

WORKING P A P E R

Methodological Innovations in Collecting Spending Data

The HRS Consumption and Activities Mail Survey

MICHAEL HURD, SUSANN ROHWEDDER

WR-646

November 2008

This paper series made possible by the NIA funded RAND Center for the Study of Aging (P30AG012815) and the NICHD funded RAND Population Research Center (R24HD050906).

This product is part of the RAND Labor and Population working paper series. RAND working papers are intended to share researchers' latest findings and to solicit informal peer review. They have been approved for circulation by RAND Labor and Population but have not been formally edited or peer reviewed. Unless otherwise indicated, working papers can be quoted and cited without permission of the author, provided the source is clearly referred to as a working paper. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. RAND® is a registered trademark.



LABOR AND POPULATION

**Methodological Innovations in Collecting Spending Data:
The HRS Consumption and Activities Mail Survey**

Michael Hurd
RAND, NBER and NETSPAR

Susann Rohwedder
RAND and NETSPAR

November 2008

Keywords: Survey methods, spending data, consumption, recall bias
JEL Codes: C81, C83, D91

Research support from the National Institute on Aging (P01AG08291) is gratefully acknowledged.

Abstract

It has traditionally been believed that collecting survey measures of total spending necessarily involved asking a large number of questions, too many for inclusion of a comprehensive spending measure in a general-purpose survey. In this paper we report on a supplemental survey to the Health and Retirement Study that took up this challenge. We discuss issues that arise designing a survey module to collect spending data with strict time constraints, describe how the implementation in the Consumption and Activities Mail Survey (CAMS) played out, and elicit anomalies that more detailed analysis of data quality revealed. We report how we addressed some of these anomalies in subsequent waves of CAMS. Other anomalies required conducting additional randomized experiments to find what explains the observed patterns. The results highlight the tension between asking about spending using a long time frame, which exacerbates recall bias, versus using a short time frame, which risks relying on an unrepresentative snapshot of a household's spending to proxy the total for the last 12 months. An important complicating factor in deciding which goods should be put into which time frames is that there is substantial heterogeneity in the frequency of spending across households even for the same category of spending.

Michael D. Hurd
RAND
P.O. Box 2138
1776 Main Street
Santa Monica, CA 90407
U.S.A.
mhurd@rand.org

Susann Rohwedder
RAND
P.O. Box 2138
1776 Main Street
Santa Monica, CA 90407
U.S.A.
susannr@rand.org

1. Introduction

A large body of economic theory is concerned with consumption by individuals and households. Empirical applications have often been limited in scope due to lack of adequate data on spending. Traditionally it has been thought that to obtain good spending measures one needs to ask a large number of questions, too many to be included in general-purpose surveys.¹ In the U.S. the primary data set collecting information about household spending is the Consumer Expenditure Survey (CEX). To measure spending, it asks about hundreds of spending items. CEX maintains two samples: First, an interview sample is asked about spending each month over the last three months in very detailed categories, e.g., sleeping garments for children under the age of two or repair of lawnmowers. For some categories of more frequent purchases, respondents are asked about spending in a typical week. Second, a diary sample, which is independent of the interview sample, is asked to record each day for a week *all* spending for more regular items such as food, as well as for some irregular purchases such as clothing.

While the CEX affords the opportunity to collect comprehensive spending data for the general population, measures of spending would be valuable additions to other household surveys. In such surveys, the respondent burden imposed by the CEX is likely to be infeasible, but even a brief set of questions might be helpful. Browning et al. (2003) give an overview of the variety of major uses of consumption data in applied economic research and argue that even partial measures of total consumption – based on as few as three or four questions - could serve as valuable proxies in a number of applications. Consider, for example, the Health and Retirement Study (HRS), a general-purpose longitudinal household survey of the U.S. population age 51 and older. Issues of wealth and income are of central interest in this survey, so measures of consumption would also be useful.

In this paper we discuss the major survey design issues and trade-offs that arise when developing a module for collecting household spending data in a general-purpose survey like the HRS that requires a much shortened format from the CEX, but is considerably longer than just a

¹ Economic theory models consumption, not spending. However, it is spending that is measured in household surveys. This difference is not a matter of great concern provided we have good methods of finding the flow of consumption services from consumer durables. Thus in surveys of household consumption such as the U.S. Consumer Expenditure Survey (CEX) no distinction is made between consumption and spending. That is the point of view in this paper and we will use consumption and spending interchangeably.

few questions. The objective is to elicit total consumption of the household in a survey that is of moderate length. We then report how the implementation in the HRS Consumption and Activities Mail Survey (CAMS) played out. It was first fielded in 2001 and has been repeated every two years since. Early data quality assessments showed that at the population level spending totals were quite close to those measured in the CEX, and that age patterns of wealth change implied by the comparison of spending with after-tax income were similar to actual wealth change observed in HRS panel data. However, more detailed analysis at the level of single categories showed some anomalies like outliers and striking patterns related to the choice of reference period. In response we designed additional data experiments in an Internet survey to test our hypotheses of what might explain the patterns in CAMS. We report the results of those experiments for additional insights into determinants of data quality resulting from different survey designs in spending questions.

2. Methodological Issues in Collecting Consumption Data

Whether designing a survey collecting consumption data with or without a strict survey time constraint, most pros and cons of specific design choices apply to both. In this section we discuss the most important ones, including survey mode; who responds (each person in the household or one designated household respondent); level of aggregation; and reference period.

