

Consumer Expenditure Survey Measurement Error Study Phase 1

Report on the State of Knowledge on Measurement Error in the CE

Roger Tourangeau

Westat

Scott Fricker, Brandon Kopp, and Lucilla Tan

Bureau of Labor Statistics

Submitted: January 7, 2013

Revised: February 19, 2013

I. Objective

The Consumer Expenditure Survey (CE) provides information on the spending patterns of consumers living in the United States. Data from the CE are used to update the expenditure weights in the computation of the Consumer Price Index (CPI). Data are also used for analyses of the relationship between income and consumption, and the development of experimental poverty thresholds. Given the importance of CE data, it is vital that they be accurate. This is easier said than done, however.

The broad range of products and services covered by the CE, as well as the increasing number of outlet and payment method options, makes it difficult for respondents to remember any single purchase. In any survey that relies so heavily upon a respondent's memory, there is bound to be some degree of measurement error (Crossley & Winter, 2012; Neter & Waksberg, 1964). The task for survey administrators and methodologists is to minimize that error; an important first step in that process is to understand the magnitude, direction, and causes of measurement error.

The purpose of this report is to describe the current state of knowledge about measurement error in the CE. It is based on a review of previous work on both the CE Interview Survey (CEQ) and the CE Diary Survey (CED), the two surveys that make up the Consumer Expenditure Surveys program. While the CE also collects information on demographics, income, financial assets, and liabilities, this report focuses on measurement error in the reporting of expenditures.

The definition of measurement error is the difference between the respondent's reported expenditure for an item in the survey instrument and the true value of the expenditure incurred. Past studies of measurement error in the CE have used a variety of external benchmarks for estimating the errors in the CE expenditure reports; these include aggregate benchmarks from external sources such as the Personal Consumption Expenditure (PCE) component of the National Income and Product Accounts (NIPA) and comparisons involving the CE data themselves such as differences in reported expenditures across waves of the CE or among subgroups within a wave, and between CEQ and CED expenditures.

To guide our review of the previous studies, we formulated a series of questions that we believed should be answered to provide a basic understanding of the nature of measurement error in reported expenditures for the CE:

1. What methods or metrics have been developed to study and monitor measurement error for the CE in previous research?
2. What is the nature of reporting errors for the different expenditure categories? By reporting errors, we mean errors in:
 - a) The occurrence of an expenditure (i.e., failing to report a purchase, reporting a purchase that was never made, or reporting a purchase that was not made within the reference period); or
 - b) The amount of the expenditure (a lower or higher reported value than the actual amount of the expenditure).
3. What are the magnitudes of these reporting errors?

The research articles on measurement error in the CE vary in their comprehensiveness in addressing these questions. Still, our effort to integrate the findings from prior research can assist the CE Program in sharpening its understanding of measurement errors in expenditure reports. In addition, this report can help inform planning and set priorities for future research to monitor changes in measurement error over time. This review can also provide input useful in evaluating the long-held belief that “higher reported expenditures indicate better quality data;” this assumption traditionally has guided the CE Program in the comparison of treatment conditions that vary survey design features.

The rest of this report is organized into four sections. In Section II, we briefly describe our approach to summarizing the state of knowledge about measurement error in the CE. Then in Section III, we provide an overview of the methods and objectives of prior measurement error analyses and attempt to identify the relative strengths and weaknesses of each method. Section IV summarizes findings about measurement error from these previous studies. Finally in Section V, we summarize the current state of knowledge about measurement error, identify gaps in that knowledge, and suggest possible ways to addressing these gaps.¹

II. Approach to review of previous studies on measurement error for the CE

The Study Team identified past studies on measurement error in the CE by reviewing the *CE Methods Research* listing of articles and reports from the CE Research Library and the *CE Data Comparisons and Presentations* listing on the CE public website (<http://www.bls.gov/ce/cecomparison.htm>); in addition, the team examined more general research that seemed relevant to an assessment of the errors in the CE.

Due to time constraints, the Study Team selected only a subset of the previous measurement error studies for careful review. The selection of these previous studies was based on the following criteria:

1. How recently the study was done (or published),
2. The item categories it covered (to help us identify what categories have been more or less studied),
3. The analytic approaches it used to study measurement error (to help us determine whether there were consistent trends in findings across methods for a particular item category), and
4. Whether the CE was the primary focus of the measurement error research.

That is, we tried to give priority to more recent studies, to studies that covered a variety of types of items, to studies that used different methods to examine measurement errors in reported expenditures, and to studies that focused specifically on the CE. The References section lists the papers the team has reviewed to date.

III. Methods used in previous studies on measurement error for the CE

Previous work on measurement error in the CE has used a variety of methods. These have ranged from small scale cognitive and records validation studies, to comparison with other data sources, to multivariate models of varying statistical complexity. Strictly speaking, only studies based on some form

¹ The next deliverable of the project will develop a more detailed proposal for developing methods and metrics to monitor changes in measurement error.

of validation of expenditure reports met our definition of a measurement error study. The other studies we examined focused on distributional features of measurement error, attempted to identify characteristics of respondents or items that are correlated with misreporting expenditures, and examined respondent reporting behavior that might be diagnostic of misreporting (such as recall effects, panel conditioning, and the classification of respondents into “quality groups”). Table 1 provides a classification of the prior research the Study Team reviewed, by study method and objective.

This section provides descriptions of the analytical methods listed in Table 1, and a summary of the Study Team’s assessment of the relative strengths and weaknesses of each. The interested reader is referred to the original studies for more details about the methods used in these studies.

1. Small Scale Studies

There have been only a few studies that have sought to determine the accuracy of a respondent’s reported expenditures. Unfortunately, all of these studies have been small in scale (n = 115 being the largest among them). Still, these studies call into question the “higher reported expenditures mean better quality” assumption long held by CE. We distinguish two approaches here—validation studies (which compare expenditure reports to records, such as household receipts) and balance edit checks (which check the consistency of expenditure reports with reports of assets and income).

Validation study. A study by Geisen and colleagues (Geisen et al., 2011) used the validation approach to study errors in expenditure reporting. In their study, 115 participants were interviewed twice. The second interview attempted to validate expenditures reported in the first interview using household records (such as receipts and credit card statements). Of the 3,039 expenditures reported in the initial interviews, records were provided for 36%. The study authors identify four primary factors affecting the availability of records: demographic characteristics of the participant; the date or frequency of purchase; significance of the item; and existence of online records. Participants who were non-Hispanic white, married, and home owners were more likely to have records than those who were minority group members, single, or renters. Participants were more likely to have records for items that were purchased or paid for on a regular basis or in recent months, and they were more likely to have them for more expensive items than inexpensive ones. The strength of a validation approach is that it is the closest to a ‘gold standard’ for verifying a respondent’s expenditure report. The major limitation of this study was its small sample size and the use of a convenience sample, largely due to time and financial resource constraints. We view it as definitely worth pursuing in the future.

For items reported in the first interview where a corresponding record was provided in the second, the reported amount matched the record for just over half (53%) of the items (range: 36% - 80%, depending on section; see Table 3-1 in Geisen et al., 2011). A “match” in this study was defined as a report that was within 10% of the record amount if the amount was less than \$200 or 5% of the record amount if the purchase was \$200 or more. On average, participants misreported the dollar values of the items they reported by an average of 36%; 37% of items were underestimated and 33% were overestimated and the remaining 30% exactly matched the record value. Underreporting and overreporting of purchases were relatively rare (3% and 1%, respectively; p.4 in Geisen et al., 2011). That is, almost all of the errors in this study involved the amount of the purchase and not whether the purchase had actually occurred.

The financial record-check approach offers a significant advantage over many other methods outlined in this report in that a comprehensive set of household financial records would accurately reflect the majority of their spending. That is, record validation is the closest approximation of a gold standard at a household level. That said, it is difficult to collect a comprehensive set of records, as is demonstrated in the Geisen et al. (2011) study where records were collected for only 36% of purchases. There are also a small, but significant set of expenditure situations that do not yield a record (e.g., street vendors).

Balance edit checks. Building on methodology used by Statistics Canada's Survey of Household Spending (SHS) and the RAND's American Life Panel (ALP), Fricker, Kopp, and To (2011) explored the feasibility of measuring respondents' expenditures, income, assets, and liabilities and using the (im)balance of those totals to select probes that might be used to improve reporting during the interview. This methodology assumes that balance between these items (i.e., expenditures and income being within 15-20% of one another, once savings and debt incurred have been taken into account) indicates higher data quality because all of the money coming into or going out of the household is accounted for. Study participants were given a modified CEQ CAPI interview that asked a brief set of demographic questions about the household, global expenditure questions for 34 categories (covering all of the CEQ section topics), and questions about household income and changes in assets and liabilities for a one month reference month. Only five of the 19 participants in this small-scale laboratory study were able to achieve balance, even after being asked questions aimed at reducing imbalances. The authors report that the approach of conducting real-time calculations based on respondent reports and providing interviewers and respondents with feedback useful for improving survey reports does not appear viable.

Still, using balance between respondents' reports of expenditures, income, assets, and liabilities as an indicator of measurement error (or data quality) may be useful. Statistics Canada found that, when they applied the balance edit *at the processing stage* (after data was collected and with no opportunity for follow-up), 29.4% of households were considered out of balance. Conducting a similar, post-data collection analysis with data from the CEQ, Meekins and Kopp found that 87% of households were out of balance. This large percentage of households out of balance is likely due to a number of factors. For example, in the CEQ, the reference period for expenditures (three months) does not match the reporting period for income, assets, and liabilities (one year). Disaggregating yearly income reports to correspond with a three-month reference period or aggregating quarterly reports to align with a yearly income total is likely to introduce a significant amount of error. Furthermore, this method requires comprehensive measures of each of the elements (i.e., expenditures, income, assets, and liabilities), but as the name suggests, the Consumer Expenditure Survey is primarily focused on expenditures-

2. Comparison with Other Data Sources

A second major strategy for assessing the accuracy of CE reports involves comparing estimates from the CE to those derived from other household surveys or the national accounts; in addition, some studies compare estimates from the two CE survey instruments with each other. The metric used in these comparisons is the ratio of aggregate or mean expenditures from the two data sources. There is no universally accepted error-free standard for comparison to the CE, and several external sources have been used.

