RESEARCH REPORT SERIES
(*Survey Methodology #2001-03)*

Using Alternatiave Question Strategies to Reduce Income Nonresponse

Jeffrey C. Moore and Laura S. Loomis

Statistical Research Division
U.S. Bureau of the Census
Washington D.C. 20233

Report Issued: November 29, 2001

Using Alternative Question Strategies to Reduce Income Nonresponse

by

Jeffrey C. Moore and Laura S. Loomis[1]
U.S. Census Bureau

key words: income nonresponse, QDERS, questionnaire design, response brackets

Abstract:
Item nonresponse is a continuing issue for surveys which attempt to collect income data. As with any survey topic, elevated levels of nonresponse necessitate additional (and often costly and time-consuming) repair measures, both in the field and during data processing, and increase the uncertainty surrounding income estimates derived from survey data. The importance of income data for social policy analysis adds an extra impetus for developing effective techniques to bring income nonresponse under better control. Recently, researchers have begun to report some success in reducing income nonresponse through the use of "unfolding brackets" and a variety of other types of closed-ended income range reporting options.

This paper reports the results of an experimental test of a new form of income range reporting, the intent of which was to address both cognitive and motivational barriers to income reporting, without adding to the tedium of the interview interaction. The experiment was conducted as part of the Census Bureau's 1999 Questionnaire Design Experimental Research Survey (QDERS), an RDD/telephone survey which served as the vehicle for several methodological experiments. The test of relevance here compared a standard annual income amount reporting task for several common types of assets to a new form of income range reporting which we term "implicit brackets." In essence, the implicit bracket approach first asks the respondent whether the amount exceeds some minimum threshold ($10 to $100, depending on the asset type), and then, if yes, asks for a report of the income amount "to the nearest X dollars," with X varying from (in this case) $5 to $50, depending on the type of asset.

Results of the experiment suggest that the implicit brackets approach had small but consistently positive effects on nonresponse, primarily through a reduction in "don't know" nonresponse. We find insignificant and inconsistent effects on average income amounts reported, and similarly inconsistent effects on response distributions. We find no reduction in the precision of income reports generated by the implicit brackets approach; in fact, those reports were actually less likely to be rounded than were control treatment reports. Interviewers perceived significant benefits to the implicit brackets approach, in terms of respondent ease and the accuracy of reports. We acknowledge the limitations of the QDERS test, but nevertheless view these results as a promising first step toward the development of improved survey procedures for capturing income data.

---

[1]Statistical Research Division/Center for Survey Methods Research, Room 3133-4, Washington DC 20233-9150. This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau publications. This report is released to inform interested parties of research and to encourage discussion. We gratefully acknowledge the contributions of the following colleagues to this research effort: Jennifer Rothgeb, whose vision and energy were essential to the development of QDERS and its initial launch; Tommy Wright, for providing initial QDERS funding; Chet Bowie for providing extra funding to support essential enhancements; John Bushery, Keith Albright, and Jennifer Reichert for development and implementation of the response variance reinterview and assistance with its analysis; the Census Bureau's Hagerstown Telephone Center interviewers and staff; the volunteer behavior coders from our colleagues in the Census Bureau's Center for Survey Methods Research; and the helpful comments of Julia Klein Griffiths and Ed Welniak, who reviewed an earlier draft of this paper.

1. Introduction and Background

1.1 Income Nonresponse

Item nonresponse is a continuing issue for surveys that attempt to collect income data. Despite their generally high overall response rates, U.S. government surveys are by no means immune from this problem. Moore, Stinson, and Welniak (1999), for example, citing data from the Current Population Survey (CPS), report income nonresponse rates ranging from a minimum of 20% up to almost 50% for some types of income, with questions about income from assets showing the highest rates of nonresponse. As with any survey topic, elevated levels of nonresponse necessitate additional (and often costly and time-consuming) repair measures both in the field and in data processing, and increase the uncertainty surrounding income estimates derived from survey data. The importance of income data for social policy analysis adds an extra impetus for developing effective techniques to bring income nonresponse under better control.

Various reasons have been advanced to explain the high levels of income item nonresponse and other reporting problems: the complexity of the reporting task, confusion and uncertainty about income source labels and other terms, sensitivity, etc. (Moore, Stinson, and Welniak, 1999). However, when it comes to asset income (e.g., interest and dividends), one of the most important reasons may simply be a lack of knowledge, or genuine uncertainty about the accurate amount, due to lack of salience and recall difficulties (Cantor, Brandt, and Green, 1991).

1.2 Questionnaire Design Solutions

The problems of respondent knowledge and income question sensitivity have proven difficult to combat using questionnaire design strategies. Recently, however, researchers have begun to report some success in reducing income nonresponse using a technique called "unfolding brackets." In general, this design strategy involves asking respondents a series of closed-ended income range questions (e.g., "Is it $10,000 or more?"), after the respondent has refused or said "don't know" to an open-ended income question asking for the exact income amount. Using this technique, which slowly "unfolds" a reasonably tight range containing the amount in question, several studies have shown that relatively large proportions of respondents who initially refuse or don't know the exact answer to an income question *will* answer follow-up income range questions (Juster and Smith, 1997; Kennickell, 1997; Ross and Reynolds, 1996; Hippler and Hippler, 1986; Bell, 1984). In fact, Juster and Smith (1997) and Heeringa, Hill, and Howell (1995) find that the amount of completely missing data on asset income questions is usually reduced to less than 10 percent through the use of such bracket questions. Another specific benefit of the bracket strategy appears to be the increased willingness of relatively wealthy households to answer these bracket questions, thus improving measurement of income at the upper end of the income distribution. When researchers have used information from bracketed questions in imputation procedures, much higher estimates of wealth and asset amounts have resulted, which is generally assumed to mean much <u>better</u> estimates (Juster and Smith, 1998, 1997; Kennickell, 1997).

Thus, the unfolding brackets technique appears to have clear benefits with regard to reducing nonresponse, and with regard to improving estimates for relatively wealthy households. Those benefits, however, are not cost-free. One disadvantage of unfolding brackets has simply to do with the tedium of reading long lists of categories, or the painfully slow "unfolding" of the correct income category. In-person interviews can make use of show-cards to display income brackets, but telephone interviews are constrained to reading the series of categories or unfolding bracket questions. This is particularly problematic for surveys like the Census Bureau's Survey of Income and Program Participation (SIPP), and the March Income Supplement to the CPS, which focus great attention on income details, and thus ask a great number of amount questions about a great number of income sources. Also, the unfolding bracket technique is implemented after the respondent has already indicated that he or she cannot or will not answer the question, running the risk of badgering the respondent. Again, this problem is exacerbated in income-focused surveys – for example, interviewers in Kennickell's (1997) study, which had a very large number of income questions, complained that the follow-up bracketed questions were "too pushy."