2.1. Survey Mode

The choice of survey mode has important implications for data quality when asking respondents to recall their spending on certain categories. For most spending items respondents are likely to need to think for a while to retrieve the answer, because they may need to go through a whole sequence of recall phases: for example, did anyone in the household spend anything on [category X]? If so, who made the purchase(s), how many during the reference period; were they of about equal size; what did they add up to and so on. Self-administered interview modes such as paper-and-pencil or internet based surveys are preferable in this situation, because the respondent can take however much time he or she would like to answer the question. Interview modes that involve an interviewer, be it over the phone or face-to-face,

would tend to lead respondents to take less time to think about the answer to avoid unpleasant moments of silence in the interview leading to lower data quality. Another advantage of the self-administered survey mode is that respondents can – if they are so inclined – consult records or ask other household members to help with information gathering.

2.2. One or Multiple Respondents per Household

Spending questions can be asked of just one respondent in the household who would report on the spending of all household members, or each household member could be interviewed separately about his or her spending. Depending on the composition of the household and on whether spending decisions are coordinated among household members each method has its advantages and disadvantages. In a large household with limited coordination about spending one single household respondent is unlikely to be aware of all the spending undertaken by household members and therefore would be bound to underreport the spending of the household. However, if one were to interview each person in the household there is the risk of double-counting, because many consumption items cannot be uniquely allocated to a single person (e.g., utilities, food). Double-counting of this kind could be reduced by adding instructions that only the person who physically paid for the item should report it. Interviewing each household member adds substantial logistic complications to the survey operation, however, and increases the risk of missing data due to non-response if one or more other household members cannot be reached. Alternatively one might assign a single respondent to report on spending and encourage that person to consult family members for additional input. This approach should avoid double-counting and counteract underreporting due to the household respondent not being aware of the spending of other members in the household. However, it could be implemented only in self-administered survey modes.

2.3. Level of Aggregation

Dedicated expenditure surveys such as the CEX ask about spending in several hundred different spending categories. This level of detail is not an option in a general-purpose survey. Instead respondents are queried about more aggregated categories of spending. Among survey specialists the view is that more categories will produce a larger total, and a larger total is likely

to be closer to the true value because respondents are likely to omit spending on small categories. It is plausible that increasing the number of categories will increase the total up to some point, but it is not obvious that this will continue. In fact the total could even decrease. For example, many respondents would not be able to recall how much they spent on peaches or on lettuce (or whether they purchased those items some days ago) whereas they may have a good idea of total spending at the grocery store. On the other hand, there could also be overshooting when the number of categories becomes large in that respondents may be tempted to affirm that they spent money on a particular item in the recall period, because it is something they usually purchase.

2.4. Reference Period

Suppose the objective is to measure spending in the 12 months before the survey is filled out. The challenge is to design the survey to minimize the reliance on respondents' memories of what happened many months prior and so minimize recall bias. For measurement, the ideal situation would be that a respondent regularly spends the same amount on each item during some relatively short reference period such as a week or month. Then we would ask about spending during that reference period (which should be easily remembered) and convert that to an annual amount. However, while spending on some items may be fairly regular, that is not the case with many items. A rough categorization might run as shown in Table 1.

Obviously, choice of a meaningful reference period will vary from category to category. However, the choice will also vary *within* category across households. Some people buy clothing often, while others may buy none in a year. What is regular for some is irregular for others. Some take prescription drugs on a regular basis and make the same purchases each month. Others, not suffering from chronic conditions, take drugs only when they are sick.

Given the difficulty in selecting a single reference period, why not, if the survey is concerned with annual expenditures, fall back to asking about spending in the preceding 12 months? There are at least two reasons why this would be problematic. First, as indicated above, it is difficult for respondents to remember spending from 8 or 10 months previously, at least for items bought only irregularly. Second, it is easier for respondents to report accurately spending on regular, medium- to high-frequency items such as rent when they are asked about

the associated natural reference period (which for rent is typically per month, not per year). On such frequent expenditures, the respondent has to do the math required to convert naturally remembered amounts per week or per month to annual amounts. This may lead to inaccuracies or to high rates of item nonresponse.

The survey designer must also consider whether the objective is to devise population estimates or individual household estimates. The latter are required for studying variation in spending across households or households' variation in spending across time.

Population and individual household estimates call for different survey designs, especially with respect to which reference period to ask about. For example, in the case of irregular purchases, respondents may be asked about their purchases, say, in the preceding month. Then, for an annual average population estimate, these amounts may be multiplied by 12 and the average calculated. Zeroes are counted just like any other number. The result should be an accurate average for the population. But this method will be quite wrong if the objective is a set of accurate household estimates. For those who happened to purchase nothing in the past month, annual spending will be estimated as zero. The method will also give mistaken results for those who did make an irregular purchase in the preceding month: Multiplying by 12 will yield a very high annual spending estimate.

Both population and individual household estimates may be of interest to researchers drawing their data from the same general-purpose survey like HRS. Thus, either trade-offs must be made between survey design methods, or different methods must be used in the same survey. Of course, in making such survey design choices in a general-purpose survey, respondent burden must be borne in mind.

