

Using Data Analytics for Early Prediction of Response Rate Changes in GSS

Colm O'Muircheartaigh¹, Holly Hagerty², Chang Zhao², Ned English²

¹University of Chicago and NORC at the University of Chicago; ²NORC at the University of Chicago

Abstract

For the past 15 years, researchers at NORC have been predicting final response rates for face-to-face studies by building a model based on detailed field disposition histories from previous projects. In addition to providing the overall prediction of response rate, the model permits more informed case releases, early warning of potential production shortfalls, and the potential to test remedies in real time. Projects have included Making Connections [Annie E Casey Foundation]; the National Social Life, Health, and Aging Project [National Institute on Aging], the General Social Survey [GSS; National Science Foundation], the Survey of Consumer Finances [Federal Reserve] and the National Longitudinal Survey of Youth [Bureau of Labor Statistics]. The GSS has experienced a significant decline in response rates since the 2014 round, paralleling the general sectoral decline in response rates. We demonstrate the use of the model to predict the response rate for the 2018 round of GSS. We show the strengths and weaknesses of the approach during the fieldwork period. Our research is relevant both to those who would benefit from early warnings of response rate issues in field surveys and to those who wish to test and assess remedies in the field.

Key Words: response rate prediction, GSS, disposition histories, predictive analytics

1. Introduction

1.1 Concern with the implications of increasing and unpredictable survey nonresponse

There has been a secular decrease in response rates for surveys in general, and for face-to-face surveys in particular, over the past fifty years (Groves et al., 2009). This constitutes both a business problem and a science problem for survey research. From a business perspective, early warning of the ultimate outcome of the survey is crucial. A major project objective is to achieve a target response rate at or under budget; many contracts require advance notification to the client (in particular the US Government) when targets may not be met. It has traditionally been extremely difficult to predict the final level of field success until the late stages of fieldwork. A tool is needed that can provide this predictive capacity, one that will provide early and, as the survey proceeds, increasingly robust estimates of outcome.

1.2 The NORC Response Rate Prediction Model

Many projects at NORC have used predictive analytics to provide early projections of response rates for face-to-face surveys, including MC, NSHAP, GSS, SCF, and NLSY. The approach requires the following steps:

- Identify a completed survey for which the target population and the field burden are similar to the survey currently being executed; this is the *base survey*.
- Access the full record of calls made in the base survey, including all contact attempts and their outcomes.
- At each time point in the base survey – typically each week – the current status of each case is determined (the *interim status*).
- The final outcome for each case is classified as *Complete (C)*, *Out of Scope (OOS; not eligible)*, or *Final Nonresponse (NIR)*.
- For each interim status (defined by week and category) the observed number of outcomes of each type [C_b , OOS_b , and NIR_b] is calculated. From these, yield rates for each class are calculated: completion yield rate, out of scope yield rate, and nonresponse yield rate. These are the *observed yield rates* for each interim status for the base survey.
- Each week the predicted outcomes for the *target survey* [C_t , OOS_t , and NIR_t] are calculated by multiplying the frequencies in each cell for the week by the observed yield rates in the base survey.
- From these predicted outcomes, two key measures can be calculated: the predicted number of completes C_t and the predicted response rate RR_t [$C_t / (C_t + NIR_t)$].

1.3 The GSS

Since 1972, the General Social Survey (GSS) has studied the growing complexity of American society. It is the only full-probability, personal-interview survey designed to monitor changes in both social characteristics and attitudes currently being conducted in the United States.

1.3.1 The GSS Sample Release: Phase I, Phase II and Phase III

A two-phase sampling approach for GSS, with subsampling out of cases at the midpoint of the fieldwork, was introduced in 2004. In Phase I, a larger sample of potential respondents was selected than would be necessary to generate the required number of completes. At the midpoint of the field period, a subsample of the non-completed cases was removed from the field, and the remaining cases were worked more intensively; this is Phase II. This process was intended to provide efficiency gains by retaining the more competent field staff to work the more difficult cases that remained after the first phase. Typically, about 40% of the non-completed cases were retained in the second phase.

The two-phase approach prioritizes number of completes over effective sample size. The weighted response rate is unaffected in principle by the subsampling, though additional resources, high-performing interviewers, and refusal conversion efforts can be focused on remaining subsample and may increase the weighted response rate.

In 2016 and 2018 an additional subsample (already worked in Phase I but not Phase II) was released later in the fieldwork. This is referred to as Phase III.

