A CONTROLLED EXPERIMENT FOR TESTING RESPONDENT-GENERATED INTERVALS METHODOLOGY IN SURVEYS

Judith M. Tanur, State University of New York at Stony Brook; S. James Press, University of California, Riverside; Diane Miller, University of California, Riverside; and Diana Petitti, Kaiser Permanente of Southern California

Judith M. Tanur, Department of Sociology, State University of New York at Stony Brook, Stony Brook, NY 11794-4356

Key Words: Recall, Cognitive Aspects of Surveys

This paper presents an experiment that will extend the knowledge of the Respondent-Generated Interval (RGI) procedure. The RGI procedure is a method of questioning in surveys, proposed by Press (2000), that is expected to increase accuracy of estimators of population parameters from data obtained from respondents' recall of factual numerical quantities such as amounts, dates, frequencies, and so forth. In this paper we ask how various approaches to the procedure compare in accuracy and response rate.

The technique as used so far requests respondents to supply not only a point assessment of the recalled quantity, but also assessments of the smallest and largest values that quantity could possibly have taken. Press (2000) proposed using the best recollection of each respondent to supply information about the center of his/her recall distribution, and then, in addition, asking for the smallest and largest values the true value might have been. Six point estimators of the population mean were compared including the sample mean and median, a midpoint estimator, a weighted average estimator, and two distinct Bayesian estimators obtained by hierarchical modeling. Several interval estimators were proposed as well. Press and Tanur (2000a), working with parallel record check surveys on two college campuses, showed that in that case these point and interval estimates of the population mean are quite accurate. Further, using the same campus surveys, we showed that some respondents choose not to offer the best recollection at all, but did offer the bounds. On the assumption that these respondents might not have responded at all if they were not been given the option to offer bounds, we concluded that the RGI procedure reduces item nonresponse, at least in this context (Press and Tanur, 2000b).

Asking two questions instead of one, however, adds to respondent burden and to the time and cost of interviews. So one of the major questions addressed in this new research is whether requesting bounds alone achieves accuracy similar to that achieved in the campus surveys. The cognitive processes involved in answering just the bounds question differ from those involved in answering both questions (notably, anchoring the bounds estimate on the point assessment is not an obvious answering strategy when the point estimate is not explicitly enunciated), which may affect accuracy. Will respondents be willing to answer questions in this unusual format? We will also address several other questions in this experiment. Will more accuracy result if respondents may choose whether to give a point assessment or bounds? Which are they more likely to choose? Does the order of the options affect choice if choice is offered, or accuracy whether or not there is a choice?

One of these questions has been addressed in earlier research. Marquis and Press (2000) conducted a telephone survey asking respondents questions about several types of income and changes in income over a five-year interval. While a quarter of the sample were given the same question ordering as were the respondents in the campus experiments (that is, point estimate first, then bounds), three-quarters of the respondents were asked for the bounds estimates before they were asked for the point estimate. Those data are currently being analyzed, and we hope to have some firm results before the experiment described here is in the field.

We should note that surveys sometimes ask respondents to bound their responses, that is, to answer within an interval. Commonly this is done when great precision is

not required for the uses contemplated for the data (for example, age to the nearest decade is often sufficiently precise for many applications), or when the question is sensitive and respondents are expected to feel more comfortable giving an approximate answer rather than an exact one (for example, respondents are sometimes asked to give their income within a range, often with branching depending on the respondent's answer in order to narrow the original range). But most often the intervals used in these procedures are supplied by the survey's author. The novelty of the procedure applied here is that the respondent him/herself generates the interval. We are aware of only one earlier use of intervals generated by the respondent, in the 1995 Survey of Consumer Finances (see Kennickell (1997). Here respondents were questioned about the dollar amount of some 479 variables. In cases in which the respondent gave either a "don't know" or a refusal response, a choice of interviewer-supplied or respondent-generated ranges was offered.