2.5. Wording of Reference Period: “typical month” versus “last month”

It appears at first sight that using the wording “typical month” rather than “last month” would be a way of taking advantage of the smaller recall bias of a short reference period while encouraging the respondent to make adjustments to any unusual spending patterns of recent

months. However, how respondents are supposed to accomplish this task is unclear. Take the example of prescription drugs for a person who does not take medications on a regular basis, but who was sick over the last couple of months. When asked about spending on prescription drugs in a “typical month” should this person answer zero? Or do we expect the respondent to review the household’s spending over the last 12 months and calculate monthly average spending? When interested in total spending over the last 12 months we would want the respondent do the latter. This suggests that for the case of irregular spending patterns asking about “typical” month is a masked way of asking about a much longer time period after all with the only difference that the respondent is supposed to perform some averaging to arrive at the requested answer which one would expect to affect data quality in a negative manner. The same issues apply with respect to asking about a typical week versus last week.

3. Overview of CAMS

The Consumption and Activities Mail Survey (CAMS) was designed to collect a reliable measure of total annual spending in a general-purpose, population-representative survey. In October 2001, CAMS wave 1 was mailed to 5,000 households selected at random from households that participated in HRS 2000. In households with couples it was sent to one of the two spouses at random. A total of 3,866 questionnaires were returned, which corresponds to an overall unit response rate of 77.3 percent.

CAMS is a panel study. In October 2003 a somewhat modified questionnaire was sent to the same households with exceptions for death, loss to follow-up, or participation in another HRS study (to reduce respondent burden). The response rate to CAMS wave 2 was 78%. CAMS waves 3 and 4 were mailed in October 2005 and 2007. The focus of this paper will be CAMS waves 1 and 2, because of changes we made to the questionnaire between those waves and because of some experiments we conducted in response to the wave 1 data.

CAMS wave 1 consisted of three parts. In Part A the respondent was asked about the amount of time spent in a number of activities. Part B collected information on actual spending in each of 32 categories, as well as anticipated and recollected spending change at retirement.

Part C asked about prescription drug spending and current labor force status. In this paper we will only discuss data collection on actual spending in Part B.

One of the earliest choices that had to be made in designing CAMS was the data collection format for the information on expenditures. A paper-and-pencil survey format was chosen in preference to a telephone interview, because recall of spending amounts by respondents is likely to be more accurate if they can take their time to review their spending habits and those of household members. For the more difficult items, respondents can consult records or other household members to help in arriving at the answer (see prior discussion in Section 2.1 for further advantages).

The choice of a self-administered paper-and-pencil survey meant that the survey instrument had to be very simple to follow; complex skip-patterns had to be avoided. We also wanted it to be brief – about 20 minutes – to ensure that respondents did not give up in the middle of it.

A number of other design choices had to be considered carefully. The principal ones were these:

- Level of aggregation: how many spending categories should be used?
- Reference period: should it be specified in the survey, or should the respondent be allowed to choose? If specified, how long should it be? A week, a month, a year? And how should it be worded? "Last week," "weekly," "in a typical week," "last month," "monthly," "in a typical month"?

In deciding which design choices might yield the most reliable estimates, at the population and household levels, we reached the conclusion that it depends on the specific spending category. Levels of aggregation, reference periods, and question wording might differ between very irregular, low-frequency items such as furniture and quite regular, high-frequency items such as food. Allowances would also have to be made for within-category differences in households' spending habits, e.g., the example of medications, mentioned above.

For starters, then, CAMS Wave 1 used 32 spending categories – the same ones used in CEX. It seemed that this grouping would be natural for the respondent, it promoted

comparability with CEX, and it took advantage of many years experience with CEX. In Wave 2, the number was increase to 38 categories, to include several small categories that we omitted from Wave 1 because of time constraints.²

For big-ticket items, CAMS asked whether there was a purchase in the last 12 months and what the price was. For regularly purchased items, CAMS asked for the “amount spent monthly” or the “amount spent yearly.” For many categories, the respondent could choose the reference period from “last week,” last month,” or “the last 12 months,” as shown in the extract from a Wave 1 questionnaire displayed in Table 2. The motivation behind allowing respondents to choose the reference period was that households differ in the frequency and regularity of spending habits in these categories and that item non-response would be reduced if respondents can choose whichever reference period seems most natural given their particular situation. Note that the instructions for the section with three possible reference periods to choose from encouraged respondents to use a longer reference period for items that they only purchase occasionally:

[...] If you bought an item only occasionally or on an as-needed basis, then please give your best estimate of what you spent in the last 12 months.

What have been the implications of the CAMS design for the quality of spending estimates at the population and household levels? Our general assessment is favorable. Item nonresponse rates have been low, in the single digits for most items. Spending estimates have been quite close to those in CEX: after making some adjustments to the first two waves, CAMS 2005 totals exceeded CEX totals by 4.5 percent.³ The difference between CAMS and CEX is larger at older ages. One difficulty with CEX comparisons is that the HRS and CAMS do not maintain the concept of the household head as the CEX does so that classifications by age may not refer to strictly comparable populations.⁴ To gauge whether CAMS spending levels appear

² A copy of the paper-and-pencil questionnaire for wave 1 is available online at: <http://hrsonline.isr.umich.edu/meta/2001/cams/qnaire/cams01abc.pdf>. The questionnaire for wave 2 can be found at: <http://hrsonline.isr.umich.edu/meta/2003/cams/qnaire/cams2003.pdf>.

³ Total spending in the first two waves of CAMS exceeded CEX totals for 2003 by 16 and 15 percent.

