Response Rate Projections for Household Screeners vs. Questionnaires: Can the Same Model Be Used for Both?

Katie Archambeau[1], Rebecca Curtis[1], Ned English[1], and Colm O'Muircheartaigh[1,2]

[1]NORC at the University of Chicago, 55 E. Monroe St. Ste. 3100, Chicago, IL 60603
[2]University of Chicago Harris School of Public Policy, 1155 E 60th Street, Chicago, IL 60637

Abstract
In the past, survey completion rates were assumed to be relatively steady across the data collection period. That assumption was soon modified to a simplistic curve, which assumed that completion rates would peak several weeks into production, then taper off until the end of data collection. Over the past few years researchers at NORC have employed field disposition histories for multiple projects in order to more accurately predict final response rates that can consider mid-project fluctuations in productivity. The National Social Life, Health, and Aging Project (NSHAP) does this by monitoring case dispositions, including whether there has been a refusal, and comparing these cases to the final response status of cases with similar weekly dispositions in previous studies. This application is used to model final response rates for the current study, permitting more informed case releases and early warning of potential production shortfalls. In this paper, we examine whether these questionnaire response rate predictions can be applied to household screening.

This research reports on the use of response rate projections for NSHAP Wave 3, which includes both household screening and a questionnaire. We will look to see how well the model performs in predicting household screener response rates at early points in data collection. Does the model need to be tweaked when moving between household screening and questionnaire completion projections? If adjustments need to be made, how is the screening model similar to the questionnaire model, and where are the differences most pronounced? The NSHAP data will serve as our guide to understanding similarities and differences between screener and questionnaire response, and we will use this information to inform the audience on how to best utilize this information on future projects. Our research has value for other field data collection projects that would benefit from early indicators of ultimate production and response rates.

Key Words: response rates, projections, monitoring, screening, indicators

1. Introduction

The National Social Life, Health, and Aging Project (NSHAP) is a longitudinal study designed to understand the roles that social support and personal relationships play in health and aging (O'Muircheartaigh et al. 2014). Led by a team of principal investigators at the University of Chicago and funded by multiple grants from the National Institute on Aging (NIA), NSHAP is a nationally representative sample of more than 3,000 older adults and their spouses/partners who are interviewed every five years. The most recent round of NSHAP returned to interview baseline cohort members, now ages 67 through 95, plus any spouses/partners from Wave 2, for the third wave of data collection. Original cohort members, frequently referred to as Returning Respondents, were eligible for Wave 3 if they participated in any of the prior waves. Because of the aging nature of this cohort, investigators added a new baby boomer cohort in Wave 3. The baby boomer cohort is often referenced as the New Cohort, and it contains individuals aged 50 through 67. New Cohort respondents were recruited via a "Screen and Go" method, where interviewers would roster households to identify eligible respondents and their spouses/partners during a brief screening process and then have the capability to immediately go into the full interview with any eligible respondents.

NSHAP employs an in-person data collection effort, a practice with which NORC has a rich history with flagship studies including the General Social Survey (GSS) and Survey of Consumer Finances (SCF). Project management is mindful of the fact that a major objective of any project with a field component is achieving a target response rate at or under budget. In order to meet this objective, early warning of the ultimate project outcome is critical. In the last decade, Eckman and O'Muircheartaigh (2008) began to develop response rate prediction models that would harness historic patterns in field outcomes from generally related and unrelated projects in an effort to refine predictive capabilities. Originally developed for the Making Connections project for the Annie E Casey Foundation in 2003, and modified for Wave 1 of NSHAP (2005) and the General Social Survey (GSS), the model has been modified over time and has proven effective in providing project management with projections on final response rate outcomes at early stages of data collection. NSHAP has successfully used this model in projecting response rate outcomes for the interviews of Returning Respondents, but it has been uncertain whether or not the same model could be used in providing estimates for screener completion rates. Our paper focuses on how one can successfully predict outcomes for screeners based on past experience.

2. Data and Methods

The objective of the response rate model is to start with imperfect approximations, based on some data and the judgment of field operations staff, and use them to develop a model to predict ultimate field performance. At the end of each project, we can improve the model by comparing projected and actual outcomes. The model has the ability to be refined according to special needs and characteristics of a given project.

