Analyzing the Impact of Events Through Surveys

Online Appendix

Contents

1 Quota Sampling

In this design, we survey two groups of people before and after the event, selecting participants based on covariates to try to make the two groups similar to each other and to the total population. Let n be the total number of people who we consider surveying, with n_a denoting the number in Wave 2 and n_b denoting the number in Wave 1. As in the paper, let $r_{ia} \in \{0,1\}$ denote whether individual i completed the survey in Wave 2, and let $r_{ib} \in \{0,1\}$ denote whether individual i completed the survey in Wave 1. In addition, let $q_i \in \{0,1\}$ denote whether individual i is in our quota group for either Wave 1 or Wave 2.

Building off this notation, we can let g_a denote the number of people in the Wave 2 quota group and g_b denote the number of people in the Wave 1 quota group. We can also define $g = g_a + g_b$ as the total number of people in our quota sample. Individuals were not randomized to be contacted in Wave 1 or Wave 2, so g, g_a, g_b are all parameters, not random variables. In fact, there is no randomization in this design that could make any of these values a random variable.

1.1 Proof of Proposition 1

Proposition 1. *Bias in the quota sampling design is given by*

$$Bias(\hat{\tau}_{ka|r_a=1,q=1}) = Bias_{\mathbf{X}}(\hat{\tau}_{ka|r_a=1,q=1}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka|r_a=1,q=1}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka|r_a=1,q=1}) + Bias_{\mathbf{M}}(\hat{\tau}_{ka|r_a=1,q=1})$$

where

$$\begin{aligned}
Bias_{\mathbf{X}}(\hat{\tau}_{ka|r_a=1,q=1}) &= \bar{y}_{ikb|r_a=1,q=1} - \bar{y}_{ikb|r_b=1,q=1} \\
Bias_{\mathbf{T}}(\hat{\tau}_{ka|r_a=1,q=1}) &= \bar{y}_{ikc|r_a=1,q=1} - \bar{y}_{ikbc|r_a=1,q=1} \\
Bias_{\mathbf{A}}(\hat{\tau}_{ka|r_a=1,q=1}) &= \bar{y}_{ikbc|r_a=1,q=1} - \bar{y}_{ikb|r_a=1,q=1} \\
Bias_{\mathbf{M}}(\hat{\tau}_{ka|r_a=1,q=1}) &= \bar{\epsilon}_{kt|r_a=1,q=1} - \bar{\epsilon}_{kb|r_b=1,q=1}
\end{aligned}$$

Proof. The causal parameter that we want to estimate is the average causal effect of the event on the Wave 2 quota group's truthful responses to question k of the survey:

$$\bar{\tau}_{ka|r_a=1,q=1} = \frac{1}{g_a}\sum_{i=1}^{g_a}(y_{ikt|r_a=1,q=1} - y_{ikc|r_a=1,q=1})$$

The statistic that we will use to estimate this parameter is the average difference between the reported answers of the g_a respondents who completed our survey in Wave 2 and the g_b respondents who completed it in Wave 1.

$$\hat{\tau}_{ka,r_a=1,q=1} = \frac{1}{g_a}\sum_{i=1}^{g_a} y_{ikto|r_a=1,q=1} - \frac{1}{g_b}\sum_{i=1}^{g_b} y_{ikbo|r_b=1,q=1}$$

Following the same procedures from the analysis in Section 2 of the paper, the bias in this estimator can be rewritten as

$$Bias(\hat{\tau}_{ka|r_a=1,q=1}) = Bias_{\mathbf{X}}(\hat{\tau}_{ka|r_a=1,q=1}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka|r_a=1,q=1}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka|r_a=1,q=1}) + Bias_{\mathbf{M}}(\hat{\tau}_{ka|r_a=1,q=1})$$

where

$$Bias_{\mathbf{X}}(\hat{\tau}_{ka|r_a=1,q=1}) = \bar{y}_{ikb|r_a=1,q=1} - \bar{y}_{ikb|r_b=1,q=1}$$

$$Bias_{\mathbf{T}}(\hat{\tau}_{ka|r_a=1,q=1}) = \bar{y}_{ikc|r_a=1,q=1} - \bar{y}_{ikbc|r_a=1,q=1}$$

$$Bias_{\mathbf{A}}(\hat{\tau}_{ka|r_a=1,q=1}) = \bar{y}_{ikbc|r_a=1,q=1} - \bar{y}_{ikb|r_a=1,q=1}$$

$$Bias_{\mathbf{M}}(\hat{\tau}_{ka|r_a=1,q=1}) = \bar{\epsilon}_{kt|r_a=1,q=1} - \bar{\epsilon}_{kb|r_b=1,q=1}$$

□

The difference between this overall bias term and the $Bias(\hat{\tau}_{ka|r_a=1})$ that we derived in Section 3 is that this term restricts the focus to our quota sample.

1.2 Proof of Proposition 2

Proposition 2. *Quota designs reduce bias if and only if*

$$|Bias(\hat{\tau}_{ka|r_a=1,q=1})| < |(\tfrac{g_a}{n_a})Bias(\hat{\tau}_{ka|r_a=1,q=1}) + (\tfrac{e_a}{n_a})Bias(\hat{\tau}_{ka|r_a=1,q=0}) + (\tfrac{g_a}{n_a} - \tfrac{g_b}{n_b})(\bar{y}_{ikbo|r_b=1,q=1} - \bar{y}_{ikbo|r_b=1,q=0})|$$

When the inequality if flipped, quota sampling ***amplifies*** *bias.*

Proof. Whether quota sampling improves on our baseline model depends in part on whether the bias in the excluded group is smaller or in the opposite direction as the bias in the quota group. It also depends, to some extent, on external validity considerations, since using the quota sampling estimator changes the parameter that we are estimating.

Focusing just on the potential bias reduction, the difference in bias between the baseline model and quota sampling can be written as

$$
\begin{aligned}
|Bias(\hat{\tau}_{ka|r_a=1})| - |Bias(\hat{\tau}_{ka|r_a=1,q=1})| =& |Bias_{\mathbf{X}}(\hat{\tau}_{ka|r_a=1}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka|r_a=1}) + \\
& Bias_{\mathbf{A}}(\hat{\tau}_{ka|r_a=1}) + Bias_{\mathbf{M}}(\hat{\tau}_{ka|r_a=1})| - \\
& |Bias_{\mathbf{X}}(\hat{\tau}_{ka|r_a=1,q=1}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka|r_a=1,q=1}) + \\
& Bias_{\mathbf{A}}(\hat{\tau}_{ka|r_a=1,q=1}) + Bias_{\mathbf{M}}(\hat{\tau}_{ka|r_a=1,q=1})|
\end{aligned}
$$

We can decompose $Bias(\hat{\tau}_{ka|r_a=1})$ into a weighted average of the bias in the sub-sample who we would have interviewed if we had done quota sampling and the bias in the sub-sample who we would have been excluded in the quota sampling design. We will denote the number of Wave 2 individuals who would have been excluded under quota sampling as $e_a = n_a - g_a$. Likewise, we will denote the number of Wave 1 individuals who would have been excluded under quota sampling as $e_b = n_b - g_b$.

We then have

$$
\begin{aligned}
Bias(\hat{\tau}_{ka|r_a=1}) =& Bias_{\mathbf{X}}(\hat{\tau}_{ka|r_a=1}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka|r_a=1}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka|r_a=1}) + Bias_{\mathbf{M}}(\hat{\tau}_{ka|r_a=1}) \\
Bias(\hat{\tau}_{ka|r_a=1}) =& \tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikb|r_a=1} - \tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1} + \\
& \tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikbc|r_a=1} - \tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikb|r_a=1} + \\
& \tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikc|r_a=1} - \tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikbc|r_a=1} + \\
& \bar{\epsilon}_{kt|r_a=1} - \bar{\epsilon}_{kb|r_b=1}
\end{aligned}
$$

which we can separate into

$$
\begin{aligned}
Bias(\hat{\tau}_{ka|r_a=1}) =& \frac{1}{n_a}\sum_{i=1}^{g_a} y_{ikb|r_a=1,q=1} + \frac{1}{n_a}\sum_{i=1}^{e_a} y_{ikb|r_a=1,q=0} - (\frac{1}{n_b}\sum_{i=1}^{g_b} y_{ikb|r_b=1,q=1} + \frac{1}{n_b}\sum_{i=1}^{e_b} y_{ikb|r_b=1,q=0}) + \\
& \frac{1}{n_a}\sum_{i=1}^{g_a} y_{ikbc|r_a=1,q=1} + \frac{1}{n_a}\sum_{i=1}^{e_a} y_{ikbc|r_a=1,q=0} - (\frac{1}{n_a}\sum_{i=1}^{g_a} y_{ikb|r_a=1,q=1} + \frac{1}{n_a}\sum_{i=1}^{e_a} y_{ikb|r_a=1,q=0}) + \\
& \frac{1}{n_a}\sum_{i=1}^{g_a} y_{ikc|r_a=1,q=1} + \frac{1}{n_a}\sum_{i=1}^{e_a} y_{ikc|r_a=1,q=0} - (\frac{1}{n_a}\sum_{i=1}^{g_a} y_{ikbc|r_a=1,q=1} + \frac{1}{n_a}\sum_{i=1}^{e_a} y_{ikbc|r_a=1,q=0}) + \\
& \frac{1}{n_a}\sum_{i=1}^{g_a} y_{ikto|r_a=1,q=1} + \frac{1}{n_a}\sum_{i=1}^{e_a} y_{ikto|r_a=1,q=0} - (\frac{1}{n_a}\sum_{i=1}^{g_a} y_{ikt|r_a=1,q=1} + \frac{1}{n_a}\sum_{i=1}^{e_a} y_{ikt|r_a=1,q=0}) - \\
& (\frac{1}{n_b}\sum_{i=1}^{g_b} y_{ikbo|r_b=1,q=1} + \frac{1}{n_b}\sum_{i=1}^{e_b} y_{ikbo|r_b=1,q=0} - (\frac{1}{n_b}\sum_{i=1}^{g_b} y_{ikb|r_b=1,q=1} + \frac{1}{n_b}\sum_{i=1}^{e_b} y_{ikb|r_b=1,q=0}))
\end{aligned}
$$

