



Building a Real-Time Attrition Risk Score for Probability Panelists

Mickey Jackson

VP of Data Science and Innovation

CIPHER 2025

February 26-28, 2025



- Probability-based
- Nationally Representative of U.S. Adults 18 and Older
- Recruited via ABS with Prepaid Cell RDD Supplement
- Multi-mode Capability
 - Most respondents participate online
 - Live telephone interviewing available to those who don't use the internet or who prefer phone surveys
- English and Spanish

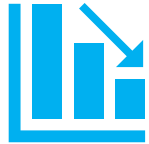
GET MORE INFO

[SSRS.COM/OPINION-PANEL](https://ssrs.com/opinion-panel)

ssrs.com/opinion-panel | [@ssrs_research](https://twitter.com/ssrs_research)



Motivation



Probability Panels **Dramatically Reduce** the Coverage and Data Quality Issues Inherent to Online Opt-in Panels



BUT: Recruitment and Replenishment are **Expensive!**



Accurately Tracking Attrition Risk is Critical in Probability Panel Management:



Accurately Plan Replenishment Needs



Intervene with Panelists at Risk of Attrition (Often Cheaper than Recruiting New Panelists)

Current Approach and Research Question

The Balancing Act: We Want To...

Retire the
“never responders”
(to increase study-level
completion rates)

Keep the
“sometimes responders”
(to reduce risk of bias)

Current SSRS Opinion Panel Approach

**Count of consecutive
nonresponses is primary
attrition indicator**

10 – 15 consecutive
nonresponses triggers a
mailed follow-up survey

Nonrespondents to
follow-up survey are
scrubbed from Panel

Research Question

Would a
multivariate model
more accurately project
attrition risk,
especially early in the
panelist’s tenure?

Methodology

Model Type



Random Forest

Dependent Variable



Response to **any of next 5** survey invitations

Primary Evaluation Criterion



Negative predictive value (NPV)



$P(\text{no response to any of next 5} \mid \text{predicted no response to any of next 5})$



Goal is to obtain higher NPV than the current approach

Predictors

Panelist-level Demographics

- Educational attainment
- Age
- Voter registration status
- Gender
- Party
- Ideology
- Religion
- Nativity
- Employment status
- Household size
- Parental status
- Volunteering

Panelist-by-survey-level Paradata

- Survey number
- Days since last survey
- Month
- Day of week
- Years since joined Panel
- Disposition: non-open
- Disposition: breakoff
- Disposition: terminates
- Response latency (days to respond)

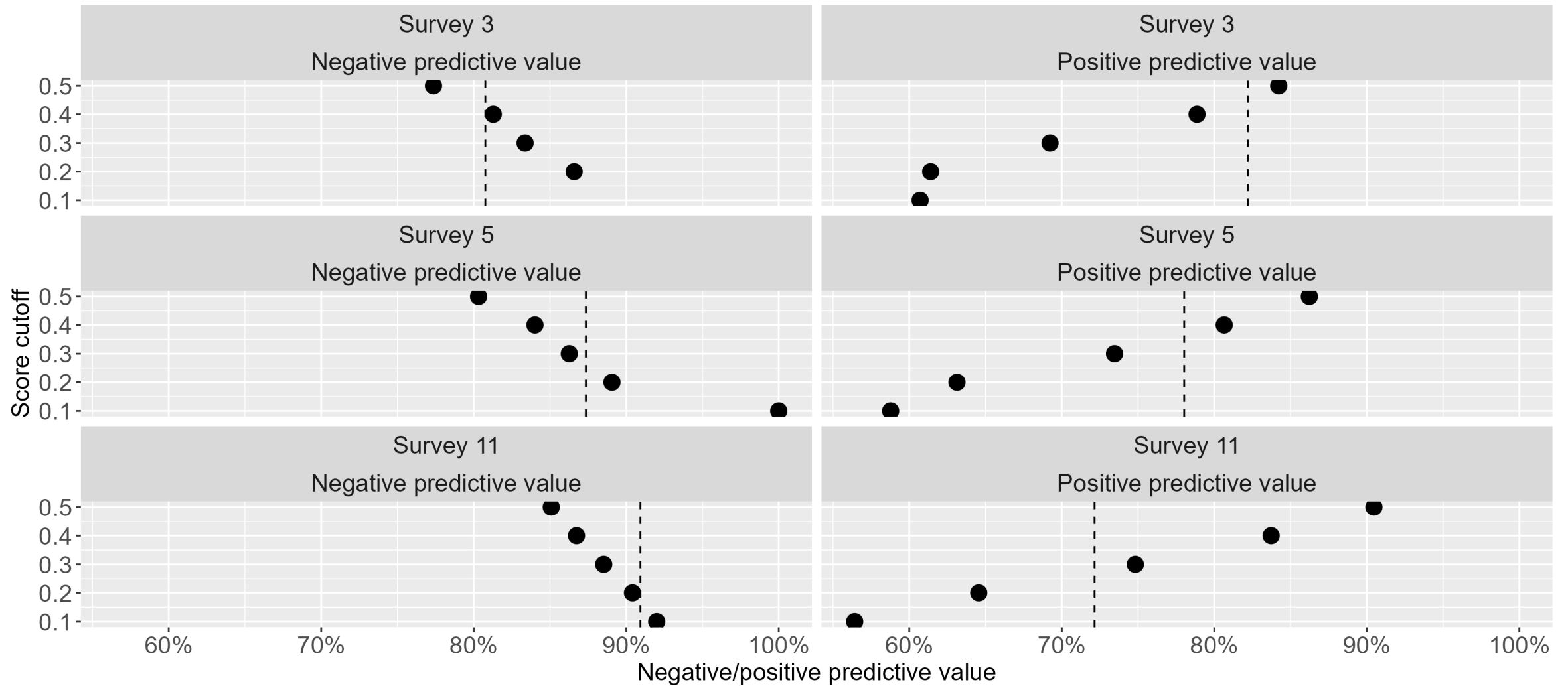
Panelist-level Registration Survey Paradata

