PANEL CONDITIONING IN A RANDOM DIGIT DIAL SURVEY

Kevin Wang, Urban Institute, David Cantor, Westat, Adam Safir, Urban Institute
Kevin Wang, Urban Institute, 2100 M St. NW, Washington DC, 20037

Key Words: attrition, response bias, panel effects, National Survey of America's Families

The National Survey of America's Families (NSAF) is a survey of the well-being of children, adults under age 65 and their families with an emphasis on the low income population. Round 1 was done in 1997, while Round 2 was completed in 1999. As part of the Urban Institute's project Assessing the New Federalism, the purpose of the NSAF is to estimate changes in the well-being of children, adults and families as authority for social programs devolves from the federal to state and local governments.[1]

In both rounds, the NSAF sample consists of a random digit dial (RDD) sample of telephone households that is supplemented with an area probability sample of nontelephone households. The NSAF features large, state-representative samples in each of 13 states and the balance of the nation. Two-thirds of the telephone numbers that were used in Round 2 were also used in Round 1. The remaining one-third of the Round 2 sample was a new RDD sample drawn to represent the study population for the NSAF in 1999. Re-using numbers in Round 2 was done to increase the precision of estimates of change between the two rounds. For details on the NSAF sample design, see Judkins, et al. (in press).

This paper assesses whether re-using the same telephone numbers on the NSAF led to panel conditioning. In the first section we briefly review previous research on panel conditioning. The second section describes the NSAF and the analytic strategy used to assess panel conditioning. The third section presents the results of the analysis and the final section provides a summary of the results.

Previous Research

Panel conditioning refers to when an earlier interview influences how a respondent reports at subsequent

interviews. Conditioning can take several forms. At one extreme, an interview may influence the behavior of interest. For example, Traugott and Katosh (1979) report that pre-election surveys have a positive influence on turning out to vote in the election. A second example would be asking about participation in government programs at the first interview. This line of questioning might then motivate a respondent to actually apply to one of the programs mentioned during the interview. Other types of conditioning occur when exposure to the questionnaire influences how the respondent interprets and answers questions in subsequent administrations of the questionnaire. One such effect is <u>burden avoidance</u>, where respondents learn at the first questionnaire administration how to avoid being asked additional questions by answering key filter questions in particular ways. The result is to generally reduce the number of reports of phenomena of interest in later interviews (Bailar, 1975).

A second variant of this type of conditioning is when exposure to the questionnaire provides respondents with a better understanding of the intent of the survey. This, for example, is how some interpret the effects of "bounding" interviews on surveys that require retrospective recall (e.g., Biderman and Cantor, 1984). During the initial interview respondents are anxious to report phenomena of interest, but may not totally understand the need for precision with respect to placement in time. Once exposed to the questionnaire, where detailed questions are asked about events (e.g., who, what, where, when), the respondent is more aware of the precision needed and subsequently applies this at future administrations.

Both behavioral and conceptual conditioning are possible on the NSAF. Behavioral change may occur, for example, if respondents become aware of particular government programs through the questioning related to program participation. Respondent burden and increases in respondent understanding of survey concepts may be affected for a number of items on the NSAF.

When discussing panel conditioning, it is useful to make the distinction between items that ask about factual phenomena and those items that require subjective judgment. Factual items ask for estimates of numbers or well-defined statuses. This would include, for example, asking about having health insurance, income

received by family members or employment status. Items requiring subjective judgment are those that refer issues or concepts that are not as clearly defined. For the NSAF this might include assessments of personal health (e.g., excellent, good, poor, etc..), experiencing unmet medical needs or descriptions of a respondent's mood (e.g., "down-hearted and blue"). One might hypothesize that given the differences in how respondents interpret and report across factual and subjective items, the effects of panel conditioning may also differ. In fact, one might expect that factual items should be less subject to conditioning, since they may not be as subject to idiosyncratic interpretations of the item's scope, purpose and terminology.

The NSAF Design and Analytic Strategy

The NSAF interview ranges from 25 to 45 minutes, depending on whether the questions are intended to ask about adults or about children in the household. For households with children, up to two children were sampled, under the age of 6 and the other between the ages of 6 and 17. Interviews were conducted with the person in the household who knows the most about the health and well-being of the sampled children (Most Knowledgeable Adult or MKA). For households without children, up to two adults under age 65 were sampled for interviewing. Among the topics included in the interview are demographics, health status, health insurance and health care utilization, employment and earnings, income, welfare, program participation, child care arrangements, social service needs and child/family well-being measures.