⁴ The “age” of the CEX household is the age of the reference person who is the owner or renter of the dwelling. Thus, a 80 year-old living with her 54 year-old son in his house would be classified as age 54 in the CEX. In the HRS each person is a sample member so we can study consumption by the household in which 80 year-olds live.

to be too high or too low, we compared them with after-tax income for the same households and assessed the implied saving rates, which seemed reasonable in magnitude and accorded qualitatively with predictions from a standard life-cycle model.⁵ However, it is difficult to obtain reliable outside estimates of what these saving rates should be. A test on independent data is to compare saving rates calculated from CAMS spending data and HRS after-tax income with implicit saving rates based on HRS panel wealth change. In Hurd and Rohwedder (2008) we found that – if anything – spending is underestimated in CAMS, because the saving rates out of after-tax income are too high compared to what is implied from data on wealth change. Yet, the saving rates out of after-tax income implied by CEX totals are even higher, suggesting even more underestimation of total spending in CEX and casting doubt on whether the CEX spending rates are a valid standard of comparison.

After the rather favorable assessment of the CAMS data based on population-level estimates we turned to a more detailed analysis of data quality down to the level of single spending categories. We found that there were some outliers, some individual observations in the CAMS data that were very large, which occurred at somewhat higher frequency in the shortest reference period. We also observed that high-frequency reference periods (“last week,” “last month”) yielded much higher means than reports for lower-frequency periods (“last 12 months”).

For some usually infrequently incurred spending categories such as “trips and vacations” or “vehicle repair and maintenance” it was clear that those responding “last week” did not actually spend the reported amounts every week on those spending categories. So we changed the design in CAMS wave 2 for categories like these and asked only about the amount spent in the last 12 months. We give a more detailed account in Section 4.

In other spending categories differences between annualized spending as a function of which reference period the households chose for reporting were also large, but the amounts were not unreasonable when judging the overall resulting distribution in those spending categories.

⁵ HRS income measures have been shown to be closely comparable to those measured in the Current Population Survey (CPS). For details see Table 3 in Hurd and Rohwedder (2006).

The patterns could be real if they resulted from higher spenders self-selecting into shorter reference periods. We conducted additional experiments to shed light into this issue (see Section 5).

Finally, two kinds of shifts occurred between Waves 1 and 2 in some expenditure categories: First, the frequency with which respondents chose reference periods changed (typically towards shorter periods) even though the same choices were presented. Second, where the frequency of chosen reference periods remained about the same, the means changed. In the following section, we explore such data quality issues in more detail.

4. Data Quality Issues

The association of higher-frequency reference periods with greater annual spending may be seen in Table 3 which summarizes expenditures on vehicle maintenance (parts, repairs, and servicing). The average annual spending among those using a reference period of "last week" (6.6% of the sample) was estimated to be about \$10,000. The median among the same group was about \$2,700. The very large difference between the mean and the median indicates the presence of high-end outliers. But for the 49% of the sample who reported on an annual basis, the mean and median were \$601 and \$400.

Can these differences be correct? We did not think so. However, to find out one would have liked to ask those who answered "last month" to also give us their report for spending in the last 12 months; and similarly for those who reported last week and last 12 months. We conducted additional experiments following this design, which we discuss in Section 5. In the meantime we eliminated the "last week" option for most categories and the "last month" option for some categories in the design of CAMS Wave 2. Table 4 compares the Wave 1 and 2 data for vehicle maintenance, which in Wave 2 was reportable only for the last 12 months. The reference period of "last 12 months" yielded about the same mean and median in both waves. Meanwhile, reports of no spending declined by a third. We have no explanation for that. The

overall median was about the same in both waves, while the mean was substantially lower in Wave 2.

Recall bias may explain some portion of the lower means and medians for “last 12 months” questions, compared to those with shorter reference periods. Recall bias is higher the longer the reference period. While eliminating shorter reference periods successfully reduced the occurrence of extreme outliers (from multiplying “last week” amounts by 52 or “last month” amounts by 12), this comes at the price of introducing recall bias. For comparison, the CEX average for vehicle maintenance and repair is \$577 for the population 55 and older, which is 13.5 percent higher than the CAMS measure in wave 2.

For the categories like “vehicle maintenance and repair” or “trips and vacations” it is obvious that households who reported their spending for the week did not actually spend the same amount every week of the year and the same is true for the amounts reported for last month. Multiplying by 52 or by 12 to arrive at annual amounts clearly yields inflated spending measures for these categories. However, for other spending categories it is less clear whether the higher spending observed in shorter reference periods is unreasonable.

For food and beverages, “food and drinks, including alcoholic, that you buy in grocery or other stores,” we did not change the choice of reference period from one wave to the next. Again, the mean annual spending estimate was much higher when the last week was selected as the reference period (see Table 5). The difference was smaller when the last month was selected.

What might explain these differences? There are several possible mechanisms. First, bigger spenders may be selecting themselves into shorter reference periods. That is, because high spenders tend to spend more frequently, the shorter reference periods are a more natural choice for them, while more modest spenders tend to select the longer periods. If that is true, the differences shown are real.

A second possibility is recall error leading to underestimation for longer reference periods. That is, the longer the period, the less the respondent remembers. If that is true, the

differences are not real but are due to biased estimates in the case of the 12-month reference period. The higher estimates from the shorter reference period are more accurate.

Third, some respondents may mistakenly enter the annual amount into the wrong reference period slot. That would lead to some extreme outliers resulting from multiplying what is really an annual amount by 52 (if entered as “last week”) or by 12 (if “last month”). The resulting extreme outliers would especially affect the estimate of the mean.