In order to understand the model as it relates specifically to screening, one must first understand the general structure of the model. At its core, projections are based on the record of call history, which is an event level file noting the outcome of each attempted contact with a study participant. Each time an event occurs related to a specific case, the field interviewer logs a call record, which contains a disposition to indicate the result of the contact event. We group the call history dispositions into the following categories: Virgin (never been worked), No Contact (such as "no answer at door"), Refusal, Other Non-Interviewed Respondent (NIR) (such as "unavailable during field period"), Appointment, Complete, and Out of Scope (such as "language barrier"). These categories are placed in a hierarchy, as seen in Figure 1. Once a case enters a "higher" disposition class, it cannot revert to a previous disposition. An example of a potential call history for a specific case can be found in Figure 2. This example illustrates that a case never falls back in disposition categories; once the case reached Appointment status in contact attempt four, the max disposition never fell below 500, even when there were Refusal and No Contact attempts in later visits.

MAX Disposition	Category	Components
100	Virgin	Virgin
200	No Contact	Mover/No Contact
300	Refusal	Refused (Initial or Final)
400	Other NIR	Other (Initial or Final)
500	Appointment	Appointment
600	Complete	Complete/Partial
700	Out of Scope	Deceased/Temp./Permanent

Figure 1: Record of Call Disposition Categories

Case ID	Attempt	Week	Outcome	MAX Disp.
154	1	1	No Contact	200
154	2	1	Refusal	300
154	3	3	No Contact	300
154	4	3	No Contact	300
154	5	4	Appointment	500
154	6	6	Refusal	500
154	7	7	Refusal	500
154	8	8	No Contact	500
154	9	8	No Contact	500
154	10	9	Complete	600

Figure 2: Example Call History Log

Given complete call history records, one can produce a matrix containing probabilities of success based on prior dispositions. Each case is also flagged to indicate whether or not there is a refusal in the case history. This matrix provides week-by-week probabilities that a case sitting in a given disposition class will end up completing the survey by the end of data collection. Figure 3 illustrates an extract of four weeks in the matrix. Using these figures, one can see that in week two of data collection, a case that has never made contact with a respondent has a 60% chance of completion. In week eight, for example, a case with an appointment as its max disposition, with no refusals in the case history, is projected to complete 88% of the time; a week eight appointment with one or more refusals in the case history completes about 68% of the time. As a final example, a case with a refusal as the highest disposition in week 22, but with no previous refusals (so, the only refusal associated with the case was the most recent contact) will complete 25% of the time, whereas a refusal case with multiple refusals in the case history at this point in time will likely complete 27% of the time. We are able to use the record of call outcomes from the previous round of data collection to inform the current round.

		Week 2		Week 8		Week 16		Week 22	
		NR	R	NR	R	NR	R	NR	R
Highest Disposition	No Contact	0.6	n/a	0.6	n/a	0.5	n/a	0.27	n/a
	Refusal	0.65	0.59	0.61	0.47	0.31	0.37	0.25	0.27
	Appointment	0.97	n/a	0.88	0.68	0.74	0.51	0.53	0.35
	Complete	1.00	n/a	1.00	1.00	1.00	1.00	1.00	1.00

*NR = No Refusal in History
R = Refusal in History

Figure 3: Probabilities of Success Matrix

3. Results and Discussion

As previously mentioned, the model described here has proven very successful in projecting final response rates early in data collection. Before testing the model on the NSHAP screener, we applied the model to our Returning Respondent panelists in Wave 3 of data collection. Figure 4 shows the actual and project response rates for the Returning Respondent Cohort.