2. Methods

2.1 Transition Matrix Creation

Using the base survey, we define the transition matrix as the set of yield rates calculated for each category of case at a given date in the fieldwork (the time point is defined in terms of the number of weeks since the beginning of fieldwork). Thus if we have six classification cells and 40 weeks of data collection in the base survey, we will have a 40 x 6 matrix of transition probabilities. For example, for each cell in the transition matrix, the probability that a case will become a final complete is determined by the rate at which such cases become final completes in the base survey:

$$Rate_Complete_{abcdef} = \frac{\text{number of eventual complete}_{abcdef}}{\text{number of cases}_{abcdef}}$$

Where:

- (a): Project Week of data collection
- (b): Case Week of data collection
- (c): Current highest disposition
- (d): Ever have a refusal or not
- (e): Have more than 2 refusals or not
- (f): Have more than 10 attempts or not

A similar calculation is carried out for OOS (out of scope) and NIR (nonresponse) cases in the base survey.

2.2 Response Rate Prediction in the Target Survey

In each week of the target survey, the rates calculated above for the base survey are applied to the frequencies that are observed in the target survey. The predicted numbers of completes (C_t) and nonrespondents (NIR_t) in the target survey are calculated directly. The predicted response rate is calculated as $RR_t = 100[C_t / (C_t + NIR_t)]$

3. Results

3.1 2018 GSS Phase I Projection

For this prediction, the base survey is GSS 2016, and the target survey is GSS 2018. This is the ideal situation in which to use the model, as (i) the sample design is the same for the two years; (ii) the survey questionnaire is also very similar from year to year; (iii) NORC interviewers carried out the fieldwork in both cases. This maximizes the appropriateness (and thus portability) of the data from the base survey for the target survey.

Figure 1 shows the predicted number of completes (upper line) and the actual number of completes (lower line) for each week of the ten weeks of Phase I in the 2018 GSS. During this ten week period, all of the cases selected into the sample are included in the field work; no subsampling occurs until after this phase. The prediction is excellent. The projection is

very stable¹, and by week 5, the projected number of completes is within 5 percentage points of the actual number at the end of phase I. The predicted (and achieved) response rates for Phase I differ substantially between 2016 (not shown) and 2018. The Phase I response rate was 32% in 2016, while it was 40% in 2018. The most likely explanation for this is in the change in field strategy in 2018. In 2018, incentives were used extensively from the beginning of the fieldwork whereas in 2016 incentives were held back until Phase II. It is notable that despite this, the model produces a very good prediction of the response rate.

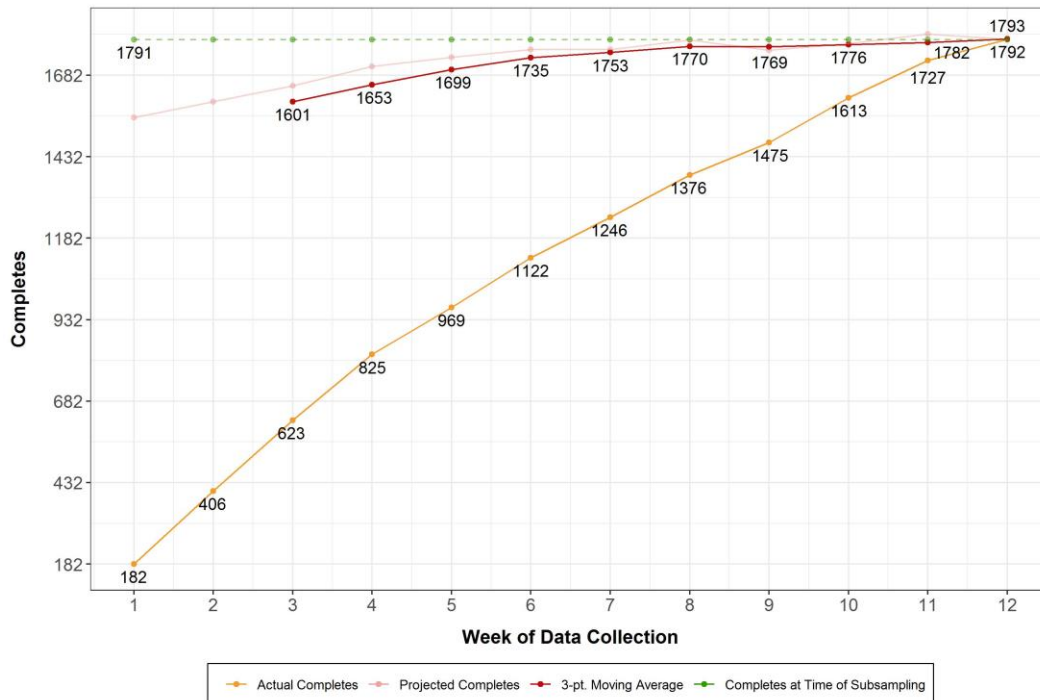


Figure 1: 2018 GSS Phase I: actual and projected completes based on 2016 GSS Phase I case week matrix

¹ In each diagram, we present also the graph of the predicted values using a three-point moving average. This is an approach that may be desirable when the prediction is based on small numbers of cases, but does not have any significant impact in the current analyses.

