

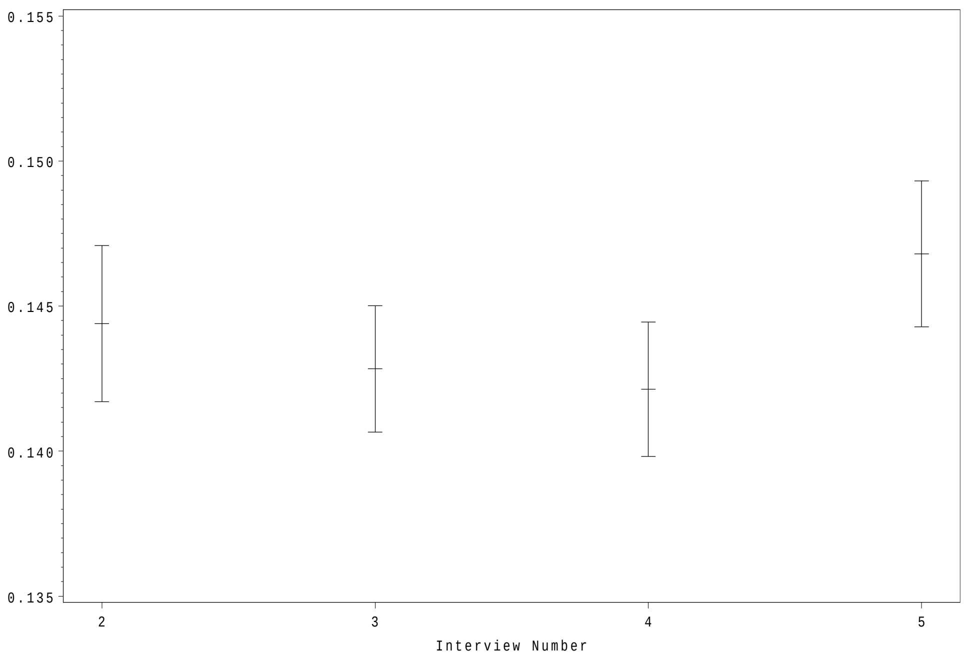

Figure 1: Weighted Reporting Rates and Associated 95% Confidence Intervals for Each Interview Number