We can simplify this equation as follows

$$
\begin{aligned}
Bias(\hat{\tau}_{ka|r_a=1}) =& (\frac{g_a}{n_a})\bar{y}_{ikb|r_a=1,q=1} + (\frac{e_a}{n_a})\bar{y}_{ikb|r_a=1,q=0} - ((\frac{g_b}{n_b})\bar{y}_{ikb|r_b=1,q=1} + (\frac{e_b}{n_b})\bar{y}_{ikb|r_b=1,q=0}) + \\
& (\frac{g_a}{n_a})\bar{y}_{ikbc|r_a=1,q=1} + (\frac{e_a}{n_a})\bar{y}_{ikbc|r_a=1,q=0} - ((\frac{g_a}{n_a})\bar{y}_{ikb|r_a=1,q=1} + (\frac{e_a}{n_a})\bar{y}_{ikb|r_a=1,q=0}) + \\
& (\frac{g_a}{n_a})\bar{y}_{ikc|r_a=1,q=1} + (\frac{e_a}{n_a})\bar{y}_{ikc|r_a=1,q=0} - ((\frac{g_a}{n_a})\bar{y}_{ikbc|r_a=1,q=1} + (\frac{e_a}{n_a})\bar{y}_{ikbc|r_a=1,q=0}) + \\
& (\frac{g_a}{n_a})\bar{y}_{ikto|r_a=1,q=1} + (\frac{e_a}{n_a})\bar{y}_{ikto|r_a=1,q=0} - ((\frac{g_a}{n_a})\bar{y}_{ikt|r_a=1,q=1} + (\frac{e_a}{n_a})\bar{y}_{ikt|r_a=1,q=0}) - \\
& ((\frac{g_b}{n_b})\bar{y}_{ikbo|r_b=1,q=1} + (\frac{e_b}{n_b})\bar{y}_{ikbo|r_b=1,q=0} - ((\frac{g_b}{n_b})\bar{y}_{ikb|r_b=1,q=1} + (\frac{e_b}{n_b})\bar{y}_{ikb|r_b=1,q=0})) \\
Bias(\hat{\tau}_{ka|r_a=1}) =& (\frac{g_a}{n_a})\bar{y}_{ikb|r_a=1,q=1} - (\frac{g_b}{n_b})\bar{y}_{ikb|r_b=1,q=1} + (\frac{e_a}{n_a})\bar{y}_{ikb|r_a=1,q=0} - (\frac{e_b}{n_b})\bar{y}_{ikb|r_b=1,q=0} + \\
& (\frac{g_a}{n_a})\bar{y}_{ikbc|r_a=1,q=1} - (\frac{g_a}{n_a})\bar{y}_{ikb|r_a=1,q=1} + (\frac{e_a}{n_a})\bar{y}_{ikbc|r_a=1,q=0} - (\frac{e_a}{n_a})\bar{y}_{ikb|r_a=1,q=0} + \\
& (\frac{g_a}{n_a})\bar{y}_{ikc|r_a=1,q=1} - (\frac{g_a}{n_a})\bar{y}_{ikbc|r_a=1,q=1} + (\frac{e_a}{n_a})\bar{y}_{ikc|r_a=1,q=0} - (\frac{e_a}{n_a})\bar{y}_{ikbc|r_a=1,q=0} + \\
& (\frac{g_a}{n_a})\bar{y}_{ikto|r_a=1,q=1} - (\frac{g_a}{n_a})\bar{y}_{ikt|r_a=1,q=1} - ((\frac{g_b}{n_b})\bar{y}_{ikbo|r_b=1,q=1} - (\frac{g_b}{n_b})\bar{y}_{ikb|r_b=1,q=1}) + \\
& (\frac{e_a}{n_a})\bar{y}_{ikto|r_a=1,q=0} - (\frac{e_a}{n_a})\bar{y}_{ikt|r_a=1,q=0} - ((\frac{e_b}{n_b})\bar{y}_{ikbo|r_b=1,q=0} - (\frac{e_b}{n_b})\bar{y}_{ikb|r_b=1,q=0}) \\
Bias(\hat{\tau}_{ka|r_a=1}) =& (\frac{g_a}{n_a})\bar{y}_{ikb|r_a=1,q=1} - (\frac{g_b}{n_b})\bar{y}_{ikb|r_b=1,q=1} + (\frac{e_a}{n_a})\bar{y}_{ikb|r_a=1,q=0} - (\frac{e_b}{n_b})\bar{y}_{ikb|r_b=1,q=0} + \\
& (\frac{g_a}{n_a})Bias_{\mathbf{T}}(\hat{\tau}_{ka|r_a=1,q=1}) + (\frac{e_a}{n_a})Bias_{\mathbf{T}}(\hat{\tau}_{ka|r_a=1,q=0}) + \\
& (\frac{g_a}{n_a})Bias_{\mathbf{A}}(\hat{\tau}_{ka|r_a=1,q=1}) + (\frac{e_a}{n_a})Bias_{\mathbf{A}}(\hat{\tau}_{ka|r_a=1,q=0}) + \\
& (\frac{g_a}{n_a})\bar{\epsilon}_{kt|r_a=1,q=1} - (\frac{g_b}{n_b})\bar{\epsilon}_{kb|r_b=1,q=1} + (\frac{e_a}{n_a})\bar{\epsilon}_{kt|r_a=1,q=0} - (\frac{e_b}{n_b})\bar{\epsilon}_{kb|r_b=1,q=0}
\end{aligned}
$$

$$
\begin{aligned}
Bias(\hat{\tau}_{ka|r_a=1}) =& (\tfrac{g_a}{n_a})\bar{y}_{ikb|r_a=1,q=1} - (\tfrac{g_b n_a}{n_a n_b})\bar{y}_{ikb|r_b=1,q=1} + \\
& (\tfrac{e_a}{n_a})\bar{y}_{ikb|r_a=1,q=0} - (\tfrac{e_b n_a}{n_a n_b})\bar{y}_{ikb|r_b=1,q=0} + \\
& (\tfrac{g_a}{n_a})Bias_{\mathbf{T}}(\hat{\tau}_{ka|r_a=1,q=1}) + (\tfrac{e_a}{n_a})Bias_{\mathbf{T}}(\hat{\tau}_{ka|r_a=1,q=0}) + \\
& (\tfrac{g_a}{n_a})Bias_{\mathbf{A}}(\hat{\tau}_{ka|r_a=1,q=1}) + (\tfrac{e_a}{n_a})Bias_{\mathbf{A}}(\hat{\tau}_{ka|r_a=1,q=0}) + \\
& (\tfrac{g_a}{n_a})\bar{\epsilon}_{kt|r_a=1,q=1} - (\tfrac{g_b n_a}{n_a n_b})\bar{\epsilon}_{kb|r_b=1,q=1} + \\
& (\tfrac{e_a}{n_a})\bar{\epsilon}_{kt|r_a=1,q=0} - (\tfrac{e_b n_a}{n_a n_b})\bar{\epsilon}_{kb|r_b=1,q=0} \\
Bias(\hat{\tau}_{ka|r_a=1}) =& (\tfrac{g_a}{n_a})(\bar{y}_{ikb|r_a=1,q=1} - (\tfrac{g_b n_a}{g_a n_b})\bar{y}_{ikb|r_b=1,q=1}) + \\
& (\tfrac{e_a}{n_a})(\bar{y}_{ikb|r_a=1,q=0} - (\tfrac{e_b n_a}{e_a n_b})\bar{y}_{ikb|r_b=1,q=0}) + \\
& (\tfrac{g_a}{n_a})Bias_{\mathbf{T}}(\hat{\tau}_{ka|r_a=1,q=1}) + (\tfrac{e_a}{n_a})Bias_{\mathbf{T}}(\hat{\tau}_{ka|r_a=1,q=0}) + \\
& (\tfrac{g_a}{n_a})Bias_{\mathbf{A}}(\hat{\tau}_{ka|r_a=1,q=1}) + (\tfrac{e_a}{n_a})Bias_{\mathbf{A}}(\hat{\tau}_{ka|r_a=1,q=0}) \\
& (\tfrac{g_a}{n_a})(\bar{\epsilon}_{kt|r_a=1,q=1} - (\tfrac{g_b n_a}{g_a n_b})\bar{\epsilon}_{kb|r_b=1,q=1}) + \\
& (\tfrac{e_a}{n_a})(\bar{\epsilon}_{kt|r_a=1,q=0} - (\tfrac{e_b n_a}{e_a n_b})\bar{\epsilon}_{kb|r_b=1,q=0}) \\
Bias(\hat{\tau}_{ka|r_a=1}) =& (\tfrac{g_a}{n_a})(\bar{y}_{ikb|r_a=1,q=1} - \bar{y}_{ikb|r_b=1,q=1} + \bar{y}_{ikb|r_b=1,q=1} - (\tfrac{g_b n_a}{g_a n_b})\bar{y}_{ikb|r_b=1,q=1}) + \\
& (\tfrac{e_a}{n_a})(\bar{y}_{ikb|r_a=1,q=0} - \bar{y}_{ikb|r_b=1,q=0} + \bar{y}_{ikb|r_b=1,q=0} - (\tfrac{e_b n_a}{e_a n_b})\bar{y}_{ikb|r_b=1,q=0}) + \\
& (\tfrac{g_a}{n_a})Bias_{\mathbf{T}}(\hat{\tau}_{ka|r_a=1,q=1}) + (\tfrac{e_a}{n_a})Bias_{\mathbf{T}}(\hat{\tau}_{ka|r_a=1,q=0}) + \\
& (\tfrac{g_a}{n_a})Bias_{\mathbf{A}}(\hat{\tau}_{ka|r_a=1,q=1}) + (\tfrac{e_a}{n_a})Bias_{\mathbf{A}}(\hat{\tau}_{ka|r_a=1,q=0}) + \\
& (\tfrac{g_a}{n_a})(\bar{\epsilon}_{kt|r_a=1,q=1} - \bar{\epsilon}_{kb|r_b=1,q=1} + \bar{\epsilon}_{kb|r_b=1,q=1} - (\tfrac{g_b n_a}{g_a n_b})\bar{\epsilon}_{kb|r_b=1,q=1}) + \\
& (\tfrac{e_a}{n_a})(\bar{\epsilon}_{kt|r_a=1,q=0} - \bar{\epsilon}_{kb|r_b=1,q=0} + \bar{\epsilon}_{kb|r_b=1,q=0} - (\tfrac{e_b n_a}{e_a n_b})\bar{\epsilon}_{kb|r_b=1,q=0}) \\
Bias(\hat{\tau}_{ka|r_a=1}) =& (\tfrac{g_a}{n_a})Bias_{\mathbf{X}}(\hat{\tau}_{ka|r_a=1,q=1}) + (\tfrac{g_a}{n_a})(\bar{y}_{ikb|r_b=1,q=1} - (\tfrac{g_b n_a}{g_a n_b})\bar{y}_{ikb|r_b=1,q=1}) + \\
& (\tfrac{e_a}{n_a})Bias_{\mathbf{X}}(\hat{\tau}_{ka|r_a=1,q=0}) + (\tfrac{e_a}{n_a})(\bar{y}_{ikb|r_b=1,q=0} - (\tfrac{e_b n_a}{e_a n_b})\bar{y}_{ikb|r_b=1,q=0}) + \\
& (\tfrac{g_a}{n_a})Bias_{\mathbf{T}}(\hat{\tau}_{ka|r_a=1,q=1}) + (\tfrac{e_a}{n_a})Bias_{\mathbf{T}}(\hat{\tau}_{ka|r_a=1,q=0}) + \\
& (\tfrac{g_a}{n_a})Bias_{\mathbf{A}}(\hat{\tau}_{ka|r_a=1,q=1}) + (\tfrac{e_a}{n_a})Bias_{\mathbf{A}}(\hat{\tau}_{ka|r_a=1,q=0}) + \\
& (\tfrac{g_a}{n_a})Bias_{\mathbf{M}}(\hat{\tau}_{ka|r_a=1,q=1}) + (\tfrac{g_a}{n_a})(\bar{\epsilon}_{kb|r_b=1,q=1} - (\tfrac{g_b n_a}{g_a n_b})\bar{\epsilon}_{kb|r_b=1,q=1}) + \\
& (\tfrac{e_a}{n_a})Bias_{\mathbf{M}}(\hat{\tau}_{ka|r_a=1,q=0}) + (\tfrac{e_a}{n_a})(\bar{\epsilon}_{kb|r_b=1,q=0} - (\tfrac{e_b n_a}{e_a n_b})\bar{\epsilon}_{kb|r_b=1,q=0})
\end{aligned}
$$

$$
\begin{aligned}
Bias(\hat{\tau}_{ka|r_a=1}) =& (\tfrac{g_a}{n_a})Bias(\hat{\tau}_{ka|r_a=1,q=1}) + (\tfrac{e_a}{n_a})Bias(\hat{\tau}_{ka|r_a=1,q=0}) + \\
& (\tfrac{g_a}{n_a})\bar{y}_{ikb|r_b=1,q=1} - (\tfrac{g_b}{n_b})\bar{y}_{ikb|r_b=1,q=1} + (\tfrac{e_a}{n_a})\bar{y}_{ikb|r_b=1,q=0} - (\tfrac{e_b}{n_b})\bar{y}_{ikb|r_b=1,q=0} + \\
& (\tfrac{g_a}{n_a})\bar{\epsilon}_{kb|r_b=1,q=1} - (\tfrac{g_b}{n_b})\bar{\epsilon}_{kb|r_b=1,q=1} + (\tfrac{e_a}{n_a})\bar{\epsilon}_{kb|r_b=1,q=0} - (\tfrac{e_b}{n_b})\bar{\epsilon}_{kb|r_b=1,q=0} \\
Bias(\hat{\tau}_{ka|r_a=1}) =& (\tfrac{g_a}{n_a})Bias(\hat{\tau}_{ka|r_a=1,q=1}) + (\tfrac{e_a}{n_a})Bias(\hat{\tau}_{ka|r_a=1,q=0}) + \\
& (\tfrac{g_a}{n_a})\bar{y}_{ikbo|r_b=1,q=1} - (\tfrac{g_b}{n_b})\bar{y}_{ikbo|r_b=1,q=1} + (\tfrac{e_a}{n_a})\bar{y}_{ikbo|r_b=1,q=0} - (\tfrac{e_b}{n_b})\bar{y}_{ikbo|r_b=1,q=0} \\
Bias(\hat{\tau}_{ka|r_a=1}) =& (\tfrac{g_a}{n_a})Bias(\hat{\tau}_{ka|r_a=1,q=1}) + (\tfrac{e_a}{n_a})Bias(\hat{\tau}_{ka|r_a=1,q=0}) + \\
& (\tfrac{g_a}{n_a})\bar{y}_{ikbo|r_b=1,q=1} - (\tfrac{g_b}{n_b})\bar{y}_{ikbo|r_b=1,q=1} + (1 - \tfrac{g_a}{n_a})\bar{y}_{ikbo|r_b=1,q=0} - (1 - \tfrac{g_b}{n_b})\bar{y}_{ikbo|r_b=1,q=0} \\
Bias(\hat{\tau}_{ka|r_a=1}) =& (\tfrac{g_a}{n_a})Bias(\hat{\tau}_{ka|r_a=1,q=1}) + (\tfrac{e_a}{n_a})Bias(\hat{\tau}_{ka|r_a=1,q=0}) + \\
& (\tfrac{g_a}{n_a} - \tfrac{g_b}{n_b})\bar{y}_{ikbo|r_b=1,q=1} - (\tfrac{g_a}{n_a} - \tfrac{g_b}{n_b})\bar{y}_{ikbo|r_b=1,q=0} \\
Bias(\hat{\tau}_{ka|r_a=1}) =& (\tfrac{g_a}{n_a})Bias(\hat{\tau}_{ka|r_a=1,q=1}) + (\tfrac{e_a}{n_a})Bias(\hat{\tau}_{ka|r_a=1,q=0}) + && (1) \\
& (\tfrac{g_a}{n_a} - \tfrac{g_b}{n_b})(\bar{y}_{ikbo|r_b=1,q=1} - \bar{y}_{ikbo|r_b=1,q=0}) && (2)
\end{aligned}
$$

Therefore, we have shown that the bias in the standard estimator in the baseline model is simply the weighted average of the bias in the estimate from a quota sample and the bias for the sub-sample that would be excluded, along with a residual correction factor.