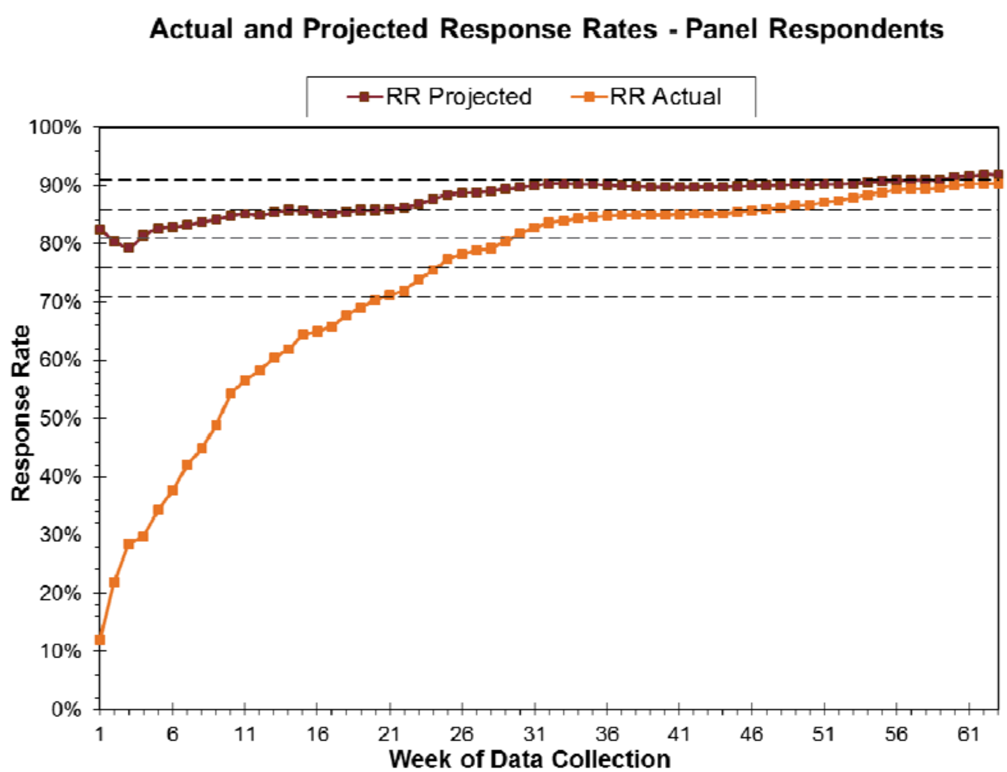

Figure 4: NSHAP Returning Respondent Projections

The bottom line depicts the actual response rate by week, which begins at approximately 12% in week one and gradually climbs to the final response rate of 91% in week 63. The upper line tracks the projected final response rate by week, using the matrix probabilities applied to each case. It can be noted that the projected response rate takes several weeks to level off as there is known volatility in the first few weeks of data collection, due to factors such as interviewers getting familiar with the project and the staggered nature of interviewer trainings. However, by about week 10, which is quite early in the data collection period, our final response rate estimates had already leveled off around 85%. One will notice a jump in projected final response rate around week 22, which is when the project implemented a substantial increase in incentives. Soon after this increase, the projected final response rate leveled off around 90%, where it stayed until the final weeks of data collection. The last slight bump in projections came around week 56, when we executed the final incentive increase. Overall, Figure 4 illustrates just how successful the model was in projecting final response rates for the Returning Respondent cohort interview, as it was able to use outcomes from NSHAP Wave 2 in projecting outcomes for Wave 3.

Given the success of the model when applied to interview response rates, we wanted to see if the model would have similar utility when applied to screener response. NSHAP had impressive goals of screener rates exceeding 90%, so it was important to know early on if this goal was realistic. Because we had limited call history data relating specifically to screeners, we decided to use the same questionnaire response rate matrix and apply it to NSHAP screening. Figure 5 shows the result of attempts to project final screener response rates using this model.