Comparison to other household surveys. Several recent research efforts have compared data from the CE to data from other household surveys, including the Panel Survey on Income Dynamics (PSID; Li et al., 2010), the Medical Expenditure Panel Survey/Household Component (MEPS; Foster, 2010), the Residential Energy Consumption Survey (RECS; BLS, 2010b), the American Community Survey (ACS; BLS, 2010a), and the Survey of Consumer Finances (SCF; Johnson & Li, 2009).

These other household surveys differ from the CE in the definition of their target populations, survey designs, frequency in the conduct of the survey, survey objective(s), and expenditure categories covered. Nonetheless, the data collected in these other national surveys are based on reports by household respondents (rather than national accounts data). In addition, some of these household surveys have built-in validation features that might justify their use as external benchmarks for specific expenditure categories in the CE. For example, the MEPS uses medical provider information, wherever possible, and this information is regarded as less prone to error than household reports, although not all providers are compliant or provide complete information. The provider data supplement (or replace) respondent reports on hospital, physician, and prescription drug spending. With the respondent permission, the RECS obtains the household's energy billing data from the energy provider. In other instances, the survey is mandatory, such as the ACS with a response rate in the mid-90% range since its inception in 2000.

In the studies reviewed, the authors attempt to make adjustments to render estimates from the CE and the other household survey source more comparable in scope and definition. While these adjustments cannot completely align the two sources of data and the data from other surveys are clearly not error-free, the strengths of these other household surveys make comparisons with them informative about measurement errors in the CE, particularly when considered in the context of other findings about the CE.

Comparison to National Accounts. The PCE of the National Income Product Accounts, produced by the Bureau of Economic Analysis (BEA), is the data source most often used as a validation measure for the CE. Among external data sources, the PCE offers the broadest range of expenditure categories for comparison. While the comparison of CE estimates to the PCE dates back to the 1980s (see http://www.bls.gov/cex/pce_compare_9091.pdf), efforts to make the two sources compatible with one another, in terms of their scope, definition of items, and aggregation across items, were not implemented until 2006 (Garner et al., 2006). That work is still ongoing (see <http://www.bls.gov/cex/cecomparison.htm>; Passero, 2012). A further refinement in this CE-PCE data comparison work was introduced by Bee et al. (2012), who compared CE estimates from the CEQ and CED survey instruments against the PCE separately (using the concordance methodology of Garner et al. (2006)). Prior to the study by Bee and colleagues (2012), comparison of CE to the PCE was done using integrated CE data. That is, where expenditure information for an item was collected in both the CE Interview Survey and Diary Survey instruments, the 'best' estimate (based on CE's source selection criteria) was selected for the CE-PCE comparison.

CE estimates are generally lower than PCE estimates, but, in several instances, not by much (for example, rents and utilities). Imputed rents of own-farm dwellings for the CE are even higher than for the PCE. In cases where the CE estimates are lower than those of the PCE, it is usually assumed that the primary reason for the difference is due to underreporting in the CE. However, it has been found that expenditure estimates from household surveys are generally lower than estimates from National Accounts in many other countries as well (e.g., Deaton, 2005); this suggests that perhaps there are reasons other than underreporting that contribute to the differences.

Notwithstanding the item concordance work begun by Garner et al. (2006) used in ongoing CE-PCE comparisons, the multiple sources of industry production data (from censuses to surveys) used to produce the National Accounts data mean that the errors associated with each of these data sources will also affect the accuracy of the PCE estimates. In addition, the estimation of the final value of items in the PCE entails many adjustments (including wholesale and retail trade margins, taxes, transportation costs); for many goods and services, values are apportioned to households as a residual after allocating the total value to other users of the item (government, exporters, and industry). For these reasons, the PCE cannot be seen as a definitive gold standard for the CE, as has been repeatedly pointed out (e.g., Garner, McClelland, & Passero, 2009; the National Research Council Panel on Redesigning the BLS Consumer Expenditure Surveys, 2012). Thus, although CE-PCE comparisons provide some sense of the magnitude and direction of possible measurement errors in the CE, these comparisons cannot be regarded as conclusive.

In light of the differences that confound the CE and PCE comparisons, Maki and Garner's (2010) attempt to quantify the fraction of the gap in the CE/PCE ratio that can be attributed to reporting error (as opposed to other factors) was a useful contribution to the literature on comparing the CE and PCE. Their work showed that misreporting is not the only factor producing differences between the CE and the PCE. They examined the demand for 10 "clusters of items" based on the 1994 PCE classification, with socio-economic and demographic characteristics as covariates in probit and tobit regression models. The authors differentiated between reporting errors that arise from respondents underreporting (i.e., failing to report a purchase that was made) and errors of underestimation (incorrectly reporting the monetary value of the purchase).² If the ratio CE/PCE was less than 1 for a category, there is underreporting for the category and the authors attempted to trace that discrepancy to reporting errors and other sources:

- If the estimated average probability of underreporting across households for a category of item from the model *equals* the CE/PCE ratio, the underreporting in the CE is fully explained by misreporting and is equal to the proportion given by the average gap from the model to the CE/PCE ratio. This *gap* ($= 1 - \text{CE/PCE ratio}$) represents the probability of underreporting across the sample households.
- However, if the estimated average probability of underreporting *does not equal* the ratio CE/PCE, the discrepancy is attributable to other factors in addition to misreporting.

For example, the CE/PCE ratio for *TVs, radios, and sound equipment* was 0.6 (see Maki and Garner, Table 8), or equivalently a gap of 0.4 ($= 1 - 0.6$). The average gap from the authors' model was 0.11, indicating that 28% ($100 * 0.11 / 0.40$) of that CE/PCE gap was attributable to misreporting.

The authors found that the probability of not reporting a purchase was associated with respondent characteristics, but the association varied by item categories. They also found that on average households correctly reported expenditure amounts. The number of households that failed to report purchases was small (ranging from 1 to 4 percent). The authors concluded that while the

²Maki and Garner (2010) use different terminology. In their paper, they refer to failing to report an expenditure as misreporting (Type 1 misreporting) and incorrectly reporting a lower expenditure than the actual amount as underreporting (Type 2 misreporting).

prevalence of misreporting (the failure to report purchases made) was low, it significantly contributed to the underreporting observed in the CE/PCE ratio.

The contribution of this method is its attempt to separate out the measurement error due to misreporting from the other sources of discrepancies between the CE and the PCE. However, as with other modeling approaches, the inferences are limited by the model assumptions and complexity. In addition, this approach did not address the possibility of reporting a purchase that had not been made (or that had not been made during the reference period), or reporting an amount higher than the actual value. Thus, though worthwhile, this approach may not offer great promise for the future.

Comparison of CE Diary and CE Interview Reports. In many studies, diary reports are treated as more accurate than reports based on respondents' memories (e.g., Thompson, 1982). Silberstein and Scott (1992) found that this is not necessarily the case in the CE surveys, at least under the assumption that higher levels of expenditure reports are better. They compared monthly totals for apparel, home furnishings, and entertainment by collection mode (i.e., Diary Survey or Interview Survey). Their analysis also took year and month into account. They examined differences in reported expenditures between the Diary and Interview for apparel, home furnishings, and entertainment as the dependent variables. They found that, for apparel, Diary Survey expenditure values are higher than the Interview Survey in every month except December, where there is a noticeable spike in Interview reports for apparel expenditures. For entertainment, reported expenditures are slightly higher in the Interview Survey than the Diary Survey in 10 out of 12 months (including December), but the difference between Diary and Interview estimates becomes much wider in December. The authors cite several reasons why Interview Survey values may be higher than Diary Survey values in December, including fear of gift disclosure if purchases are written into the Diary and a decrease in the amount of time and energy respondents are willing to devote to making entries in the Diary.

Food is one category in which the CE Interview Survey and the CE Diary Survey use similar global questions; in addition, food expenditures are recorded as individual entries in the Diary form. Battistin and Padula (2009) investigated whether the classical assumption about measurement error (i.e., the error is uncorrelated with the true value of the underlying variable) was valid in the CE. They assumed the Diary Survey reports were true and utilized a variety of regression models to test this classical assumption. They concluded that measurement error varied systematically with respondent characteristics (including family type, ethnicity, and education).

Battistin and Padula (2009) also examined distributions from the two surveys for other categories. They assumed that some categories are better measured via recall questions (*the Recall group*: Housing and public services; heating fuel; light and power; transportation; clothing; footwear and services) and other categories are better measured via diary entries (*the Diary group*: Food at home; food away; alcohol; tobacco; housekeeping services; personal care; and entertainment services). They found that the distributions of expenditures for many of these categories differed between the Interview Survey and Diary Survey. From this, they inferred that the use of recall questions for the Diary group of categories would have led to underestimates of expenditures. Similarly, the use of diary entries to collect expenditures for the Recall group would also have led to underestimates of expenditures for these categories. Aside from the complexity of this approach, the major drawback of this methodology is that it depends on the assumption that one survey instrument collects accurate data on specific expenditure categories. There is currently insufficient validation information to say definitively which CE survey instrument collects more accurate data on what categories.

Taking a computationally simpler approach, Henderson (2012) compared the ratio of food expenditure estimates collected as detailed entries in the CE Diary Survey with those collected from the global food questions asked in the CE Interview Survey. After adjusting for comparability in the categories between the two sources, the author found that over the period 1998-2011, the Diary/Interview ratio was consistently less than 1 for food at home and consistently greater than 1 for food away.³ While this approach is useful for flagging potential problems with the survey instrument when estimates from the two CE components diverge significantly, its major limitation is the lack of comprehensive validation information to say definitively which CE survey instrument collects more accurate data for food.