In Section 5 we present some experiments that we designed to shed light into what is driving the observed differences in spending patterns as a function of the chosen reference period.

Notice the large increase between waves in the percentage selecting the last week as a reference period. This may have been the result of how the categories were grouped. In Wave 1, the respondent was allowed three reference periods (a week, a month, and 12 months) for 15 categories. The query about spending on food in stores was embedded in queries about categories for which the one-week reference period was not natural. (It followed “home repairs and maintenance.”) Switching reference periods frequently when answering the CAMS questionnaire may have been cognitively challenging for respondents. When categories were not sorted by whether they were more or less frequent (on average), respondents might have been reluctant to switch to “last week” if they had just been thinking about another category in terms of “the last 12 months.” The design was changed for Wave 2: the last-week option was eliminated for all but three frequently purchased categories – food and beverages purchased for consumption at home, food purchased away from home, and gasoline. These frequently purchased categories were grouped in a separate block with additional instructions: The language, which follows, drew attention to the possibility of using a reference period of a week.

For the items on this page we have included three time periods so that you can estimate your spending in the way that is easiest for you for each category. For example, if it is easiest for you to think about what you spent on food and beverages last week, then please enter the amount in the first column.

As shown in Table 6, the selection of the last-week reference period increased sharply – by about 10 percentage points – for all three categories between Waves 1 and 2. The higher level remained in later waves, which retained the same question order and structure.

5. Experiments about the Variation in Annual Spending as a Function of Reference Period

As discussed above, annual spending has been reported as higher in CAMS when a high-frequency reference period has been chosen. How would we know if these apparently biased reports are actually true, e.g., because of self-selection of high spenders into the short reference periods? We would need to know the total spending in the last 12 months by those who chose the shorter reference period, relative to those who did not. We would also like to know what households reporting “last year” would have reported if they had been asked about “last month.”

To find out, we conducted randomized experiments in the American Life Panel (ALP), an Internet survey run by RAND. The ALP is a panel of approximately 1,500 respondents who, at the time of the data collection reported here, were age 40 or over. Respondents in the panel log on to the Internet using either their own computer or a web TV, which provides access to the Internet via a television and a telephone line and thus permits participation by those previously lacking Internet access. At least once a month respondents receive an email with a request to fill out questionnaires on the Internet. Typically an interview will take less than 30 minutes. Respondents are paid an incentive of about \$20 per thirty minutes of interviewing. The ALP is modeled after the CentER panel maintained at the University of Tilburg, The Netherlands, which has been in existence since 1990. Participants in the ALP are recruited from respondents to the Monthly Survey conducted by the Survey Research Center at the University of Michigan.

Our experiments were included in the third wave of the ALP, which was fielded in the summer of 2005. A total of 1,067 respondents completed the interview. Should the experiments show evidence of bias resulting from allowing respondents to choose the reference period we would have to contemplate an alternative design for CAMS. Therefore we added to the

objectives of the experiments to compare the effect on population estimates of spending when respondents could choose a reference period as opposed to being limited or directed to one specified by CAMS.

We conducted the experiments for a total of ten spending categories that were also elicited in CAMS, using the same wording as in CAMS for the category cues. The sample was randomized into two groups with group A (527 respondents) being asked about spending in a single specified reference period and group B (540 respondents) being allowed to choose among several reference periods, just as in CAMS. Everybody, irrespective of whether they were in group A or in group B, received follow-up questions about their spending in an alternative reference period. The specified reference period for group A was chosen to be the most natural one for the spending category queried and the reference period one might consider using if one were to shift to a design of CAMS that would not allow respondents to choose among different time frames. The ten spending categories queried included five items that tend to be irregularly purchased, three more regular ones and two most regular ones. The experimental design varied for irregular, more regular and most regular categories. Table 7 gives an overview of the experimental design.

Irregular Purchases

The five spending categories of irregularly purchased items included clothing, trips and vacations, home repairs, health care, and gifts. Respondents in group A would first be asked about their spending on particular categories during the “last 12 months,” followed by a question about their spending in the “last month.” Group B was asked about the same spending categories but was offered a choice of reference periods – “last 12 months,” “last month,” or “no money spent during last 12 months.” In the first two cases, respondents were secondly asked to report using the alternative. That is, a respondent choosing “last 12 months” was given a follow-up question as to what was spent “last month,” and vice-versa.

Between Groups A and B, we thus collected data on six measures of annual purchases: specified annual, followed by specified last month (which we converted to an annual amount);

respondent-chosen annual, followed by last month (converted to annual); and respondent-chosen last month (converted to annual), followed by annual. In the absence of selection and recall bias, all the medians and means should be the same. In the presence of a selection effect, the short-period median and mean should be higher for those choosing the shorter period than for those not doing so. In the presence of recall bias, the short-period median and mean should be higher than the long-period ones. (Those who picked “No money spent in last 12 months” were left out of the analysis because they could not be assigned to a reference period.)

As is apparent from the results in Table 8, for clothing, the “last month” medians and means were all greater than the “last year” medians and means, suggesting that people do not recall very well their purchases over a year’s time (i.e., recall bias is important). Also, those choosing “month” had higher medians and means both for the amount last month and for the amount last year. This suggests that they did, indeed, spend more.

As mentioned above, there were five categories of irregular spending. Table 9 gives information about each.

The fraction choosing “last month” first varied considerably: For example, in the case of trips and vacations, 21% chose “last month” first, whereas 48% did so for clothing. This seems to accord with frequency of purchasing. Many go on just one trip per year, whereas many buy clothing at least several times a year.