The Planned Experiment

We plan to field an experiment embedded in a survey (Fienberg and Tanur, 1988) involving seven groups that are asked questions about recall of various aspects of health. We are working with organizations that will verify the accuracy of answers from patient records (anonymously, or course). Questions will deal with duration of membership in the organization, frequency of visits, dating of most recent treatments, tests, and immunizations, and the numerical values of such results as cholesterol level. We will vary whether both the bounds and the point estimate are asked versus only one, the order of requesting the bounds and the point estimates, and whether or not the respondent has a choice of answering the bounds or point estimate question. Most of the experimental design is diagrammed in Figure 1. As can be seen, choice is completely crossed with order of presentation of the questions in a 2 by 2 factorial design. In addition, there are two further groups. The point-estimate-only group is a control group, reflecting standard survey practice. The bounds-only group will indicate if the RGI procedure can be administered with only one question, thus saving interview time and respondent burden, and still be acceptable to respondents and generate accurate data. The seventh group, not shown in Figure 1, asks the questions in the special form currently used by the organizations, so that we can answer the operational question of whether any of the new procedures are better than their current one.

Hypotheses and Speculation

There are a variety of dependent variables and hypotheses about them that can be considered in this work. The following is a long, but hardly exhaustive, list.

Item Nonresponse

First of all, following the finding of Press and Tanur (2000b), we expect item nonresponse to be lower when the bounds option is presented. According to this reasoning, we expect less item nonresponse in the bounds-only condition than in the control group, although there is a possibility that the strangeness of the bounds-only condition may puzzle some respondents and thus depress response rates. Further, we expect that when respondents are offered a choice of the form of the question they will answer, they will feel that they have the opportunity to tailor the question to their own situation and thus be more likely to respond. Thus we expect item nonresponse, as measured by the proportion of respondents failing to answer at least one of the questions (bounds question and point-estimate question) to be lower in the choice conditions than in the non-choice conditions.

Choice

We also expect there to be a main effect of choice on accuracy. If the respondent can choose the question he or she will answer, presumably he or she will choose the form that most closely mirrors the memory storage of his or her response. Thus we would expect the respondents choosing to answer the bounds question in the choice conditions to display greater accuracy than displayed on the bounds question in the non-choice conditions or when the bounds question appears alone. We hope that the degradation of accuracy in the bounds-only condition will be sufficiently small that the question will retain its utility while reducing respondent burden. Similarly, we expect that respondents choosing to answer the point-estimate question in the choice conditions will display more accuracy than is displayed in the non-choice or point-estimate only conditions.

Accuracy

In replication of Press and Tanur (2000a), we expect the Bayesian estimators derived from the RGIs to be closer to the truth than the usual sample means. Thus we expect that the Bayesian estimator in the bounds-only condition will be closer to the truth than the sample mean in the point-estimate-only condition. In the non-choice condition, we also expect the Bayesian estimator derived from the RGIs to be closer to truth than the sample mean derived from the point estimates. In the choice conditions, because we expect respondents to choose the mode of answering that more closely matches the way the information is coded and stored in

memory, we expect little difference in closeness to truth between the Bayesian estimator derived for those respondents who choose the bounds version and the sample mean derived for those respondents who choose the point estimate version.

Effects of ordering of question form

We expect that the ordering of question form (bounds question first or point-estimate question first) will affect responses through an anchoring and adjustment mechanism (Tversky and Kahnemann, 1974). When the point estimate version of the question is presented first to respondents, we expect them to anchor their bounds on the point estimate they choose, more strongly when they have to answer the point estimate question (non-choice conditions) than when they merely read it or listen to it first (choice conditions). We expect this effect to be evident in the length of the RGI. That is, we expect the length of the average RGI to be greatest in the bounds-only condition and smallest in the non-choice, point-estimate-first condition. Intermediate lengths of the average RGI, in the following ascending order, should appear in the point-estimate-first, choice condition; the bounds-estimate-first, non-choice condition; and the bounds- estimate-first, choice condition. Of course, no RGI will be available from the point-estimate only condition.

Symmetry of Intervals

The formulation of the Bayesian hierarchical model in Press (2000) assumed an approximating normal recall distribution for each respondent. In a Census Bureau survey of income recall that used the RGI protocol, Marquis and Press (2000) studied the issue of the symmetry of the respondents' recall distribution. Because the questions asking for the bounds ask for equal uncertainty on either side of the point estimate (e.g. What is the highest dollar amount you think this could have been? What is the lowest dollar amount you think this could have been?) our naïve belief was that respondents would give intervals symmetric around their point estimate. The authors instead found that about half of the respondents do give symmetric intervals, and about a quarter give intervals that are right skew (that is, the point response is between the midpoint of the given interval and the lower bound) and about a quarter that are left skew.