3. Comparisons within the CE Interview Survey

In addition to comparison to other data sources, there have been data comparisons within the CE Interview itself. These have involved examining reporting patterns within the Interview survey across quarters of the survey or across months within a single quarter. These methods of internal comparison, in general, have several advantages over other approaches discussed in Section III. Unlike comparisons with other household surveys and the PCE, these methods are not contingent upon the continued and consistent implementation of the other survey over time. That is, if the comparison survey is cancelled or will no longer provide data, then it could no longer serve as a measurement error indicator. In addition, if the scope or methodology is changed for the comparison survey program, then changes to the measurement error time series would be difficult to interpret.

Similarly, methods involving within-CE comparisons are also useful because, to the extent that the constructs and methodology of the CE have remained stable over time, they can be carried backward to establish a measurement error time series. CE measurement error trends (as indicated by these measures) could be extrapolated as far as is deemed prudent given the changing nature of the CE surveys.

Finally, these methods are promising because the CE content and analysis knowledge already exists at BLS. Making these comparisons requires extensive preparation of the data and knowledge of the underlying constructs (i.e., expenditure categories) in order to ensure that like concepts are being compared. CE production staff and BLS mathematical statisticians are well versed in both of these areas.

These methods, however, have drawbacks as well; those will be discussed when each one is described in more detail below. The most notable shortcoming of all of the methods based on internal comparisons is that the measures chosen to indicate the underlying “true” values (e.g., the most recent month of the reference period, the first wave of a panel survey, etc.) are likely to suffer from their own measurement errors and, thus, do not represent an objective true value. Also, the use of internal data comparisons to track measurement error before and after a redesign is problematic because the changes made to the survey could affect both the “true” value and the comparison, making any changes in the apparent level of measurement error difficult to interpret.

³ In 2007, the reference period for the global question for Food Away in the CE Interview Survey was changed to “average weekly” from “average monthly”.

Comparisons of reporting across waves. Several studies have attempted to measure error in the CE Interview Survey by comparing the incidence of reporting for trips in Wave 1 (Shields & To, 2005) or the level of overall expenditure reporting in Wave 2 (Yan & Copeland, 2010) with later waves of the survey. These studies considered the earlier wave in their analysis to be the true (or at least the truer) value and any difference from that value in later waves to be indicative of measurement error.

Shields and To (2005) compared the percentage of respondents who reported having taken a trip or vacation during the first interview to those reporting trips during subsequent interviews. For two out of five types of trips—recreational trips and visiting relatives—the percentage of respondents reporting a trip was slightly higher in the first interview than in subsequent interviews ($\approx 4\%$ and $\approx 7\%$ higher, respectively). The decrease in the incidence of reporting leveled off after Wave 2. The major limitation of this study was its focus on an expenditure category that is likely a relatively rare event and this fact was not accounted for in their analyses

Yan and Copeland (2010) examined data for respondents completing their second, third, fourth, or fifth interview during the second quarter of 2008. They compared the overall mean level of expenditures (in dollars) across different waves of the interview, with the assumption that any statistically significant decrease between waves would be evidence of panel conditioning. Using this methodology and dataset, they found no evidence of panel conditioning across the full sample or among demographic subgroups; they examined subgroups they thought might have a higher propensity for decreased reporting in later waves (e.g., based on the household size, cooperativeness, etc.). The authors speculate that, by excluding the first interview, they may have missed a significant drop in expenditure reporting between the first and second interviews. Respondents may learn the procedures of the survey in the first interview and then suppress their levels of reporting in all subsequent interviews. This fits with what Shields and To (2005) found in their analyses of the vacation section. First wave reports are not directly comparable to later waves, however, because they only cover a one-month reference period and are not bounded by a previous interview.

Several studies have approached the comparison of interview waves from the opposite direction, treating later waves as true reports and differences from these values in earlier waves as indicative of measurement error (Shields & To, 2005; Silberstein, 1990). Silberstein (1990) compared recall from the first, unbounded interview in the CEQ to monthly means that combined the first (most recent) and second month of the recall period in Waves 2-5. Her assumption was that significantly higher estimates in the first, unbounded interview (relative to the later bounded interviews) were a sign of telescoping (i.e., the reporting of items purchased prior to start of the reference period). Silberstein found that for clothing, overall, first wave expenses were 14.5% higher than subsequent waves; for home furnishings, first wave expenses were 48.6% higher than subsequent waves. Examining categories within these sections, Silberstein concluded that this effect, which she attributed to telescoping, was only significant for larger purchases (e.g., coats, jackets, furs, and suits; major/other appliances; furniture; large/other household entertainment equipment). This is generally consistent with Neter and Waksberg (1964), who found higher levels of telescoping for larger home repairs.

In addition to the analysis described earlier, Shields and To (2005) also compared response patterns in which respondents reported a trip in any of the first four interviews and then subsequently reported no trips in the fifth interview to the response pattern in which respondents reported a trip only during the fifth and final interview. This latter response pattern was used as a baseline (that is, as a true value), since these respondents would not be aware of the follow-up questions and would thus not have experienced the conditioning effect. For respondents who reported a vacation in only one interview,

there were slightly (but significantly) more respondents who reported having taken a vacation in one of the first four interviews and then subsequently reported no vacations than there were respondents who reported a vacation only in the fifth interview. Shields and To (2005) cite this as evidence of panel conditioning.

Since the Shields and To (2005) analyses were performed only on the vacations section of the CEQ, it is unclear whether the findings are evidence of panel conditioning or of the natural variations in the prevalence of vacations. That is, it is reasonable that a person who reports having taken a trip in one quarter would not report a trip for several subsequent quarters. The expense, in terms of time and money, may warrant that they wait some time before taking another trip. The same could be true for other large expenses (e.g., furniture). Similar analyses should be done with sections where regular expenses across waves would be expected (e.g., clothing, miscellaneous items, etc.) to see if the underlying assumptions of Shields and To hold for such expenditures.

Comparing earlier to later months in the reference period. Silberstein (1989) examined the issue of recall error; that is, that respondents' memory for expenditures fades as the time between the purchase and the interview increases. Silberstein treated the most recent month of the reference period as the true value because it was least subject to this memory decay. For her analyses, she first took the value of expenditures in a category for the most recent month in the three month recall period and divided it by total expenditures in that category for all three months (e.g., (first month apparel expenditures/total apparel expenditures) * 100). Silberstein considered values greater than 35% indicative of recall effects, under the assumptions that the most recent month should be easiest to recall and, across respondents, one would expect an even distribution of expenditures across the three months of the recall period (i.e., 33% reported for each month). Silberstein found a number of factors contributed to respondents being classified as showing moderate or large recall effects. These factors include how many persons were present at the time of the interview, whether the respondent used records to aid their memory, the respondent's age and education, the number of other people in the household, the respondent's relationship to the other members of their household, and reports of expenditures in previous waves.

Mode effects. In their study on the effect of interview mode (face-to-face versus telephone) on expenditure reporting, Safir and Goldenberg (2008) found that respondents observed to use the information booklet or records or who completed the interview face-to-face visit reported higher expenditures than those who did not. While this approach does not provide a concrete metric for quantifying measurement error, it suggests the use of recall aids differentially affects reporting of across the different expenditure categories.

4. Multivariate models

Latent class analysis. In a series of papers, Tucker and his colleagues have used latent class analysis (LCA) to examine measurement error in the CE (e.g., Tucker, Biemer, and Meekins, 2003, 2004, 2008, 2009, 2010, 2011a, 2011b). This methodology assumes an error model for the CE data and uses maximum likelihood estimation techniques to estimate parameters of the model. The basic approach the authors have taken in this program of research has been (1) to identify a set of variables believed to be related to reporting error, (2) apply LCA to these data to extract a small number of response error latent classes (e.g., 'poor,' 'fair,' or 'good' reporters), (3) assign consumer units (CUs) to one of these classes based on the results of the model, and (4) examine key performance measures (e.g., expenditure

means) for different CE commodities across the latent classes. For example, in one of the first applications of this method to CE data, the authors examined four indicators of underreporting behavior in the CE Diary: the difference between 1st and 2nd week expenditure reports; the difference between respondents' average weekly expenditures for groceries as reported in the diary vs. their reported 'usual' weekly grocery expense; a measure of respondent style based on interviewer assessments (i.e., 'complainers,' 'misleaders,' and 'resisters'); and the extent to which diary expenditure information was collected by recall (vs. recorded diligently prior to diary pick up). The results of this study suggested that the response error latent variable was a better measure of underreporting than any of the observed indicators taken individually and that diary expenditure means decreased monotonically across the three latent classes (i.e., \$94.22, \$65.50, and \$44.40 for the low error, moderate error, and high error groups, respectively).

Subsequent research by these authors primarily has focused on data from the CEQ, exploring a range of LCA models using a broad array of potential response error indicators (see Table 6 below for our summary), and incorporating information about reporters and non-reporters from multiple interview waves. The studies appear to have some promise in identifying potential predictors of CEQ reporting error. In particular, variables capturing income missingness, record use, interview length, number of completed interviews, reluctance due to time constraints, and average number of attempts consistently have been found to be associated with response error latent classes. However, model performance in many of these studies – in terms of their ability to identify latent classes that are predictive of level of underreporting – is inconsistent and generally poor. Although some studies have found the expected monotonic increase in expenditures as one moves from the 'poor' to 'good' latent class, others failed to find this pattern or showed that results varied depending the level (e.g., totals vs. category-level) and type of expenditures examined (see Tucker et al., 2011a, for a review).