We have calculated the weighted average of the five medians and five means, using the number of observations as weights. This is just a way of summarizing overall patterns. It does not represent the spending of any population, because different people are in the means and medians of each spending category. For example, one person may choose “last month” first for clothing but “annual” first for trips.

For the medians, when the reference period is specified as annual, there is little difference between expenditure reports for the last month and those for the last 12 months. Apparently, for the relatively small values near the median, there is little recall bias in the population. There is a

large difference between last-month expenditures when the last month is chosen as the reference period, versus when the last 12 months is either chosen or specified. These differences suggest that those who select that reference period had unusually large purchases in the last month—and/or that they were possibly subject to recall bias, as reflected in their reports of smaller amounts when they were asked for a 12-month estimate. Among those who chose the 12-month reporting period, the opposite seems to have been the case: they reported unusually small purchases in the last month. This is especially true in the case of trips and vacations.

For the means, the pattern is different. All groups reported higher amounts when responding about the last month than they did for the last 12 months. This suggests substantial recall bias among those who are more frequent purchasers and so contribute substantially to the mean. Furthermore, spending on each of the five goods follows this pattern. The difference is greatest among those who chose the last month and who may thus include higher spenders.

More Regular Purchases

We conducted similar experiments with more regularly purchases, such as prescription drugs, dining out, and telephone, cable, and Internet services. Drugs are of particular interest because, as mentioned above, they may be regularly purchased by some (with chronic conditions) but irregularly purchased by others.

As with experiments for irregularly purchased spending categories, Group A received questions with a specified reference period, that is, respondents in this group had no choice. They were asked to report their spending in a “typical month.” They received a follow-up question asking about “last month.”

Respondents in Group B were again offered multiple ways of reporting their spending. They could choose among reporting it “during the last 12 months” or “during a typical month,” or they could check a box saying “no money spent during the last 12 months.” If they chose the “last 12 months” or a “typical month,” they were given a follow-up question asking them about their spending “last month.”

Table 10 reports the results for prescription drugs. Note that the last two columns under “Median” and under “Mean” are the first responses, specified or chosen, depending on the group. “Last month” was the follow-up question for all groups.

Those choosing the shorter reference period, i.e. typical-month, reported higher amounts when asked about the last month than did those who chose an annual reference period – \$720 on an annual basis at the median versus \$480. They also reported higher amounts than did those for whom a typical-month reporting period was specified – \$720 at the median versus \$540. The same qualitative patterns hold for the means. While we do not have the last 12 months-measure for all groups, this evidence is at least suggestive of those choosing the shorter reference period having higher spending on prescription drugs.

Comparing the summary statistics for typical month and last month we find no systematic relationship between expenditures reported in a typical month versus those in the last month. This is not surprising considering the discussion we provided in Section 2.5, arguing that asking about typical month or typical week creates substantial ambiguity for respondents as to how to report irregular expenses.

Most-Regular Purchases

A third set of experiments was conducted for high-frequency purchases (e.g., gasoline, food). As was the case for the preceding set of experiments, Group A was first asked about spending in a “typical month” and then about spending “last month.” Group B was again offered a choice, but this time the options included spending in a “typical week,” as well as spending in a typical month and over the last 12 months. They were also allowed to check a box for “no money spent during last 12 months.” Follow-up questions depended on the respondent’s choice of reference period: Those who chose a “typical month” or the “last 12 months” were asked about their spending “last month”, then “last week.” Those who chose to answer for a “typical week” were then asked about spending “last week.”

Table 11 reports the results for food. There was no difference in the median, and little in

the mean, between amounts reported for the last week and for a typical week (when a typical week was the respondent's preferred reference period). The same was true for last month and the typical month, when the respondent's preference was for the typical month (though not when a typical month was specified as the reference period).

As in the previous experiments, respondents choosing shorter reference periods reported higher expenditures. Compare the medians for those who chose a typical week, a typical month, and the last 12 months: \$5214, \$4171 and \$3780, respectively.

In summary we draw the following conclusions from the experiments that we conducted in the ALP:

- Respondents' choice of reference period is related to their household's frequency and level of spending in a particular category: more respondents choose a longer reference period for less frequently purchased items while very few do so for high frequency items; and those who choose a shorter reference period also report higher spending in follow-up questions using longer reference periods.
- Recall bias is important when using longer reference periods such as "last 12 months." Therefore, even if the measure of interest is spending during the last 12 months, specifying this reference period for all spending categories – frequent or infrequent – would lead to substantial underestimates of total spending, limiting the use of the resulting consumption measure. Instead longer reference periods should be used sparingly with relatively frequently purchased items.
- Short reference periods risk obtaining an unrepresentative snapshot of household spending. Encouraging respondents to make adjustments to any unusual recent spending patterns is not a viable way of taking advantage of the smaller recall bias of shorter reference periods.

6. Conclusions

Measuring spending by households is difficult. They are heterogeneous in the goods they purchase and especially in the frequency with which they purchase similar goods; some households may purchase a given good frequently, others irregularly, still others not at all.

None of this would be a problem if it was not for recall bias. Without recall bias one could ask all respondents about their household's spending during the last 12 months. However, we showed that recall bias is important for such a long reference period. Recall bias combined with the large heterogeneity in household spending patterns makes it difficult to know which goods should be put into which time frames. The reference period for a particular category should be short enough to keep recall bias small and long enough to minimize the risk of obtaining an unrepresentative snapshot of a household's spending.