Quota sampling will then decrease bias if and only if

$$
|Bias(\hat{\tau}_{ka|r_a=1,q=1})| < |Bias(\hat{\tau}_{ka|r_a=1})|
$$

or (utilizing Lines 1-2)

$$
\begin{aligned}
|Bias(\hat{\tau}_{ka|r_a=1,q=1})| < |& (\tfrac{g_a}{n_a})Bias(\hat{\tau}_{ka|r_a=1,q=1}) + (\tfrac{e_a}{n_a})Bias(\hat{\tau}_{ka|r_a=1,q=0}) + \\
& (\tfrac{g_a}{n_a} - \tfrac{g_b}{n_b})(\bar{y}_{ikbo|r_b=1,q=1} - \bar{y}_{ikbo|r_b=1,q=0})|
\end{aligned}
$$

□

Examples of How Quota Sampling Could Amplify Bias

We will begin with a hypothetical phone survey that was carried out in a town before and after an important event. In Wave 1, the survey firm was able to meet its quotas without needing to call anyone twice. However, before the end of Wave 2, the survey firm ran out of people to call and began redialing numbers. Without the quota constraints, the firm may have been able to contact a sufficient number of people without making multiple attempts to reach a single individual. However, the quota constraints in this example would lead to a Wave 2 sample with (on average) harder-to-reach individuals than the Wave 1 sample. These harder-to-reach individuals might differ in many ways from the easier-to-reach individuals, even after conditioning on the covariates that were balanced through the quotas. As such, quota sampling could either reduce or amplify bias compared to convenience sampling, depending on the relationship between the covariates and the potential outcomes.

We can next consider a hypothetical example involving an online survey. In Wave 1, the survey company is able to meet its quotas without an issue. However, in Wave 2, the quota constraints make it difficult for the survey company to obtain a sufficiently large sample. For this reason, the survey company has to work harder, either by advertising the survey more broadly or by offering to pay respondents more. This change in sampling procedures could lead to large demographic differences between the Wave 1 and Wave 2 respondents, for instance on unobservables. Meanwhile, convenience sampling would have resulted in some imbalance on the factors that quota sampling did balance. However, balance on other factors might be much better under convenience sampling than quota sampling. Whether quota sampling or convenience sampling would lead to greater bias would depend on the relationship between the imbalanced factors in each design and the potential outcomes.

2 Rolling Cross-Sections

2.1 Proof of Proposition 3

Proposition 3. *When estimating $\bar{\tau}_{ka|u=1}$, the bias in $\hat{\tau}_{ka|u=1}$ can be written as*

$$Bias(\hat{\tau}_{ka|u=1}) = Bias_{\mathbf{S}}(\hat{\tau}_{ka|u=1}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka|u=1}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=1}) + Bias_{\mathbf{M}}(\hat{\tau}_{ka|u=1})$$

where

$$Bias_{\mathbf{S}}(\hat{\tau}_{ka|u=1}) = \frac{m_a^*}{N_a^*}(\bar{y}_{ikt|w_1=0} - \bar{y}_{ikt|u=1}) + \frac{m_b^*}{N_b^*}(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|w_2=0})$$

$$Bias_{\mathbf{T}}(\hat{\tau}_{ka|u=1}) = \bar{y}_{ikc|u=1} - \bar{y}_{ikbc|u=1}$$

$$Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=1}) = \bar{y}_{ikbc|u=1} - \bar{y}_{ikb|u=1}$$

$$Bias_{\mathbf{M}}(\hat{\tau}_{ka|u=1}) = \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}$$

Proof. Under this design, researchers start with a large group of individuals and randomly assign them to be asked to complete the survey in either Wave 1 or Wave 2. Some of the individuals complete the survey and others do not, sometimes because they are never successfully contacted. We can think about our sample as including a group of always-responders who will complete the survey if asked in either Wave 1 or Wave 2, as well as a group of sometimes-responders who would complete the survey in either Wave 1 or Wave 2 but not both. There may also be some never-responders, but we will put them aside for this analysis since they are inaccessible to us. Let N denote the total number of always-responders and sometimes-responders in the large group that we initially select to be in our study. Further, we can denote the number who actually complete the survey in Wave 1 as n_b and the number who actually complete the survey in Wave 2 as n_a. Among the n_b Wave 1 respondents, we will use n_b^* to denote the number of Wave 1 always-responders and m_b to denote the number of Wave 1 sometimes-responders (who would not have completed the survey if we had tried to ask them in Wave 2). Likewise, among the n_a Wave 2 respondents, we will use n_a^* to denote the number of Wave 2 always-responders and m_a to denote the number of Wave 2 sometimes-responders (who would not have completed the survey if we had tried to ask them in Wave 1). We will denote the total number of respondents by n and the total number of always-responders by n^*.

For clarity, we have these relationships:

$$n^* \leq N$$

$$n_b + n_a = n$$

$$n_b^* + m_b = n_b$$

$$n_a^* + m_a = n_a$$

$$n_b^* + n_a^* = n^*$$

Also note that N and n^* are parameters that do not depend on the randomization. Meanwhile, n_a, n_b, n_a^*, n_b^*, m_a, and m_b are all random variables that depend on the randomization.

We can denote whether individual i is an always-responder (instead of a sometimes-responder) by $u_i \in \{0, 1\}$. We can also denote whether an individual completed the survey in Wave 1 by $r_{ib} \in \{0, 1\}$, and whether they completed the survey in Wave 2 by $r_{ia} \in \{0, 1\}$. Further, for the sometimes-responders ($u = 0$), we can denote whether they would complete the survey in Wave 1 by $w_{i1} \in \{0, 1\}$ and whether they would complete the survey in Wave 2 by $w_{i2} \in \{0, 1\}$.

We might be tempted to think that the causal parameter of interest is the average treatment effect for the Wave 2 respondents:

$$\frac{1}{n_a}\sum_{i=1}^{n_a}(y_{ikt|r_a=1} - y_{ikc|r_a=1}).$$

However, this value is a random variable, not a parameter, since n_a is a random variable. Instead, there are two causal parameters that we might want to estimate. The first is the average treatment effect for the always-responders from Waves 1 and 2:

$$\bar{\tau}_{ka|u=1} = \frac{1}{n^*}\sum_{i=1}^{n^*}(y_{ikt|u=1} - y_{ikc|u=1})$$

The second is the average treatment effect for the combined group of always-responders, sometimes-responders who we sampled, and sometimes-responders who we might have sampled but did not due to the randomization:

$$\bar{\tau}_{ka} = \frac{1}{N}\sum_{i=1}^{N}(y_{ikt} - y_{ikc})$$

The statistic that we will use to estimate both parameters is the average difference in reported answers between the n_a respondents who complete our survey in Wave 2 and the n_b respondents who complete it in Wave 1.

$$\hat{\tau}_{ka|u=1} = \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikto|r_a=1} - \frac{1}{n_b}\sum_{i=1}^{n_b} y_{ikbo|r_b=1}$$

Starting with $\hat{\tau}_{ka|u=1}$, the bias in this estimator is

$$\begin{aligned} Bias(\hat{\tau}_{ka|u=1}) &= E[\hat{\tau}_{ka|u=1}] - \bar{\tau}_{ka|u=1} \\ &= E[\frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikto|r_a=1} - \frac{1}{n_b}\sum_{i=1}^{n_b} y_{ikbo|r_b=1}] - \frac{1}{n^*}\sum_{i=1}^{n^*}(y_{ikt|u=1} - y_{ikc|u=1}) \end{aligned}$$

Taking advantage of the fact that $y_{ikto} = y_{ikt} + \epsilon_{ikt}$ and $y_{ikbo} = y_{ikt} + \epsilon_{ikb}$, we get

$$\begin{aligned}
Bias(\hat{\tau}_{ka|u=1}) &= E[\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikto|r_a=1} - \tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikbo|r_b=1}] - \tfrac{1}{n^*}\sum_{i=1}^{n^*}(y_{ikt|u=1} - y_{ikc|u=1}) \\
&= E[\tfrac{1}{n_a}\sum_{i=1}^{n_a}(y_{ikt|r_a=1} + \epsilon_{ikt|r_a=1}) - \tfrac{1}{n_b}\sum_{i=1}^{n_b}(y_{ikb|r_b=1} + \epsilon_{ikb|r_b=1}))] - \tfrac{1}{n^*}\sum_{i=1}^{n^*}(y_{ikt|u=1} - y_{ikc|u=1}) \\
&= E[\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikt|r_a=1}] + \bar{\epsilon}_{ikt|w_2=1} - E[\tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}] - \bar{\epsilon}_{ikb|w_1=1} - \tfrac{1}{n^*}\sum_{i=1}^{n^*}(y_{ikt|u=1} - y_{ikc|u=1})
\end{aligned}$$

In the last line, we utilize the fact that the n_a values of $y_{ikt|r_a=1}$ that we draw are a random sample from the $y_{ikt|w_2=1}$ values, and likewise the n_b values of $y_{ikbo|r_b=1}$ that we draw are a random sample from the $y_{ikt|w_1=1}$ values.

We can further rewrite the overall bias term as

$$\begin{aligned}
Bias(\hat{\tau}_{ka|u=1}) &= E[\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikt|r_a=1}] - E[\tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}] - \tfrac{1}{n^*}\sum_{i=1}^{n^*}(y_{ikt|u=1} - y_{ikc|u=1}) + \\
&\bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
&= E[\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikt|r_a=1}] - E[\tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}] - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} + \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} + \\
&\bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
&= (E[\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikt|r_a=1}] - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1}) + (\tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - E[\tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}]) + \qquad (3) \\
&\bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \qquad (4)
\end{aligned}$$

We can begin by focusing on the first of the two differences in Equation 3. We can break this expression down as follows:

$$\begin{aligned}
E[\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikt|r_a=1}] - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} &= E[\tfrac{1}{n_a}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1} + \tfrac{1}{n_a}\sum_{i=1}^{m_a} y_{ikt|r_a=1,u=0}] - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} \\
&= E[\tfrac{1}{n_a}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}] + E[\tfrac{1}{n_a}\sum_{i=1}^{m_a} y_{ikt|r_a=1,u=0}] - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} \\
&= E[\tfrac{1}{n_a}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}] - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} + E[\tfrac{1}{n_a}\sum_{i=1}^{m_a} y_{ikt|r_a=1,u=0}] \qquad (5)
\end{aligned}$$

Focusing on the first term in Line 5, note that

$$
\begin{aligned}
E[\tfrac{1}{n_a}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}] &= E[\tfrac{n_a^*}{n_a n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}] \\
&= E[\tfrac{n_a^*+(n_a-n_a)}{n_a n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}] \\
&= E[\tfrac{n_a^*-n_a}{n_a n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}] + E[\tfrac{n_a}{n_a n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}] \\
&= E[\tfrac{-m_a}{n_a n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}] + E[\tfrac{1}{n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}] \qquad (6)
\end{aligned}
$$

The second term in Line 6 is the expected value of a sample of n_a^* draws from the $y_{ikt|u=1}$ values (the y_{ikt} of the always-responders). Therefore, it equals the mean of the $y_{ikt|u=1}$ values.

$$
E[\tfrac{1}{n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}] = \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} \qquad (7)
$$