Another potential cost of this technique is the risk that early exposure to the bracket questions might affect respondents' willingness to provide accurate exact answers to later income questions. The evidence on this issue is mixed. Interestingly, Juster and Smith (1997) found that respondents did *not* tend to exploit the bracket options, but rather, after being exposed to them early on in the interview, were actually *more* likely to respond to the initial exact amount questions later in the interview. Juster and Smith speculate that the bracket questions may have cued respondents to the idea that approximate answers were acceptable, and that approximate answers were then given later to the exact income questions. In contrast, other researchers have found that the use of ranges, once offered as an option, persists throughout the interview (Kennickell, 1997), and that *any* exposure (early or late in the interview) is associated with lower quality income reports, due to self-selection of less accurate reporters into the bracket questions (Heeringa, Hill, and Howell, 1995)

A related concern focuses on interviewers' potential over-reliance on the income brackets, and the risk that they will be invoked too readily in place of traditional probing techniques in response to reluctance to provide an exact answer. Kennickell (1997) observed that exact answers to some dollar amount income questions declined sharply after introducing bracketed question techniques in the Survey of Consumer Finances, and suggested that this might have been due to interviewers' reliance on the follow-up questions rather than probing for exact answers. It was also observed that, when confronted with nonresponse to an income amount, interviewers in this study preferred to offer a show card to respondents, rather than read the series of unfolding bracket questions.

A final concern about the unfolding brackets technique is potential for anchoring effects – the dollar amount used as the initial entry point into the sequence of bracketing questions can influence the distribution of responses (Hurd and Rodgers, 1998). When higher dollar amounts are used as entry points, there is evidence that the distribution of responses tends to be more skewed to the right than if lower dollar amounts are used as entry points.

1.3 Goals of the Current Research

This paper presents the results of research which attempts to build on the unfolding brackets procedure, and to continue to develop and refine questionnaire design solutions to the problem of incomplete reporting of income amounts. We report on a test of a new, but related, form of income range reporting, which we label "implicit brackets." One goal of the new procedure was to reduce sensitivity concerns, by not forcing exact amount reports. Another was to reduce cognitive burden, both by abandoning almost all efforts to obtain information about very small amounts, and by allowing reasonable approximations ("How much was it to the nearest $50?"), while at the same time allowing precise, point-estimate reports, if respondents were willing and able to provide them. This procedure also avoided the need to present a long series of income categories for each amount question, or multiple instances of a rather arduous, multi-step question sequence to zero-in on an appropriate amount bracket. (See Attachment 2 for specific questionnaire procedures, which are described in more detail in Section 2.3, below.)

2. Methods and Procedures

In this section we describe briefly the design of the research. We begin with a description of the Questionnaire Design Experimental Research Survey (QDERS), the Census Bureau's new methodological research survey in which the “implicit brackets” experiment was embedded. We follow that with a description of the experimental design, and details of the two approaches to asking income amount questions that comprised our experimental and control treatments.

2.1 The Questionnaire Design Experimental Research Survey (QDERS)

The research presented here was embedded in the initial launch of the Census Bureau's Questionnaire Design Experimental Research Survey (QDERS), a special vehicle developed by Bureau staff for conducting questionnaire design research outside of the cognitive laboratory but "off-line" from the agency's ongoing production surveys. The goal of QDERS is to allow Census researchers an opportunity to conduct questionnaire design field experiments in a flexible environment, without risking impacts on important statistics or placing additional burdens on the Census Bureau's already-overburdened production survey staffs. The first QDERS, fielded in April 1999, included several experiments on alternative questionnaire design strategies for collecting information about functional limitations (disabilities), health insurance coverage, non-wage income sources, asset ownership, asset income amounts, and within-household relationships. (See U.S. Census Bureau (1999) for a description of QDERS in general and the 1999 QDERS implementation specifically.) This paper focuses on the asset income amount component of the 1999 QDERS experiment.

2.2 Sampling and Experimental Design

QDERS was a split-sample controlled experiment, using paper-and-pencil questionnaires in a telephone interview. It used a nationally representative (excluding Alaska and Hawaii) RDD sample, with independent sub-samples for each of the treatments. (See GENESYS (1997) for a more complete description of the QDERS RDD sample.) Once an interviewer reached an eligible residential phone number, he or she conducted an interview with one household respondent, who was asked to report for himself/herself and up to five other adult household members. The initial sample consisted of 5,870 telephone numbers, which had been pre-screened to identify working residential banks of numbers. This sample was expected to yield approximately 1,800 completed interviews – 900 in each treatment group for the income amount reporting experiment – a goal dictated primarily by budget constraints. (In fact, as indicated below, the sample proved considerably less productive than anticipated.)

2.3. Questionnaires

In this section we describe the two[2] questionnaires used in the income amounts reporting experiment – a standard, control treatment, and the experimental "implicit brackets" approach. As previously noted, these were paper-and-pencil questionnaires, administered in a telephone interview.

The basic questionnaire content for the income questions in each treatment was identical; only the manner in which the income amount questions were asked differed. Both treatments consisted of an initial set of asset ownership questions, which were asked of all adult household members. For no other reason than that they are among the more commonly-owned types of asset holdings, we selected the following five asset types for inclusion in the QDERS interview: interest-earning checking accounts, savings accounts, certificates of deposit (CDs), mutual funds, and stocks. Respondents were to report ownership of these assets regardless of whether they were held individually or jointly.

QDERS procedures called for income amounts to be collected for any assets reported for the household "reference person," the primary adult owner/renter of the sample unit who was to be listed first on the household roster. In fact, due no doubt to the complexity of the QDERS questionnaires necessitated by other QDERS experiments, interviewers mistakenly collected this information for many people other than the "reference person" as well. In the analyses to follow, we include all income

[2]In fact four different QDERS questionnaires were used in the income amount reporting experiment. Another QDERS experiment was implemented to evaluate a person-by-person approach to gathering demographic, health, and income source information, versus an approach which employed household-level screening questions ("Does anyone in the household [have characteristic X]?") and then followed up with individual questions as needed (see Hess, et al., 2000). The income amount reporting experiment was crossed with the person/household experiment, and so, for purposes of this evaluation we assume that the additional questionnaire sub-designs are irrelevant, and can be ignored in this paper.

reporters, regardless of whether they were the intended income reporter or others in the household to whom the amount questions were administered in error[3].

The distinctions between the questionnaire treatments are described below.

2.3.1 Control treatment

The control treatment amounts questions were quite straightforward. For each asset type that had been reported earlier, in the "asset ownership" section of the interview, a question of the following format was asked: "How much interest did [NAME] earn [from asset type] in 1998?" Specific questions are shown in Attachment 1.

2.3.2 Experimental, "implicit brackets" treatment

The income amounts question format for the experimental treatment consisted of two basic parts. The first part asked whether the annual income amount for 1998 was "more or less than $X?" where X was a minimum amount which varied according to asset type ($10 for checking and savings account interest, $50 for interest on CDs, and $100 for mutual fund or stock dividends). The goal of this part of the question was to eliminate almost all burden associated with the reporting of trivial income amounts, since the question itself was assumed to present a very easy judgment task, and there were no followup questions to a "less" response[4].