In addition to using this method to assign units to certain response error latent classes and then examining expenditure reporting within those classes, the authors have applied this approach to estimate the magnitude of the bias in CEQ expenditure reports. Tucker et al. (2011b) modeled the accuracy of 20 expenditure commodities using data for CUs that completed all five interview waves. The model included several response error indicators (CU size, age of respondent, education of respondent, income percentile, type and frequency of record use during the interview, and length of interview), as well as patterns of purchases for a given commodity over the CUs' panel life. Using Markov LCA, the authors then compared the percentage of CUs that reported a purchase with an estimated 'true' (model-based) purchase prevalence, computed an accuracy rate (reported/true), and then examined those accuracy rates against CE/PCE ratios for the same commodity (see Table 6, Tucker et al., 2011b). The results show a wide variance in accuracy across expenditure categories, ranging from near 100% for items that are purchased regularly (e.g., electricity, gas, cable TV, trash collection), to less than 50% for infrequent or non-salient purchases (e.g., minor vehicle services, eye care, clothing accessories, furniture). Moreover, the commodities with the highest estimated accuracy rates also were those that had the highest CE/PCE ratios; the lower accuracy rates estimated from the model tended to match CE/PCE ratios less well, but generally were in the same direction (i.e., lower ratios for lower accuracy items).

In sum, LCA models may offer an attractive approach for assessing the measurement properties of survey items in the absence of an external benchmark. Applications of this method to CE data have been successful in identifying groups of covariates (respondent, survey design, and interview-level variables) associated with reporting error (specifically, with underestimation), but have been less effective in predicting the level of reporting error across latent classes. Moreover, these models often

rely on very strong assumptions that may be invalid or difficult to satisfy, and there is some evidence that quantitative estimates from LCA may not always align with those from more traditional analyses involving record check data (e.g., Kreuter, Yan, and Tourangeau, 2008).

Multilevel models. Since each respondent reports on multiple expenditures, the measurement errors for these reports are likely correlated within respondents. Using data from the Geisen et al. (2011) small-scale validation study, Strohm et al. (2012) use multilevel modeling to analyze the relative contribution of respondent characteristics and item characteristics to measurement error. The measurement error metric was the difference between the reported value in the survey and record value (i.e., the value of the expense from a record of the expense, such as a receipt, credit card statement, etc.) as a proportion of the record value. They found that almost all the variance in the sample's measurement error was attributable to the item level (93%), but that individual item characteristics (amount, category, time elapsed since purchase) in the model only accounted for 4% of the item-level variation. In contrast, total variance attributable to the respondent level was 7% but respondent characteristics in the model accounted for 46% of the respondent-level variation. The potential advantage of this approach when validation data are available is that it shows which types of items are most susceptible to error and the item characteristics that can account for the variation in measurement error. However, as pointed out by the authors, the findings of this study are limited by its small scale, use of convenience sample, and the limited number of categories examined.

IV. Findings from the Studies Reviewed

In this section, we summarize findings across the studies we reviewed for this report. First, we highlight quantitative results about the magnitude and direction of measurement error; then, we review findings relating to indicators and characteristics predictive of respondents' reporting quality.

The approach we took to the task of summarizing both types of findings has been to focus on item categories that have been examined in more than one of the studies we reviewed. We adopted this approach because the inferences about an item category or reporting behavior based on multiple analytical methods are likely to be more "robust" than conclusions based on any single method; this is because no single method is without its limitations.

We highlight findings for the CE Interview Survey and Diary separately, although studies on the Interview Survey are far more prevalent than studies on the Diary.

1. Summary of quantitative findings about the magnitude and direction of measurement error. Table 2 displays the general item categories examined by studies that compared expenditures collected in the CE to some other source besides the PCE. We intentionally left the CE-PCE studies out of Table 2 because we wanted to understand the range of categories that had been analyzed for measurement error by other methods in prior research. (Table 4 covers the CE-PCE comparisons.) As Table 2 indicates, the range of distinct expenditure categories that have been previously analyzed for measurement error is limited relative to the number of categories published in official CE tables. Looking down the column of Table 2 shows that among these categories, only *Food, Energy-Utilities, Rent-Mortgage, Health, Appliance-Furniture-Furnishings, and Clothing* have been examined in two or more studies. We report on quantitative findings across the studies for these six categories later. Table 3 offers another perspective of expenditure category coverage across the studies we reviewed by the method of analysis; it highlights the CE categories that have and have not been researched.

The PCE is the single external data source that has the most categories available for comparison with the CE, and this may partly explain why the PCE is the source of data against which the CE is most often compared. Table 4 shows the 50 most comparable item categories for the CE-PCE comparisons based on the item concordance methodology of Garner et al. (2006) and updated by Passero (2012). The information presented in Table 4 is based on Passero (2012), highlighting these comparable categories classified into three broad groups — Durables, Nondurables, and Services. The categories within each classification were sorted in descending magnitude of the CE-PCE ratio. Among these 50 categories, only *Sewing items* and *Imputed rental of owner-occupied nonfarm housing* had a CE-PCE ratio of one or more. The prevalence of ratios less than 1.0 gives the clear impression that underreporting may be common in the CE.

Findings about the magnitude and direction of the six broad categories in Table 2 that have been the focus of at least two studies are summarized in Tables 5a through 5f, with one broad category appearing in each table in this series. Each table identifies whether the CE Interview, Diary, or integrated survey is the focus of analysis, the method of study, the metric used, the item subcategory (other than “total”, the labels used for this column are the same as the labels that appeared in the individual studies), the result of analysis and data period, and the citation to the study. If a study analyzed multiple data periods, we report the results for the most recent data period in these tables. While findings from comparisons to PCE are shown in these tables, our summary description of the results highlights categories or subcategories that have been analyzed by other methods as well.

Food & nonalcoholic beverages for off-premises consumption (see Table 5a). When compared with PCE data, this category is underestimated in both the CEQ and CED. However, the CEQ overestimated this category when compared to the PSID, another household survey.

Clothing (see Table 5b). When compared with PCE data, this category and its subcategories (*Women and girls excluding shoes, Men and boys excluding shoes, Shoes and footwear*) are underestimated in both the CEQ and CED. There was no comparison with other household surveys for this category. Evidence of recall effects from examining levels of reported expenditure over the 3-month reference period and latent class analyses in the CEQ also suggest underestimation in this category.⁴ In contrast, the small scale validation study indicates overestimation in this category and a subcategory (*sewing services*).

Mortgage & rent (see Table 5c). The labeling of subcategories across studies for this category was less uniform than that of the preceding two categories. There were no comparisons to PCE data for mortgages. Compared to other household surveys (such as ACS, SCF, and PSID), the (*primary*) *mortgage* is underestimated in the CEQ; there were no comparisons for the CED in this category. However, the small scale validation study indicates overestimation in *Mortgage/loans*, and ACS indicates overestimation of the 2nd *mortgage* in the CEQ. When compared with PCE data, *Rent & utilities* is (slightly) underestimated in both the CEQ and CED. However, the small scale validation study indicates overestimation of *Rent*.

⁴ In Tucker et al. (2011b) and Maki and Garner (2010), the authors assume there is no telescoping forward, so all errors from their analyses were in the direction of underreporting.

Utilities-phone-internet (see Table 5d). The labeling of subcategories across studies for this category is also less uniform than it is for the first two categories. For *utilities* across household surveys, there were findings of both underestimation (ACS) and overestimation (PSID) in the CEQ. The small-scale validation study also indicates overestimation of *utilities*. *Electricity* and *natural gas* show overestimation in the integrated CE compared to RECS, but underestimation in the CEQ compared to the ACS and by latent class analyses. *Fuel oil* and *liquefied petroleum gas* was found to be underestimated in the integrated CE compared to RECS, but also the CEQ was found to be overestimated compared to the ACS. *Telephone lines* and *Internet service* were found to be underestimated by small scale validation study.

Household appliances, furniture, and furnishings (see Table 5e). When compared with PCE data, *household appliances, furniture and furnishings* are underestimated in both the CEQ and CED. There were no comparisons to other household surveys in this category. However, the small-scale validation study indicates overestimation in *major appliances*, but underestimation in *minor appliances*. Latent class analyses indicated underestimation of *kitchen accessories* in the CEQ.

Healthcare (see Table 5f). All the findings for this category apply to the CEQ. When compared with the National Health Expenditure Accounts (NHEA), a component of the national accounts data, as well as other household surveys (PSID, MEPS), *total health care* is underestimated in the CEQ. In the subcategory *physician services*, both the NHEA and the MEPS household survey indicate underestimation in the CEQ. However, for the subcategory *prescription drugs*, MEPS indicates underestimation but NHEA indicates overestimation in the CEQ. For the subcategory *health insurance*, there was no comparison to national accounts data, but comparison with PSID and the small-scale validation study indicated underestimation in the CEQ. Latent class analyses indicated underestimation of *drugs and medical supplies* in the CEQ.

2. Factors associated with reporting quality. An indirect approach to measurement error analysis is to identify indicators of reporting quality. In Table 6, we summarize findings on characteristics of respondents, items, and the data collection process that were found to be associated with various measurement error indicators in each study. A few patterns emerge from Table 6. Looking across each row of Table 6, past research has found that respondent age, race/ethnicity, education attainment, and home ownership are related to multiple indicators of measurement error. In addition, the mode of interview has also been a common predictive factor. Respondent age, race/ethnicity, marital status, education attainment, and household size were also associated with the use of recall aids. Among data collection process characteristics, contact with respondent by personal visit and contact attempt strategies of setting appointments and use of refusal letter were found to be associated with reporting quality as well.

V. Conclusions about the State of Knowledge of Measurement Error in the CE

1. What methods or metrics have been developed to study and monitor measurement error for the CE in previous research?

The fundamental difficulty with quantifying the measurement error for the expenditures reported in the CE is that some record that documents the actual expenditure incurred is the only source for a true value for that expenditure. In addition, there is also the error that occurs when incurred

expenditures are not reported at all, although at least one study suggests that such errors are rare (Geisen et al., 2011).