Combining Lines 3-4, 5, 6, and 7, we get

$$
\begin{aligned}
Bias(\hat{\tau}_{ka|u=1}) = {} & (E[\tfrac{-m_a}{n_a n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}] + \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} + E[\tfrac{1}{n_a}\sum_{i=1}^{m_a} y_{ikt|r_a=1,u=0}])+ \\
& (\tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - E[\tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}]) + \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
= {} & (E[\tfrac{1}{n_a}\sum_{i=1}^{m_a} y_{ikt|r_a=1,u=0}] - E[\tfrac{m_a}{n_a n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}])+ \qquad (8) \\
& (\tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - E[\tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}]) + \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \qquad (9)
\end{aligned}
$$

We can now focus on the first difference in Line 9. We can rewrite this expression as

$$
\begin{aligned}
\tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - E[\tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}] &= \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - (E[\tfrac{1}{n_b}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}] + E[\tfrac{1}{n_b}\sum_{i=1}^{m_b} y_{ikb|r_b=1,u=0}]) \\
&= \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - E[\tfrac{1}{n_b}\sum_{i=1}^{m_b} y_{ikb|r_b=1,u=0}] - E[\tfrac{1}{n_b}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}] \qquad (10)
\end{aligned}
$$

The last term in Line 10 can be written as

$$
\begin{aligned}
E[\tfrac{1}{n_b}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}] &= E[\tfrac{n_b^*}{n_b n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}] \\
&= E[\tfrac{n_b^*+(n_b-n_b)}{n_b n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}] \\
&= E[\tfrac{n_b^*-n_b}{n_b n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}] + E[\tfrac{n_b}{n_b n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}] \\
&= E[\tfrac{-m_b}{n_b n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}] + E[\tfrac{1}{n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}] \qquad (11)
\end{aligned}
$$

Similar to what we did in Line 7, we can note that the second term in Line 11 is the expected value of a sample of n_b^* draws from the $y_{ikb|u=1}$ values (the y_{ikb} of the always-responders). Therefore, it equals the mean of the $y_{ikb|u=1}$ values.

$$
E[\tfrac{1}{n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}] = \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} \qquad (12)
$$

Combining Lines 8-9, 10, 11, and 12, we get

$$
\begin{aligned}
Bias(\hat{\tau}_{ka|u=1}) =&(E[\tfrac{1}{n_a}\sum_{i=1}^{m_a} y_{ikt|r_a=1,u=0}] - E[\tfrac{m_a}{n_a n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}]) + \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - \\
&E[\tfrac{1}{n_b}\sum_{i=1}^{m_b} y_{ikb|r_b=1,u=0}] - (E[\tfrac{-m_b}{n_b n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}] + \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1}) + \\
&\bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
=&(E[\tfrac{1}{n_a}\sum_{i=1}^{m_a} y_{ikt|r_a=1,u=0}] - E[\tfrac{m_a}{n_a n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}]) + \qquad (13) \\
&\tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} + \qquad (14) \\
&(E[\tfrac{m_b}{n_b n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}] - E[\tfrac{1}{n_b}\sum_{i=1}^{m_b} y_{ikb|r_b=1,u=0}]) + \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \qquad (15)
\end{aligned}
$$

The expression in Line 14 is the average difference between the Wave 2 truthful answers of the always-responders in the counterfactual world where the event did not happen and their Wave 1 truthful answers in the world where the event did happen. As we did in Section 2, we can decompose this term into the bias caused by temporal factors between Waves 1 and 2 and the bias caused by anticipatory factors.

$$\frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} = \bar{y}_{ikc|u=1} - \bar{y}_{ikb|u=1} \tag{16}$$

$$= Bias_{\mathbf{T}}(\hat{\tau}_{ka|u=1}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=1}) \tag{17}$$

where

$$Bias_{\mathbf{T}}(\hat{\tau}_{ka|u=1}) = \bar{y}_{ikc|u=1} - \bar{y}_{ikbc|u=1} \tag{18}$$

$$Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=1}) = \bar{y}_{ikbc|u=1} - \bar{y}_{ikb|u=1} \tag{19}$$

Therefore, we can write the overall bias term as

$$Bias(\hat{\tau}_{ka|u=1}) = (E[\frac{1}{n_a}\sum_{i=1}^{m_a} y_{ikt|r_a=1,u=0}] - E[\frac{m_a}{n_a n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}]) + Bias_{\mathbf{T}}(\hat{\tau}_{ka|u=1}) + \tag{20}$$

$$Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=1}) + (E[\frac{m_b}{n_b n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}] - E[\frac{1}{n_b}\sum_{i=1}^{m_b} y_{ikb|r_b=1,u=0}]) + \tag{21}$$

$$\bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \tag{22}$$

We will start with the first difference in Line 20. Note that

$$E[\frac{1}{n_a}\sum_{i=1}^{m_a} y_{ikt|r_a=1,u=0}] - E[\frac{m_a}{n_a n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}] = E[\frac{m_a}{n_a}(\frac{1}{m_a}\sum_{i=1}^{m_a} y_{ikt|r_a=1,u=0}) - \frac{m_a}{n_a}(\frac{1}{n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1})]$$

$$= E[\frac{m_a}{n_a}(\frac{1}{m_a}\sum_{i=1}^{m_a} y_{ikt|r_a=1,u=0} - \frac{1}{n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1})] \tag{23}$$

So the expression inside the parentheses in Line 23 is just the difference of two averages. The outside weight $\frac{m_a}{n_a}$ is the proportion of Wave 2 respondents who are sometimes-responders. Inside the parentheses, the first average is the mean y_{ikt} value of the Wave 2 respondents who are sometimes-responders, and the second average is the mean y_{ikt} value of the Wave 2 respondents who are always-responders.

To progress further, let us consider a new parameter m_a^*, which will denote the number of sometimes-responders who would complete the survey if they were assigned to take it in Wave 2. We can then let $N_a^* = n^* + m_a^*$ denote the total number of possible Wave 2 respondents who might be in our sample. Likewise, we can denote the number of sometimes-responders who would only complete the survey if they were assigned to take it in Wave 1 as m_b^*.

The number of possible Wave 1 respondents who might be in our sample can then be denoted by $N_b^* = n^* + m_b^*$. These new parameters relate back to N, the total number of always-responders and sometimes-responders in the large group that we initially select to be in our study, by the equation $n^* + m_b^* + m_a^* = N$.

Therefore, the first average inside the parentheses in Line 23 is the average of a random draw of the m_a^* sometimes responders who would only complete the survey in Wave 2. We can use p to denote the probability that an individual will initially be randomized to be contacted in Wave 1, making $1 - p$ their initial probability of being contacted in Wave 2. Then the expected numbers of always-responders and sometimes-responders who will complete the survey in Wave 1 and Wave 2 are

$$
\begin{aligned}
E[n_b^*] &= pn^* \\
E[n_a^*] &= (1-p)n^* \\
E[m_b] &= pm_b^* \\
E[m_a] &= (1-p)m_a^*
\end{aligned}
$$

Similarly, the expected numbers of Wave 1 and Wave 2 respondents are

$$
\begin{aligned}
E[n_b] &= E[n_b^* + m_b] = E[n_b^*] + E[m_b] = p(n^* + m_b^*) \\
E[n_a] &= E[n_a^* + m_a] = E[n_a^*] + E[m_a] = (1-p)(n^* + m_a^*)
\end{aligned}
$$

Also, the expected proportion of sometimes-responders in Waves 1 and 2 is

$$
\begin{aligned}
E[\tfrac{m_b}{n_b}] &= \tfrac{m_b^*}{N_b^*} \\
E[\tfrac{m_a}{n_a}] &= \tfrac{m_b^*}{N_a^*}
\end{aligned}
$$

Returning to the expression in Line 23, the proportion of Wave 2 respondents who are sometimes-responders is statistically independent of the mean y_{itk} value of these sometimes-responders. Likewise, it is statistically independent of the mean y_{itk} value of the Wave 2 respondents who are always-responders. We therefore have

$$
\begin{aligned}
E[\tfrac{m_a}{n_a}(\tfrac{1}{m_a}\sum_{i=1}^{m_a} y_{ikt|r_a=1,u=0}) - \tfrac{1}{n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1})] &= E[\tfrac{m_a}{n_a}]E[\tfrac{1}{m_a}\sum_{i=1}^{m_a} y_{ikt|r_a=1,u=0} - \tfrac{1}{n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}] \\
&= \tfrac{m_a^*}{N_a^*}(E[\tfrac{1}{m_a}\sum_{i=1}^{m_a} y_{ikt|r_a=1,u=0}] - E[\tfrac{1}{n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}]) \quad (24)
\end{aligned}
$$

Inside the parentheses of Line 24, the first term is the average of m_a random draws from the y_{ikt} values of the m_a^* sometimes-responders who would only complete the survey in Wave 2. Similarly, the second term is the average of n_a^* random draws from the y_{ikt} values of the n^* always-responders. Therefore, we have

$$E[\frac{1}{m_a}\sum_{i=1}^{m_a} y_{ikt|r_a=1,u=0}] = \frac{1}{m_a^*}\sum_{i=1}^{m_a^*} y_{ikt|u=0,w_2=1}$$

and

$$E[\frac{1}{n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}] = \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1}$$

We can then write

$$\frac{m_a^*}{N_a^*}(E[\frac{1}{m_a}\sum_{i=1}^{m_a} y_{ikt|r_a=1,u=0}] - E[\frac{1}{n_a^*}\sum_{i=1}^{n_a^*} = y_{ikt|r_a=1,u=1}]) = \frac{m_a^*}{N_a^*}(\frac{1}{m_a^*}\sum_{i=1}^{m_a^*} y_{ikt|u=0,w_2=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1})$$

Substituting this expression into the overall bias term, we get

$$Bias(\hat{\tau}_{ka|u=1}) = \frac{m_a^*}{N_a^*}(\frac{1}{m_a^*}\sum_{i=1}^{m_a^*} y_{ikt|u=0,w_2=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka|u=1}) + \tag{25}$$

$$Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=1}) + (E[\frac{m_b}{n_b n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}] - E[\frac{1}{n_b}\sum_{i=1}^{m_b} y_{ikb|r_b=1,u=0}]) + \tag{26}$$

$$\bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \tag{27}$$

We can now turn to the difference in Line 26. Similar to before, we can begin by noting that

$$\begin{aligned} E[\frac{m_b}{n_b n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}] - E[\frac{1}{n_b}\sum_{i=1}^{m_b} y_{ikb|r_b=1,u=0}] &= E[\frac{m_b}{n_b}(\frac{1}{n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}) - \frac{m_b}{n_b}(\frac{1}{m_b}\sum_{i=1}^{m_b} y_{ikb|r_b=1,u=0})] \\ &= E[\frac{m_b}{n_b}(\frac{1}{n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1} - \frac{1}{m_b}\sum_{i=1}^{m_b} y_{ikb|r_b=1,u=0})] \end{aligned} \tag{28}$$

As in Line 23, the expression inside the parentheses in Line 28 is just the difference of two averages. The outside weight $\frac{m_b}{n_b}$ is the proportion of Wave 1 respondents who are sometimes-responders. Inside the parentheses, the first average is the mean y_{ikt} value of the Wave 1 respondents who are always-responders, and the second average is the mean y_{ikt} value of the Wave 1 respondents who are sometimes-responders. Since in this context the weight is statistically independent of the averages, as explained earlier, we can write

$$E[\frac{m_b}{n_b}(\frac{1}{n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1} - \frac{1}{m_b}\sum_{i=1}^{m_b} y_{ikb|r_b=1,u=0})] = E[\frac{m_b}{n_b}]E[\frac{1}{n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1} - \frac{1}{m_b}\sum_{i=1}^{m_b} y_{ikb|r_b=1,u=0}]$$

$$= \frac{m_b^*}{N_b^*}(E[\frac{1}{n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}] - E[\frac{1}{m_b}\sum_{i=1}^{m_b} y_{ikb|r_b=1,u=0}]) \quad (29)$$

Inside the parentheses of Line 29, the first term is the average of n_b^* random draws from the y_{ikt} values of the n^* always-responders. Likewise, the second term is the average of m_b random draws from the y_{ikt} values of the m_b^* sometimes-responders who would only complete the survey in Wave 1. Therefore, we have

$$E[\frac{1}{n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}] = \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1}$$

and

$$E[\frac{1}{m_b}\sum_{i=1}^{m_b} y_{ikb|r_b=1,u=0}] = \frac{1}{m_b^*}\sum_{i=1}^{m_b^*} y_{ikb|u=0,w_1=1}$$

We can then write

$$\frac{m_b^*}{N_b^*}(E[\frac{1}{n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}] - E[\frac{1}{m_b}\sum_{i=1}^{m_b} y_{ikb|r_b=1,u=0}]) = \frac{m_b^*}{N_b^*}(\frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} - \frac{1}{m_b^*}\sum_{i=1}^{m_b^*} y_{ikb|u=0,w_1=1})$$