The "implicit bracket" procedure was invoked if the response to the initial "more or less than" question was "more." The procedure was simply to ask a question of the following form: "How much was it to the nearest $Z?" The intent was to create implicit brackets for respondents' use, of width Z, without having to present those brackets overtly. As with the minimum amount (see above), the specific value of Z varied with different asset types – $5 for checking accounts, $10 for savings accounts and CDs,

---

[3]The additional, unplanned income reports derived almost exclusively from the QDERS treatment which used household screener techniques to identify characteristics of interest, including asset ownership. As noted, that experiment was fully crossed with the income reporting experiment of interest here, and so we choose to exploit the procedural errors for the increased statistical power they serendipitously provide. In fact, throughout our analyses we examine reference-person-only reports separately from all-income-reporters reports and find that, beyond occasional changes in the significance levels of some comparisons, the inclusion or exclusion of these cases has no substantive impact on the results. (Note that a more careful implementation of procedures in the reinterview eliminated any unplanned "extra" amount reports; thus, the reliability analyses described in Section 3.5 only use data from the one intended income reporter, the household reference person.)

[4]In fact, there were two subcomponents to this initial "more/less" task. The first was a lead-in question which was identical to the control treatment amount question (e.g., "How much interest did you earn from all CDs in 1998?"). Interviewers were instructed to pause briefly following the reading of this lead-in question, in case respondents were willing and able to supply an amount without any hesitation. The second part was the "more or less than" question itself (e.g., "Would you say it was more or less than $50?"). See Attachment 2.

$50 for mutual funds, and $25 for stocks. The specific questions used in the experimental approach are shown in Attachment 2.

2.4 Data Collection

2.4.1 Interviewers and interviewer training

A staff of 22 experienced telephone interviewers received approximately five hours of initial QDERS training, and a slightly reduced refresher training session midway through the field period[5]. Initial training for all interviewers included training on both income amount reporting treatments. From the outset of the field period, and throughout interviewing, interviewers' assignments included a mix of both control and experimental income reporting formats, so there is no confounding of experimental treatments with interviewer characteristics.

2.4.2 Response rates

As is often the case with RDD telephone surveys, we can identify the upper and lower bounds of the QDERS response rate, but the presence of a substantial number of "ring, no-answer" cases, with unknown eligibility, prevents us from being able to provide a precise point estimate. Using American Association for Public Opinion Research (AAPOR, 1998) guidelines, the "near minimum" response rate overall for QDERS (including partial interviews as completes, and including all cases of unknown eligibility in the denominator) was 36%, and the "maximum" response rate (also including partial interviews as completes, but excluding from the denominator cases of unknown eligibility) was 46%. Excluding eligible non-contact cases from the denominator yields a cooperation rate among contacted households of 52%. QDERS procedures did not include any special refusal conversion attempts, and as a result refusals accounted for approximately half of the observed non-response, or about 30% of all cases. For all of these statistics – the upper and lower bounds on the response rate, the completion rate, and the refusal rate – the differences between the two income amount reporting treatments were trivial and nonsignificant. The final number of completed interviews (households) was 1,304, of which 13 were subsequently excluded due to missing data, for a final total of 1,291 completed interviews.

Regardless of the necessary imprecision of the response rate estimate, it is nevertheless quite clear that the true QDERS response rate fell substantially short of the typical rate for Census Bureau and other government surveys. Since our goal was to look for differences associated with experimental treatments, we are perhaps somewhat more justified in ignoring the biasing effects of nonresponse than we would be had we intended to use these data to make precise estimates of population parameters.

---

[5]The refresher training session coincided with a shift in interviewers' assignments – those who had been working exclusively on QDERS' person-level questionnaire format switched to the household screener treatment, and vice-versa. The primary purpose of the refresher training was to acquaint interviewers with the new instrument format.

The fact that the response rate estimates for our two treatments were essentially identical offers some additional comfort in this regard. On the other hand, while we have no reason to believe that the propensity to respond to the QDERS survey would interact with the propensity to be affected by our questionnaire design treatments, the low rate of response represents a real limitation on confidence in the reliability of our findings.

3. Results

In this section we summarize the results of the experiment. First, we examine the impact of the experimental treatments on item nonresponse. Next we compare the average dollar amount estimates obtained by each treatment, and the distributions of amount reports. We also use the frequency of amount rounding to compare the precision of reporting behavior under each treatment. Next, a response variance reinterview in QDERS permits an assessment of the reliability of amount reports under each type of reporting procedures. And finally, we examine the assessments of interviewers, through their responses to an evaluation questionnaire.

3.1 Item Nonresponse

Table 1 summarizes the results of the amount reporting experiment with regard to item nonresponse. The conclusion is quite clear: for all five of the asset income sources included in the QDERS interview, the observed nonresponse rate for the experimental treatment is lower than the observed rate for the control treatment. The difference is statistically significant for only one of the five individual comparisons[6]; however, a simple sign test (Snedecor and Cochran, 1967) suggests that the complete, five-out-of-five consistency of the direction of differences across the individual comparisons is itself statistically significant (p=.0625), indicating that rejection of the null hypothesis is appropriate.

Table 2 makes clear, first of all, that the primary nonresponse problem for asset income amounts is "don't know's" (DK), which generally outnumber refusals by a four- or five-to-one ratio or more. More importantly for present purposes, Table 2 also makes clear the fact that the improvement in nonresponse in the experimental treatment is due to a reduction in DK nonresponse, and not to any improvement with regard to refusals. In all five comparisons, the control treatment suffered a higher rate of DK nonresponse than the experimental treatment (significantly so in one instance). Although the differences are often quite small, we also note a mirror image effect for refusal rate differences, which consistently indicate higher refusal nonresponse for the experimental treatment. To the extent, then, that the different forms of nonresponse are meaningful – with DK nonresponse the manifestation of some cognitive difficulty, and refusals likewise of motivational issues surrounding sensitivity – it appears that

[6]As noted above, we focus here and throughout these analyses on the full set of available data, which are summarized in the left-most "All Persons..." columns. In this case – and in general throughout these analyses – the results for the procedurally pure "Reference Person" data are completely consistent with the larger data set, although the smaller n's yield fewer statistically significant comparisons.

the implicit brackets approach made important inroads on resolving cognitive barriers, but did not have the desired impact with regard to reducing sensitivity concerns, and may even have had the opposite effect.

3.2 Comparison of Estimates and Distributions

Our goal in implementing the implicit brackets approach was to reduce nonresponse without negatively affecting the quality of income amount reports. Absent specific validity assessment procedures, our tools for evaluating quality differences are rather blunt, at best, and our small n's only add to imprecision. Our approach is to assume that the experimental treatment was quality neutral to the extent that it produced similar data outcomes when compared to the control treatment. In this section we compare the two treatments with regard to two very basic outcomes – the central tendency of the reports, and their distribution.

3.2.1 Means and medians

Table 3 compares the two treatments in terms of the average reported annual income amount for each type of asset. We present both mean amounts[7] and medians, since the former are strongly affected by outlier amounts, and by decisions regarding the recoding of the "less than X" reports in the experimental treatment. In addition, to avoid the latter difficulty and ensure greater comparability between the treatments, we also show (in panel B of Table 3) the results excluding all amounts in both treatments that were below the "bracket" treatment's minimum value.