To date, there has been only one small scale records-based validation study conducted for the CE Interview Survey (Geisen et al., 2011, leaving aside Fricker and Edgar's, 2010, preliminary study). Other methods to study measurement error in the CE have ranged from comparison of CE expenditure estimates to other data sources (predominantly to the National Accounts, but also for a much smaller group of spending categories to other household surveys) to more complex modeling methods that attempt to provide evidence of misreporting behavior or to identify characteristics of respondents, items, and the data collection process that are associated with misreporting. Table 1 provides a summary of the range of measurement error methods used by the studies reviewed for this report and Table 6 provides a summary of the characteristics associated with misreporting..

2. What is the nature of reporting error of the expenditure categories? By reporting errors, we mean errors in [a] the occurrence of an expenditure (i.e., either the failure to report a purchase or as reporting a purchase that was never made or was not made within the reference period); or [b] the amount of the expenditure (a lower or higher reported value than the actual amount of the expenditure).

We found that with the exception of Maki & Garner (2010) and Geisen et al. (2011), the other studies did not explicitly differentiate between error in reporting the occurrence of a purchase (under/over-reporting) and errors in reporting the amount paid (under/over-estimation).

Relative to the number of expenditure categories that appear in official CE published tables, we found that only a very small subset of the official CE published spending categories have been the focus of past measurement error research, and this subset of categories is even smaller when comparisons of CE estimates to National Accounts data (PCE, NHEA) are excluded (see Table 2 and Tables 5a-5f).

3. What are the direction and magnitudes of these reporting errors?

Given our lack of understanding of the error components in other data sources used for data comparison, the lack of uniformity in the data periods covered, and the strong assumptions behind the modeling methods, we decided against summarizing the magnitude information that already appear in Tables 4 through 5e.

We found that for the categories examined by more than one method, findings about the direction of measurement error were mixed.

- Maki and Garner (2010, see Table 9) estimated error in reporting incidence (failure to report incurred spending) to be “relatively small” — between one to four percent across the ten consumer durable categories they examined.
- Whereas comparisons of CE estimates (Diary and Interview surveys) with PCE data predominantly indicated underestimation or underreporting in the CE, findings from comparisons (for a very limited number of categories) with other household surveys were mixed (see Tables 5a-5f).

- The small-scale validation study (Geisen et al., 2011; Table 4-1) suggests that magnitudes of error in reporting incidence (underreporting) is relatively less severe than errors in magnitudes of reported amounts and that overestimation is more likely than underestimation in nine of the 16 item categories studied.
- Among the 6 categories in Tables 5a through 5f, the most consistent findings across the methods were for *health care*: comparisons to National Health Expenditure Accounts and household survey (PSID, MEPS) indicate *total health care* is underestimated in the CEQ. For *health insurance*, both comparison with the PSID and the small scale validation study also indicated underestimation in the CEQ. Findings for the remaining five categories were not consistent across the methods. Given CE's historic concern with underreporting, it is worth noting that the small-scale validation study indicated overestimation in *Clothing, Sewing services, Mortgage/loans, Rent, Utilities, and Minor appliances*.

Summary

Estimates from the CE most often have been compared to estimates from the National Accounts, particularly the PCE. These comparisons almost uniformly suggest underreporting in the CE. This picture changes when we look at comparisons between the CE and other household surveys, where often the CE estimates are nearly equal to or higher than other surveys measuring comparable expenditures. Only one study compares CE reports directly with respondent records and, although this is a small study with a convenience sample, it suggests that overestimation is just as common as underestimation in the CE. Model-based analyses of extant CE data suggest that the direction and magnitude of measurement error varies considerably by expenditure category and respondent characteristics. Thus, the conclusions to date seem strongly dependent on the method and external comparison data source for estimating measurement errors in the CE.

VI. Next Steps

The next step following work on this report is to write a proposal for developing methods and metrics to monitor measurement error. Based on the information garnered from work done for this report, the following are preliminary thoughts for further consideration in the development of the proposal.

Comparison with Personal Consumption Expenditures (PCE)

There is a compelling case for using the PCE for ongoing comparisons with the CE data as there is no other single source of nationally representative data that covers the breadth of expenditure categories the PCE does. The relative ease of becoming familiar with one set of data and setting up a process for periodic comparisons makes the use of the PCE compelling. However, if the PCE is selected for comparison with the CE surveys, a better understanding of the magnitude and direction of measurement errors in the PCE is warranted.

1. Publish the degree of item or category concordance between the CE and PCE for comparable categories. Future comparisons could include only those expenditure categories that meet a certain threshold of item concordance.

2. Investigate the measurement error properties of the PCE to assess its appropriateness as the primary external benchmark for the CE.

Comparison with household surveys

3. Look for additional surveys that cover a broader range of CE expenditure categories than the ones we found, including private organization and state government surveys and industry sales data (e.g., loyalty cards, "Google Price Index").
4. Determine the level of alignment and coverage of expenditure categories. As comparable expenditure concepts are found, a methodology for comparing estimates with CE estimates will need to be developed.

Records Validation

5. Conduct validation studies with a larger, nationally-representative sample to draw firmer conclusions about the nature of measurement error in the CE. The CE Records Study (Geisen et al., 2011) demonstrated the feasibility of a records check study.
6. Methods should be developed, for both the production CE survey and for a potential record validation sub-sample, to collect more records (or all records) within a household. Participants in the CE Records Study provided records for only 36% of their purchases. Further, Strohm et al. (2012) found that there is a relationship between what records participants were willing and able to provide and the accuracy of their reports for those expenditures. Geisen et al. (2011) recommended prospective record collection, though this may cause participants to alter their behavior.
7. Develop a methodology for analyzing records validation data. That is, would respondents in the record check sample be matched up with similar respondents in the non-records sample (i.e., current CE procedures) or would respondents be asked to make memory-based reports that are then compared with their records, as was done in Geisen et al. (2011)? Also, some respondents in the non-records sample would provide records on their own or based on the subtle prompts included in the CEQ instrument. It is not clear how this would be accounted for in analyses.

Internal Data Comparisons

8. Additional internal comparisons to see whether these methods could be used with a broader range of expenditure categories. To date, the work using these methods has focused on a narrow set of consumption categories. For example, work comparing the most recent month of the reference period to the overall total has been done only for clothing and home furnishings.

Overall

1. Compare CE with several outside sources simultaneously rather than just one. Each comparison survey or data source will have its own measurement errors (i.e., noise), but looking across

several “flawed indicators” may provide a more comprehensive picture of measurement error that any single indicator could.

2. To that end, additional research time and funding should focus on finding multiple data sources that align with CE expenditure categories (e.g., PCE, ACS, MEPS, and RECS) and conducting comparison analyses for the same or similar data publication periods.

Table 1. Previous studies on Measurement Error in the CE, by Methods and Objectives

Method	OBJECTIVE OF ANALYSIS			
	Magnitude and/or direction of error	ME distribution	Evidence of respondent behavior suggestive of misreporting (e.g. recall effects, panel conditioning, use of recall aids, mode effects, reporting quality classification)	Predictive Respondent & Item Characteristics
Small Scale Studies				
a. Records validation	Geisen et al. (2011)			
b. Balanced edit check	Fricker et al. (2011)			
Data Comparison				
a. Household surveys: ACS, MEPS, PSID, RECS, SCF	Foster (2010) Johnson & Li (2009) Li et al. (2010) BLS(2010 a, b)			
b. National Accounts	Bee et al. (2012) Foster (2010) Maki & Garner (2010) Garner et al. (2006)			Maki & Garner (2010)
c. CEQ with CED	Henderson (2012)	Battistin & Padula (2009)	Silberstein & Scott (1991) Silberstein & Scott (1992)	
c. within CEQ			Shields & To (2005); Silberstein (1989; 1990) Yan & Copeland (2010) Neter & Waksberg (1963) Safir & Goldenberg (2008)	Silberstein (1989) Yan & Copeland (2010)
Multivariate Models				
a. Latent class			Tucker et al. (2010; 2011b)	Tucker et al. (2010; 2011b)
b. Multilevel				Strohm et al. (2012)
c. Other			Olson (2011)	Olson (2011)

Table 2. Coverage of Item Categories, by Studies Reviewed

Study with quantitative findings for specific expenditure categories	Food (off premises consumption)	Transportation	Housing	Energy/ Utilities	Rent	Mortgage	Other loans	Health	Appliance	Furnishings, furniture	Clothing	Subscriptions	Misc	Education	Childcare	Entertainment	Trips and vacations
BLS (2010a; ACS)				X	X	X											
BLS (2010b; RECS)				X													
Foster(2010)								X									
Fricker(2011)																	
Geisen (2011) & Strohm (2012)				X	X	X		X	X	X	X	X	X				
Henderson (2012)	X																
Johnson (2009)						X	X										
Li (2010)	X	X	X					X						X	X		
Shields & To (2005)																	X
Silberstein (1989)										X	X						
Silberstein (1990)										X	X						
Silberstein (1992)										X	X					X	
Tucker et al. (2011b)		X		X				X	X	X	X						

Note: This table covers studies comparing the CE with data sources other than the PCE; the PCE comparisons are described in Table 4.