A broad range of variables were selected to test for panel conditioning. The intent was to select variables that play a prominent role in NSAF analyses, as well as cutting across "factual" and "subjective" domains. The final list of variables included in the analysis are shown in Table 1. Factual variables include health insurance coverage and program participation. Measures that require more subjective judgment are health status, confidence in the medical care the family receives and the degree of school engagement of the child. The latter asked questions such as:

> .....please tell me if it describes (CHILD) all of the time, some of the time or none of the time....
>
> - cares about doing well in school
> - only works on schoolwork when forced to
> - does just enough schoolwork to get by

- always does homework

Each of these items requires respondents to make judgments about the child's performance in ways that can differ across respondents.[2]

The analysis used several different comparison groups to test for panel conditioning. As noted above, the Round 2 sample was divided into two parts. Two-thirds of the sample re-used telephone numbers included in the Round 1 sample. This is the group that is subject to panel conditioning, since respondents could have been interviewed in both rounds. The other third of the sample was drawn as a fresh RDD sample. Theoretically, the only difference between the two is that one was interviewed twice and the other was interviewed only once. However, in practice, this simple comparison is confounded because the overlap sample does not consist entirely of households that were interviewed during Round 1. Some portion of the Round 2 overlap sample were not interviewed during Round 1 because: 1) the household associated with the telephone number during Round 1 was not the same household associated with the number during Round 2, 2) the households were the same between rounds but different persons within the household were interviewed, 3) no household was associated with the telephone number during Round 1, and 4) the household did not respond during Round 1.

Given this problem, two separate "overlap" groups were analyzed. The "Overlap-Same Phone" group restricts the overlap sample to those persons that reported during Round 2 as having the same telephone number in 1997. This group is still subject to the problems noted above except it eliminates persons that could not have been called during Round 2 because they had a different telephone number.

The "Matched" group is restricted to just those persons identified as being subjects of interviews for Round 1. This identification was made by matching persons in the two data files by telephone number and demographic characteristics. Unlike the "Overlap-Same Phone" group, these respondents only include those in households that were interviewed in both rounds. However, this is a more restricted group than the "Overlap-Same Phone" group, since it only includes persons that were in the same household over the two rounds and cooperated with the survey at both times.

---

[2] This does not imply that measures using items such as these are not reliable. Ehrle and Moore (1999) report that these and similar items from the NSAF show high degrees of reliability and validity.

The analysis compared each of these groups to households that were part of the new RDD sample drawn for Round 2. An analysis group (the "New Sample-Same phone") was created by retaining from the entire new RDD sample only those persons in households which reported on the survey that they had the same phone number since 1997. Any differences between the "Matched" and "New Sample-Same Phone" groups or between the "Overlap-Same Phone" and "New Sample-Same Phone" groups are attributed to respondent conditioning.

Table 2 provides the total sample sizes for the child and adult samples for each of the different groups used in the analysis. The overlap adult sample includes 44,252 sampled adults (under the age of 65). The "Overlap-Same phone" group is created by retaining the 36,836 adults in households that reported having the same phone number in 1997. This is further reduced to 15,636 adults when restricting it to just those who were matched to sampled adults from Round 1. There are 27,057 adults in the entire new RDD sample. This drops to 19,451 in the "New Sample-Same Phone" group, when restricting it to the those reporting the same telephone number in 1997.

Use of two different groups in the comparisons provides for a way to bound the estimates of conditioning. Comparisons between the "Overlap-Same Phone" and the "New Sample-Same Phone" group should provide an underestimate of the presence of conditioning, since there are a number of persons in the "Overlap-Same Phone" group that had not been interviewed during Round 1. Comparisons between the "Matched" and "New Sample-Same Phone" groups has the potential of either under- or over-estimating the effects of conditioning, since there are likely to be differences that can be accounted for by the differential composition between the groups. If the effects of conditioning are in the same direction as the compositional differences (e.g., the "New Sample-Same Phone" group is both poorer and is more likely to report being without insurance), then the estimates of the effects of conditioning would be over-estimated. Underestimates would occur if the effects of conditioning are in the opposite direction of those implied by compositional differentials.

For space reasons, only the second set of comparisons are discussed below (i.e., "Matched" vs. "New Sample-Same Phone"). The analysis comparing the "Overlap-Same Phone" and "New Sample-Same Phone" groups resulted in fewer significant differences than those discussed below.

Results

Table 3 provides demographic comparisons between the entire new sample, the "New Sample-Same Phone" and "Matched" groups. This provides an idea of the magnitude and direction of the differences in composition between these samples. As expected, there are quite a few differences between the groups. The "Matched" group has fewer minorities, fewer married couples, fewer people working and significantly fewer persons that are under poverty. These differences indicate that the persons in the "New Sample-Same Phone" group tend to be more likely to be in lower socioeconomic status. Note also that the effect of removing from the entire new sample cases that did not report having the same phone number two years ago serves to make the comparison groups more similar to each other.