The conclusions are the same regardless of whether one examines all reports or just the above-minimum-value reports. Neither the means nor the medians show any significant treatment differences in any cell of Table 3. The mean comparisons, in addition, show no consistent pattern of differences, offering further support to the conclusion that the null hypothesis – that the mean reported income amount reports do not differ – cannot be rejected. The story is not quite the same, however, for the other measure of central tendency, the median income amounts, where we note a general consistency in the direction of the observed differences. The control treatment medians are never smaller than the experimental treatment values, and often larger (there are also some "ties"). Although the extent of the consistency lacks sufficient statistical potency to justify rejection of the null hypothesis, the pattern of results warrants a closer examination of the distribution of reports. One obvious possibility, which might account for the observed pattern, is that the two questionnaire treatments produced different outcomes at the low end of the response distribution – one of the intended features of the experimental

[7]Recall that, in the "bracket" treatment, amounts could simply be reported as less than some initial minimum value, which varied by type of asset, with no attempt to pin down a precise amount. For purposes of these analyses, such reports were recoded to the mid-point between 0 and the minimum value – i.e., $5 for "less than $10" reports (checking accounts and savings accounts), $25 for "less than $50" reports (CDs), and $50 for "less than $100" reports (mutual funds and stocks).

treatment was, after all, reduced burden for reporting very small amounts. Next we examine this possibility.

3.2.2 Categorical distributions

To assess the similarity of amount report distributions we recoded the continuous amount reports into four categories, and compared the categorical distributions using chi-square. We assigned categories by determining the approximate quartile values – 25%, 50%, and 75% – for the full array of reports, including both treatments. The results of this categorization, with "category 1" the lowest quartile, and "category 4" the highest, are summarized, by treatment, in Table 4[8].

According to a chi-square test, for only one asset type, stocks, do the distributions of amount reports differ significantly by questionnaire treatment. (We note a similar effect for mutual funds in the "reference person only" analyses.) However, within this larger finding there is no indication of a greater tendency for the proportion of experimental treatment cases in the lowest category to significantly exceed the comparable proportion in the control treatment – in fact, for the lowest quartile of responses the significant difference is in the opposite direction. The other asset types show a mix of non-significant treatment differences in the lowest amount category. In sum, the experimental treatment does not seem to have elicited more very low amount reports than the control treatment, and in general the response distributions do not seem to differ by questionnaire treatment. Thus, the implicit brackets approach does not seem prone to the "anchoring" phenomenon that has been identified with the use of unfolding brackets – the tendency for reports to be affected by the dollar amount used as the initial entry point into the sequence of bracketing questions, and in particular the tendency for lower "entry" values to elicit lower amount reports (e.g., Hurd and Rodgers, 1998).

The results summarized in Table 4 do suggest a different kind of treatment effect, one that operates at the high end of the response distribution, rather than the low end. Although the observed individual differences are often small, and in no case confirmed as real differences by statistical significance tests, it is impossible to ignore the fact that for all five asset types the proportion of control treatment cases in the highest category exceeds the comparable figure in the experimental treatment. This fact, combined with the presence of some suspiciously high outliers, led us to additional analyses focused on high-end amounts, which we summarize below.

3.2.3 High amount outliers

As noted above, "eyeball" analysis of the distribution of amount reports revealed occasional unbelievably high amounts – e.g., reports of checking accounts that had produced multiple thousands of

[8]In response to a reviewer's suggestion, we repeated the categorical/distribution analysis, using only the control treatment responses as the "baseline" for establishing the quartile category break-points. This change had no important impact on the results, and no impact at all on the conclusions to be drawn from them.

dollars in interest income. We suspect that such reports might in fact signal respondents who misinterpreted the survey task, and thought they were being asked to report the asset balance or value, as opposed to the annual income it produced. We further wondered whether the "more or less than [minimum value]" component of the implicit bracket approach might have served as a clearer cue to the question's intent to respondents in that treatment, and thus reduced the frequency of erroneous and very high "balance" reports. One manifestation of this process would be a higher proportion of outliers in the control treatment reports.

Table 5 summarizes the results of an analysis of treatment differences in the frequency of outlier reports at the high end of the response distribution, for three definitions of "outlier" – reports at or beyond the 90th percentile, the 95th percentile, and the 99th percentile. These analyses are hampered somewhat by very small cell n's, especially for the more extreme outlier definitions, which makes statistical significance difficult to detect. We do find a completely consistent tendency for the least extreme (90th percentile) outlier analysis to show a higher proportion of very high amounts in the control treatment; however, this pattern disappears entirely at the more extreme (95th and 99th percentile) outlier levels. Thus, the control treatment's tendency to produce more upper quartile amounts, as shown in the previous section, extends to even more extreme high amounts, but not to such extreme levels as to signal with near certainty the presence of overreporting errors.

Without validating data, we find these results difficult to interpret. On the one hand, if the high reports are generally true, they could indicate an important advantage of the control treatment with regard to eliciting better reports from high income asset holders. But on the other, if the high reports are generally false, they could indicate important problems with regard to respondents' understanding of the response task.

3.3 The Precision of Amount Reports

One consequence of the unfolding bracket technique, at least as experienced by some researchers, is a reduction in precise, point-estimate reports, and an increase in the use of ranges, even as a first-choice response (e.g., Kennickell, 1997). Although the implicit brackets procedure does not use ranges explicitly, and is not a fallback response option which attempts to salvage information from an initial nonresponse, it is still of concern that the implicit invoking of ranges ("... to the nearest $10?") might reduce the precision of respondents' answers. In fact, an impetus for this procedure was just this; we hoped to signal to respondents with sensitivity issues concerning the reporting of exact amounts, or who were uncertain as to an exact amount, that we were not necessarily after extreme precision, and that it was acceptable to report a rounded amount. In this section we report the results of our examination of the precision of reported amounts in the implicit brackets experiment, using the frequency of rounded reports as our operational definition of precision. We find that reduced precision is not a byproduct of the experimental approach – just the opposite, in fact.

Table 6 contrasts the two treatments with regard to the proportion of amount reports that were rounded. We show two levels of rounding – a mild form (amounts that are divisible by 10) and a more extreme form (amounts divisible by 100). Again, we show separate results for each asset type, and separate analyses both including and excluding below-minimum amounts. Regardless of the details of the analysis, it is clear that the experimental treatment did not yield a higher proportion of less precise (rounded) amount reports. In fact, the experimental treatment seems to have actually increased report precision. With one exception ($10 rounding, savings accounts, excluding below-minimum amounts), all of the several significant differences are in the direction of increased rounding in the control treatment, and similarly with one exception ($10 rounding, mutual funds, including all amount both above and below the"more/less" minimum amounts), all of the nonsignificant differences are also of that type. It appears that the experimental treatment may have had an effect opposite to what was intended, and sent stronger signals for precision rather than weaker ones, (which may explain its failure to reduce refusal nonresponse). Regardless of the mechanism, it is clear from these results that reduced precision is <u>not</u> a byproduct of the implicit brackets approach as implemented in this experiment.

3.5 Response Reliability

An important indicator of survey data quality is the reliability of responses – the extent to which the consistent administration of a survey question, under consistent conditions, leads to consistent replies. Although high reliability does not automatically mean high data quality (a consistently incorrect response may be perfectly reliable), high reliability is nevertheless a necessary condition for high data quality, and it is generally assumed that increases in reliability indicate increases in overall data quality.

