Social Research Centre

Life in Australia™

Using propensity scores for adaptive in-field management

Benjamin Phillips, PhD; Danny Fryer, PhD; Dinah Lope, PhD; Dina Neiger, PhD; Jack Barton, MStat&OpRes; Dale VanderGert; Kinto Behr, MStats

CIPHER 2025 Conference, Washington, D.C., February 28

Land Acknowledgement

We acknowledge the Boonwurrung and Wurundjeri Woi-wurrung peoples of the Kulin Nation as the Traditional Custodians of the lands on which our company is located and the Nacotchtank and Piscataway people on whose traditional lands we are today, and pay our respects to Elders, past and present.

The Social Research Centre is committed to honoring First Nations peoples' ongoing unique cultural and spiritual connections to the land, water and seas and their rich contribution to society.

We extend our respects to all First Nations peoples.

Life in Australia™ Social Research Centre

Thanks

John Collins, PhD candidate,
University of Mannheim

Life in Australia™ Social Research Centre

Agenda

01 Panel overview

02 Problem statement

03 Approach

04 Challenges and responses

05 Conclusion

Life in Australia™ Social Research Centre

01

Panel overview

Evolution of Life in Australia™

Then (2016)

- One wave per month
- Two week field period
- One survey per wave
- c. 3,322 panelists, c. 2,500 completes
- Full panel each wave
- Always general population (18+, Australian residents)
- No stratification

Now (2025)

- Two waves per month, occasional third wave
- Two week field period
- Multiple independent surveys per wave
- c. 10,000 panelists, max. c. 7,950 completes
- Full panel waves rare
- Subpopulation samples very common (women, age groups, states, etc.)
- Stratification on age, gender, use language other than English at home (Australian proxy for race/ethnicity), and education to yield completes approximately in proportion to population (for smaller *n*'s)

02

Problem statement

The wicked problem of commercial field management

- Framing as commercial field management, but likely much applies in non-profit and academic panels, *mutandis mutatis*
- The problem:
  - Clients require $n_{achieved} \geq n_{target}$
  - $n_{achieved} > n_{target}$ incurs unbudgeted costs, building in buffer will make less cost competitive
  - Thus, $n_{achieved} \cong n_{target}$ is imperative
- Limited options for $n_{achieved} \cong n_{target}$
  - Invite more panelists than we think we will need—abruptly closing field once reached quota may impact panelist relations, harm sample composition
- Invite just enough panelists, add more sample if needed at last minute—may harm sample composition, increases operational complexity and may harm panelist relations
- Invite *slightly* more panelists than estimated required and manipulate contact attempts (add, subtract, delay) and, where needed, slightly extend close of field (i.e., hours, not days) in an effort to come as close to n_{target} as possible

The centrality of the propensity score

- Central to any of this is the propensity score (Dutwin & Bilgen, 2023)
- Needed for
  1. Predicting feasibility / maximum achievable *n* for given target populations, survey characteristics (primarily length), and field dates
  2. Accurate pre-field estimate of invitations required for n_{target} given survey characteristics
  3. Accurately estimate $n_{achieved}$ at dth day in field

- Not discussed in this presentation:
  - Retrospective response propensity models used as part of the weighting process