To initially test for conditioning effects, comparisons between the "New Sample-Same Phone" and "Matched" groups were made along the 32 variables listed in Table 1. The results of these comparisons found 7 to be statistically significant at the .05 level or less.[3] These, along with several other comparisons are shown in Table 4. The direction of these differences are such that the "New Sample-Same Phone" group reports lower levels of well-being than the "Matched" group. For example, for the adult sample, the "New Sample-Same Phone" has a higher percentage of persons that are not insured (13.4% vs. 10.9%) and are more likely to report food insecurity (17.8% vs. 13.8%). Similarly, for the child estimates, the "New Sample-Same Phone" group is more likely to receive food stamps, to not have a usual source of health care and to be uninsured.

Taken at face value this would imply that conditioning generally leads to people's reported well-being to improve at later interviews. However, given that the "New Sample-Same Phone" is of lower economic status, these results have to be viewed with some skepticism. Having more uninsured persons in the "New Sample-Same Phone" group may simply be because there are more poorer respondents in this group relative to the "Matched" group.

To test this idea, 7 logistic regressions were estimated that re-estimated the panel conditioning effect, once

---

[3] All estimates were produced using final adult and child weights. All test statistics and significance levels reported in this paper are based on variance estimates using a jackknife replication approach.

controlling for important socioeconomic and demographic characteristics. In these regressions, the dependent variables were the outcome variables in Table 4 where statistically significant differences were observed between the "Matched" and "New Sample-Same Phone" cases. Each variable was regressed on a set of independent variables including age, gender, employment status, family income as a percentage of the poverty threshold, marital status, region of the country, and race/ethnicity, as well as a dummy variable indicating whether a particular case was in the "Matched" or "New Sample-Same Phone" group. The coefficient on this dummy variable serves as an indicator of group differences on the variable of interest, after controlling for the other independent variables. These regressions found that three of the seven differences shown in Table 4 were not significant at the .05 level --- receipt of food stamps, participation in volunteer activities and insurance status. For the other four variables, however, the coefficient for sample group (panel conditioning) was statistically significant at the .05 level.

Discussion

This analysis provides very mixed evidence of panel conditioning for the NSAF. Of the 32 simple comparisons made, 7 were significant at the .05 level. Four of these differences remained significant (at the .05 level) while controlling for some compositional differences between the two groups.

The four variables that remained significant were: 1) child participation in extracurricular activities, 2) Adults reporting food insecurity, 3) confidence in getting medical care and 4) whether the respondent heard of Medicaid. These exhibit each type of conditioning discussed above. Child participation in extracurricular activities is, to a large degree, a factual question. Confidence about getting health care is a subjective question. Whether or not the respondent had heard about Medicaid is analogous to the behavioral effects discussed above. That is, the first interview may have informed the respondent about Medicaid so that at the second interview the respondent who might have answered "no" would answer "yes" at the second interview. The food security items are a bit more ambiguous with respect to the "factual" – "subjective" dimension. These questions asked respondents to agree or disagree with the following statements:

> We worried whether our food would run out before we got money to buy more.

> The food that we bought just didn't last and we didn't have enough money to get any more.

> In the last 12 months,...did you ever cut the size of your meals or skip meals because there wasn't enough money for food? How often did this happen?

The first statement refers to a subjective judgment (e.g., "..worried whether our food would run out..."). However, the other two statements are not as subjective, at least in the sense that they refer to relatively specific situations. Nonetheless, by asking these questions at the initial interview, it is possible that it provoked additional thought on the part of some respondents who then reconsidered their answers at the second interview. The topic is clearly a salient one and could be of some concern for persons who are borderline with respect to food security.

The limitations of this analysis preclude us from definitively concluding that the four variables discussed above exhibit a conditioning effect. There are too many differences between the two comparison groups that may not have been accounted for, such as variables more directly related to the items of interest or measures of the degree of effort it took to contact and interview the respondent. As noted above, the Matched group is of higher socio-economic status than the "New Sample-Same Phone" group. The differences shown above are in the direction one would expect from these compositional differences (i.e., the "Matched" group exhibits higher levels of well-being than the "New Sample-Same Phone" group). Consequently, there is still the possibility that the differences are due to compositional differences, rather than to conditioning. We attempted to mitigate this problem by estimating logistic regressions that control for socioeconomic and demographic variables. This eliminated three of the comparisons that were significant at a bivariate level. However, it is unclear if these controls were sufficient to fully equalize the two groups.

With only four differences remaining after controlling for socio-economic characteristics, one has to wonder whether these significant effects appeared by chance, given the large number of comparisons made (32 total). Even if one assumes that there is a conditioning effect, the size of this effect appears to be relatively small, especially once controlling for the compositional differences between the two groups.