The QDERS experiment included a response variance reinterview to permit a comparison of the reliability of the data produced by the two forms of the amount report questions. For the most part, reinterview procedures mimicked the original interview procedures. The same facility and staff were used, the introductory script was modified slightly for reinterview, and interviewers collected all the information that was gathered in the original interview, in the same manner, with the exception of the household roster. An office procedure carried out prior to reinterviewing transcribed the household members' names from the roster of the original interview onto the reinterview roster; at the start of the reinterview, interviewers verified that those listed still lived in household, and anyone no longer in residence was simply dropped from the reinterview roster. Interviewers were allowed to conduct the reinterview with any eligible household member, regardless of who had served as the original interview household respondent. The field period for the reinterview began about two weeks after the original interview field period ended, and ran for about two weeks. All 1,291 originally-interviewed households were eligible for reinterview; altogether, 1,088 re-interviews were completed, or about 84% of all original interviews.

We use two measures of reliability to interpret the reinterview data: the "index of inconsistency" and the "gross difference rate" (GDR). The index of inconsistency represents the percentage of the total variance in the responses to a survey item that is due to response variance; the GDR is simply the

percentage of responses that change between the original interview and the reinterview. Because it takes into account an item's distributional properties – unlike the GDR, the index is not confounded by the frequency of the characteristic being measured – the index of inconsistency is the most commonly-used measure at the Census Bureau to evaluate the results of reinterview studies. While a large GDR is indicative of a problem, a small GDR does not necessarily mean the item is without problems. For items with few response categories, and especially for items measuring very rare characteristics, the GDR may be small but the index may be high. Despite the limitations of the GDR for producing precise reliability estimates, it is still useful for experimental evaluation purposes, and its simplicity offers an intuitive clarity and appeal that is somewhat lacking in the index of inconsistency, and thus we show both measures in the results that follow. We also note that in the present case the two measures yield highly consistent results.

Reliability results for asset amount reports (recoded into quintile categories) are summarized in Table 7. Regardless of which indicator is used, we see that most of the differences between the control and the experimental procedure are very small, and none even approaches statistical significance. (Experimental-control treatment differences for CDs appear much larger than for the other asset types, by observation, but the precision of the CDs estimates is limited by the very small n's, rendering the effects nonsignificant.) The trend in the direction of the observed differences, while somewhat compelling, is not sufficiently consistent to justify concluding that it is meaningful statistically. Thus we are forced to conclude that the two treatments were roughly equivalent with regard to the random "noisiness" of the responses they produced.

3.6 Interviewers' Evaluations of the Two Treatments

At the end of the QDERS field period we administered a brief questionnaire to the interviewers, to gather data from them which might assist the evaluation of the implicit brackets income amount reporting experiment. Most of the interviewer evaluation items concerned other aspects of the QDERS design not of relevance for present purposes. Three items, however, dealt explicitly with the income amount question experiment in the QDERS survey. Each of these items presented interviewers with parallel statements about each of the questionnaire treatments, and asked them to indicate their "level of agreement or disagreement" with each statement, using a five-point, verbal label scale – "strongly agree," "agree," "neutral," "disagree," "strongly disagree" – to which we assigned numeric values 1 (strongly agree) through 5 (strongly disagree) for analysis. Below we describe the results of these evaluations.

3.6.1 "Encouraged Rs to be careless"

One evaluation item asked interviewers to provide agree/disagree scores to two statements (one for each instrument treatment) of the following form: "[Treatment] encouraged Rs to be careless." Interviewers mostly disagreed with this statement, regardless of which treatment they were rating. The average scores assigned were 3.5 and 3.4 for the control treatment and experimental treatment,

respectively, a difference that is not statistically significant. We conclude that interviewers did not perceive in respondents' behavior any decrease in effort on the part of those to whom they administered the implicit brackets question, compared to the control treatment.

3.6.2 "Easy for Rs to answer"

A second evaluation item asked interviewers to provide agree/disagree scores to two statements of the following form: "[Treatment] was easy for Rs to answer." On this dimension, interviewers reported a clear difference between the two treatments, with an average score of 3.7 for the control treatment, versus 1.4 for the implicit brackets treatment. Not only is this difference highly significant statistically ($t=6.3$, 14df, $p<.001$), but it also seems to have substantive importance as well, putting the control treatment solidly on the "disagree" side of the neutral point, and the experimental treatment even more solidly on the "agree" side. Interviewers clearly perceived a difference between the two questions in terms of the burden they imposed on respondents.

3.6.3 "Accurate answers"

The third evaluation item used the same format as described above to capture interviewers' assessments regarding the accuracy of respondents' reports: "[Treatment] collected accurate answers." Of course, interviewers' assessments of response accuracy are somewhat difficult to defend, although it is possible that they sensed the nonresponse and precision differences, and entered these factors into their accuracy calculus. Regardless of the basis for their judgments, on this dimension, too, interviewers reported that they could perceive clear differences between the treatments, again in favor of the implicit brackets design. The average scale rating assigned to the control treatment was 3.2, versus 1.9 for the experimental treatment, a difference which is easily statistically significant ($t=4.5$, 14df, $p<.001$).

4. Summary and Conclusions

In this test, the experimental, implicit brackets approach showed some significant advantages over a more standard approach to obtaining survey reports of asset income amounts. Most prominently, using the implicit brackets approach resulted in a reduction in item nonresponse, primarily through a reduction in "don't know" nonresponse. Reports were, if anything, more rather than less precise in the experimental treatment. These two major findings in the survey data are supported by interviewers' subjective judgments that the implicit brackets approach was easier for respondents and produced more accurate reports. We find no important differences in the reliability of the income reports elicited by the two treatments, and only insignificant effects on the average dollar amounts reported, although with some indication that the standard approach elicited more reports at the high end of the response distribution. This latter finding justifies some concern about the experimental treatment, given the research evidence that asset income amounts in general suffer from underreporting (Moore, Stinson, and Welniak, 1999). On the other hand, to the extent that the high amount reports represent overreports of true income – which some of the more extreme outliers almost certainly must be – this

may indicate an advantage of the implicit brackets approach, in reducing respondent misunderstanding of the response task.

We speculate about two mechanisms which might account for reduced nonresponse. One possibility, of course, is the extreme ease of the question sequence and response task for respondents with little income to report – all they had to do was make a simple "more than/less than" judgment about the position of the true value with regard to a minimum value. This is in marked contrast to respondents in the control condition who, no matter how trivial the amount, were expected to report it accurately. This process, however, would seem very likely to produce two outcomes which we did not see in the QDERS results: more low amount reports in the experimental treatment (due to more DK's for low amounts in the control treatment), and a corresponding reduction in mean amounts. Perhaps, then, another mechanism was at work to reduce nonresponse not just at the low end of the distribution, but throughout it: the "more/less" task was also a very easy response task for respondents with larger amounts to report, and perhaps served as a "foot in the door" to the followup question requesting a more specific response. Regardless of the underlying process responsible for the reduction in DK nonresponse, our intention that the implicit brackets approach signal to respondents somewhat reduced needs for precision, and thus reduce their sensitivity concerns, and thus reduce refusals, appears not to have worked, and may in fact have backfired.

We view the QDERS results, on the whole, as a positive first step in the development of improved procedures for capturing survey reports of income amounts. The necessary next steps are, first, the refinement of procedures for reducing sensitivity concerns, and second, the replication of a test of the implicit brackets approach in a survey setting without QDERS' limitations – i.e., a larger scale, non-RDD sample survey, perhaps administered in-person, and certainly with a more satisfactory response rate.