- Mode of completion
- Mode selected for panel surveys
- Language of completion
- Language selected for panel surveys
- Consent to text messaging
- Response latency (weeks to respond)
- Month joined
- Sampling frame (ABS vs. RDD)

Aggregate Panelist-by-survey-level Paradata Over Last (Up To) 10 Surveys

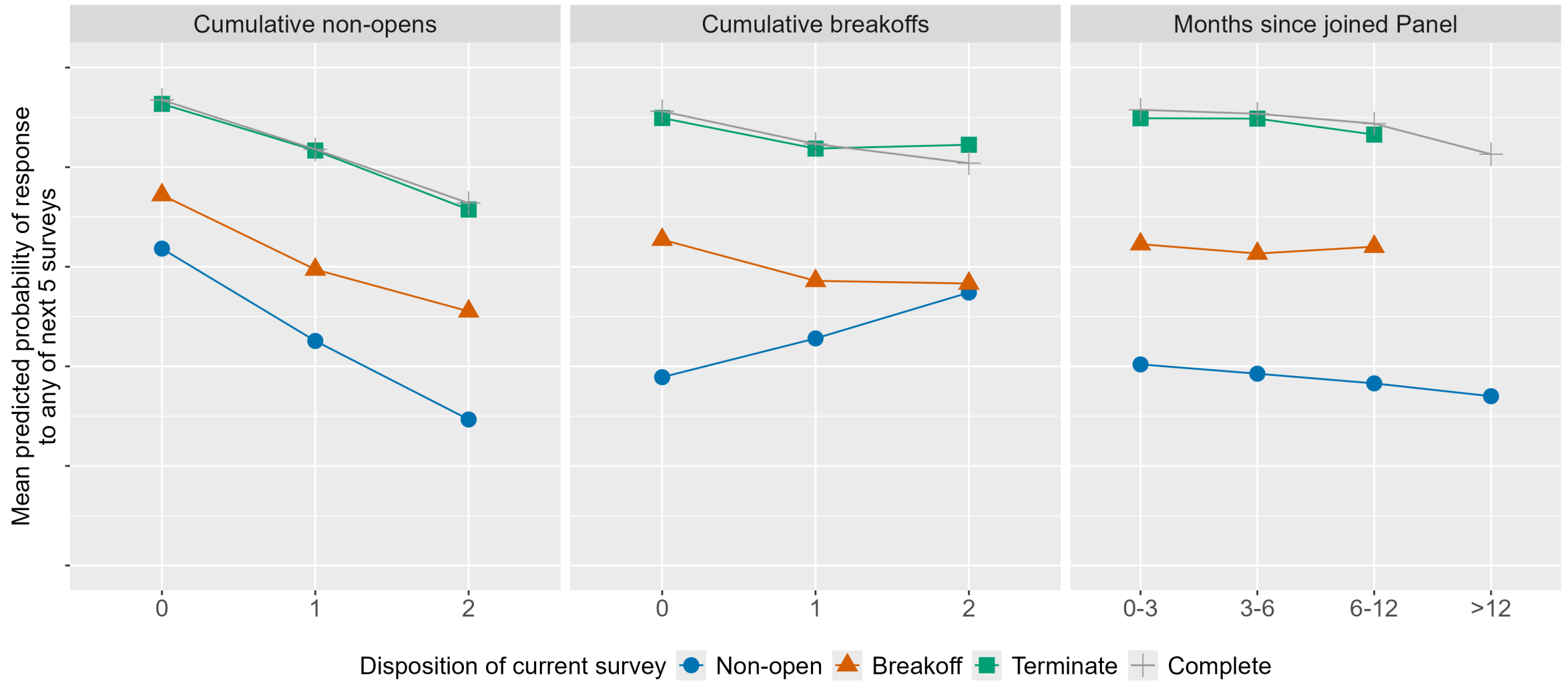
- Number of non-opens
- Number of breakoffs
- Number of terminates
- Average days between surveys
- Average response latency (days to respond)

Negative and Positive Predictive Value



Dashed lines indicate NPV/PPV of current count-based rule.

Predictions by Selected Predictors



As of survey 3.

Conclusions



For panelists with many survey invites, “nonresponse predicts nonresponse”— multivariate models **do not improve on simple count of consecutive nonresponses**



Multivariate Models do a bit Better at First Few Invites

- Early non-opens are much more predictive of attrition than breakoffs
- Infrequent invitations early in panelist tenure may predict attrition



Possible Operational “Compromise”

Continue using count-based rules but focus specifically on non-opens (rather than any nonresponse)



Thank You **CIPHER 2025**

Mickey Jackson

SSRS VP of Data Science and Innovation

mjackson@ssrs.com