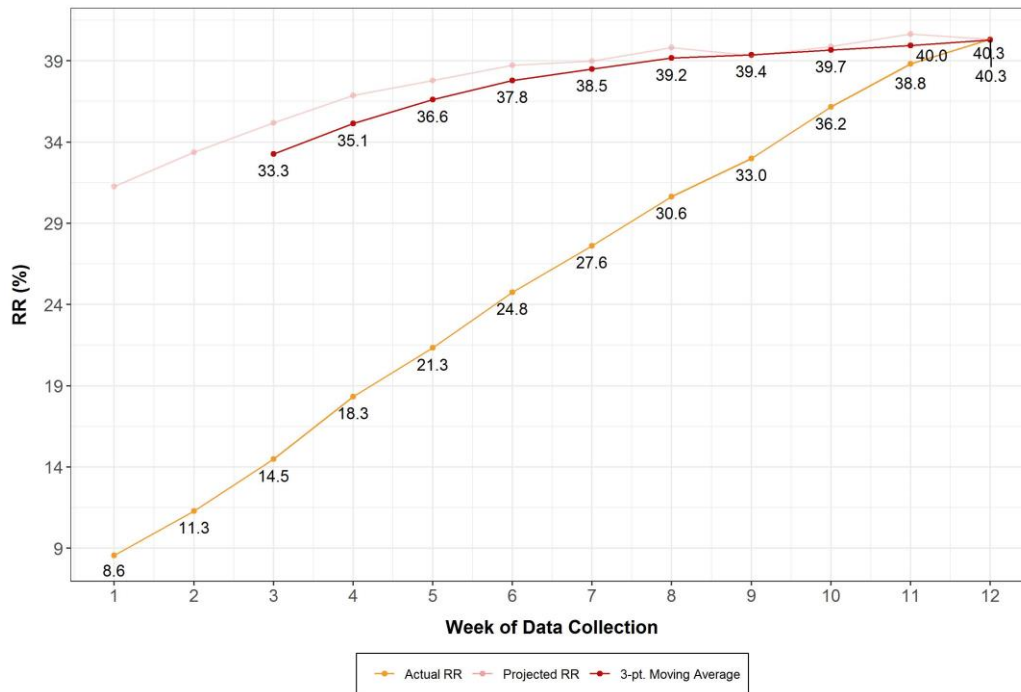


Figure 2: 2018 GSS Phase I: actual and projected response rates based on 2016 GSS Phase I case week matrix

3.2 2018 GSS Phase II Prediction

The prediction for Phase II is conditional on the outcome of Phase I; the fieldwork is briefly suspended at the end of Phase I while the subsample for follow-up is being selected, and the number of successful completes at that stage is known. Figures 3 and 4 show the performance of the model subsequent to the transition to Phase II. Again the upper line shows the predicted value, the lower line the actual value for each week. Remember that the two lines are forced to converge at the end of the fieldwork. The fieldwork lasted 32 weeks in total.

The predicted number of completes (and the response rate) are both too high until very late in the fieldwork. The overestimate is not particularly severe in terms of the final totals (at its worst about 85 cases and 4 percentage points) but is more severe if we use as a base the number of cases actually completed in Phase II (85 above the 440 actually completed in Phase II and 4 percentage points above the 19 percentage points achieved). The only sign of instability in the prediction is the decline in the predicted value that occurs beginning in week 23 and continues monotonically until the end of the fieldwork. We should note that the model was built on a longer field period in the base survey, and this may have contributed to the inaccuracy.