References

American Association for Public Opinion Research (1998), “Standard Definitions – Final Dispositions of Case Codes and Outcome Codes for RDD Telephone Surveys and In-Person Household Surveys.” *[http://www.aapor.org/ethics/stddef.html]*

R. Bell (1984), "Item Nonresponse in Telephone Surveys: An Analysis of Who Fails to Report Income." Social Science Quarterly. 65: 207-215.

Cantor, D., S. Brandt, and J. Green (1991), "Results of first wave of SIPP interviews." Unpublished Westat report to the U.S. Bureau of the Census (memorandum to Chet Bowie), February 21, 1991.

GENESYS Sampling Systems (1997), "GENESYS Sampling System Methodology." GENESYS Sampling Systems, Fort Washington, PA.

Heeringa, S., D. Hill, and D. Howell (1995), "Unfolding Brackets for Reducing Item Nonresponse in Economic Surveys." Health and Retirement Study Working Paper Series, Paper No. 94-029. Survey Research Center, Institute for Social Research, University of Michigan.

Hess, J., C. Keeley, J. Moore, J. Pascale, and J. Rothgeb (2000), “The Effects of Person-level vs. Household-level Questionnaire Design on Survey Estimates and Data Quality.” Paper presented at the annual meetings of the American Association for Public Opinion Research, Portland, OR, May 18-21, 2000.

Hippler, H., and G. Hippler (1986), "Reducing Refusal Rates in the Case of Threatening Questions: The 'Door-in-the-Face' Technique." Journal of Official Statistics. 2(1): 25-33.

Hurd, M., and W. Rodgers (1998), "The Effects of Bracketing and Anchoring on Measurement in the Health and Retirement Study." *[source?]*

Juster, F., and J. Smith (1997), "Improving the Quality of Economic Data: Lessons from the HRS and AHEAD." Journal of the American Statistical Association. 92(440): 1268-1278.

Juster, F., and J. Smith (1998), "Enhancing the Quality of Data on Income and Wealth: Recent Developments in Survey Methodology." Paper prepared for the 25th General Conference of the International Association for Research in Income and Wealth. Cambridge, England, August 23-29, 1998.

Kennickell, A. (1997), "Using Range Techniques with CAPI in the 1995 Survey of Consumer Finances." Survey of Consumer Finances Working Paper. Board of Governors of the Federal Reserve System, Washington, DC.

Moore, J., L. Stinson, and E. Welniak (1999), "Income Reporting in Surveys: Cognitive Issues and Measurement Error." Chapter 10 in Sirkin, M., D. Herrmann, S. Schechter, N. Schwarz, J. Tanur, and R. Tourangeau (eds.), Cognition and Survey Research. New York: John Wiley & Sons, 1999.

Ross, C., and J. Reynolds (1996), "The Effects of Power, Knowledge, and Trust on Income Disclosure in Surveys." Social Science Quarterly. 77(4): 899-911.

Snedecor, G. and W. Cochran (1967), Statistical Methods. Ames, IA: The Iowa State University Press.

U.S. Census Bureau (1999), "QDERS Training Guide." Unpublished U.S. Census Bureau document prepared for the 1999 QDERS project, (undated, spring 1999).

Attachment 1

Control Treatment Income Amounts Questions

The next few questions are about interest and dividends income (you/name) may have received in the last year, that is, between January and December of 1998. For these questions, we're NOT interested in interest and dividends associated with an IRA, a 401K, or any other type of retirement account.

47. How much interest did (you/name) earn on all interest-earning checking accounts in 1998?	9999.99 (Go to next source) DK 9 (Go to next source) Ref 9 (Go to next source)
48. How much interest did (you/name) earn on all savings accounts in 1998?	9999.99 (Go to next source) DK 9 (Go to next source) Ref 9 (Go to next source)
49. How much interest did (you/name) earn from all CDs in 1998?	9999.99 (Go to next source) DK 9 (Go to next source) Ref 9 (Go to next source)
50. Excluding retirement accounts, how much did (you/name) earn in mutual fund dividends in 1998?	9999.99 (Go to next source) DK 9 (Go to next source) Ref 9 (Go to next source)
51. Excluding retirement accounts, how much did (you/name) earn in stock dividends in 1998?	9999.99 DK 9 Ref 9

Attachment 2

Experimental Treatment Income Amounts Questions

The next few questions are about interest and dividends income (you/name) may have received in the last year, that is, between January and December of 1998. For these questions, we're NOT interested in interest and dividends associated with an IRA, a 401K, or any other type of retirement account.

47. How much interest did (you/name) earn on all interest-earning checking accounts in 1998? Would you say it was more or less than $10? 47b. How much was it, to the nearest $5?	More 9 (Ask 47b) DK 9 (Go to next source) Less 9 (Go to next source) Ref 9 (Go to next source) 9999.99 (Go to next source) DK 9 (Go to next source) Ref 9 (Go to next source
48. How much interest did (you/name) earn on all savings accounts in 1998? Would you say it was more or less than $10? 48b. How much was it, to the nearest $10?	More 9 (Ask 48b) DK 9 (Go to next source) Less 9 (Go to next source) Ref 9 (Go to next source) 9999.99 (Go to next source) DK 9 (Go to next source) Ref 9 (Go to next source)
49. How much interest did (you/name) earn from all CDs in 1998? Would you say it was more or less than $50? 49b. How much was it, to the nearest $10?	More 9 (Ask 49b) DK 9 (Go to next source) Less 9 (Go to next source) Ref 9 (Go to next source) 9999.99 (Go to next source) DK 9 (Go to next source) Ref 9 (Go to next source)
50. Excluding retirement accounts, how much did (you/name) earn in mutual fund dividends in 1998? Would say it was more or less than $100? 50b. How much was it, to the nearest $50?	More 9 (Ask 50b) DK 9 (Go to next source) Less 9 (Go to next source) Ref 9 (Go to next source) 9999.99 (Go to next source) DK 9 (Go to next source) Ref 9 (Go to next source)
51. Excluding retirement accounts, how much did (you/name) earn in stock dividends in 1998? Would you say it was more or less than $100? 51b. How much was it, to the nearest $25?	More 9 (Ask 51b) DK 9 (Go to next source) Less 9 (Go to next source) Ref 9 (Go to next source) 9999.99 (GO TO NEXT SOURCE) DK 9 (GO TO NEXT SOURCE) Ref 9 (GO TO NEXT SOURCE)

Table 1: Asset Income Amount Item Nonresponse, by Asset Type and Question Treatment, for All Persons Asked an Income Amount Question, and for Reference Persons Only

Asset Type:		ALL PERSONS ASKED AN INCOME QUESTION			REFERENCE PERSONS ASKED AN INCOME QUESTION		
		Question Treatment			Question Treatment		
		Control		Experimental	Control		Experimental
Checking Accounts	% nr (n)	48.8 (301)	>	43.7 (348)	46.3 (149)	>	43.9 (171)
Savings Accounts	% nr (n)	**50.6**[1/] (425)	>	**44.5**[1/] (425)	50.7 (219)	>	44.1 (202)
CDs	% nr (n)	62.2 (119)	>	57.0 (128)	66.7 (57)	>	61.3 (62)
Mutual Funds	% nr (n)	57.4 (148)	>	52.2 (136)	56.2 (73)	>	55.6 (72)
Stocks	% nr (n)	50.9 (159)	>	47.8 (136)	52.0 (75)	>	46.2 (65)