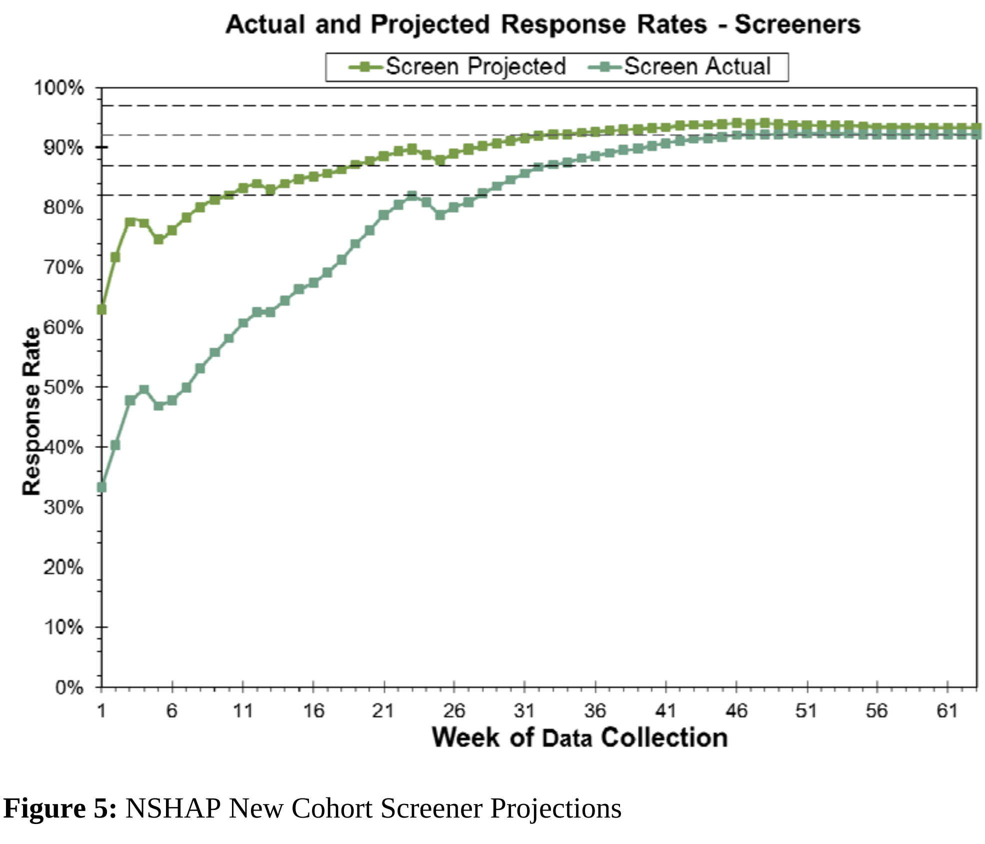

Figure 5: NSHAP New Cohort Screener Projections

As in Figure 4, the lower line represents the actual screener response rate over time, and the upper line represents projected final screener response rate by week. However, unlike the Returning Respondent Projections, this model does not have a good fit. Final response rate projections did not level off until around week 30, which is far too late in the data collection period to use this projection to make informed decisions. The final response rate projection appears to mirror actual response rates instead of the desired outcome of a relatively flat horizontal projection. Weeks 12, 20, and 37 saw screener incentive increases, but we did not see a direct link between incentive escalation and projected response rate, as we did in the questionnaire model.

The charts in Figures 6 and 7 display an attempt to better understand why the same model cannot be used for questionnaires and screeners. Both figures are visual representations of the weekly matrix probabilities, with the top line showing Screener Probabilities fitted to the current round, and the bottom line showing the Interview Probabilities from the previous round. The No Contact probabilities (Figure 6) shows that, in cases where No Contact is the highest disposition achieved, screener cases retain a higher probability of completion for a much longer period of time. Likewise, the Refusal probabilities (Figure 7) reveal a similar story, where screener cases with a first refusal as the highest disposition keep much higher probabilities of completion late in the data collection period than do interview cases.