If one knew which households tend to spend money on an item frequently and which do not one could administer a longer reference period to those not spending frequently and regularly and a shorter reference period to those with frequent and regular spending. In principle, this could be accomplished with additional screening questions for each spending category, but it would add substantially to the length of the survey, which is not a realistic option in a general-purpose survey and even less so in a paper-and-pencil format.

The CAMS survey adopted an innovative alternative approach: let respondents choose from a set of reference periods of different lengths. In a set of additional experiments we showed that respondents who choose a short reference period have indeed higher spending than those who choose a longer reference period, that is, they tend to self-select into the more appropriate time frame. However, evidence from CAMS wave 1 revealed that one can also offer too much choice in reference periods. The first wave of CAMS included the option "last week" for a number of categories, also for very irregular ones such as home repairs or trips and vacations. It was immediately clear for irregular categories that the amounts reported in the field "last week" were not amounts the household would spend every week. In response to this observation we eliminated the "last week" option for all but three high-frequency categories in wave 2. For the same reasons we also eliminated the "last month" option for some particularly infrequent spending categories like trips and vacations and home repairs.

The results from the ALP experiments highlighted the importance of recall bias, and that respondents tend to self-select into the most appropriate reference period. Based on these results, we changed the instructions in the third wave of CAMS to emphasize explicitly how we would like respondents to choose the reference period for a particular category:

The next block has items that some people do not purchase on a regular basis. Please use the time period that best reflects your spending over the last 12 months to estimate what you actually spent.

For example:

- If your household's spending on clothing in the last year was irregular or concentrated in just a few months then please report your best estimate of the total amount your household spent on clothing in the last 12 months.
- If your household's spending on clothing was fairly evenly distributed over the year, then you can choose whether to report the average monthly amount or the total amount spent in the last 12 months, whichever you find easier.

Again, if you did not spend money on a specific item or service in the last 12 months, then check the "No money spent on this in last 12 months" box.

If you bought an item only occasionally or on an as-needed basis, then please give your best estimate of what you spent in the last 12 months.

Wave 2 and later waves of CAMS have come to a compromise that seems to produce good quality data that are consistent with independently collected measures of income and wealth. We conclude that useful spending data can be collected in a fairly short self-administered survey at a low financial cost and with modest respondent burden.

References

Browning, Martin, Thomas F. Crossley and Guglielmo Weber. "Asking Consumption Questions in General Purpose Surveys." *Economic Journal*, 113:F540-F567 (November, 2003.)

Hurd, Michael D. and Susann Rohwedder. 2008. "Wealth Change and Active Saving at Older Ages," RAND manuscript.

Hurd, Michael D. and Susann Rohwedder. 2006. "Economic Well-Being at Older Ages: Income- and Consumption-based Poverty Measures in the HRS," RAND Working Paper No. WR-410.

Table 1. A classification of spending categories by regularity and frequency of expense

Regularity	Frequency	Examples
Somewhat regular	High (weekly)	Food, gasoline
Regular	Medium (monthly)	Rent, utilities, prescription drugs (for some people)
Regular	Low (annual or semiannual)	Property taxes, auto license
Irregular	Varies across households (for some quarterly, for others less than once a year)	Clothing, health care (e.g., co-payments)
Big ticket items	Usually less than once a year	Automobiles, home improvements

Table 2. Choice of reference period offered in CAMS Wave 1 questionnaire

	Amount spent last week		Amount spent last month		Amount spent in last 12 months		No money spent on this in last 12 months
B18. Housekeeping and yard supplies: <i>cleaning supplies, lawn and garden products</i>	\$ _____	OR	\$ _____	OR	\$ _____	OR	<input type="checkbox"/>
B19. Home repairs and maintenance: <i>materials plus any costs for hiring a professional</i>	\$ _____	OR	\$ _____	OR	\$ _____	OR	<input type="checkbox"/>

Table 3. Mean and median expenditures on vehicle maintenance, by self-selected reference period, in CAMS Wave 1.

Reference period	Percent distribution	Mean	Median
last week	6.6	10,196	2,711
last month	14.6	2,358	900
last 12 months	48.9	601	400
no spending	29.9	0	0
All	100.0	1,308	295

Table 4. Annual spending on vehicle maintenance, CAMS waves 1 and 2 Comparison of self-selected and CAMS-specified reference periods

Reference period	Percent distribution		Mean		Median	
	Wave 1	Wave 2	Wave 1	Wave 2	Wave 1	Wave 2
last week	6.6	--	10,196	--	2,711	--
last month	14.6	--	2,358	--	900	--
last 12 months	48.9	80.2	601	633	400	400
no spending	29.9	19.8	0	0	0	0
All	100.0	100.0	1,308	508	295	300

Table 5. Annual spending on food and beverages, CAMS waves 1 and 2. Comparison of self-selected reference period

Reference period	Percent distribution		Mean		Median	
	Wave 1	Wave 2	Wave 1	Wave 2	Wave 1	Wave 2
last week	39.1	51.3	4,732	5,082	3,911	3,911
last month	30.4	23.9	3,323	3,843	2,400	2,700
last 12 months	28.6	21.9	2,833	2,681	2,400	2,125
no spending	1.9	3.0	0	0	0	0
All	100.0	100.0	3,669	4,110	2,868	3,128

Table 6. Percent distribution of self-selected reference period for three high-frequency categories, CAMS