1/ chi-square=3.19, 1df, p<.10

Table 2: Type of Asset Income Amount Item Nonresponse (DK vs. Refused), by Asset Type and Question Treatment, for All Persons Asked an Income Amount Question, and for Reference Persons Only

Asset Type:		ALL PERSONS ASKED AN INCOME QUESTION			REFERENCE PERSONS ASKED AN INCOME QUESTION		
		Question Treatment			Question Treatment		
		Control		Experimental	Control		Experimental
Checking Accts	% DK	40.9	>	35.6	40.3	>	36.3
	% ref	8.0	<	8.1	6.0	<	7.6
	% amt report'd	51.2		56.3	53.7		56.1
	(n)	(301)		(348)	(149)		(171)
(chi-square test)		n.s.			n.s.		
Savings Accts	% DK	**45.9**[1/]	>	**36.7**[1/]	**47.5**[3/]	>	**36.1**[3/]
	% ref	**4.7**[2/]	<	**7.8**[2/]	**3.2**[4/]	<	**7.9**[4/]
	% amt report'd	**49.4**		**55.5**	49.3		55.9
	(n)	(425)		(425)	(219)		(202)
(chi-square test)		chi-square=9.03, 2df, p<.05			chi-square=8.39, 2df, p<.05		
CDs	% DK	49.6	>	40.6	57.9	>	48.4
	% ref	12.6	<	16.4	8.8	<	12.9
	% amt report'd	37.8		43.0	33.3		38.7
	(n)	(119)		(128)	(57)		(62)
(chi-square test)		n.s.			n.s.		
Mutual Funds	% DK	49.3	>	43.4	48.0	>	45.8
	% ref	8.1	<	8.8	8.2	<	9.7
	% amt report'd	42.6		47.8	43.8		44.4
	(n)	(148)		(136)	(73)		(72)
(chi-square test)		n.s.			n.s.		
Stocks	% DK	44.0	>	39.7	45.3	>	41.5
	% ref	6.9	<	8.1	6.7	>	4.6
	% amt report'd	49.1		52.2	48.0		53.9
	(n)	(159)		(136)	(75)		(65)
(chi-square test)		n.s.			n.s.		

1/ t=2.73, 848df, p<.01
2/ t=1.85, 848df, p<.10
3/ t=2.37, 419df, p<.05
4/ t=2.10, 419df, p<.05

Table 3: Average Dollar Amounts Reported, by Asset Type and Question Treatment, for All Persons Who Reported an Income Amount, and for Reference Person Amount Reporters Only

Asset Type:		ALL PERSONS WHO REPORTED AN INCOME AMOUNT			REFERENCE PERSON AMOUNT REPORTERS ONLY		
		Question Treatment			Question Treatment		
		Control		Experimental	Control		Experimental
A. Including all amounts, both above and below the "more/less" minimum amount							
Checking Accounts	mean	$316	>	$242	$289	>	$138
	median	40	>	30	38	>	30
	(n)	(154)		(196)	(80)		(96)
Savings Accounts	mean	254	<	255	329	>	291
	median	50		50	50		50
	(n)	(210)		(236)	(108)		(113)
CDs	mean	847	>	779	1018	>	664
	median	400	>	200	500	>	190
	(n)	(45)		(55)	(19)		(24)
Mutual Funds	mean	1286	<	1310	821	<	1110
	median	400		400	400	>	300
	(n)	(63)		(65)	(32)		(32)
Stocks	mean	1098	>	1047	1250	>	777
	median	300	>	100	330	>	50
	(n)	(78)		(71)	(36)		(35)
B. Excluding amounts below the "more/less" minimum amount							
Checking Accounts	mean	$389	>	$330	$361	>	$195
	median	80	>	65	100	>	65
	(n)	(125)		(143)	(64)		(67)
Savings Accounts	mean	321	<	329	399	>	390
	median	100		100	100	>	95
	(n)	(166)		(182)	(89)		(84)
CDs	mean	1055	<	1061	1135	>	927
	median	550	>	450	500	>	300
	(n)	(36)		(40)	(17)		(17)
Mutual Funds	mean	1717	>	1541	1243	<	1306
	median	600	>	500	800	>	500
	(n)	(47)		(55)	(21)		(27)
Stocks	mean	1779	<	2022	2242	>	1875
	median	1000	>	825	1750	>	340
	(n)	(48)		(36)	(20)		(14)

Note: No difference is statistically significant.

Table 4: Percent Distributions of Reported Dollar Amounts (4 Category Recode), by Asset Type and Question Treatment, for All Persons Who Reported an Income Amount, and for Reference Person Amount Reporters Only

Asset Type:		ALL PERSONS WHO REPORTED AN INCOME AMOUNT			REFERENCE PERSON AMOUNT REPORTERS ONLY		
		Question Treatment			Question Treatment		
		Control		Experimental	Control		Experimental
Checking Accts	category 1	18.8	<	27.0	18.8	<	30.2
	category 2	29.9	>	24.0	30.0	>	21.9
	category 3	26.0	>	24.0	23.8	<	25.0
	category 4	25.3	>	25.0	27.5	>	22.9
	(n)	(154)		(196)	(80)		(96)
(chi-square test)		n.s.			n.s.		
Savings Accts	category 1	21.0	<	22.9	24.1	<	31.9
	category 2	30.5	>	30.1	26.9	>	21.2
	category 3	24.8	>	23.7	22.2	<	25.7
	category 4	23.8	>	23.3	26.9	>	21.2
	(n)	(210)		(236)	(108)		(113)
(chi-square test)		n.s.			n.s.		
CDs	category 1	20.0	<	27.3	10.5	<	33.3
	category 2	24.4	<	29.1	21.1	<	29.2
	category 3	24.4	>	23.6	31.6	>	20.8
	category 4	31.1	>	20.0	36.8	>	16.7
	(n)	(45)		(55)	(19)		(24)
(chi-square test)		n.s.			n.s.		
Mutual Funds	category 1	28.6	>	26.2	**34.4**[3/]	>	**15.6**[3/]
	category 2	22.2	<	27.7	**9.4**[4/]	<	**40.6**[4/]
	category 3	20.6	<	26.2	28.1	>	21.9
	category 4	28.6	>	20.0	28.1	>	21.9
	(n)	(63)		(65)	(32)		(32)
(chi-square test)		n.s.			chi-square=9.00, 3df, p<.05		
Stocks	category 1	**35.9**[1/]	>	**11.3**[1/]	**38.9**[5/]	>	**14.3**[5/]
	category 2	**6.4**[2/]	<	**46.5**[2/]	**2.8**[6/]	<	**45.7**[6/]
	category 3	30.8	>	22.5	19.4	<	31.4
	category 4	26.9	>	19.7	**38.9**[7/]	>	**8.6**[7/]
	(n)	(78)		(71)	(36)		(35)
(chi-square test)		chi-square=34.49, 3df, p<.001			chi-square=25.50, 3df, p<.001		