Substituting this expression into the overall bias term allows us to write

$$Bias(\hat{\tau}_{ka|u=1}) = \frac{m_a^*}{N_a^*}(\frac{1}{m_a^*}\sum_{i=1}^{m_a^*} y_{ikt|u=0,w_2=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka|u=1}) +$$

$$Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=1}) + \frac{m_b^*}{N_b^*}(\frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} - \frac{1}{m_b^*}\sum_{i=1}^{m_b^*} y_{ikb|u=0,w_1=1}) +$$

$$\bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}$$

$$= \frac{m_a^*}{N_a^*}(\bar{y}_{ikt|u=0,w_2=1} - \bar{y}_{ikt|u=1}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka|u=1}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=1}) +$$

$$\frac{m_b^*}{N_b^*}(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|u=0,w_1=1}) + \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}$$

$$= \frac{m_a^*}{N_a^*}(\bar{y}_{ikt|u=0,w_2=1} - \bar{y}_{ikt|u=1}) + \frac{m_b^*}{N_b^*}(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|u=0,w_1=1}) + \quad (30)$$

$$Bias_{\mathbf{T}}(\hat{\tau}_{ka|u=1}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=1}) + \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \quad (31)$$

The first expression in Line 30 is the proportion of possible Wave 2 respondents who are sometimes-responders multiplied by the average difference between these sometimes-responders' y_{ikt} values and the always-responders' y_{ikt} values. The second expression in Line 30 is the proportion of possible Wave 1 respondents who are sometimes-responders multiplied by the average difference between the always-responders' y_{ikb} values and these potential Wave 1 sometimes-responders' y_{ikb} values. Therefore, we can think of the sum of these two expressions as the bias caused by having sometimes-responders in the Wave 1 and Wave 2 samples.

$$Bias_{\mathbf{S}}(\hat{\tau}_{ka|u=1}) = \frac{m_a^*}{N_a^*}(\bar{y}_{ikt|u=0,w_2=1} - \bar{y}_{ikt|u=1}) + \frac{m_b^*}{N_b^*}(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|u=0,w_1=1})$$

We now have

$$Bias(\hat{\tau}_{ka|u=1}) = Bias_{\mathbf{S}}(\hat{\tau}_{ka|u=1}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka|u=1}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=1}) + \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \tag{32}$$

The expression $\bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}$ is the potential bias caused by differential misreporting in the answers of the possible Wave 1 and Wave 2 respondents. We can define it as

$$Bias_{\mathbf{M}}(\hat{\tau}_{ka|u=1}) = \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}$$

In sum, we can write write the overall bias as

$$Bias(\hat{\tau}_{ka|u=1}) = Bias_{\mathbf{S}}(\hat{\tau}_{ka|u=1}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka|u=1}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=1}) + Bias_{\mathbf{M}}(\hat{\tau}_{ka|u=1}) \tag{33}$$

where

$$Bias_{\mathbf{S}}(\hat{\tau}_{ka|u=1}) = \frac{m_a^*}{N_a^*}(\bar{y}_{ikt|u=0,w_2=1} - \bar{y}_{ikt|u=1}) + \frac{m_b^*}{N_b^*}(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|u=0,w_1=1})$$

$$Bias_{\mathbf{T}}(\hat{\tau}_{ka|u=1}) = \bar{y}_{ikc|u=1} - \bar{y}_{ikbc|u=1}$$

$$Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=1}) = \bar{y}_{ikbc|u=1} - \bar{y}_{ikb|u=1}$$

$$Bias_{\mathbf{M}}(\hat{\tau}_{ka|u=1}) = \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}$$

□

2.2 Proof of Proposition 4

Proposition 4. *When estimating $\bar{\tau}_{ka}$, the bias in $\hat{\tau}_{ka}$ is*

$$Bias(\hat{\tau}_{ka}) = Bias_{\mathbf{S}}(\hat{\tau}_{ka}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka}) + Bias_{\mathbf{M}}(\hat{\tau}_{ka})$$

where,

$$\begin{aligned}
&Bias_{\mathbf{S}}(\hat{\tau}_{ka}) = \tfrac{m_b^*}{N}(\bar{y}_{ikt|w_2=1} - \bar{y}_{ikt|w_2=0}) + \tfrac{m_a^*}{N}(\bar{y}_{ikb|w_1=0} - \bar{y}_{ikb|w_1=1}) \\
&Bias_{\mathbf{T}}(\hat{\tau}_{ka}) = \bar{y}_{ikc} - \bar{y}_{ikbc} \\
&Bias_{\mathbf{A}}(\hat{\tau}_{ka}) = \bar{y}_{ikbc} - \bar{y}_{ikb} \\
&Bias_{\mathbf{M}}(\hat{\tau}_{ka}) = \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}
\end{aligned}$$

Proof. We can now turn to estimating the bias when our target parameter is the average treatment effect of the combined group of always-responders, sometimes-responders who we sampled, and sometimes-responders who we might have sampled but did not due to the randomization:

$$\bar{\tau}_{ka} = \tfrac{1}{N}\sum_{i=1}^{N}(y_{ikt} - y_{ikc}) \tag{34}$$

To estimate this parameter, we will use the same estimator as before: the average difference in reported outcomes between the n_a respondents who complete our survey in Wave 2 and the n_b respondents who complete it in Wave 1.

$$\hat{\tau}_{ka} = \tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikto|r_a=1} - \tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikbo|r_b=1}$$

The bias in this estimator is therefore

$$
\begin{aligned}
Bias(\hat{\tau}_{ka}) &= E[\hat{\tau}_{ka}] - \bar{\tau}_{ka} \\
&= E[\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikto|r_a=1} - \tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikbo|r_b=1}] - \tfrac{1}{N}\sum_{i=1}^{N}(y_{ikt} - y_{ikc}) \\
&= E[\tfrac{1}{n_a}\sum_{i=1}^{n_a}(y_{ikt|r_a=1} + \epsilon_{ikt|r_a=1}) - \tfrac{1}{n_b}\sum_{i=1}^{n_b}(y_{ikb|r_b=1} + \epsilon_{ikb|r_b=1}))] - \tfrac{1}{N}\sum_{i=1}^{N}(y_{ikt} - y_{ikc}) \\
&= E[\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikt|r_a=1}] + \bar{\epsilon}_{ikt|w_2=1} - E[\tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}] - \bar{\epsilon}_{ikb|w_1=1} - \tfrac{1}{N}\sum_{i=1}^{N}(y_{ikt} - y_{ikc}) \\
&= E[\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikt|r_a=1}] - E[\tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}] - \tfrac{1}{N}\sum_{i=1}^{N} y_{ikt} + \tfrac{1}{N}\sum_{i=1}^{N} y_{ikc} + \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
&= (E[\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikt|r_a=1}] - \tfrac{1}{N}\sum_{i=1}^{N} y_{ikt}) + (\tfrac{1}{N}\sum_{i=1}^{N} y_{ikc} - E[\tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}]) + \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \qquad (35)
\end{aligned}
$$

Note that, like before, $w_1 \in \{0, 1\}$ denotes whether an individual would complete the survey if assigned to Wave 1 and $w_2 \in \{0, 1\}$ denotes whether they would complete the survey if assigned to Wave 2. Focusing on the first difference in Line 35, we can rewrite the expression as

$$
\begin{aligned}
E[\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikt|r_a=1}] - \tfrac{1}{N}\sum_{i=1}^{N} y_{ikt} &= \tfrac{1}{N_a^*}\sum_{i=1}^{N_a^*} y_{ikt|w_2=1} - \tfrac{1}{N}\sum_{i=1}^{N} y_{ikt} \\
&= \bar{y}_{ikt|w_2=1} - \bar{y}_{ikt} \qquad (36)
\end{aligned}
$$

Similarly, the middle difference in Line 35 can be rewritten as

$$
\begin{aligned}
\tfrac{1}{N}\sum_{i=1}^{N} y_{ikc} - E[\tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}] &= \tfrac{1}{N}\sum_{i=1}^{N} y_{ikc} - \tfrac{1}{N_b^*}\sum_{i=1}^{N_b^*} y_{ikb|w_1=1} \qquad (37) \\
&= \tfrac{1}{N}\sum_{i=1}^{N} y_{ikc} - \tfrac{1}{N_b^*}\sum_{i=1}^{N_b^*} y_{ikb|w_1=1} \qquad (38) \\
&= \tfrac{1}{N}\sum_{i=1}^{N} y_{ikc} - \tfrac{1}{N_b^*}\sum_{i=1}^{N_b^*} y_{ikb|w_1=1} + \tfrac{1}{N}\sum_{i=1}^{m_a^*} y_{ikb|w_1=0} - \tfrac{1}{N}\sum_{i=1}^{m_a^*} y_{ikb|w_1=0} \qquad (39) \\
&= \tfrac{1}{N}\sum_{i=1}^{N} y_{ikc} - (\tfrac{1}{N_b^*}\sum_{i=1}^{N_b^*} y_{ikb|w_1=1} + \tfrac{1}{N}\sum_{i=1}^{m_a^*} y_{ikb|w_1=0}) + \tfrac{1}{N}\sum_{i=1}^{m_a^*} y_{ikb|w_1=0} \qquad (40)
\end{aligned}
$$

We can rewrite the expression inside the parentheses in Line 40 as

$$
\begin{aligned}
\frac{1}{N_b^*}\sum_{i=1}^{N_b^*} y_{ikb|w_1=1} + \frac{1}{N}\sum_{i=1}^{m_a^*} y_{ikb|w_1=0} =& \frac{N}{N_b^* N}\sum_{i=1}^{N_b^*} y_{ikb|w_1=1} + \frac{1}{N}\sum_{i=1}^{m_a^*} y_{ikb|w_1=0} \\
=& \frac{N+(N_b^*-N_b^*)}{N_b^* N}\sum_{i=1}^{N_b^*} y_{ikb|w_1=1} + \frac{1}{N}\sum_{i=1}^{m_a^*} y_{ikb|w_1=0} \\
=& \frac{N-N_b^*}{N_b^* N}\sum_{i=1}^{N_b^*} y_{ikb|w_1=1} + \frac{N_b^*}{N_b^* N}\sum_{i=1}^{N_b^*} y_{ikb|w_1=1} + \frac{1}{N}\sum_{i=1}^{m_a^*} y_{ikb|w_1=0} \\
=& \frac{N-N_b^*}{N_b^* N}\sum_{i=1}^{N_b^*} y_{ikb|w_1=1} + \frac{1}{N}\sum_{i=1}^{N_b^*} y_{ikb|w_1=1} + \frac{1}{N}\sum_{i=1}^{m_a^*} y_{ikb|w_1=0} \\
=& \frac{N-N_b^*}{N_b^* N}\sum_{i=1}^{N_b^*} y_{ikb|w_1=1} + \frac{1}{N}\sum_{i=1}^{N} y_{ikb}
\end{aligned}
$$

Combining Lines 36, 40, and 41, we get

$$
\begin{aligned}
Bias(\hat{\tau}_{ka}) =& (\bar{y}_{ikt|w_2=1} - \bar{y}_{ikt}) + \frac{1}{N}\sum_{i=1}^{N} y_{ikc} - \Big(\frac{N-N_b^*}{N_b^* N}\sum_{i=1}^{N_b^*} y_{ikb|w_1=1} + \frac{1}{N}\sum_{i=1}^{N} y_{ikb}\Big) + \frac{1}{N}\sum_{i=1}^{m_a^*} y_{ikb|w_1=0} + \\
& \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
=& (\bar{y}_{ikt|w_2=1} - \bar{y}_{ikt}) + \frac{1}{N}\sum_{i=1}^{N} y_{ikc} - \frac{1}{N}\sum_{i=1}^{N} y_{ikb} + \frac{N_b^*-N}{N_b^* N}\sum_{i=1}^{N_b^*} y_{ikb|w_1=1} + \frac{1}{N}\sum_{i=1}^{m_a^*} y_{ikb|w_1=0} + \\
& \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
=& (\bar{y}_{ikt|w_2=1} - \bar{y}_{ikt}) + (\bar{y}_{ikc} - \bar{y}_{ikb}) + \frac{N_b^*-N}{N_b^* N}\sum_{i=1}^{N_b^*} y_{ikb|w_1=1} + \frac{1}{N}\sum_{i=1}^{m_a^*} y_{ikb|w_1=0} + \\
& \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}
\end{aligned}
$$