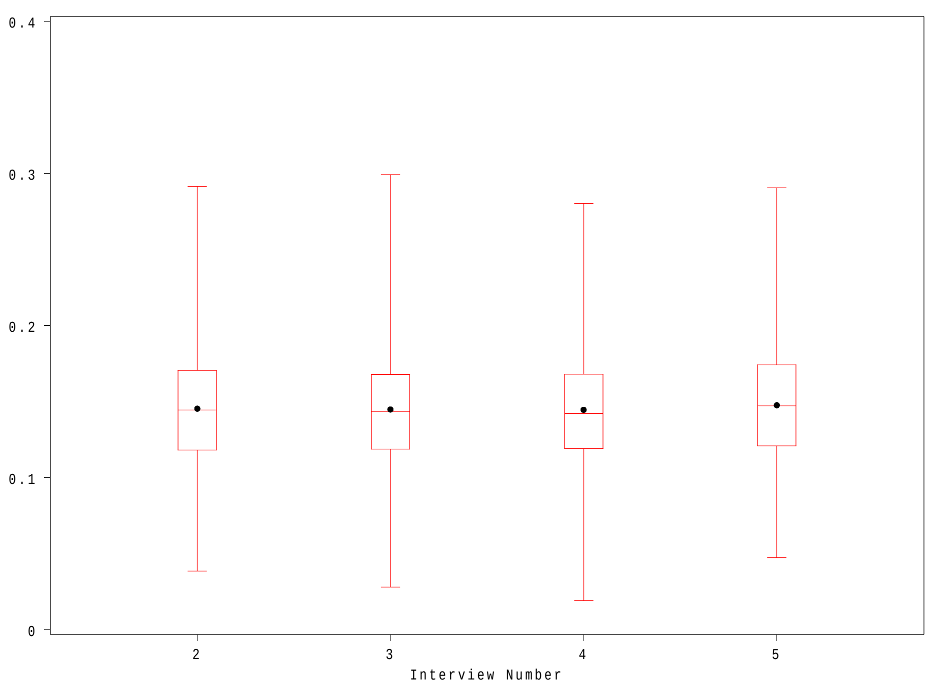

Figure 2: Box Plot of Interviewer Reporting Rates for Each Int Num

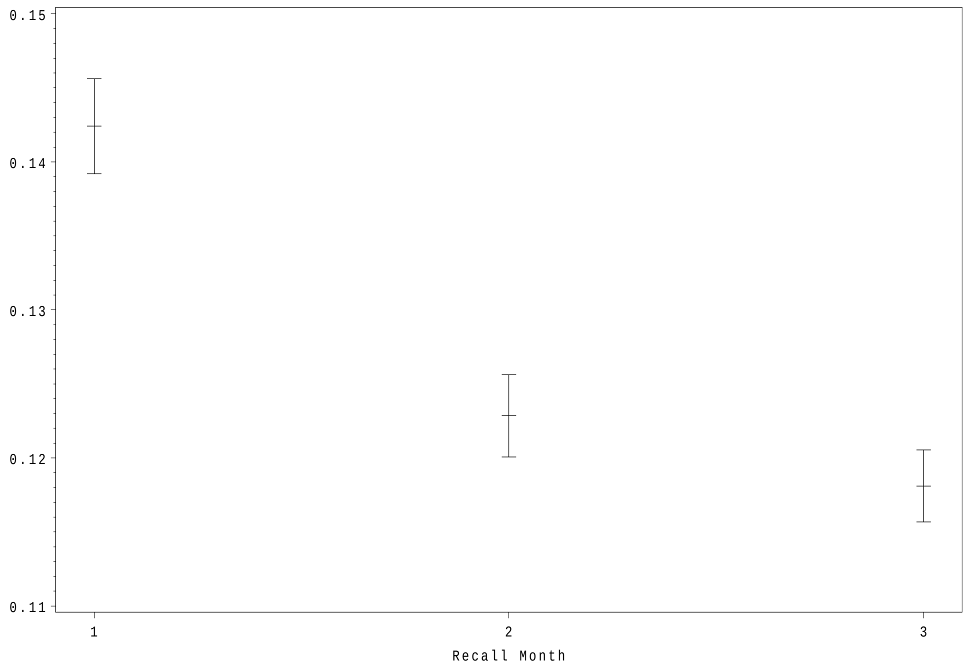

Figure 3: Confidence Interval of Weighted Reporting Rate for Selected UCCs for Each Recall Month

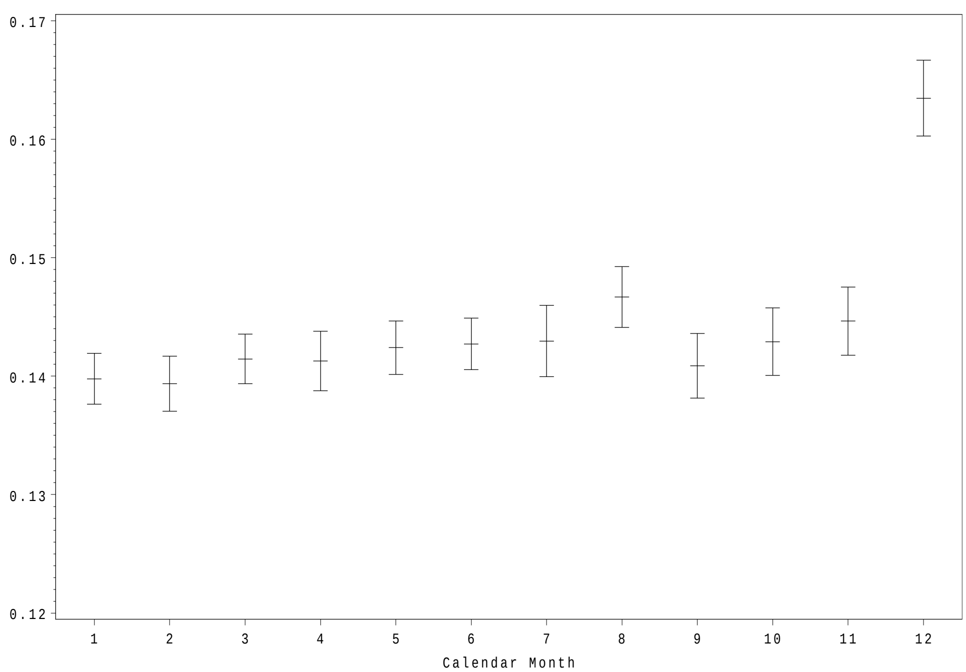

Figure 4: Confidence Interval of Weighted Reporting Rate for Each Calendar Month