03

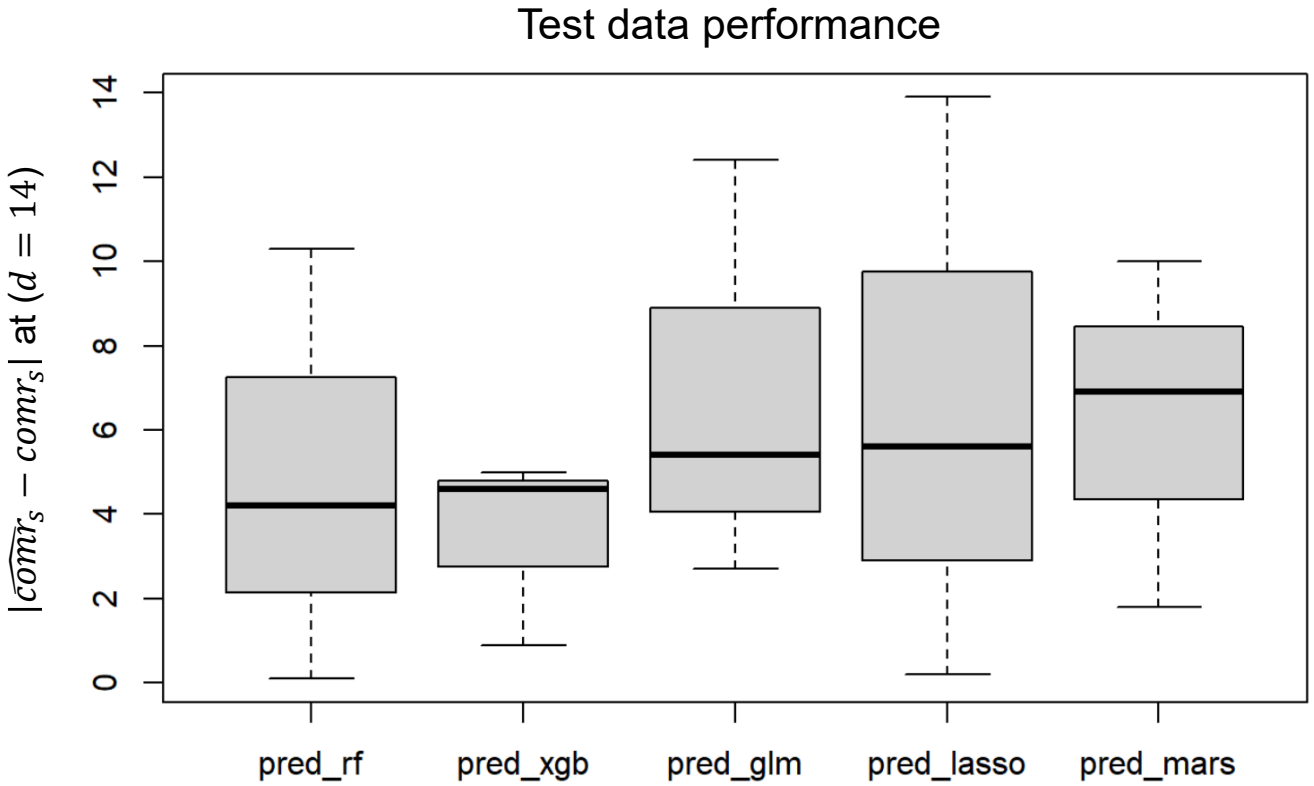
Approach

Current approach (“conditional day”)

- Group of 15 models (classification random forests), corresponding to pre-field (“day zero”) through 14th and final day of fieldwork (“day 14”) ($d = 0, \ldots, D$)
- Rationale for day models is that the ith panelist’s non-completion by day d should inform their estimated propensity ($\hat{p}_i$)
- Model features include:
  - Survey characteristics: expected length, points value
  - Recruitment characteristics: time/mode of recruitment
  - Demographics: age, sex, geography, education, area-level SES, labor force status, voting, etc.
  - Panelist behavior: cumulative completion rate ($comr_i$), time since last completed survey, time since last selected for a survey, online/offline status
  - For $d > 0$: survey-level $comr$ as of day d
- Models retrained every two weeks

Model development

- Initial development tested random forests (RF) (Brieman, 2001), XGBoost (Chen & Guestrin, 2016), standard logit, penalized logit (Friedman, Tibshirani, & Hastie, 2010), and multivariate adaptive spline model (MARS) (Friedman, 1991)
- Tree-based approaches outperformed regression approaches
- RF preferred over the slightly better-performing XGBoost due to being simpler and broadly more tractable



$comr_s = \sum_{i=1}^{n_s} y_{i,s}/n_s$, where $y_{i,s}$ is completion of the sth survey by the ith panelist

$\widehat{comr_s} = \sum_{i=1}^{n_s} \hat{p}_{i,s}/n_s$ where $\hat{p}_{i,s}$ is the response propensity for the sth survey and ith panelist

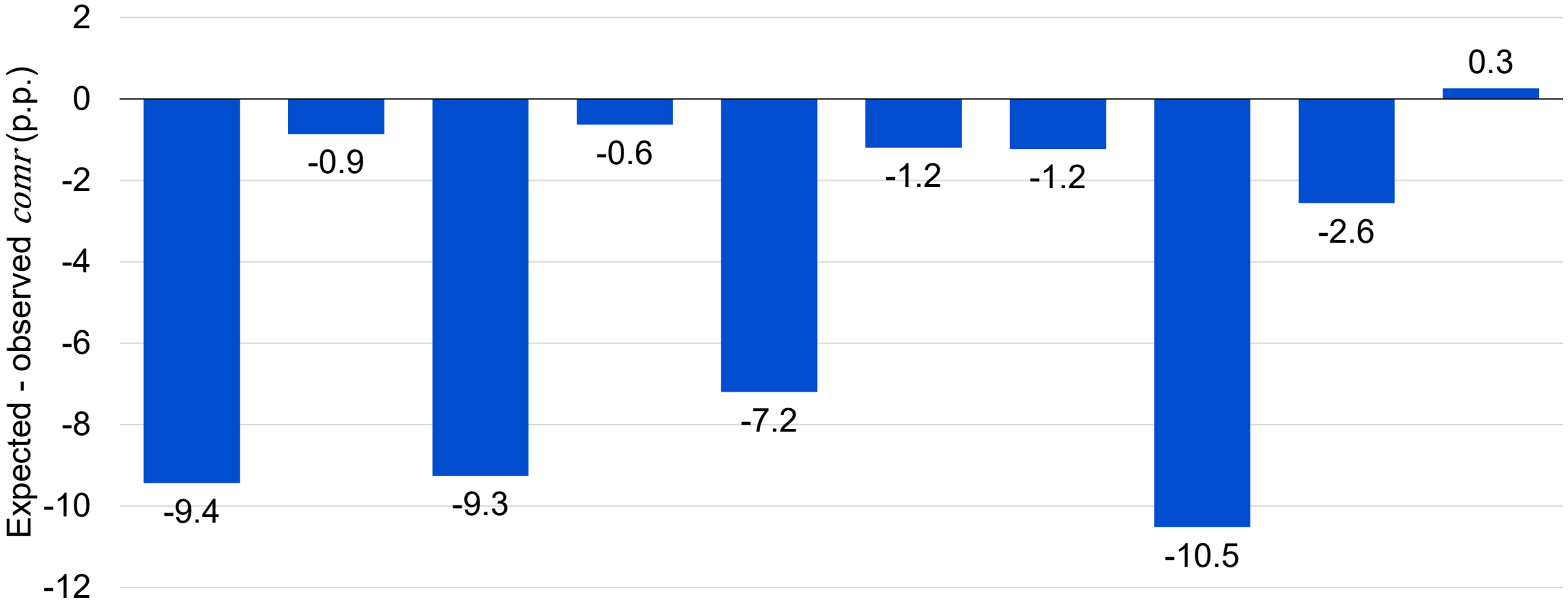
Life in Australia™ Social Research Centre