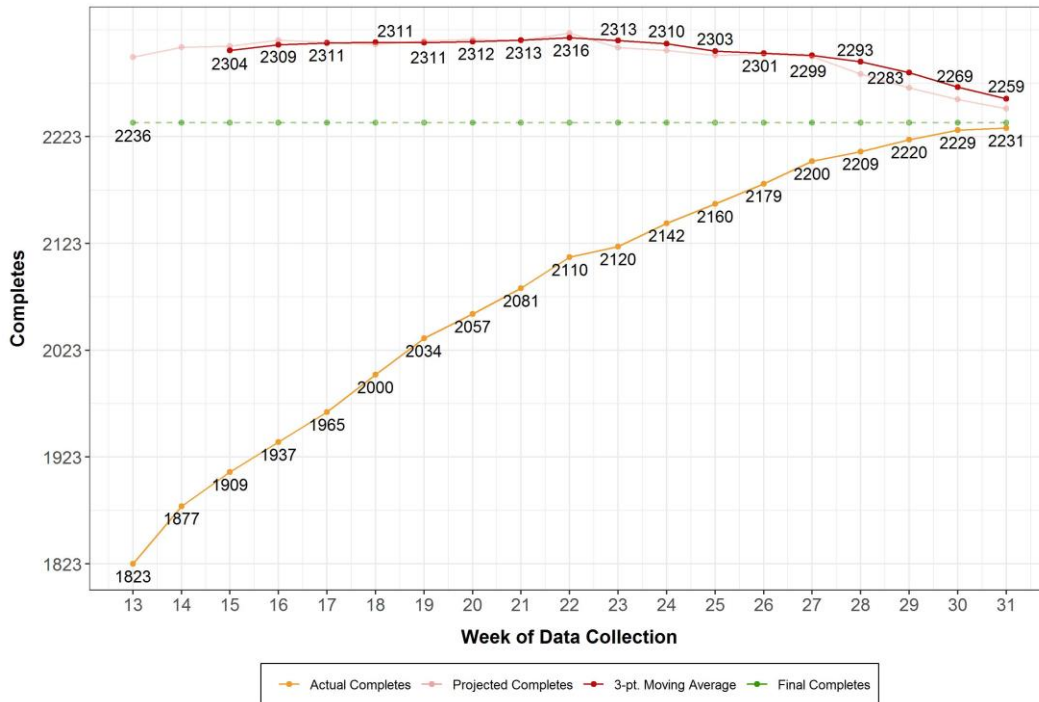


Figure 3: 2018 GSS Phase II without re-released cases: actual and projected completes based on 2016 GSS Phase II case week matrix

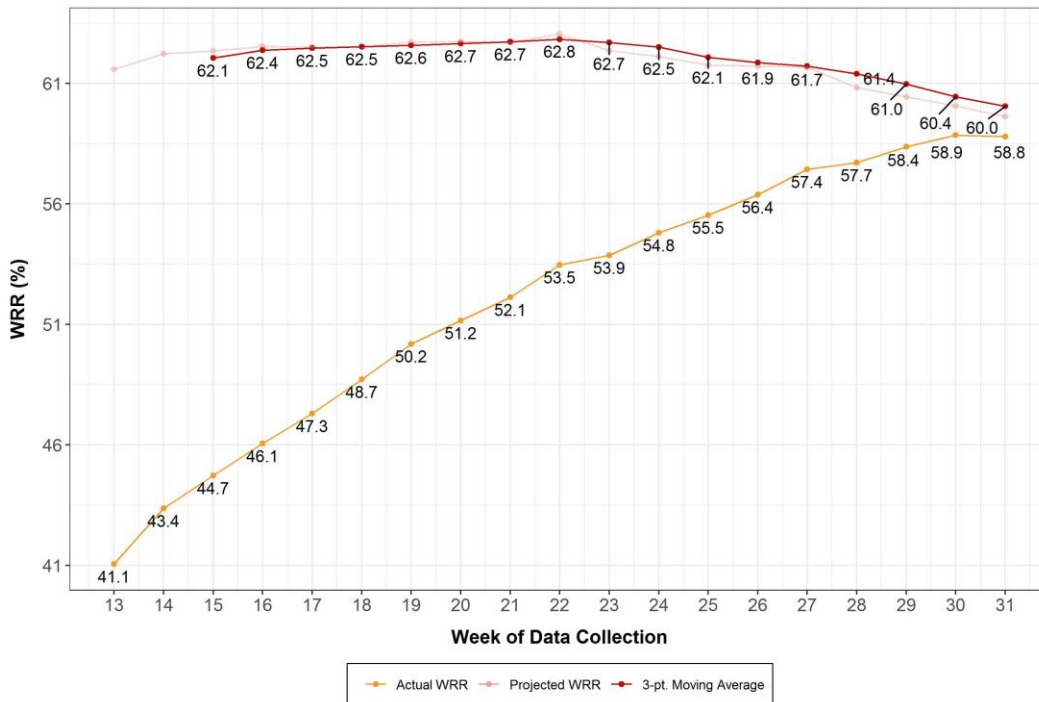


Figure 4: 2018 GSS Phase II without re-released cases: actual and projected weighted response rates based on 2016 GSS Phase II case week matrix