Similar to our discussion in Section 3 of the paper, this difference $\bar{y}_{ikc} - \bar{y}_{ikb}$ can be decomposed into the bias from temporal factors and the bias from anticipatory factors. Thus, we can write

$$\bar{y}_{ikc} - \bar{y}_{ikb} = Bias_{\mathbf{T}}(\hat{\tau}_{ka}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka})$$

where

$$Bias_{\mathbf{T}}(\hat{\tau}_{ka}) = \bar{y}_{ikc} - \bar{y}_{ikbc}$$

$$Bias_{\mathbf{A}}(\hat{\tau}_{ka}) = \bar{y}_{ikbc} - \bar{y}_{ikb}$$

We can now write the overall bias term as:

$$Bias(\hat{\tau}_{ka}) = (\bar{y}_{ikt|w_2=1} - \bar{y}_{ikt}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka}) + \tag{41}$$

$$\left(\frac{N_b^* - N}{N_b^* N} \sum_{i=1}^{N_b^*} y_{ikb|w_1=1} + \frac{1}{N} \sum_{i=1}^{m_a^*} y_{ikb|w_1=0}\right) + \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \tag{42}$$

The first difference in Line 42 can be written as

$$\begin{aligned}
\frac{N_b^* - N}{N_b^* N} \sum_{i=1}^{N_b^*} y_{ikb|w_1=1} + \frac{1}{N} \sum_{i=1}^{m_a^*} y_{ikb|w_1=0} &= \frac{1}{N} \sum_{i=1}^{N_b^*} y_{ikb|w_1=1} - \frac{1}{N_b^*} \sum_{i=1}^{N_b^*} y_{ikb|w_1=1} + \frac{1}{N} \sum_{i=1}^{m_a^*} y_{ikb|w_1=0} \\
&= \frac{1}{N} \sum_{i=1}^{N_b^*} y_{ikb|w_1=1} + \frac{1}{N} \sum_{i=1}^{m_a^*} y_{ikb|w_1=0} - \frac{1}{N_b^*} \sum_{i=1}^{N_b^*} y_{ikb|w_1=1} \\
&= \frac{1}{N} \sum_{i=1}^{N} y_{ikb} - \frac{1}{N_b^*} \sum_{i=1}^{N_b^*} y_{ikb|w_1=1} \\
&= \bar{y}_{ikb} - \bar{y}_{ikb|w_1=1}
\end{aligned}$$

We can now write the overall bias term as

$$\begin{aligned}
Bias(\hat{\tau}_{ka}) &= (\bar{y}_{ikt|w_2=1} - \bar{y}_{ikt}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka}) + (\bar{y}_{ikb} - \bar{y}_{ikb|w_1=1}) + \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
&= (\bar{y}_{ikt|w_2=1} - \bar{y}_{ikt}) + (\bar{y}_{ikb} - \bar{y}_{ikb|w_1=1}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka}) + \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \qquad (43)
\end{aligned}$$

The differences within the two sets of parentheses in Line 43 come from not being able to see any y_{ikto} values for the Wave 1 sometimes-responders nor any of the y_{ikbo} values for the Wave 2 sometimes-responders. We can think of this bias as arising from having sometimes-responders in our sample who differ in systematic ways from the always-responders and the sometimes-responders who answer the survey in the other wave. We can rewrite this bias term as as

$$(\bar{y}_{ikt|w_2=1} - \bar{y}_{ikb|w_1=1}) - (\bar{y}_{ikt} - \bar{y}_{ikb}) = (\bar{y}_{ikt|w_2=1} - \bar{y}_{ikb|w_1=1}) - \left(\frac{N_a^*}{N}\right)\bar{y}_{ikt|w_2=1} - \left(\frac{m_b^*}{N}\right)\bar{y}_{ikt|w_2=0} + \tag{44}$$

$$\left(\frac{N_b^*}{N}\right)\bar{y}_{ikb|w_1=1} + \left(\frac{m_a^*}{N}\right)\bar{y}_{ikb|w_1=0} \tag{45}$$

$$= \frac{m_b^*}{N}(\bar{y}_{ikt|w_2=1} - \bar{y}_{ikt|w_2=0}) + \frac{m_a^*}{N}(\bar{y}_{ikb|w_1=0} - \bar{y}_{ikb|w_1=1}) \tag{46}$$

We can label this bias

$$Bias_{\mathbf{S}}(\hat{\tau}_{ka}) = \frac{m_b^*}{N}(\bar{y}_{ikt|w_2=1} - \bar{y}_{ikt|w_2=0}) + \frac{m_a^*}{N}(\bar{y}_{ikb|w_1=0} - \bar{y}_{ikb|w_1=1})$$

We can now write the overall bias term as

$$Bias(\hat{\tau}_{ka}) = Bias_{\mathbf{S}}(\hat{\tau}_{ka}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka}) + Bias_{\mathbf{M}}(\hat{\tau}_{ka}) + \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}$$

As in the previous proof, $\bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}$ is the potential bias caused by differential misreporting in the answers of the possible Wave 1 and Wave 2 respondents. We can define it as

$$Bias_{\mathbf{M}}(\hat{\tau}_{ka|u=1}) = \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}$$

In sum, we can write the overall bias as

$$Bias(\hat{\tau}_{ka}) = Bias_{\mathbf{S}}(\hat{\tau}_{ka}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka}) + Bias_{\mathbf{M}}(\hat{\tau}_{ka}) \tag{47}$$

where

$$\begin{aligned}
Bias_{\mathbf{S}}(\hat{\tau}_{ka}) &= \frac{m_b^*}{N}(\bar{y}_{ikt|w_2=1} - \bar{y}_{ikt|w_2=0}) + \frac{m_a^*}{N}(\bar{y}_{ikb|w_1=0} - \bar{y}_{ikb|w_1=1}) \\
Bias_{\mathbf{T}}(\hat{\tau}_{ka}) &= \bar{y}_{ikc} - \bar{y}_{ikbc} \\
Bias_{\mathbf{A}}(\hat{\tau}_{ka}) &= \bar{y}_{ikbc} - \bar{y}_{ikb} \\
Bias_{\mathbf{M}}(\hat{\tau}_{ka}) &= \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}
\end{aligned}$$

□

Before proceeding, we can consider the special case where there are no always-responders. We can think of this scenario as the baseline model but when the parameter that we are estimating is the ATE, not the ATT. The $Bias_{\mathbf{S}}(\hat{\tau}_{ka})$ term becomes a weighted average of the baseline and post-event differences between the Wave 1 and Wave 2 respondents' truthful answers. Clearly, bias when estimating the ATT in the baseline model is more straightforward to comprehend than bias when estimating the ATE in the baseline model.

We can now examine how the bias in the estimators $\hat{\tau}_{ka|s=1}$ and $\hat{\tau}_{ka}$ from the rolling cross-section design compare to the bias in the baseline model from the paper. The rolling cross-section design trades the bias in demographic differences between Wave 1 and Wave 2 respondents for the bias caused by sometimes-responders. Focusing on Equation 33, we can better understand $Bias_{\mathbf{S}}(\hat{\tau}_{ka|u=1})$ by considering the special case where the initial numbers of sometimes-responders in Waves 1 and 2 are the same ($m_a^* = m_b^*$). In that case, $\frac{m_a^*}{N_a^*} = \frac{m_b^*}{N_b^*}$, which we will denote as $\alpha < 1$. This symmetry allows us to rewrite $Bias_{\mathbf{S}}(\hat{\tau}_{ka|u=1})$ as

$$\begin{aligned} Bias_{\mathbf{S}}(\hat{\tau}_{ka|u=1}) &= \tfrac{m_a^*}{N_a^*}(\bar{y}_{ikt|u=0,w_2=1} - \bar{y}_{ikt|u=1}) + \tfrac{m_b^*}{N_b^*}(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|u=0,w_1=1}) \\ &= \alpha(\bar{y}_{ikt|u=0,w_2=1} - \bar{y}_{ikt|u=1}) + \alpha(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|u=0,w_1=1}) \\ &= \alpha(\bar{y}_{ikt|u=0,w_2=1} - \bar{y}_{ikb|u=0,w_1=1} - \bar{y}_{ikt|u=1} - \bar{y}_{ikb|u=1}) \\ &= \alpha[(\bar{y}_{ikt|w_1=0} - \bar{y}_{ikb|w_2=0}) - (\bar{y}_{ikt|u=1} - \bar{y}_{ikb|u=1})] \end{aligned} \tag{48}$$

Note that the terms inside both sets of parentheses in Line 48 resemble our estimator in the baseline model of Section 2 of the paper. In fact, they are equivalent to that estimator in the special case where there is no measurement error. The first term is simply the standard estimator $\hat{\tau}_{ka|r_a=1}$ without measurement error on a sample consisting entirely of sometimes-responders. The second term is the same estimator on a sample consisting entirely of always-responders, except in a world where the Wave 1 and Wave 2 individuals are identical on demographic characteristics. Since there is no randomness in the baseline model, we can think of both estimators as the average treatment effect for that sub-sample combined with the corresponding bias term (following from $Bias(\hat{\tau}) = E[\hat{\tau}] - \bar{\tau}$). We can therefore write

$$\begin{aligned} Bias_{\mathbf{S}}(\hat{\tau}_{ka|u=1}|m_a^* = m_b^*) &= \alpha[(\bar{\tau}_{ka|u=0} + Bias_{\mathbf{X}}(\hat{\tau}_{ka|u=0}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka|u=0}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=0})) - \\ &\quad (\bar{\tau}_{ka|u=1} + Bias_{\mathbf{X}}(\hat{\tau}_{ka|u=1}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka|u=1}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=1}))] \\ &= \alpha[(\bar{\tau}_{ka|u=0} + Bias_{\mathbf{X}}(\hat{\tau}_{ka|u=0}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka|u=0}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=0})) - \\ &\quad (\bar{\tau}_{ka|u=1} + Bias_{\mathbf{T}}(\hat{\tau}_{ka|u=1}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=1}))] \end{aligned}$$

Substituting this expression into Equation 33, we obtain

$$Bias(\hat{\tau}_{ka|u=1}|m_a^* = m_b^*) = \alpha Bias_{\mathbf{X}}(\hat{\tau}_{ka|u=0}) + (\alpha Bias_{\mathbf{T}}(\hat{\tau}_{ka|u=0}) + (1-\alpha) Bias_{\mathbf{T}}(\hat{\tau}_{ka|=1})) + \tag{49}$$

$$(\alpha Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=0}) + (1-\alpha) Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=1})) + Bias_{\mathbf{M}}(\hat{\tau}_{ka|u=1}) + \tag{50}$$

$$\alpha(\bar{\tau}_{ka|u=0} - \bar{\tau}_{ka|u=1}) \tag{51}$$

The expression involving bias from temporal factors is just a weighted average of the temporal bias for the sometimes-responders and always-responders. The same logic holds for the expression involving the bias from anticipatory factors. We also add a new bias term involving the difference in average treatment effect between the sometimes-responders and always-responders.

In sum, the rolling cross section estimator reduces demographic bias by a factor of α, but it also complicates the rest of the overall bias term in ways that could either decrease or enlarge the total bias in this design.

If we instead use Equation 47 and consider the special case where the initial numbers of possible Wave 1 and Wave 2 sometimes-responders are the same ($m_a^* = m_b^*$), then the weights we obtain $\frac{m_b^*}{N}$ and $\frac{m_a^*}{N}$ will again be equal. In this situation, we can define $\lambda = \frac{m_b^*}{N} = \frac{m_a^*}{N}$. We can then rewrite the equation for $Bias_{\mathbf{S}}(\hat{\tau}_{ka}|m_a^* = m_b^*)$ as

$$Bias_{\mathbf{S}}(\hat{\tau}_{ka}|m_a^* = m_b^*) = \lambda[(\bar{y}_{ikt|w_2=1} - \bar{y}_{ikb|w_1=1}) - (\bar{y}_{ikt} - \bar{y}_{ikb})]$$

Here we are utilizing the way that we wrote $Bias_{\mathbf{S}}(\hat{\tau}_{ka})$ at the beginning of Line 44.