Table 3. Coverage of expenditure categories by analysis method

		National Accounts		Comparison with household surveys					Record Validation / Multilevel models	Internal comparison			Latent class analysis
		PCE	NHEA	ACS	MEPS	RECS	SCF	PSID		CED & CEQ	Month 1 vs. 3-month	Across waves	
CE interview Survey Expenditure category													
Section 2	Rented Living Quarters	X		X					X				
Section 3	Owned Living Quarters	X		X			X		X				
Section 4	Utilities and Fuels	X		X		X		X	X				X
Section 5	Constructions, Repairs, Alterations												
Section 6	Appliances, Household Equipment	X							X				X
Section 7	Household Item Repairs and Service Contracts												
Section 8	Home Furnishings	X							X	X	X		
Section 9	Clothing	X							X	X	X		X
Section 10 & 11	Rented, Leased, and Owned Vehicles												
Section 12	Vehicle Operating Expenses												X
Section 13	Insurance Other Than Health												
Section 14	Hospitalization and Health Insurance		X		X				X				
Section 15	Medical and Health Expenditures		X		X			X					X
Section 16	Educational Expenses							X					
Section 17	Entertainment Expenses								X	X			
Section 18	Trips and Vacations											X	
Section 19	Miscellaneous Expenses							X	X				X
Section 20	Selected Services and Goods	X											
Diary	Food at Home	X						X		X			
Diary	Food Away From Home								X				

Note: This table is intended to give an overview of what sections of the CE surveys have been covered in past research. The section labels in the left-hand column do not necessarily match the expenditure categories that were included in the various analyses. In some cases, subsections were used. An "X" indicates that prior research studies have attempted to assess measurement error in at least one expenditure (sub)category in that section.

Table 4. Most Comparable Item Categories on the Basis of Concepts and Comprehensiveness for CE-to-PCE Comparisons, 2010

	DURABLE GOODS		NONDURABLE GOODS		SERVICES	
CE/PCE	0.63		0.63		0.86	
Range of CE/PCE	0.10 - 0.92		0.05 - 1.0		0.1 - 1.1	
No. of items	17		18		15	
ITEMS						
1	Pleasure boats	(0.92)	Sewing items	(1.00)	Imputed rental of owner-occupied nonfarm housing	(1.10)
2	Motor vehicles & parts	(0.89)	Pets and related products	(0.84)	Rent and utilities	(0.94)
3	Household appliances	(0.80)	Pharmaceutical products	(0.82)	Communication	(0.83)
4	Photographic equipment	(0.56)	Gasoline and other energy goods	(0.78)	Audio-video, photographic, & information processing equipment services	(0.82)
5	Furniture & furnishings	(0.54)	Food purchased for off-premises consumption	(0.66)	Veterinary & other services for pets	(0.71)
6	Personal computers & peripheral equipment	(0.54)	Nonalcoholic beverages purchased for off-premises consumption	(0.64)	Accounting & other business services	(0.68)
7	Bicycles & accessories	(0.45)	Shoes & other footwear	(0.62)	Other motor vehicle services	(0.65)
8	Other recreational vehicles	(0.41)	Household paper products	(0.51)	Purchased meals and beverages	(0.61)
9	Televisions	(0.39)	Household cleaning products	(0.50)	Household maintenance	(0.58)
10	Musical instruments	(0.39)	Women's and girls' clothing	(0.50)	Repair & hire of footwear	(0.42)
11	Glassware, tableware, & household utensils	(0.36)	Men's and boys' clothing	(0.49)	Funeral & burial services	(0.40)
12	Telephone & facsimile equipment	(0.34)	Tobacco	(0.46)	Personal care services	(0.37)
13	Jewelry & watches	(0.30)	Personal care products	(0.42)	Child care	(0.31)
14	Sporting equipment, supplies, guns, & ammunition	(0.28)	Household linens	(0.41)	Food supplied to civilians	(0.27)
15	Audio equipment	(0.27)	Clothing materials	(0.31)	Gambling	(0.10)
16	Recreational books	(0.24)	Alcoholic beverages purchased for off-premises consumption	(0.26)		
17	Outdoor equipment & supplies	(0.10)	Newspapers & periodicals	(0.17)		
18			Film and photographic supplies	(0.05)		

Source: Integrated CE and PCE comparisons based on 2002PCE Benchmark from Passero (2012, Table 1).

Note: Numbers in parentheses within each classification indicate an item category's CE/PCE ratio.

Table 5a. Food & Nonalcoholic Beverages for Off-Premise Consumption

CE Survey	Method	Metric	Subcategory	Finding	Data (latest shown in source)	Source /Study	Notes
CED	Comparison to National Accounts: PCE	Ratio of aggregate expenditure	Total	CED/PCE=0.66	2010	Bee et al. (2012, Table 1)	Compatible categories between the 2 sources (Garner et al., 2006 methodology)
CEQ	Comparison to National Accounts: PCE	Ratio of aggregate expenditure	Total	CEQ/PCE=0.86	2010	Bee et al. (2012, Table 1)	
CEQ and CED	Comparison: global question in CEQ to itemized entries in CED	Ratio of aggregate expenditures	Total	CEQ/CEQ = 0.78; consistent trend of lower CED estimate between 1998 - 2011	2011	Henderson (2012, Table 1)	Adjustment made for comparable estimates. In 2011, within Diary global and itemized comparison also found lower estimate from itemized entries.
Integrated CED and CEQ	Comparison to National Accounts: PCE	Ratio of aggregate expenditures	Food	CE/PCE=0.63 for food (not shown for non-alcoholic beverages)	2002	Garner et al. (2006, Table 2)	Compatible categories between the 2 sources (Garner et al., 2006 methodology)
Integrated CED and CEQ	Comparison to National Accounts: PCE	Ratio of aggregate expenditures	Food	CE/PCE=0.66	2010	Passero (2012, Table 1)	Compatible categories between the 2 sources (Garner et al., 2006 methodology)
			Nonalcoholic beverages	CE/PCE=0.64			
			Alcoholic beverages	CE/PCE=0.26			
CEQ	Comparison to HH survey: PSID	category as % of total expenditures		CE=15.1%; PSID=15.3%	2001	Li et al. (2009, Table 4)	PSID data collected using expenditures concept, not consumption.
CEQ	Comparison to HH survey: PSID	Ratio of mean expenditures		CE/PSID =1.10 (2003)	2003	Li et al. (2010, Table 4)	

Table 5b. Clothing

CE Survey	Method	Metric	Subcategory	Finding	Data (latest shown in source)	Source /Study	Notes
CED	Comparison to National Accounts: PCE	Ratio of aggregate expenditure	Total	CE/PCE=0.49	2010	Bee et al. (2012, Table 1)	Compatible categories between the 2 sources (Garner et al. 2006 methodology)
CEQ	Comparison to National Accounts: PCE	Ratio of aggregate expenditure	Total	CEQ/PCE=0.32	2010	Bee et al. (2012, Table 1)	
CEQ	Validation: records feasibility study	% Matching reports for the category	Total (n=96)	45%	2011	Geisen et al. (2011, Table 4-1)	Over- or underestimation indicates the average directional difference between the respondent report and record for purchases in the category. The magnitude of difference is the average absolute difference between the respondent report and record for purchases in that category.
		% Overestimation	Total	5%			
		Size of difference	Total	55%			
CEQ	Compare most recent month of recall period to other 2 months by taking ratio of first month to three month total	Underreporting = month 1 reported expenditures / (total 3 month reported expenditures)	Total	29% of HH had Moderate recall effect (35% to 75%); 27% had Great recall effect (75% to 100%)	1984	Silberstein (1989, Table 2)	Silberstein considers values greater than 35% indicative of "recall effects" because across households one would expect monthly expenditures to be about equal across the reporting period (i.e., 33%/month). Higher reported totals in the most recent month indicate underreporting due to poor recall for the previous two months.
CEQ	Latent class analysis	% reported / % true purchase	Total	88.8%	1996-1998	Tucker et al. (2011b, Table 6)	Authors assume that probability of reporting a purchase when none made is 0.
Integrated CED and CEQ	Comparison to National Accounts: PCE	Ratio of aggregate expenditure	Women and girls excluding shoes	CE/PCE=0.59	2002	Garner et al. (2006, Table 2)	Compatible categories between the 2 sources (Garner et al. 2006 methodology)
Integrated CED and	Comparison to National Accounts: PCE	Ratio of aggregate expenditure	Women and girls clothing	CE/PCE=0.50	2010	Passero (2012, Table 1)	Compatible categories between the 2 sources (Garner et al. 2006

Table 5b. Clothing

CE Survey	Method	Metric	Subcategory	Finding	Data (latest shown in source)	Source /Study	Notes
CEQ							methodology)
Integrated CED and CEQ	Comparison to National Accounts: PCE	Ratio of aggregate expenditure	Man and boys excluding shoes	CE/PCE=0.49	2002	Garner et al. (2006, Table 2)	
Integrated CED and CEQ	Comparison to National Accounts: PCE	Ratio of aggregate expenditure	Man and boys clothing	CE/PCE=0.49	2010	Passero (2012, Table 1)	
Integrated CED and CEQ	Comparison to National Accounts: PCE	Ratio of aggregate expenditure	Shoes	CE/PCE=0.71	2002	Garner et al. (2006, Table 2)	
CEQ	Latent class analysis	% reported / % true purchase	Shoes	61.6%	1996-1998	Tucker et al. (2011b, Table 6)	
Integrated CED and CEQ	Comparison to National Accounts: PCE	Ratio of aggregate expenditure	Shoes and other footwear	CE/PCE=0.62	2010	Passero (2012, Table 1)	
CEQ	Validation: records feasibility study	% Matching reports for the category	Services/Sewing (n=32)	50%	2011	Geisen et al. (2011, Table 4-1)	
		% Overestimation	Services/Sewing	8%			
		Size of difference	Services/Sewing	12%			