04

Challenges and responses

Pervasive underestimates of completion rates

$\widehat{comr}_s - comr_s$ for surveys fielded September 2024 – January 2025



Life in Australia™ Social Research Centre

Potential causes of poor performance

- Feedback loop between day 0 predictions and model
  - When the model underestimates propensity, too much sample is drawn
  - This requires interventions (i.e., omitting contact attempts), reducing completion rates
  - We are, in other words, forcing $comr_s$ to match $\widehat{comr}_s$
  - We are effectively training the model to expect an intervention
- Aggregation effects
  - Optimized for predicting $\hat{p}_{i,s}$ not $\widehat{comr}_s$
  - Assumes $\hat{p}_{i,s}$ independent and additive, which may not hold
  - At risk of abnormalities in distribution of $\hat{p}_{i,s}$ and boundary effects close to 0,1 (bimodal distribution with modes close to these boundaries)
- Lack of specialization
  - Use of the same model for day 0 (planning, sampling) and in-field monitoring
- Model misspecification
  - May not be capturing evolving panelist behavior

Model adjustments

Streamlined model

- Revisited predictor selection for conditional day model based on a framework that covers 6 main categories (Australian Bureau of Statistics, 2022; Black et al., 2010) and their corresponding predictor importance
- Included survey-level characteristic
- Excluded problematic surveys
- Excluded predictors with large amounts of missing data (> 5%)
- Fine tuned parameters and predictors
- Slightly improves model performance over non-streamlined model (c. 0.5 p.p.)

Conditional reminder model

- Change from:

  "*What is the probability panelist i will complete the survey within 14 days, given they have not completed in d days?*"

  to

  "*What is the probability panelist i will complete the survey after receiving r_T reminders, given they have already received r_t reminders?*"

- This *should* flexibly accommodate fieldwork interventions
- Still being implemented

Other model adaptations being trialed / to be trialed

- Removing surveys with unusual methods or target populations from the models (results not shown, only very recently implemented)
- Comparing performance of various models in live surveys (ongoing)
- Switch from current imputation approach (reimpute separately for each model) to use standard imputed values used in survey weighting (to be made)
- Adding refusal and unsubscribe from specific mode of communications history as model features (to be trialed)
- Case weighting: weight more recent surveys more highly (to be trialed)

05

Conclusion

Conclusion

- Still very much a work in progress
- Main finding: adaptive fieldwork + classical propensity model make a bad combination due to feedback loop
- Slight improvements from not using a “kitchen sink” model / streamlining model features
- Thoughts and suggestions very welcome!

Thank you

Benjamin Phillips, PhD
benjamin.phillips@srcentre.com.au

References

Australian Bureau of Statistics. (2022, April 27). *Raising survey response rates using machine learning to predict gold providers*. https://www.abs.gov.au/statistics/research/raising-survey-response-rates-using-machine-learning-predict-gold-providers

Black, M., Brent, G., Bell, P., Starick, R. and Zhang, M. (2010). *Empirical models for survey cost, response rate and bias using paradata* (cat. no. 1352.0.55.113). Australian Bureau of Statistics. https://www.abs.gov.au/ausstats/abs@.nsf/mf/1352.0.55.113

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Dutwin, D., & Bilgen, I. (2023, January 26). *Everything you need to know when utilizing probability panels: Best practices in planning, fielding, & analysis* [webinar]. American Association for Public Opinion Research. https://vimeo.com/aapor/jan2023

Friedman, J. H. (1991). Multivariate adaptive regression splines. *Annals of Statistics*, *19*(1), 1–67. https://doi.org/10.1214/aos/1176347963

Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22. https://doi.org/10.18637/jss.v033.i01