3.3 Disaggregation of Phase II Prediction

At the end of Phase I the cases are placed in three categories: (i) those identified by the Field Management as very promising; (ii) those identified by the model as promising; and (iii) the remainder of the cases. On the basis of past experience, those in categories (i) and (ii) are all retained in the field. A simple random subsample (about 40%) of category (iii) is retained (subsampled in); the remaining cases are not worked any further in Phase II². Figures 5, 6, and 7 show the predicted and actual numbers of completes and the response rates for each of the categories.

3.3.1 Category (i): Most Promising Cases Identified by the Field Staff

Figure 6 shows the predicted completes and predicted response rate for the 93 cases identified by the field staff as most promising; these comprised cases with live appointments and other cases where the interviewer felt there was a high probability of successfully completing an interview. The model produced a very accurate and stable prediction for these cases. The final conditional response rate in this category was 83.5%.

² As described earlier, late in the fieldwork some of the rejected cases in category (iii) were reintroduced into the fieldwork (Phase III); we do not present the results for those cases here.

Fig. 5(a)

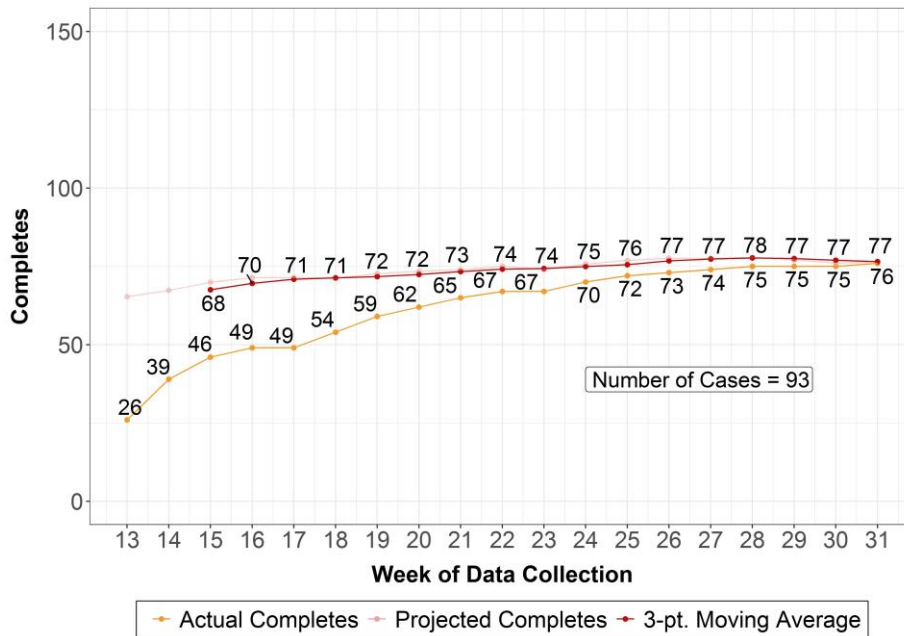


Fig. 5(b)