Table 5c. Mortgage & Rent

CE Survey	Method	Metric	Subcategory	Finding	Data (latest shown in source)	Source /Study	Notes
CEQ	Validation Study	% records match	Mortgage/loans (n=38)	69%	2011	Geisen et al. (2011, Table 4-1)	
		% overestimation	Mortgage/loans	1%			
		Magnitude of difference	Mortgage/loans	11%			
CEQ	Comparison to HH survey: ACS	Ratio of aggregate expenditure	All mortgages	CEQ/ACS=0.85	2007	BLS(2010a)	
CEQ	Comparison to HH survey: SCF	Ratio of aggregate expenditure	Primary mortgage	CEQ/SCF=0.87	2007	Johnson et al. (2009), Table 1	
CEQ	Comparison to HH survey: ACS	Ratio of aggregate expenditure	Mortgage	CEQ/ACS=0.81	2007	BLS(2010a)	
CEQ	Comparison to HH survey: PSID	Ratio of mean expenditure	Mortgage	CEQ/PSID=0.85	2003	Li et al. (2010)	PSID/CEQ=1.17
CEQ	Comparison to HH survey: ACS	Ratio of aggregate expenditure	2nd mortgage	CEQ/ACS=1.26	2007	BLS(2010a)	
CEQ	Comparison to HH survey: SCF	Ratio of aggregate expenditure	Other mortgage	CEQ/SCF=0.77	2007	Johnson et al. (2009), Table 1	
CEQ	Validation Study	% records match	Rent (n=60)	76%	2011	Geisen et al. (2011, Table 4-1)	
		% overestimation	Rent	16%			
		Magnitude of difference	Rent	27%			
CEQ	Comparison to National Accounts: PCE	Ratio of aggregate expenditure	Rent & Utilities	CEQ/PCE=0.95	2010	Bee et al. (2012, Table 1)	Compatible categories between the 2 sources (Garner et al., 2006 methodology)
CED	Comparison to National Accounts: PCE	Ratio of aggregate expenditure	Rent & Utilities	CED/PCE=0.80	2010	Bee et al. (2012, Table 1)	
Integrated CED and CEQ	Comparison to National Accounts: PCE	Ratio of aggregate expenditures	Rent & Utilities	CE/PCE=0.94	2010	Passero (2012, Table 1)	Compatible categories between the 2 sources (Garner et al., 2006 methodology)
Integrated CED and CEQ	Comparison to National Accounts: PCE	Ratio of aggregate expenditures	Rent & Utilities, excluding telephone	CE/PCE=0.98	2007	Garner et al. (2009, Table 1b)	Compatible categories between the 2 sources (Garner et al., 2006 methodology)

Table 5d. Utilities, Phone & Internet

CE Survey	Method	Metric	Subcategory	Finding	Data (latest shown in source)	Source /Study	Notes
CEQ	Validation Study	% records match	Utilities (n=91)	36%	2011	Geisen et al. (2011), Table 4-1	
		% overestimation	Utilities	1%			
		Magnitude of difference	Utilities	44%			
CEQ	Comparison to HH survey: ACS	Ratio of aggregate expenditure	Utilities	CEQ/ACS=0.88	2007	BLS(2010a)	
CEQ	Comparison to HH survey: PSID	Ratio of mean expenditure	Utility	CEQ/PSID=1.05	2003	Li et al. (2010)	PSID/CEQ=0.95
CEQ	Comparison to National Accounts: PCE	Ratio of aggregate expenditure	Rent & Utilities	CEQ/PCE=0.95	2010	Bee et al. (2012, Table 1)	Compatible categories between the 2 sources (Garner et al., 2006 methodology)
CED	Comparison to National Accounts: PCE	Ratio of aggregate expenditure	Rent & Utilities	CED/PCE=0.80	2010	Bee et al. (2012, Table 1)	
Integrated CED and CEQ	Comparison to National Accounts: PCE	Ratio of aggregate expenditures	Rent & Utilities, excluding telephone	CE/PCE=0.98	2007	Garner et al. (2009, Table 1b)	
Integrated CE	Comparison to HH survey: RECS	Ratio of aggregate expenditure	Total energy	CE/RECS=1.07	2005	BLS (2010b)	Housing unit's consumption and expenditure data provided by energy providers.
Integrated CE	Comparison to HH survey: RECS	Ratio of aggregate expenditure	Electricity	CE/RECS=1.11	2005	BLS (2010b)	
CEQ	Comparison to HH survey: ACS	Ratio of aggregate expenditure	Electricity	CEQ/ACS=0.90	2007	BLS(2010a)	
CEQ	Latent class analysis	% reported / % true purchase	Electricity	99.4%	1996-1998	Tucker et al. (2011b, Table 6)	Authors assume that probability of reporting a purchase when none made is 0.
Integrated CE	Comparison to HH survey: RECS	Ratio of aggregate expenditure	Natural gas	CE/RECS=1.12	2005	BLS (2010b)	
CEQ	Comparison to HH survey: ACS	Ratio of aggregate expenditure	Natural gas	CEQ/ACS=0.74	2007	BLS (2010b)	
CEQ	Latent class analysis	% reported / % true purchase	Gas (housing unit)	99.3%	1996-1998	Tucker et al. (2011b, Table 6)	
Integrated	Comparison to HH	Ratio of aggregate	Fuel oil & liquefied	CE/RECS=0.71	2005	BLS (2010b)	

Table 5d. Utilities, Phone & Internet

CE Survey	Method	Metric	Subcategory	Finding	Data (latest shown in source)	Source /Study	Notes
CE	survey: RECS	expenditure	petroleum gas				
CEQ	Comparison to HH survey: ACS	Ratio of aggregate expenditure	Fuel oil & other fuels	CEQ/ACS=1.17	2007	BLS(2010a)	
CEQ	Comparison to National Accounts: PCE	Ratio of aggregate expenditure	Gasoline & other energy goods	CEQ/PCE=0.78	2010	Bee et al. (2012, Table 1)	
CED	Comparison to National Accounts: PCE	Ratio of aggregate expenditure	Gasoline & other energy goods	CED/PCE=0.73	2010	Bee et al. (2012, Table 1)	
CEQ	Comparison to HH survey: ACS	Ratio of aggregate expenditure	Water	CEQ/ACS=0.99	2007	BLS(2010a)	
Integrated CEQ and CED	Comparison to National Accounts: PCE	Ratio of aggregate expenditure	Communication	CE/PCE=0.83	2010	Passero (2012, Table 1)	
CEQ	Comparison to National Accounts: PCE	Ratio of aggregate expenditure	Communication	CEQ/PCE=0.80	2010	Bee et al. (2012, Table 1)	
CED	Comparison to National Accounts: PCE	Ratio of aggregate expenditure	Communication	CED/PCE=0.69	2010	Bee et al. (2012, Table 1)	
CEQ	Validation Study	% Records match	Telephone lines(n=98)	61%	2011	Geisen et al. (2011), Table 4-1	
		% Underestimation	Telephone lines	11%			
		Magnitude of difference	Telephone lines	36%			
CEQ	Validation Study	% Records match	Telephone other(n=19)	75%	2011	Geisen et al. (2011), Table 4-1	
		% Overestimation	Telephone other	83%			
		Magnitude of difference	Telephone other	83%			
CEQ	Validation Study	% Records match	Internet service(n=78)	66%	2011	Geisen et al. (2011), Table 4-1	
		% Underestimation	Internet service	8%			
		Size of difference	Internet service	17%			

Table 5e. Household Appliances

CE Survey	Method	Metric	Subcategory	Finding	Data (latest shown in source)	Source /Study	Notes
Integrated CEQ and CED	Comparison to National Accounts: PCE	Ratio of aggregate expenditure	Household appliances	CE/PCE=0.80	2010	Passero(2012, Table 1)	
CEQ	Validation Study	% Records match	Major appliances (n=18)	80%	2011	Geisen et al. (2011), Table 4-1	
		% Overestimation	Major appliances	6%			
		Magnitude of difference	Major appliances	10%			
CEQ	Validation Study	% Records match	Minor appliances (n=88)	62%	2011	Geisen et al. (2011), Table 4-1	
		% Underestimation	Minor appliances	7%			
		Magnitude of difference	Minor appliances	20%			
Integrated CEQ and CED	Comparison to National Accounts: PCE	Ratio of aggregate expenditure	Furniture and furnishings	CE/PCE=0.54	2010	Passero(2012, Table 1)	
CEQ	Comparison to National Accounts: PCE	Ratio of aggregate expenditure	Furniture and furnishings	CEQ/PCE=0.44	2010	Bee et al. (2012, Table 1)	
CED	Comparison to National Accounts: PCE	Ratio of aggregate expenditure	Furniture and furnishings	CED/PCE=0.43	2010	Bee et al. (2012, Table 1)	
CEQ	Latent class analysis	% reported / % true purchase	Kitchen accessories	66.8%	1996-1998	Tucker et al. (2011b, Table 6)	Authors assume that probability of reporting a purchase when none made is 0.
CEQ	Validation Study	% Records match	Home furnishings (n=73)	52%	2011	Geisen et al. (2011), Table 4-1	
		% Underestimation	Home furnishings	0%			
		Size of difference	Home furnishings	69%			

Table 5f. Healthcare

CE Survey	Method	Metric	Subcategory	Finding	Data (latest shown in source)	Source /Study	Notes
CEQ	Comparison to HH survey: PSID	Ratio of mean expenditure	Total health care	CEQ/PSID=0.88	2003	Li et al. (2010)	PSID/CEQ=1.14
CEQ	Comparison to HH survey: MEPS	Ratio of mean expenditure	Total health care	CEQ/MEPS=0.71	2006	Foster (2010)	Foster (2010): Total health = Hospital care + Physician services + Dental services + Other prof. services + Prescription drugs + Medical supplies
CEQ	Comparison to national accounts: NHEA	Ratio of mean expenditure	Total health care	CEQ/NHEA=0.74	2006	Foster (2010)	
CEQ	Comparison to HH survey: PSID	Ratio of mean expenditure	Hospital & nursing home	CEQ/PSID=0.33	2003	Li et al. (2010)	PSID/CEQ=3.03
CEQ	Comparison to HH survey: MEPS	Ratio of mean expenditure	Hospital care	CEQ/MEPS=0.98	2006	Foster (2010)	
CEQ	Comparison to national accounts: NHEA	Ratio of mean expenditure	Hospital care	CEQ/NHEA=0.86	2006	Foster (2010)	
CEQ	Comparison to HH survey: PSID	Ratio of mean expenditure	Doctor	CEQ/PSID=0.96	2003	Li et al. (2010)	PSID/CEQ=1.04
CEQ	Comparison to HH survey: MEPS	Ratio of mean expenditure	Physician services	CEQ/MEPS=0.65	2006	Foster (2010)	
CEQ	Comparison to national accounts: NHEA	Ratio of mean expenditure	Physician services	CEQ/NHEA=0.43	2006	Foster (2010)	
CEQ	Comparison to HH survey: MEPS	Ratio of mean expenditure	Prescription drugs	CEQ/MEPS=0.61	2006	Foster (2010)	
CEQ	Comparison to national accounts: NHEA	Ratio of mean expenditure	Prescription drugs	CEQ/NHEA=1.01	2006	Foster (2010)	
CEQ	Latent class analysis	% reported / % true purchase	Drugs & medical supplies	94.2%	1996-1998	Tucker et al. (2011b, Table 6)	CEQ
CEQ	Comparison to HH survey: PSID	Ratio of mean expenditure	Health insurance	CEQ/ PSID=0.92	2003	Li et al. (2010)	PSID/CEQ=1.09
CEQ	Validation Study	% Records match	Health insurance(n=57)	59%	2011	Geisen et al. (2011), Table 4-1	
		% Underestimation	Health insurance	10%			