This expression is very similar to what we saw in Line 48. Like before, we can think about each of the two differences inside the brackets as mathematically similar to the estimator from the baseline model, specifically in the case where there is no measurement error. Like before, we can think of these two estimators as the average treatment effect for that sample combined with the corresponding bias term. We can therefore write

$$Bias_{\mathbf{S}}(\hat{\tau}_{ka}|m_a^* = m_b^*) = \lambda[(\bar{\tau}_{ka|w_2=1} + Bias_{\mathbf{X}}(\hat{\tau}_{ka|w_2=1}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka|w_2=1}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka|w_2=1})) - \quad (52)$$

$$(\bar{\tau}_{ka} + Bias_{\mathbf{X}}(\hat{\tau}_{ka}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka}))] \quad (53)$$

Note that in Line 52, $Bias_{\mathbf{X}}(\hat{\tau}_{ka|w_2=1}) = \bar{y}_{ikb|w_2=1} - \bar{y}_{ikb|w_1=1}$. Since the pre-event and post-event samples in Line 53 consist of N individuals with exactly the same demographic characteristics, we can drop the $Bias_{\mathbf{X}}(\hat{\tau}_{ka})$ term.

$$Bias_{\mathbf{S}}(\hat{\tau}_{ka}|m_a^* = m_b^*) = \lambda[(\bar{\tau}_{ka|w_2=1} + Bias_{\mathbf{X}}(\hat{\tau}_{ka|w_2=1}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka|w_2=1}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka|w_2=1})) -$$
$$(\bar{\tau}_{ka} + Bias_{\mathbf{T}}(\hat{\tau}_{ka}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka}))]$$

Substituting this expression into Line 47, we get

$$Bias(\hat{\tau}_{ka}) = \lambda Bias_{\mathbf{X}}(\hat{\tau}_{ka|w_2=1}) + (\lambda Bias_{\mathbf{T}}(\hat{\tau}_{ka|w_2=1}) + (1-\lambda) Bias_{\mathbf{T}}(\hat{\tau}_{ka})) +$$
$$(\lambda Bias_{\mathbf{A}}(\hat{\tau}_{ka|w_2=1}) + (1-\lambda) Bias_{\mathbf{A}}(\hat{\tau}_{ka})) + Bias_{\mathbf{M}}(\hat{\tau}_{ka}) + \lambda(\bar{\tau}_{ka|w_2=1} - \bar{\tau}_{ka})$$

3 Panel Designs

In a panel design, we have N individuals who have the opportunity to take the same survey in Wave 1 and Wave 2. We can denote the number who take the survey in both waves by $u_i \in \{0, 1\}$ and the total number of respondents who take the survey in both waves as n^*. The causal parameter that we estimate is the average treatment effect of the event on these n^* respondents' truthful answers to question k of the survey:

$$\bar{\tau}_{ka|u=1} = \frac{1}{n^*}\sum_{i=1}^{n^*}(y_{ikt|u=1} - y_{ikc|u=1})$$

In the above line, we consider $y_{ikt|u=1}$ to be individual i's truthful answer in the world where they did not complete the survey in Wave 1. We can distinguish this value from $y_{ika|u=1}$, which we use to denote individual i's truthful answer in the world where they did complete the survey in Wave 1.

The statistic we use to estimate $\bar{\tau}_{ka|u=1}$ is

$$\hat{\tau}_{ka|u=1} = \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikao|u=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikbo|u=1}$$

In the above line, we use $y_{ikao|u=1}$ to denote individual i's reported answer in Wave 2 after having already completed the survey in Wave 1.

3.1 Proof of Proposition 5

Proposition 5. *Bias in the panel design can be written as*

$$Bias(\hat{\tau}_{ka|u=1}) = Bias_{\mathbf{C}}(\hat{\tau}_{ka|u=1}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka|u=1}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=1}) + Bias_{\mathbf{M}}(\hat{\tau}_{ka|u=1})$$

where

$$Bias_{\mathbf{C}}(\hat{\tau}_{ka|u=1}) = \bar{y}_{ika|u=1} - \bar{y}_{ikt|u=1}$$

$$Bias_{\mathbf{T}}(\hat{\tau}_{ka|u=1}) = \bar{y}_{ikc|u=1} - \bar{y}_{ikbc|u=1}$$

$$Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=1}) = \bar{y}_{ikbc|u=1} - \bar{y}_{ikb|u=1}$$

$$Bias_{\mathbf{M}}(\hat{\tau}_{ka|u=1}) = \bar{\epsilon}_{ka|u=1} - \bar{\epsilon}_{kb|u=1}$$

Proof. The bias in $\hat{\tau}_{ka|u=1}$ is just

$$\begin{aligned}
Bias(\hat{\tau}_{ka|u=1}) &= E[\hat{\tau}_{ka|u=1}] - \bar{\tau}_{ka|u=1} \\
&= E[\tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikao|u=1} - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikbo|u=1}] - \tfrac{1}{n^*}\sum_{i=1}^{n^*}(y_{ikt|u=1} - y_{ikc|u=1}) \\
&= \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikao|u=1} - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikbo|u=1} - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} + \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} \\
&= \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ika|u=1} - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} + \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} + \bar{\epsilon}_{ka|u=1} - \bar{\epsilon}_{kb|u=1} \\
&= (\tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ika|u=1} - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1}) + (\tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1}) + \bar{\epsilon}_{ka|u=1} - \bar{\epsilon}_{kb|u=1} \quad (54)
\end{aligned}$$

The first difference in Line 54 can be thought of as the average difference between the n^* always-responders' Wave 2 truthful answers in the world where they completed the survey in Wave 1 and the world where they did not. In other words, it is the average causal effect of completing the survey in Wave 1 on always-responders' true answers in Wave 2, commonly known as conditioning effects. We denote this bias as

$$Bias_{\mathbf{C}}(\hat{\tau}_{ka|u=1}) = \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ika|u=1} - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} = \bar{y}_{ika|u=1} - \bar{y}_{ikt|u=1} \quad (55)$$

Returning to Line 54, the second difference is similar to what we saw in Equation 5 from Section 2 of the paper. Following what we did in Section 2, we can decompose this expression into bias from temporal factors and bias from anticipatory factors:

$$\tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} = Bias_{\mathbf{T}}(\hat{\tau}_{ka|u=1}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=1}) \quad (56)$$

where

$$Bias_{\mathbf{T}}(\hat{\tau}_{ka|u=1}) = \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikbc|u=1} = \bar{y}_{ikc|u=1} - \bar{y}_{ikbc|u=1} \quad (57)$$

$$Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=1}) = \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikbc|u=1} - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} = \bar{y}_{ikbc|u=1} - \bar{y}_{ikb|u=1} \quad (58)$$

Finally, the last difference in Line 54 is just the potential difference in misreporting between Waves 1 and 2.

$$Bias_{\mathbf{M}}(\hat{\tau}_{ka|u=1}) = \bar{\epsilon}_{ka|u=1} - \bar{\epsilon}_{kb|u=1} \quad (59)$$

In sum, we can write the bias in the panel design as

$$Bias(\hat{\tau}_{ka|u=1}) = Bias_{\mathbf{C}}(\hat{\tau}_{ka|u=1}) + Bias_{\mathbf{T}}(\hat{\tau}_{ka|u=1}) + Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=1}) + Bias_{\mathbf{M}}(\hat{\tau}_{ka|u=1}) \tag{60}$$

where

$$Bias_{\mathbf{C}}(\hat{\tau}_{ka|u=1}) = \bar{y}_{ika|u=1} - \bar{y}_{ikt|u=1} \tag{61}$$

$$Bias_{\mathbf{T}}(\hat{\tau}_{ka|u=1}) = \bar{y}_{ikc|u=1} - \bar{y}_{ikbc|u=1} \tag{62}$$

$$Bias_{\mathbf{A}}(\hat{\tau}_{ka|u=1}) = \bar{y}_{ikbc|u=1} - \bar{y}_{ikb|u=1} \tag{63}$$

$$Bias_{\mathbf{M}}(\hat{\tau}_{ka|u=1}) = \bar{\epsilon}_{ka|u=1} - \bar{\epsilon}_{kb|u=1} \tag{64}$$

□

4 Dual Randomized Survey (DRS) Design

In the DRS design, we have n individuals who complete both surveys. We will let n_a denote the number of individuals who complete the survey in Wave 1 and n_b denote the number of individuals who complete the survey in Wave 2, such that $n_a + n_b = n$. Because of the possibility of differential attrition, we can think of our sample as consisting of "always-responders" who would complete both surveys no matter which survey they were assigned in Wave 1 and "sometimes-responders" whose participation in Wave 2 depends on which survey they receive in Wave 1. We will denote the number of always-responders by n^*, the number of possible Wave 1 sometimes-responders by m_a^*, and the number of possible Wave 2 sometimes-responders by m_b^*. We will also let N_a^* denote the total possible number of always and sometimes-responders who would complete both surveys if assigned to do Survey A in Wave 2 and N_b^* denote the total possible number of always and sometimes-responders who would complete both surveys if assigned to do Survey A in Wave 1, such that $N_a^* = n^* + m_a^*$ and $N_b^* = n^* + m_b^*$. We will use $r_{ia} \in \{0, 1\}$ to denote whether individual i completed Survey A after the event and $r_{ib} \in \{0, 1\}$ to denote whether individual i completed Survey A before the event, $u_i \in \{0, 1\}$ to denote whether individual i is an always-responder, $w_{i1} \in \{0, 1\}$ to denote whether individual i would complete Survey A if assigned to do it in Wave 1, and $w_{i2} \in \{0, 1\}$ to denote whether individual i would complete Survey A if assigned to do it in Wave 2.

We will think of each of the individuals in our sample as having a Wave 1 truthful answer (Y_{ikb}), a Wave 1 reported answer (Y_{ikbo}), a Wave 2 truthful answer in the world where they did not complete Survey B in Wave 1

(Y_{ikt}), a Wave 1 truthful answer in the hypothetical world where the event did not happen (Y_{ikc}), a Wave 2 truthful answer after having completed Survey B in Wave 1 (Y_{ika}), and a Wave 2 reported answer after having completed Survey B in Wave 1 (Y_{ikao}).

4.1 Proof of Proposition 6

Proposition 6. *Bias in the DRS design can be written as*

$$Bias(\hat{\tau}_{k|u=1}) = Bias_{\mathbf{D}}(\hat{\tau}_{k|u=1}) + Bias_{\mathbf{P}}(\hat{\tau}_{k|u=1}) + Bias_{\mathbf{T}}(\hat{\tau}_{k|u=1}) + Bias_{\mathbf{A}}(\hat{\tau}_{k|u=1}) + Bias_{\mathbf{M}}(\hat{\tau}_{k|u=1})$$

where

$$\begin{aligned}
Bias_{\mathbf{D}}(\hat{\tau}_{k|u=1}) &= \tfrac{m_a^*}{N_a^*}(\bar{y}_{ika|u=0,w_2=1} - \bar{y}_{ika|u=1}) + \tfrac{m_b^*}{N_b^*}(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|u=0,w_1=1}) \\
Bias_{\mathbf{P}}(\hat{\tau}_{k|u=1}) &= \bar{y}_{ika|u=1} - \bar{y}_{ikt|u=1} \\
Bias_{\mathbf{T}}(\hat{\tau}_{k|u=1}) &= \bar{y}_{ikc|u=1} - \bar{y}_{ikbc|u=1} \\
Bias_{\mathbf{A}}(\hat{\tau}_{k|u=1}) &= \bar{y}_{ikbc|u=1} - \bar{y}_{ikb|u=1} \\
Bias_{\mathbf{M}}(\hat{\tau}_{k|u=1}) &= \bar{\epsilon}_{ika|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}
\end{aligned}$$

Proof. The causal parameter that we are interested in estimating is the average causal effect of the event on the responses to Question K of the n^* always-responders:

$$\bar{\tau}_{k|u=1} = \tfrac{1}{n^*}\sum_{i=1}^{n^*}(y_{ikt|u=1} - y_{ikc|u=1})$$

The statistic that we use to estimate this parameter is just

$$\hat{\tau}_{k|u=1} = \tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikao|r_a=1} - \tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikbo|r_b=1}$$

The bias in this estimator is then

$$
\begin{aligned}
Bias(\hat{\tau}_{k|u=1}) &= E[\hat{\tau}_{k|u=1}] - \bar{\tau}_{k|u=1} \\
&= E[\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikao|r_a=1} - \tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikbo|r_b=1}] - \tfrac{1}{n^*}\sum_{i=1}^{n^*}(y_{ikt|u=1} - y_{ikc|u=1}) \\
&= E[\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ika|r_a=1} - \tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}] - \tfrac{1}{n^*}\sum_{i=1}^{n^*}(y_{ikt|u=1} - y_{ikc|u=1}) + \bar{\epsilon}_{ika|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
&= E[\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ika|r_a=1}] - E[\tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}] - \tfrac{1}{n^*}\sum_{i=1}^{n^*}(y_{ikt|u=1} - y_{ikc|u=1}) + \bar{\epsilon}_{ika|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
&= \tfrac{1}{N_a^*}\sum_{i=1}^{N_a^*} y_{ika|w_2=1} - \tfrac{1}{N_b^*}\sum_{i=1}^{N_b^*} y_{ikb|r_b=1} - \tfrac{1}{n^*}\sum_{i=1}^{n^*}(y_{ikt|u=1} - y_{ikc|u=1}) + \bar{\epsilon}_{ika|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
&= \tfrac{1}{N_a^*}\sum_{i=1}^{N_a^*} y_{ika|w_2=1} - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} + (\tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - \tfrac{1}{N_b^*}\sum_{i=1}^{N_b^*} y_{ikb|w_1=1}) + \bar{\epsilon}_{ika|w_2=1} + \bar{\epsilon}_{ikb|w_1=1} \\
&= \tfrac{1}{N_a^*}\sum_{i=1}^{N_a^*} y_{ika|w_2=1} - \bar{y}_{ikt|u=1} + (\bar{y}_{ikc|u=1} - \tfrac{1}{N_b^*}\sum_{i=1}^{N_b^*} y_{ikb|w_1=1}) + \bar{\epsilon}_{ika|w_2=1} + \bar{\epsilon}_{ikb|w_1=1} \qquad (65)
\end{aligned}
$$