We wish to study this phenomenon of interval asymmetry further for several reasons. First, if respondents often give asymmetric intervals, we need to model their response distributions with a probability distribution other than the normal with its implied symmetry. Second, in our earlier work we found that an ad hoc estimator which is the average of the midpoints of the RGIs performed surprisingly well. And as Tourangeau has pointed out (2000, p.49), that average midpoint RGI estimator differs from the sample mean of the point estimates only to the extent that respondents give asymmetric intervals. Finally, the findings of Fred Conrad and his colleagues (Conrad and Brown, 1994; 1996; Conrad, Brown, and Cashman, 1998.) suggest that when respondents are faced with a question asking about the frequency of a behavior, if that behavior is infrequent they attempt to count the instances; if it is frequent, they attempt to estimate. And when they count they tend to underreport, but when they estimate they tend to overreport. Perhaps if respondents' meta-cognition warns them that they may be underreporting when they count and overreporting when they estimate, then the intervals they supply in these two cases may differ in the direction of their asymmetry. Large frequencies likely to be overreported should produce intervals with left skew; small frequencies, likely to be underreported, should produce intervals with right skew. The non-choice, both-questions conditions should permit us to test this speculation with questions that deal with frequencies.

References

Conrad, Frederick G. and Brown, Norman R. (1994) Strategies for estimating category frequency: Effects of abstractness and distinctiveness. *Proceedings of the Section on Survey Research Methods.* Alexandria, VA: American Statistical Association, pp. 1345-1350.

_____ (1996) Estimating frequency: A multiple strategy approach. Pages 166-179 in Hermann, D. J., McEvoy, C., Hertzpg, C., Hertel, P. and Johnson, M.K. (Eds.) *Basic and Applied Memory Research: Practical Applications, Volume 2.* Mahwah, NJ: Lawrence Erlbaum Associates.

_____ and Cashman, Erin, R. (1998) Strategies for estimating behavioural frequency in survey interviews. *Memory, 6,* 339-366

Fienberg, Stephen E. and Tanur, Judith M. From the inside out and the outside in: Combining experimental and sampling structures. *The Canadian Journal of Statistics*, 1988, 16:135-151.

Kennickell, Arthur B. (1997) Using range techniques with CAPI in the 1995 Survey of Consumer Finances, Survey of Consumer Finances Working Paper, Jan. 1997, Board of Governors of the Federal Reserve System, Washington, DC 20551

Marquis, Kent H. and Press, S. James (1999) Cognitive design and Bayesian modeling of a Census survey of income recall. *Proceedings of the Federal Committee on Statistical Methodology Conference, Washington DC.* Nov. 16, 1999, pp. 51-64 (see http://bts.gov/fscm).

Press, S. James (2000) Respondent-generated intervals for recall in sample surveys, manuscript (see http://cnas.ucr.edu/~stat/press.htm).

Press, S. James and Marquis, Kent H. (2000) Bayesian estimation in a U.S. government survey of income using respondent-generated intervals. Submitted to Proceedings of the sixth world meeting of the International Society for Bayesian Analysis, May, 2000, Crete, Greece.

Press, S. James and Tanur, Judith M. (2000a) Experimenting with respondent-generated intervals in sample surveys, with discussion. Pages 1-18 in Monroe G. Sirken (ed.) *Survey Research at the Intersection of Statistics and Cognitive Psychology*, Working Paper Series #28, National Center for Health Statistics, U.S. Department of Health and Human Services, Center for Disease Control and Prevention.

Press, S. James and Tanur, Judith M. (2000b) Respondent-generated interval estimation to reduce item non-response. In *Proceedings of the International Conference on Applied Statistics*, Rider University, May 22-23, 1999. In press, Nova Scientific Publishers. (see http://cnas.ucr.edu/~stat/press.htm)

Tourangeau, Roger (2000) Discussion. Pages 47-51 in Monroe G. Sirken (ed.) *Survey Research at the Intersection of Statistics and Cognitive Psychology*, Working Paper Series #28, National Center for Health Statistics,
U.S. Department of Health and Human Services, Center for Disease Control and Prevention.

Tversky, Amos and Kahneman, Daniel (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.

FIGURE 1 – EXPERIMENTAL DESIGN

	No Choice of Questions	Choice of Questions	Only One Question
Interval Question First/Only	BOTH – INTERVAL FIRST Census experiment	EITHER – INTERVAL FIRST New in this experiment	INTERVAL ONLY New in this experiment
Point Question First/Only	BOTH – POINT FIRST Campus experiments, Census experiment	EITHER – POINT FIRST New in this experiment	Standard –CONTROL GROUP