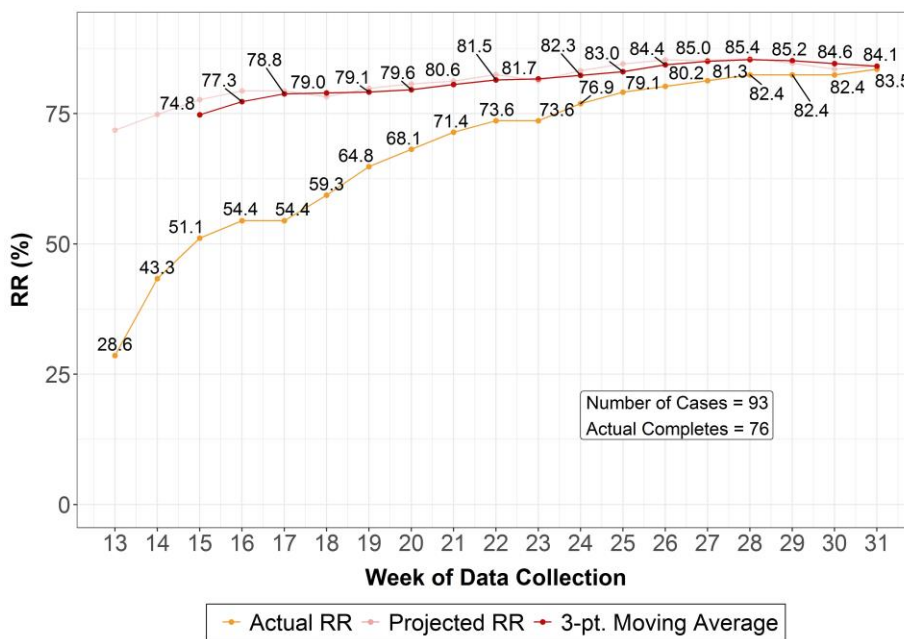


Figure 5: 2018 GSS Phase II Protected by Field cases: (a) actual and predicted completes and (b) actual and predicted response rate.

3.3.2 Category (ii): Other Promising Cases Identified by the Model

The next category comprised those additional cases where the model predicted a higher than average probability of success. Figures 6(a) and 6(b) present the results. The prediction was very stable, though too high, for the first 15 weeks of Phase II (through

week 27 of the fieldwork). It is only in the final four weeks (weeks 28-32) of the fieldwork that the prediction shows symptoms of decline, providing little advance notice of the final outcome.

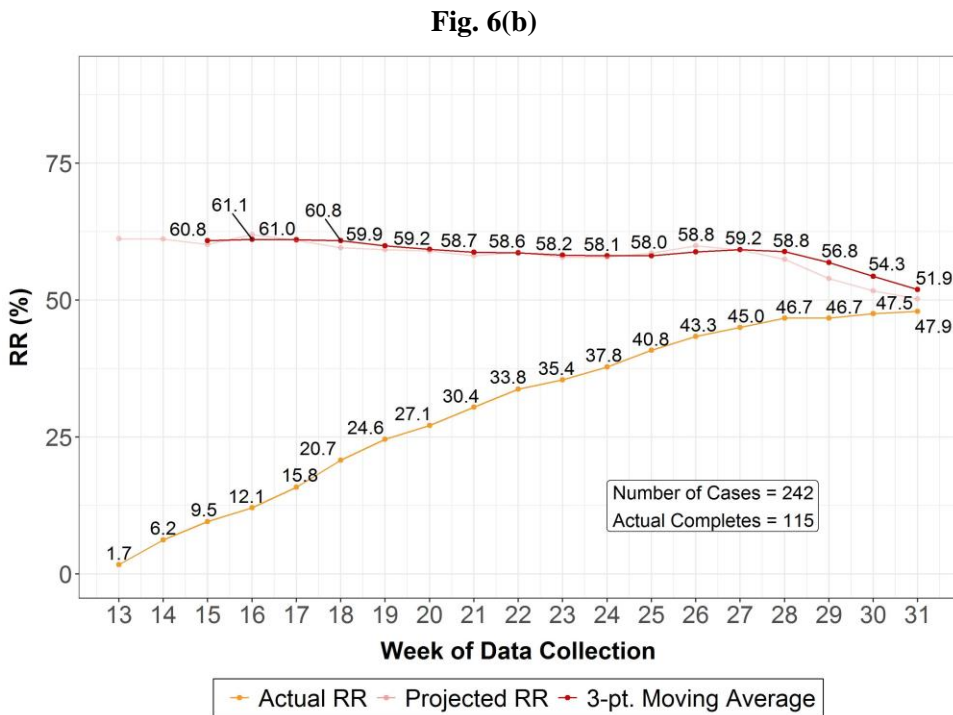
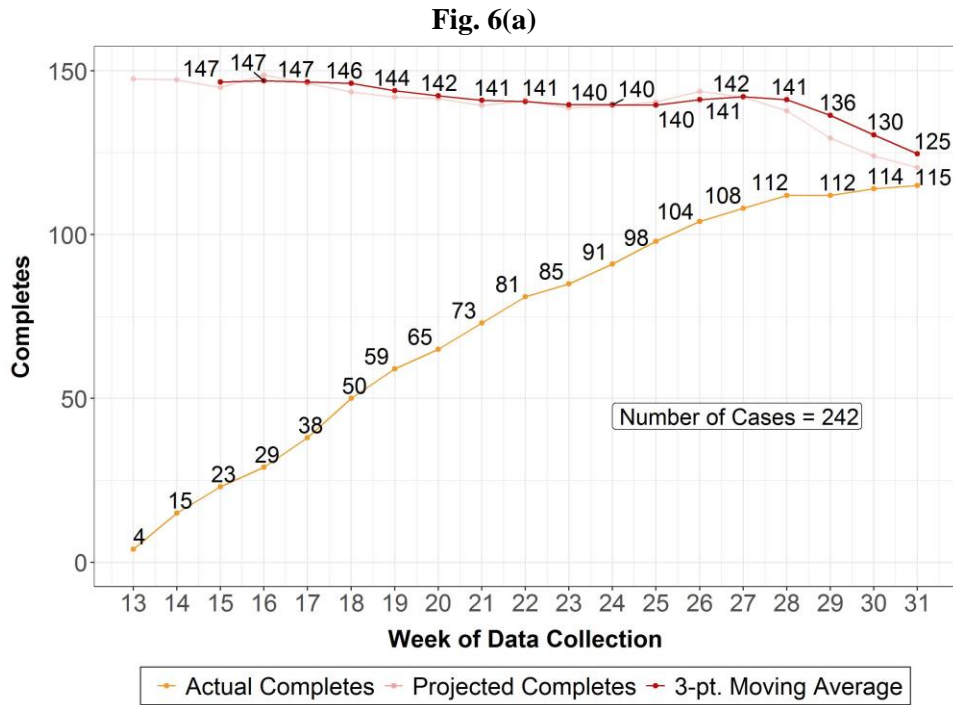