825

The negligible effect of any panel conditioning applies to a greater extent if one were to assess how it may be affecting the overall estimates produced by the Round 2 of the NSAF. This is not only because any effect of conditioning appears to be relatively small, but also because of the relatively small number of persons who were interviewed two times. As a way of illustrating this, we calculated statistics for the entire NSAF Round 2 sample assuming that the persons with a previous interview responded in the same way as those that did not have a previous interview. This actually overestimates the effects of conditioning on overall estimates, since it does not account for socio-demographic differences between the Matchers and the rest of the overlap sample. Even with this assumption, however, the change in the overall statistics are relatively small. For example, using this assumption, the percentage of adults who report food insecurity only increases from 12.3% to 12.6%.

In conclusion, this analysis found some evidence of conditioning for 4 of the 32 variables tested. There is some question whether the effects observed are due to design related problems or actual conditioning. Regardless of assumptions, the size of the effects are relatively small and do not appear to have much consequence on NSAF estimates of change.

References

Bailar, B.A. (1975) The Effects of Rotation Group Bias on Estimation from Panel Surveys. *Journal of the American Statistical Association,* 70, 23-30.

Biderman, A.D. and Cantor, D. (1984) A Longitudinal Analysis of Bounding, Respondent Conditioning, and Mobility as Sources of Panel Bias in the National Crime Survey, *Proceedings of the Section on Survey Research Methods of the American Statistical Association.*

Ehrle, J. and Moore, K.A. (1999) 1997 NSAF Benchmarking Measures of Child and Family Well-Being. Report No. 6, *NSAF Methodology Reports.* The Urban Institute, Washington, DC.

Judkins, D., Brick, J.M., Broene, P., Ferraro, D. and Strickler, T. (in press) 1999 NSAF Sample Design, Report No. 2, *NSAF Methodology Reports.* The Urban Institute, Washington, DC.

Traugott, M.W. and J.P. Katosh (1979) Response Validity in Surveys of Voting Behavior, *Public Opinion Quarterly,* 43, 359-377.

Table 1. Variables Compared in the Analysis of Panel Conditioning

Adults
Received AFDC in 1998
Received Food Stamps in 1998
No usual source of care
Unmet need, dental care
Unmet need, medical care
Unmet need, prescription drugs
Currently uninsured
Good, Very Good or Excellent Health status
Food Insecurity
Difficulties paying rent, mortgage, housing
Currently working full-time
Children
Received AFDC in 1998
Received Food Stamps in 1998
Parent asked about Medicaid
Applied for Medicaid
Parent asked about govt. child care assistance
No usual source of care
Unmet need, dental care
Unmet need, medical care
Unmet need, prescription drugs
Currently uninsured
Participates in extracurricular activities
Negative behavior problems
Negative behavior problems (6-11)
Negative behavior problems (12-17)
Positive school engagement
Most Knowledgeable Adult (MKA)
Volunteer participation–few times a month or more
Religious participation–few times a month or more
Negative (poor) mental health
High parent aggravation score
Not confident of family getting medical care
Heard of Medicaid
Heard of CHIP (Children's Health Insurance Program)

826

Table 2. Child and Adult Sample Sizes, by Sample Group

	Total	Same Phone Number 2 Years Ago	Matched Persons
Children			
Overlap sample	21,111	17,311	7,829
New sample	13,720	9,680	NA
Adults			
Overlap sample	44,252	36,836	15,636
New sample	27,057	19,451	NA

Table 3. Demographic Characteristics, by Sample Group

	Entire New Sample %	New Sample Same Phone %	Matched Persons %
Adults			
Hispanic	10.0	9.3	8.9
White	83.7	84.9	86.7
Married	59.0	61.4	66.9
Born outside U.S.	11.2	10.1	9.0
Working	77.9	78.3	79.0
Less than high school	13.1	12.4	11.0
Income below poverty	9.9	8.2	5.9
Children			
Hispanic	13.6	12.0	13.8
White	78.0	80.3	84.6
Single parent	23.8	20.9	17.3
Income below poverty	15.0	12.0	9.0

Table 4. Statistically Significant Differences Between New and Matched Sample

	New Sample (same phone) %	Matched Persons %	Chi-square (p-value)
Adults			
Currently uninsured	13.4	10.9	6.81 (p = .009)
With food insecurity	17.8	13.8	18.91 (p = .000)
With difficulties paying rent	9.5	8.3	1.97 (p = .161)
Children			
Received Food Stamps	9.7	7.7	3.88 (p = .049)
Positive extra-curricular activity	85.3	88.0	6.17 (p = .013)
No usual source of health care	5.7	4.3	3.72 (p = .054)
Currently uninsured	9.5	8.2	2.55 (p = .110)
MKA			
Volunteered few times a month or more	37.9	42.0	5.25 (p = .022)
Not confident getting medical care	8.0	6.1	8.60 (p = .003)
Heard of Medicaid	88.2	90.9	9.65 (p = .002)