		Wave 1	Wave 2	Wave 3	Wave 4
Food	last week	39.1	51.3	50.6	51.5
	last month	30.4	23.9	27.9	27.7
	last 12 months	28.6	21.9	18.0	17.0
	no money spent	1.9	3.0	3.6	3.8
	All	100.0	100.0	100.0	100.0
Dining out	last week	29.2	38.4	34.8	35.2
	last month	27.3	21.3	26.5	26.5
	last 12 months	29.4	22.2	21.9	21.0
	no money spent	14.2	18.0	16.9	17.3
	All	100.0	100.0	100.0	100.0
Gas	last week	30.5	41.7	40.4	40.5
	last month	26.8	24.2	27.3	29.6
	last 12 months	30.1	22.0	19.5	16.9
	no money spent	12.6	12.1	12.8	13.0
	All	100.0	100.0	100.0	100.0

Sample size varied between 3062 and 3698

Table 7: Overview of Experimental Design

Spending Categories	Group A specified time frame	Group B allow choice of reference period			
Irregular purchases	amount spent <i>last 12 months</i> ; follow-up: "last month"	amt spent last month follow-up: ↓ amt spent last 12 months	or	amt spent last 12 months ↓ amt spent last month	or <input type="checkbox"/> no money spent last 12 months
Mostly regular	amount you <i>usually spend in a month</i> ; follow-up: "last month" (if DK on monthly: follow-up "last 12 months")	monthly amt follow-up: ↓ amt spent last month	or	amt spent last 12 months ↓ amt spent last month	or <input type="checkbox"/> no money spent last 12 months
High frequency	amount you <i>usually spend in a month</i> ; follow-up: "last month" (if DK on monthly: follow-up "last 12 months")	weekly amt follow-up What about last week? () about same <end> () higher () lower How much last wk? \$ _____	or	monthly amt ↓ What about last month? () about the same <end> () higher () lower How much last month? \$ _____ amt spent last week \$ _____	or <input type="checkbox"/> no money spent last 12 mths amt spent last month ↓ amt spent last week \$ _____

Table 8: Estimated mean and median annual spending on clothing by reference period, conditional on positive spending in last 12 months

Annual totals calculated from reported amount last month or last 12 months						
Reference period selected	N	Median		N	Mean	
		last month	last 12 mos		last month	last 12 mos
Specified annual first	519	1,350	1,000	508	2,461	1,863
chosen annual	272	1,200	1,000	267	2,331	1,633
chosen month	256	2,400	1,200	253	3,087	1,825

Note: 19 outliers are excluded from the means but not from the medians.

Table 9: Estimated median and mean annual spending, conditional on positive spending in last 12 months by reference period for five types of irregularly purchased goods

Annual totals calculated from reported amount last month or last 12 months					
	N	Medians		Means	
		last month	last 12 months	last month	last 12 months
Clothing					
specified annual first	519	1,350	1,000	2,461	1,863
chosen annual	272	1,200	1,000	2,331	1,633
chosen month	256	2,400	1,200	3,087	1,825
Trips					
specified annual first	462	1,200	2,000	4,240	3,131
chosen annual	340	900	2,000	3,686	3,127
chosen month	93	4,800	2,000	8,631	2,779
Home repairs					
specified annual first	454	1,200	1,200	4,793	3,493
chosen annual	287	900	1,500	4,756	3,714
chosen month	131	2,400	1,100	7,871	2,424
Health care					
specified annual first	503	1,200	1,000	2,668	1,885
chosen annual	307	1,200	1,100	2,871	2,193
chosen month	184	1,800	1,000	3,811	1,823
Gifts					
specified annual first	503	1,200	1,000	2,668	1,885
chosen annual	307	1,200	1,100	2,871	2,193
chosen month	184	1,800	1,000	3,811	1,823
Weighted average					
specified annual first		1,232	1,226	3,317	2,415
chosen annual		1,076	1,360	3,314	2,591
chosen month		2,403	1,185	4,748	2,021
specified annual first		1,232	1,226	3,317	2,415
chosen		1,552	1,297	3,829	2,386

Table 10: Estimated median and mean annual spending on prescription drugs, conditional on positive spending in last 12 months by reference period

Annual totals calculated from reported amount last month, typical month or last 12 months								
reference period selected	N	Median			N	Mean		
		last month	typical month	last 12 mos		last month	typical month	last 12 mos
specified typical month	508	540	780		497	893	2,167	
chosen annual	242	480		350	242	1,073		700
chosen typical month	277	720	600		273	1,613	1,214	

Note: 15 outliers excluded from means but not from medians.

Table 11: Estimated median and mean annual spending, conditional on positive spending on food to be consumed at home in last 12 months

Annual totals calculated from reported amount						
last week, typical week, last month typical month or last 12 months						
Reference period selected	N	last week	typical week	last month	typical month	last 12 months
Median						
specified typical month first	524			4,200	4,800	
chosen annual	50	3,780		3,600		3,300
chosen typical month	195	4,171		3,600	3,600	
chosen typical week	289	5,214	5,214			
Mean						
specified typical month	513			4,758	8,796*	
chosen annual	49	4,450		4,203		4,170
chosen typical month	193	5,586		4,204	4,435	
chosen typical week	287	6,411	6,153			

Note: 15 outliers excluded from means but not from medians

* This average is affected by several high values that remained even after applying our standardized method for identifying outliers (top 1% values of the entire distribution).