Figure 6: 2018 GSS Phase II Promising Cases Identified by the Model: (a) actual and projected completes and (b) response rate.

3.3.3 Category (iii): *Subsample of Remaining Cases*

The model performance for the remaining cases, which were not identified as promising by either the field staff or the model, is presented in Figures 7(a) and 7(b). The prediction is relatively stable (rising very slightly) for the first 10 weeks of Phase II, then begins a steady decline until the end of the fieldwork. The decline begins in week 22, and could have provided some warning of the model's over-optimism. It is worth noting, however, that the most severe downturn came only in the last few weeks of the fieldwork.

One further factor, not explored in this paper, is the possible impact of the re-introduction of a number of subsampled out cases in week 20 of the fieldwork³. Whether the introduction of these cases could have distracted attention from the existing cases is a topic we intend to pursue further.

³ These are the Phase 3 cases referred to earlier. The cases were not subsampled randomly, but were chosen by the field staff as the least unpromising of the cases that had been randomly subsampled out.

Fig. 7(a)

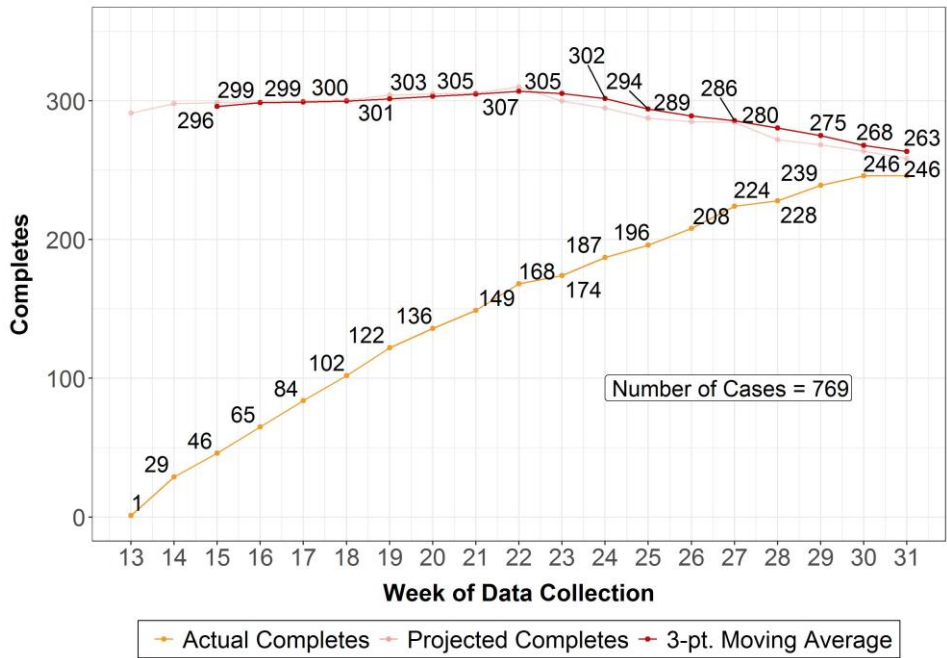


Fig. 7(b)