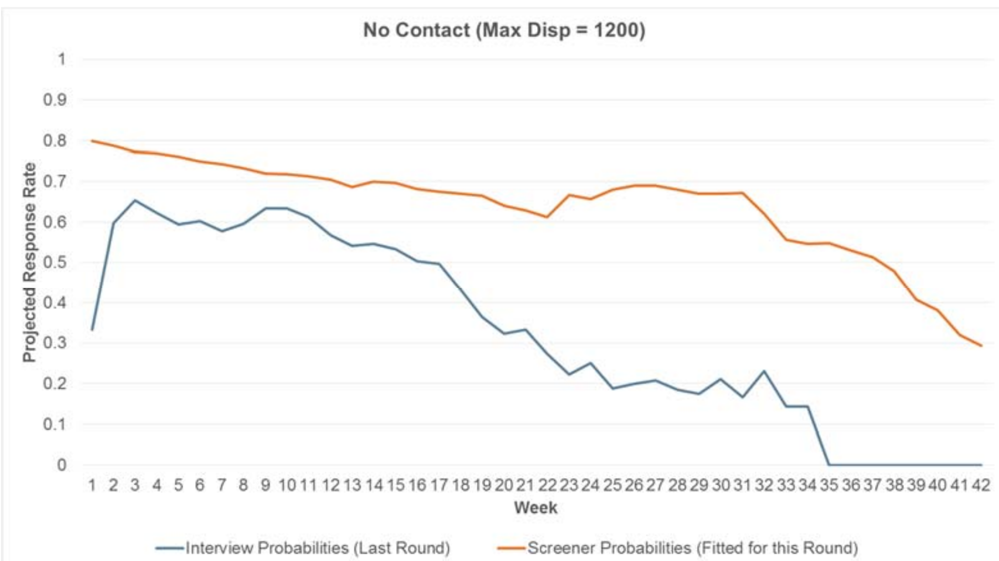

Figure 6: NSHAP No Contact Probabilities

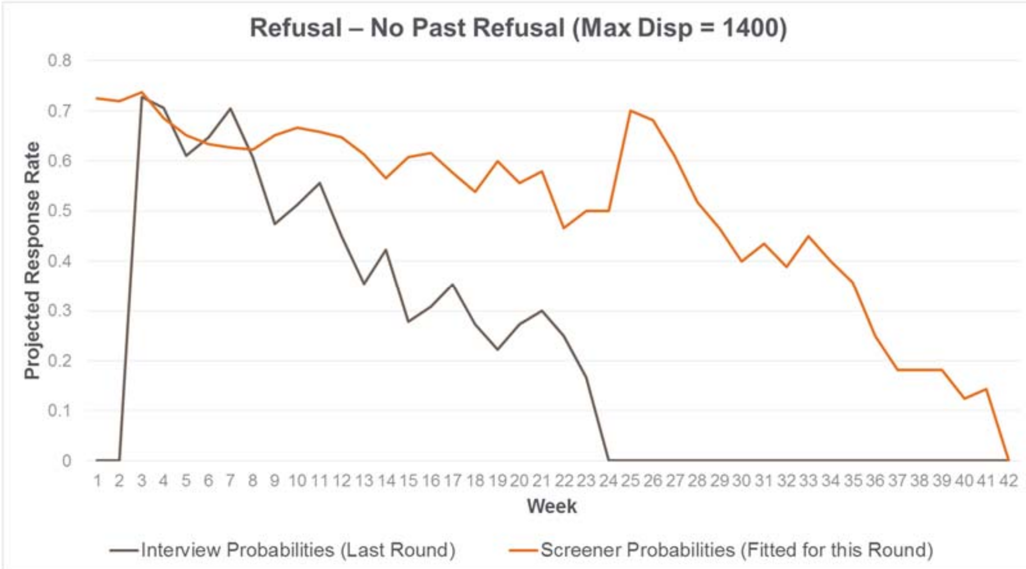

Figure 7: NSHAP Refusal Probabilities

3. Conclusions and Next Steps

NSHAP's attempt to apply the interview response rate probability model to screener response rate projections was not effective as-is. However, we learned important information about how interviewers work household screeners as distinct from member-level interviews, and this information can be applied when adjusting the model for screener use in the future. Analysis of the actual screener probabilities, fitted in retrospect for the current round, show that field staff appear to work interview cases quickly and diligently once they learn that a respondent is eligible. However, interviewers may take a more "drawn out" approach related to screeners, spreading contacts out over a longer period of time and giving potential respondents the opportunity to change their minds after a refusal. In response to learning about interviewer behavior, some simple changes to the probability matrix would likely yield much more accurate results. First, we recommend adjusting the screener probability matrix to have higher overall completion rates than the interview. Second, the screener probabilities should remain at an elevated level for a longer period of time than we see in the interview matrix.

With an updated screener probability matrix we could take future steps of interest, such as converting it to a set of probabilities or to confidence intervals. For example, we may be able to make projections such as: if the current response rate projection is 75%, what is the probability that our final response rate will be 70% or higher? Information like this would be extremely beneficial to project management staff. Similarly, using the probability matrix, we can implement exercises such as the following:

If 50% of appointments in week 22 convert to completes (on average), what is the spread of our prediction from our 200 appointments?

- 95% chance we will get between 84 and 116 interviews
- 10% chance we will get fewer than 91 interviews

As we use this model on an increasing number of projects, we can expect our ability to make predictions increase. Additional uses for this model include computing projections for metrics such as number of case completes, hours per case, criteria for cutting back on effort, and threshold for reducing or increasing sample size.

In summary, the response rate prediction model shows that we can obtain reliable early warning of challenges with field performance. However, response behavior is different for screener verses questionnaire completion, so different matrices should be used for each type of monitoring. We do have the ability to test and monitor alternative models in real time without waiting until project end, as well as designing accounting and monitoring tools to improve our knowledge base for continuous project improvement.

References

Eckman, Stephanie and Colm O'Muircheartaigh. "Using Response Rate Prediction to Tailor Contact Strategies in a Face-to-Face Survey", American Association for Public Opinion Research Conference, New Orleans, LA. May 2008.

O'Muircheartaigh, Colm, Ned English, Steven Pedlow, and Peter K. Kwok. Sample Design, Sample Augmentation, and Estimation for Wave 2 of the NSHAP. Journal of Gerontology 2014 69B:8.