1/ t=3.71, 147df, p<.001
2/ t=6.09, 147df, p<.001
3/ t=1.75, 62df, p<.10
4/ t=3.05, 62df, p<.005
5/ t=2.41, 69df, p<.05
6/ t=4.78, 69df, p<.001
7/ t=3.18, 69df, p<.005

Table 5: Percent of Reported Amounts that were Extreme (90th/95th/99th Percentile) Outliers, by Asset Type and Question Treatment, for All Persons Who Reported an Income Amount, and for Reference Person Amount Reporters Only

Asset Type, and Outlier (Percentile) Level:	ALL PERSONS WHO REPORTED AN INCOME AMOUNT			REFERENCE PERSON AMOUNT REPORTERS ONLY		
	Question Treatment			Question Treatment		
	Control		Experimental	Control		Experimental
% of reports at 90+ percentile:						
Checking Accounts	11.0	>	9.7	13.8	>	8.3
Savings Accounts	11.4	>	8.9	**13.9**[1/]	>	**7.1**[1/]
CDs	13.3	>	9.1	15.8	>	8.3
Mutual Funds	12.7	>	12.3	25.0	>	12.5
Stocks	11.5	>	8.5	**19.4**[2/]	>	**5.7**[2/]
% of reports at 95+ percentile:						
Checking Accounts	5.8	>	4.6	7.5	>	3.1
Savings Accounts	6.7	>	5.1	**10.2**[3/]	>	**4.4**[3/]
CDs	4.4	<	7.3	5.3	<	8.3
Mutual Funds	4.8	<	6.2	3.1	<	9.4
Stocks	6.4	<	7.0	8.3	>	5.7
% of reports at 99+ percentile:						
Checking Accounts	1.3	<	1.5	2.5	>	0
Savings Accounts	1.9	>	1.3	0.9	<	1.8
CDs	0	<	3.6	0	<	4.2
Mutual Funds	3.2	>	3.1	0	<	3.1
Stocks	0	<	2.8	0	<	2.9

Note: See Table 4 for n's.

1/ chi-square=2.75, 1df, p<.10
2/ chi-square=3.02, 1df, p<.10
3/ chi-square=2.73, 1df, p<.10

Table 6: Percent of Reported Amounts that were Rounded (to $10 and to $100), by Asset Type and Question Treatment, for All Persons Who Reported an Income Amount, and for Reference Person Amount Reporters Only

Asset Type, and Rounding Level:	ALL PERSONS WHO REPORTED AN INCOME AMOUNT			REFERENCE PERSON AMOUNT REPORTERS ONLY		
	Question Treatment			Question Treatment		
	Control		Experimental	Control		Experimental
A. Including all amounts, both above and below the "more/less" minimum amount						
rounded to $10						
Checking Accounts	**76.6**[1/]	>	**54.6**[1/]	**81.3**[8/]	>	**54.2**[8/]
Savings Accounts	77.1	>	71.6	**82.4**[9/]	>	**67.3**[9/]
CDs	**93.3**[2/]	>	**76.4**[2/]	**94.7**[10/]	>	**70.8**[10/]
Mutual Funds	90.5	<	93.9	90.6	<	96.9
Stocks	**89.7**[3/]	>	**78.9**[3/]	**88.9**[11/]	>	**71.4**[11/]
rounded to $100						
Checking Accounts	**40.9**[4/]	>	**25.0**[4/]	**46.3**[12/]	>	**22.9**[12/]
Savings Accounts	**44.3**[5/]	>	**31.8**[5/]	**48.2**[13/]	>	**31.0**[13/]
CDs	**82.2**[6/]	>	**47.3**[6/]	**94.7**[14/]	>	**50.0**[14/]
Mutual Funds	73.0	>	72.3	75.0	>	71.9
Stocks	**79.5**[7/]	>	**36.6**[7/]	**80.6**[15/]	>	**17.1**[15/]
B. Excluding amounts below the "more/less" minimum amount						
rounded to $10						
Checking Accounts	**85.6**[16/]	>	**73.4**[16/]	**90.6**[21/]	>	**76.1**[21/]
Savings Accounts	**85.5**[17/]	<	**91.2**[17/]	86.5	<	88.1
CDs	97.2	>	95.0	100	>	94.1
Mutual Funds	95.7	>	92.7	100	>	96.3
Stocks	**95.8**[18/]	>	**80.6**[18/]	**95.0**[22/]	>	**71.4**[22/]
rounded to $100						
Checking Accounts	41.6	>	32.9	**46.9**[23/]	>	**31.3**[23/]
Savings Accounts	44.0	>	39.6	44.9	>	39.3
CDs	**94.4**[19/]	>	**65.0**[19/]	**100**[24/]	>	**70.6**[24/]
Mutual Funds	85.1	>	83.6	90.5	>	85.2
Stocks	**89.6**[20/]	>	**69.4**[20/]	**85.0**[25/]	>	**42.9**[25/]

Notes: (1) See Table 3 for n's. (2) df=1 for all chi-square tests (below).

1/ chi-square=18.23, p<.001
2/ chi-square=5.30, p<.05
3/ chi-square=3.37, p<.10
4/ chi-square=10.03, p<.005
5/ chi-square=7.40, p<.01
6/ chi-square=12.97, p<.001
7/ chi-square=28.25, p<.001
8/ chi-square=14.36, p<.001
9/ chi-square=6.70, p<.01
10/ chi-square=4.00, p<.05
11/ chi-square=3.42, p<.10
12/ chi-square=10.66, p<.001
13/ chi-square=6.82, p<.01
14/ chi-square=10.06, p<.005
15/ chi-square=28.55, p<.001
16/ chi-square=5.98, p<.05
17/ chi-square=2.74, p<.10
18/ chi-square=5.02, p=.05
19/ chi-square=9.88, p<.005
20/ chi-square=5.41, p=.05
21/ chi-square=4.93, p<.05
22/ chi-square=3.65, p<.10
23/ chi-square=3.32, p<.10
24/ chi-square=5.86, p<.05
25/ chi-square=6.68, p<.05

Table 7: Asset Income Amount Reliability (Index of Inconsistency and Gross Difference Rate), by Asset Type and Question Treatment, for Reference Person Amount Reporters Only

Asset Type:	REFERENCE PERSON AMOUNT REPORTERS ONLY Question Treatment Control		Experimental
Checking Accounts:			
Index of Inconsistency	67.1	<	69.7
Gross Difference Rate	53.4	<	54.8
(n)	(58)		(73)
Savings Accounts:			
Index of Inconsistency	43.7	<	53.9
Gross Difference Rate	34.5	<	42.2
(n)	(84)		(109)
CDs:			
Index of Inconsistency	21.1	<	52.6
Gross Difference Rate	16.7	<	41.7
(n)	(12)		(12)
Mutual Funds:			
Index of Inconsistency	43.0	>	40.4
Gross Difference Rate	31.6	<	31.8
(n)	(19)		(22)
Stocks:			
Index of Inconsistency	52.9	<	53.5
Gross Difference Rate	37.5	>	31.8
(n)	(24)		(22)

Note: No difference is statistically significant.