Answers to Questions Asked During the Webinar

Provided by: James Wagner, Ph.D.
University of Michigan
Methods: Mind the Gap
April 15, 2020

1. **What is the study showing that more effort is needed and even more effort does not overcome the non-response?**

I have several relevant citations. The first three show that more effort has been used by surveys and that response rates continued to decline despite this increased effort.

- Williams D, Brick JM. Trends in U.S. face-to-face household survey nonresponse and level of effort. Journal of Survey Statistics and Methodology. 2017;6(2):186–211.
- Curtin R, Presser S, Singer E. The effects of response rate changes on the index of consumer sentiment. Public Opinion Quarterly. 2000;64(4):413–428. doi: 10.1086/318638.
- Curtin R, Presser S, Singer E. Changes in telephone survey nonresponse over the past quarter century. Public Opinion Quarterly. 2005;69(1):87–98.

Also of interest is the following citation which looks at the question of whether more effort is actually effective at limiting non-response bias. Or, as I alluded to in the presentation, does it tend to recruit "more of the same" type of respondent?

- Peytchev A, Baxter RK, Carley-Baxter LR. Not all survey effort is equal: Reduction of nonresponse bias and nonresponse error. Public Opinion Quarterly. 2009;73(4):785–806.

2. **What software recommendations do you have for collecting paradata, specifically key stroke data?**

This is a hard question to answer. I'll break it into two parts.

The first part will be a little about commercial software. There are two kinds of systems that collect paradata. The first type is a sample management system. Many large survey organizations develop their own software for sample management. The U.S. Census Bureau, for example, has developed a Contact History Instrument that collects paradata about interviewer-administered surveys. These systems are designed to make the workflow of interviewers and analysts be as seamless and simple as possible. A second type of paradata comes from the survey instrument administration software. There are commercial firms that specialize in these systems—Sawtooth and Blaise are two examples. These systems can provide keystroke audit trails that illuminate problems within a questionnaire. There are also commercially available web survey software solutions that combine some aspects of these two systems into a single software (e.g., SurveyMonkey, Qualtrics, Voxco). These systems allow the user to track the sample and also program and administer the questionnaire.

The second part of my answer has to do with situations that don't have the resources to invest in developing or purchasing expensive software. I think it is still possible to make an effort to track production in the field using relatively simple tools. Popular spreadsheet and database software can be used to input and track data. These data can then be summarized for use as inputs to decision-making.

3. **What is the recommended minimum sample size?**

I have a friend who jokes that the minimum sample size is two. That's because that is the minimum required to estimate variance. But seriously, I would say that there is no general recommendation on minimum sample sizes. I do think it is important to engage with the question of power during the study design stage. That is, it is important to ensure that a sample size is sufficient for the stated purpose.

4. **How do you parse what differences in your outcome measurements come from mode effects versus differences between subpopulations that respond via different modes, or response bias? Is it possible?**

That's a great question. In cases where you are considering multiple modes, it is very important to have some sense of possible mode effects. Are there sensitive questions? Is there published literature indicating that these questions are influenced by the mode in which they are asked? If so, it may require careful planning in order to mitigate those effects. One option is to not use modes that induce measurement error.

Having said that, a hot topic in survey methodology has been to focus on the question of separating measurement error from non-response error in situations where the two are confounded. I won't say much as the topic is highly statistical, but I will point you to an excellent review of several options:

- Kolenikov S, Kennedy C. Evaluating three approaches to statistically adjust for mode effects. Journal of Survey Statistics and Methodology. 2014;2(2):126–158.

5. **Am I correct in my understanding that adaptation can be based on the overall survey (phase capacity) and also on individual participant behavior (non-response after X number of days)?**

Yes, generally we choose one or the other. That is, we can define phase boundaries as a point in time. All cases cross that point in time at the same time. The alternative is to define a phase boundary as an event that can happen to each case at a different time. For example, if we attempt a case eight times and it still hasn't responded, then we might say that case has crossed the phase boundary and is in a new phase. Cases can cross this kind of boundary at different times, depending upon the pace at which they are worked.

6. **I can see how this responsive survey design could be important to reaching under-represented populations. Do you have recommendations for getting started with these methods for small organizations or community organizations?**

Yes, my first recommendation is to start with small, relatively simple actions that are easy to implement. This accomplishes two things. First, it generally requires some small steps in training staff and developing technical systems (the former might be a review of the slides for this talk or a webinar on the topic; the latter might involve designing a spreadsheet or database to track cases in the study). Assuming the experience is productive, then this can lead to further steps in this direction. Second, it gives folks some experience implementing a procedure following the steps outlined in the talk.

7. **Is there any precedent to offering different incentives among the phases or data collection methods? For example, if web surveys are more efficient, offering a larger incentive for web participants versus in-person in [the] hopes that more people will select the web option versus in-person?**

Yes, there is precedent for this. I would recommend this general introduction to incentives:

- Singer E, Ye C. The use and effects of incentives in surveys. The ANNALS of the American Academy of Political and Social Science. 2013;645(1):112–141.

I work on a study that uses different incentives across the phases. The survey is the National Survey of Family Growth. You can read about the details at the survey's website. There are several papers on how this may improve data quality. I include two citations:

- Axinn WG, Link CF, Groves RM. Responsive survey design, demographic data collection, and models of demographic behavior. Demography. 2011;48(3):1127–1149. doi: 10.1007/s13524-011-0044-1.
- Peytchev A, Peytcheva E, Groves RM. Measurement error, unit nonresponse, and self-reports of abortion experiences. Public Opinion Quarterly. 2010;74(2):319–327.

Finally, one survey recently offered higher incentives for web completion as a way to encourage that form of participation. Web responses saved the survey firm the cost of data entry. They said that they were essentially passing the savings on to the respondent.

- Biemer PP, Murphy J, Zimmer S, et al. Using bonus monetary incentives to encourage web response in mixed-mode household surveys. Journal of Survey Statistics and Methodology. 2018;6(2):240–261.

8. Do you end up using complex weights related to phases and modes, etc. than [what] might have been common 20 years ago?

In my area, complex weighting has always been a part of surveys. Most of the surveys that I work on select a person within each sampled household. That results in differential weighting between persons from small and large households. Further, today, most surveys use non-response adjustments based on some projection to the population. These adjustments further "complexify" weights.

In responsive survey design, we often use sampling between phases as a way to save costs. These savings have to generate efficiencies, or else it is not worth the complexity and loss in efficiency due to differential weighting.

9. In your experience, have institutional review boards (IRBs) been flexible with the flexible approach? Can you share any best practices in working with IRBs and minimizing protocol modifications along the way?

Yes, we normally describe our protocol along with any rule-based interventions that are made to data collection. Or, we describe our protocol flexibly enough that we can include several options. For example, we may allocate a different number of attempts we will make on different subgroups in order to place more attempts on groups that are less likely to respond. Our protocol might say "we will make up to eight attempts on each case." And then we [would] describe how we plan to allocate attempts such that low responding groups receive more of them. In general, I find that IRBs will work with us on these protocols, and the prespecification part really helps, as it is very difficult to change course midstream if we don't have an IRB approval for that.