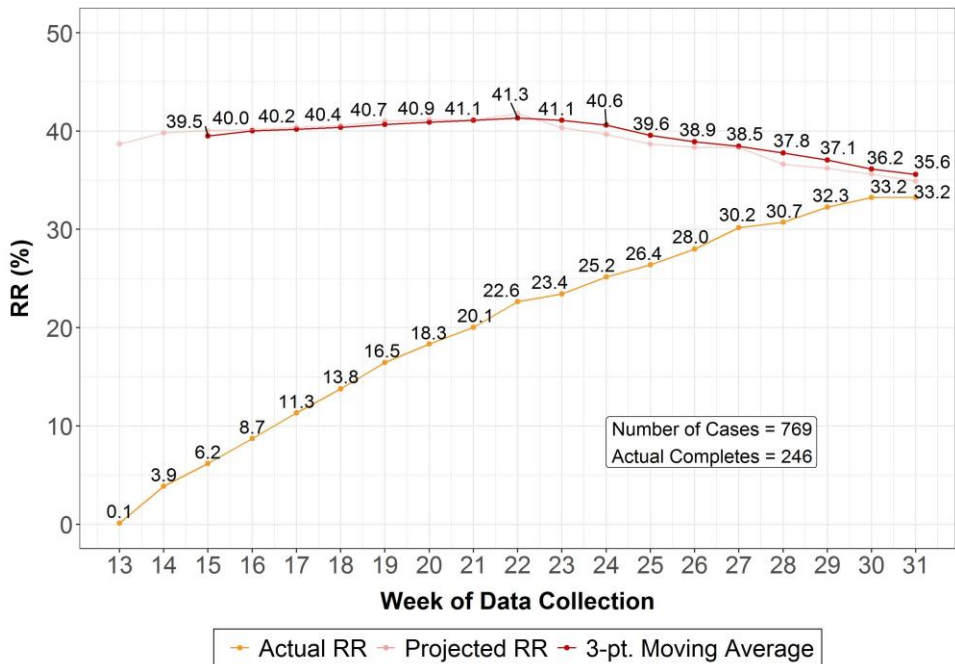


Figure 7: 2018 GSS Phase II all cases randomly subsampled in: (a) actual and predicted completes and (b) actual and predicted response rate.

4. Conclusions and Future Work

Our analytical model built on detailed field disposition histories from the previous (2016) GSS implementation provides an early prediction of response rate, and permitted an early assessment of potential production shortfalls. The stability of the predictions demonstrates the applicability of the model in this context. The model assumes that the *essential survey conditions* are familiar across the base survey and the target survey. In the case of GSS, these assumptions are not unreasonable, given that the survey organization, the field system and staff, and the survey questionnaire are quite similar between the two rounds.

Even in this case, however, there are differences that make the prediction somewhat precarious. First, the field strategy differed between the two iterations. In 2018, there was a much earlier use of incentives in the field; this changes the dynamic of the field strategy as it leaves fewer enhancements for use in the second phase, and could lead to an over-optimistic prediction of outcomes during that phase, something we see in the figures. Second, the length of the field period differed between 2016 and 2018; this presents a technical and a practical challenge. The technical problem is that we do not have a satisfactory way to extend the model to expand the transition matrix for a longer period or compress it for a shorter period as the base survey data were generated by a particular strategy and set of expectations in the field conditioned on the time available. The practical problem is that we cannot formalize the impact on the behavior of the field interviewers of these changes in the length of the fieldwork. The model is currently based entirely on the length of time the survey and the case have been in the field. We have been experimenting with ways of defining time in terms of the proximity to the beginning and end of the fieldwork as a means of splicing time periods together to generate a transition matrix for a longer or shorter time period.

Further work will include analysis of the performance of the model across other GSS iterations, appending respondent-level demographics to the field data, and incorporating area-level information from the US Census and the US American Community Survey. For the 2020 GSS, we plan to incorporate the predictions as a tool for designing experiments to examine in real time the impact of interventions in the field process.

Acknowledgements

We are delighted acknowledge the invaluable assistance provide by the NORC GSS project leaders, Tom W Smith, Rene Bautista, Lauren Doerr, and Jodi Smylie, without whose collaboration this work would not have been possible. Kirk Wolter and the leadership of the Statistics and Methodology Department at NORC provided essential funding support. This work was also funded by the National Science Foundation (NSF) under SES grant #1458922.