Table 5f. Healthcare

CE Survey	Method	Metric	Subcategory	Finding	Data (latest shown in source)	Source /Study	Notes
		Size of difference	Health insurance	24%			

Table 6. Predictors of Reporting Quality

Check mark indicates variable found to be most commonly significant (at least at the 5% level), across the categories a study examined

Respondent Characteristics and Other Predictors of Misreporting Behavior / Measurement Error	Battistin & Padula(2009, Appendix B, Table 5)	Maki & Garner (2010, p.144)	Geisen et al. (2011, Table 1)	Olson (2011, Tables 15-17)	Olson (2011, Tables 15-17)	Olson (2011, Tables 15-17)	Olson (2011, Tables 15-17)	Silberstein (1989; Tables 6-13)	Silberstein (1990; Table 4)	Silberstein & Scott (1992; Table 6)	Tucker et al. (2011b; Table 6)	Tucker et al. (2011a; Table 4)
Outcome Measure	Diary vs Interview Question effect on propensity	Likelihood of omitting report of purchase	Having records across categories examined	Use of info book during interview	Always used records during interview	Interview length	DK/RF responses	Ranking of recall effects	% difference in expn means; Reporting rates	Ratio of CEQ to CED expn mean	Accuracy rates	Reporting quality category
Age	X	X		X	X	X	X	X			X	X
Gender			X									
Race, Ethnicity	X		X	X	X	X	X					X
Marital status				X	X	X	X					
Family composition	X									X		
Education attainment	X	X		X	X	X		X			X	X
Employment status (hrs worked; employed/not)				X		X						
CU size				X	X			X			X	
Urbanicity				X			X					
Housing tenure			X			X	X	X				X
Income						X	X	X			X	X
Type											X	
Recency of purchase			X									
Size (amount) of expense			X						X			
Reports in Same Category in Previous Waves								X				
Reports in Different Category in Current Wave								X				

References

- Battistin, E. and Padula, E. (2009). Survey instruments and the reports of consumption expenditures: evidence from the Consumer Expenditure Surveys. Center for Economic Policy Research Discussion Paper No. DP8051
- Bee, A., Meyer, B., and Sullivan, J. (2012). The Validity of Consumption Data: Are the Consumer Expenditure Interview and Diary Surveys Informative? Working Paper 18308, National Bureau of Economic Research.
- Bureau of Labor Statistics, U.S. Department of Labor (2010a). Consumer Expenditure Survey Compared with the American Community Survey, in *Consumer Expenditure Survey 2006-2007*, Report 1021.
- Bureau of Labor Statistics, U.S. Department of Labor (2010b). Household Energy Spending: Two Surveys Compared. *Focus on Prices and Spending*, vol 1, no. 12. (author is Vera Crain)
- Crossley, T. and Winter, J. (2012). Asking Households about Expenditures: What Have We Learned? in *Improving the Measurement of Consumer Expenditures*, Christopher Carroll, Thomas Crossley, and John Sabelhaus, editors. Forthcoming from University of Chicago Press.
- National Research Council (2012). Measuring what we spend: towards a new Consumer Expenditure Survey. *Prepublication Draft report*. Panel on Redesigning the BLS Consumer Expenditure Surveys. Don A. Dillman and Carol C. House, Editors. Committee on National Statistics, Division of Behavioral, Social Sciences and Education. Washington DC: The National Academies Press.
- Deaton, A. (2005). Measuring Poverty in a Growing World, *The Review of Economics and Statistics*, Vol. 87, No. 1 (Feb., 2005), pp. 1-19.
- Dubreuil, G., Tremblay, J., Lynch, J., and Lemire, M. (2011). Redesign of the Canadian Survey of Household Spending. Survey Producers Workshop, June 2011.
- Foster, A. (2010). Out-of-pocket health care expenditures: a comparison. *Monthly Labor Review*, Feb 2010: 3 - 19.
- Fricker, S., Kopp, B., and To, N. (2011). Exploring a Balanced Edit Approach in the Consumer Expenditure Quarterly Survey, DCES report.
- Garner, T., Janini, G., Passero, W., Paszkiewicz, L., and Vendemia, M. (2006). The CE and the PCE: a comparison. *Monthly Labor Review*, Sept. 2006: 20 - 46.
- Garner, T., McClelland, R., and Passero, W. (2009). Strengths and Weaknesses of the Consumer Expenditure Survey from a BLS Perspective. http://www.bls.gov/cex/pce_compare_199207.pdf
- Geisen, E., Richards, A., and Strohn, C. (2011). U.S. Consumer Expenditure Records Study Final Report. DCES report.
- Johnson, K. and Li, G. (2009). Household liability data in the Consumer Expenditure Survey. *Monthly Labor Review*, Dec 2009: 18-27.
- Kreuter, F., Yan, T., and Tourangeau, R. (2008). Good item of bad item – can latent class analysis tell? The utility of latent class analysis for the evaluation of survey questions. *Journal of the Royal Statistical Society (Series A)*, 171, 723 – 738.
- Li, G., Schoeni, R., Danziger, S., and Charles, K. (2010). New expenditure data in the PSID: comparisons with the CE. *Monthly Labor Review*, Feb 2010: 29 - 39.
- Maki, A., and Garner, T. (2010). Estimation of misreporting models using micro-data sets derived from the Consumer Expenditure Survey: the gap between macro and micro economic statistics on consumer durables. *Journal of Mathematical Sciences: Advances and Applications*, 4(1): 123-152.

- Neter, J. and Waksberg, J. (1964). A Study of Response Errors in Expenditures Data from Household Interviews. *Journal of the American Statistical Association*, 59(305): 18-55.
- Olson, K. (2011). Responsive Survey Design for the Consumer Expenditure Interview Survey. Unpublished Internal Report
- Safir, A., and Goldenberg, K. (2008). Mode Effects in a Survey of Consumer Expenditures. 2008 AAPOR.
- Shields, J. and To, N. (2005). Learning to Say No: Conditioned Underreporting in an Expenditure Survey. Paper presented at 2005 AAPOR.
- Silberstein, A. (1989). Recall effects in the U.S. Consumer Expenditure Interview Survey. *Journal of Official Statistics*, 5(2): 125-42.
- Silberstein, A. (1990). First wave effects in the U.S. Consumer Expenditure Interview Survey.
- Silberstein, A. (1991). Methodological issues in the collection of consumer expenditures in the United States. Paper prepared for the Methodology Workshop at Statistics Sweden, Stockholm - February 18-20, 1991.
- Silberstein, A., and Scott, S. (1991). Expenditure diary surveys and their associated errors. Chapter 16 in *Measurement Errors in Surveys*, Biemer, P., Groves, R., Lyberg, L., Mathiowetz, N., and Sudman, S. John Wiley & Sons, Inc.
- Silberstein, A., and Scott, S. (1992). Seasonal effects in the reporting of consumer expenditures.
- Strohm, C., Geisen, E., Richards, A., and Kopp, B. (2012). Measurement error in self-reports of consumer expenditures: are errors attributable to respondents or expenditures? Draft paper of presentation at *American Association of Public Opinion Research*, Orlando.
- Thompson, C. P. (1982). Memory for unique personal events: The roommate study. *Memory and Cognition*, 10, 324-332.
- Tucker, C., Biemer, P., and Meekins, B. (2003). Latent class modeling and estimation of errors in Consumer Expenditure reports. Proceedings of the Section on Survey Methods Research, American Statistical Association.
- Tucker, C., Biemer, P., and Meekins, B. (2004). Estimating the level of under-reporting of expenditures among expenditure reporters: a micro-level latent class analysis. Proceedings of the Section on Survey Methods Research, American Statistical Association.
- Tucker, C., Biemer, P., and Meekins, B. (2008). A micro-level latent class model for measurement error in the Consumer Expenditure Interview Survey. Proceedings of the Section on Survey Methods Research, American Statistical Association.
- Tucker, C., Meekins, B., and Biemer, P. (2009). Using substantive diagnostics to evaluate the validity of micro-level latent class indicators of measurement error.
- Tucker, C., Biemer, P., and Meekins, B. (2010). Latent class analysis of Consumer Expenditure Reports. Proceedings of the Section on Survey Methods Research, American Statistical Association, 5441-52.
- Tucker, C., Biemer, P., and Meekins, B. (2011a). Latent class analysis of measurement error in the Consumer Expenditure Survey. Proceedings of the Section on Survey Methods Research, American Statistical Association, 5441-52.

Tucker, C., Biemer, P., and Meekins, B. (2011b). Estimating underreporting of consumer expenditures using Markov latent class analysis. *Survey Research Methods*, 5(2): 39-51

Yan, T., and Copeland, K. (2010). Panel Conditioning in Consumer Expenditure Interview Survey. Final Report from NORC.