We can expand the first term in the above line as follows:

$$
\tfrac{1}{N_a^*}\sum_{i=1}^{N_a^*} y_{ika|w_2=1} = \tfrac{1}{N_a^*}\sum_{i=1}^{m_a^*} y_{ika|w_2=1,u=0} + \tfrac{1}{N_a^*}\sum_{i=1}^{n^*} y_{ika|u=1} \qquad (66)
$$

Note that the second term in this expression can be rewritten as:

$$
\begin{aligned}
\tfrac{1}{N_a^*}\sum_{i=1}^{n^*} y_{ika|u=1} &= \tfrac{n^*}{N_a^* n^*}\sum_{i=1}^{n^*} y_{ika|u=1} \\
&= \tfrac{n^* + (N_a^* - N_a^*)}{N_a^* n^*}\sum_{i=1}^{n^*} y_{ika|u=1} \\
&= \tfrac{n^* - N_a^*}{N_a^* n^*}\sum_{i=1}^{n^*} y_{ika|u=1} + \tfrac{N_a^*}{N_a^* n^*}\sum_{i=1}^{n^*} y_{ika|u=1} \\
&= \tfrac{-m_a^*}{N_a^* n^*}\sum_{i=1}^{n^*} y_{ika|u=1} + \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ika|u=1} \\
&= -\tfrac{m_a^*}{N_a^* n^*}\sum_{i=1}^{n^*} y_{ika|u=1} + \bar{y}_{ika|u=1} \qquad (67)
\end{aligned}
$$

We can then use Lines 66 and 67 to rewrite the overall bias term in Line 65 as

$$Bias(\hat{\tau}_{k|u=1}) = \tfrac{1}{N_a^*}\sum_{i=1}^{m_a^*} y_{ika|w_2=1,u=0} - \tfrac{m_a^*}{N_a^* n^*}\sum_{i=1}^{n^*} y_{ika|u=1} + \bar{y}_{ika|u=1} - \bar{y}_{ikt|u=1} + \tag{68}$$

$$(\bar{y}_{ikc|u=1} - \tfrac{1}{N_b^*}\sum_{i=1}^{N_b^*} y_{ikb|w_1=1}) + \bar{\epsilon}_{ika|w_2=1} + \bar{\epsilon}_{ikb|w_1=1} \tag{69}$$

The second difference in Line 71 can be rewritten as simply the average difference between the n^* always-responders' truthful answers to question k in the world where they completed Survey B in Wave 1 and in the world where they did not. In other words, it is the average causal effect of completing Survey B in Wave 1 on the always-responders' Wave 2 truthful answers. We can think of this possible difference as a potential type of priming bias:

$$Bias_{\mathbf{P}}(\hat{\tau}_{k|u=1}) = \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ika|u=1} - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} = \bar{y}_{ika|u=1} - \bar{y}_{ikt|u=1} \tag{70}$$

Therefore, we can rewrite the overall bias as

$$Bias(\hat{\tau}_{k|u=1}) = \tfrac{1}{N_a^*}\sum_{i=1}^{m_a^*} y_{ika|w_2=1,u=0} - \tfrac{m_a^*}{N_a^* n^*}\sum_{i=1}^{n^*} y_{ika|u=1} + Bias_{\mathbf{P}}(\hat{\tau}_{k|u=1}) + \tag{71}$$

$$(\bar{y}_{ikc|u=1} - \tfrac{1}{N_b^*}\sum_{i=1}^{N_b^*} y_{ikb|w_1=1}) + \bar{\epsilon}_{ika|w_2=1} + \bar{\epsilon}_{ikb|w_1=1} \tag{72}$$

We can now turn our attention to the first difference in Line 72. We can begin by noting that

$$\tfrac{1}{N_b^*}\sum_{i=1}^{N_b^*} y_{ikb|w_1=1} = \tfrac{1}{N_b^*}\sum_{i=1}^{m_b^*} y_{ikb|w_1=1,u=0} + \tfrac{1}{N_b^*}\sum_{i=1}^{n^*} y_{ikb|u=1} \tag{73}$$

We can break the second term down further as

$$
\begin{aligned}
\frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} &= \frac{n^*}{N_b^* n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} \\
&= \frac{n^*+(N_b^*-N_b^*)}{N_b^* n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} \\
&= \frac{n^*-N_b^*}{N_b^* n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} + \frac{N_b^*}{N_b^* n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} \\
&= \frac{-m_b^*}{N_b^* n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} + \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} \\
&= -\frac{m_b^*}{N_b^* n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} + \bar{y}_{ikb|u=1} \\
&= \bar{y}_{ikb|u=1} - \frac{m_b^*}{N_b^* n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} \qquad (74)
\end{aligned}
$$

Using Lines 73 and 74, we can now rewrite the overall bias term from Lines 71-72 as

$$
\begin{aligned}
Bias(\hat{\tau}_{k|u=1}) = &\frac{1}{N_a^*}\sum_{i=1}^{m_a^*} y_{ika|w_2=1,u=0} - \frac{m_a^*}{N_a^* n^*}\sum_{i=1}^{n^*} y_{ika|u=1} + Bias_{\mathbf{P}}(\hat{\tau}_{k|u=1}) + \\
&\bar{y}_{ikc|u=1} - \Big(\frac{1}{N_b^*}\sum_{i=1}^{m_b^*} y_{ikb|w_1=1,u=0} + \bar{y}_{ikb|u=1} - \frac{m_b^*}{N_b^* n^*}\sum_{i=1}^{n^*} y_{ikb|u=1}\Big) + \bar{\epsilon}_{ika|w_2=1} + \bar{\epsilon}_{ikb|w_1=1} \\
= &\frac{1}{N_a^*}\sum_{i=1}^{m_a^*} y_{ika|w_2=1,u=0} - \frac{m_a^*}{N_a^* n^*}\sum_{i=1}^{n^*} y_{ika|u=1} + Bias_{\mathbf{P}}(\hat{\tau}_{k|u=1}) + \\
&(\bar{y}_{ikc|u=1} - \bar{y}_{ikb|u=1}) + \frac{m_b^*}{N_b^* n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} - \frac{1}{N_b^*}\sum_{i=1}^{m_b^*} y_{ikb|w_1=1,u=0} + \bar{\epsilon}_{ika|w_2=1} + \bar{\epsilon}_{ikb|w_1=1}
\end{aligned}
$$

Similar to what we did in Section 2 of the paper, we can separate the expression $\bar{y}_{ikc|u=1} - \bar{y}_{ikb|u=1}$ into the bias caused by temporal factors and the bias caused anticipatory factors:

$$
\bar{y}_{ikc|u=1} - \bar{y}_{ikb|u=1} = Bias_{\mathbf{T}}(\hat{\tau}_{k|u=1}) + Bias_{\mathbf{A}}(\hat{\tau}_{k|u=1})
$$

where

$$
\begin{aligned}
Bias_{\mathbf{T}}(\hat{\tau}_{k|u=1}) &= \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikbc|u=1} = \bar{y}_{ikc|u=1} - \bar{y}_{ikbc|u=1} \\
Bias_{\mathbf{A}}(\hat{\tau}_{k|u=1}) &= \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikbc|u=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} = \bar{y}_{ikbc|u=1} - \bar{y}_{ikb|u=1}
\end{aligned}
$$

As in Section 2, the term y_{ikbc} represents individual i's Wave 1 truthful response in the counterfactual world where the event "did not happen." Interpretation of this term depends on the counterfactual that the researcher has in mind.

The overall bias in the estimator can now be written as

$$Bias(\hat{\tau}_{k|u=1}) = \tfrac{1}{N_a^*}\sum_{i=1}^{m_a^*} y_{ika|w_2=1,u=0} - \tfrac{m_a^*}{N_a^* n^*}\sum_{i=1}^{n^*} y_{ika|u=1} + Bias_{\mathbf{P}}(\hat{\tau}_{k|u=1}) + Bias_{\mathbf{T}}(\hat{\tau}_{k|u=1}) + \tag{75}$$

$$Bias_{\mathbf{A}}(\hat{\tau}_{k|u=1}) + \tfrac{m_b^*}{N_b^* n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} - \tfrac{1}{N_b^*}\sum_{i=1}^{m_b^*} y_{ikb|w_1=1,u=0} + \bar{\epsilon}_{ika|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \tag{76}$$

We can next note that

$$\begin{aligned}\tfrac{1}{N_a^*}\sum_{i=1}^{m_a^*} y_{ika|w_2=1,u=0} - \tfrac{m_a^*}{N_a^* n^*}\sum_{i=1}^{n^*} y_{ika|u=1} &= (\tfrac{m_a^*}{N_a^*})(\tfrac{1}{m_a^*}\sum_{i=1}^{m_a^*} y_{ika|w_2=1,u=0} - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ika|u=1}) \\ &= (\tfrac{m_a^*}{N_a^*})(\bar{y}_{ika|w_2=1,u=0} - \bar{y}_{ika|u=1}) \end{aligned}\tag{77}$$

Similarly,

$$\begin{aligned}\tfrac{m_b^*}{N_b^* n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} - \tfrac{1}{N_b^*}\sum_{i=1}^{m_b^*} y_{ikb|w_1=1,u=0} &= (\tfrac{m_b^*}{N_b^*})(\tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} - \tfrac{1}{m_b^*}\sum_{i=1}^{m_b^*} y_{ikb|w_1=1,u=0}) \\ &= (\tfrac{m_b^*}{N_b^*})(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|w_1=1,u=0}) \end{aligned}\tag{78}$$

We can now rewrite the overall bias in the estimator from Lines 75-76 as

$$Bias(\hat{\tau}_{k|u=1}) = \tfrac{m_a^*}{N_a^*}(\bar{y}_{ika|u=0,w_2=1} - \bar{y}_{ika|u=1}) + Bias_{\mathbf{P}}(\hat{\tau}_{k|u=1}) + Bias_{\mathbf{T}}(\hat{\tau}_{k|u=1}) + \tag{79}$$

$$Bias_{\mathbf{A}}(\hat{\tau}_{k|u=1}) + \tfrac{m_b^*}{N_b^*}(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|u=0,w_1=1}) + \bar{\epsilon}_{ika|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \tag{80}$$

$$Bias(\hat{\tau}_{k|u=1}) = \tfrac{m_a^*}{N_a^*}(\bar{y}_{ika|u=0,w_2=1} - \bar{y}_{ika|u=1}) + \tfrac{m_b^*}{N_b^*}(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|u=0,w_1=1}) + \tag{81}$$

$$Bias_{\mathbf{P}}(\hat{\tau}_{k|u=1}) + Bias_{\mathbf{T}}(\hat{\tau}_{k|u=1}) + Bias_{\mathbf{A}}(\hat{\tau}_{k|u=1}) + \bar{\epsilon}_{ika|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \tag{82}$$

The expression in Line 81 can be labeled

$$Bias_{\mathbf{D}}(\hat{\tau}_{k|u=1}) = \tfrac{m_a^*}{N_a^*}(\bar{y}_{ika|u=0,w_2=1} - \bar{y}_{ika|u=1}) + \tfrac{m_b^*}{N_b^*}(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|u=0,w_1=1}) \tag{83}$$

This bias term captures the bias induced by differential dropout rates based on which survey respondents were assigned in Wave 1.

The final difference in Line 82 is just the potential bias caused by differential misreporting between the respondents in Waves 1 and 2. We can label this bias

$$Bias_{\mathbf{M}}(\hat{\tau}_{k|u=1}) = \bar{\epsilon}_{ika|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}$$

Therefore, we can now write the overall bias term in condensed form as

$$Bias(\hat{\tau}_{k|u=1}) = Bias_{\mathbf{D}}(\hat{\tau}_{k|u=1}) + Bias_{\mathbf{P}}(\hat{\tau}_{k|u=1}) + Bias_{\mathbf{T}}(\hat{\tau}_{k|u=1}) + Bias_{\mathbf{A}}(\hat{\tau}_{k|u=1}) + Bias_{\mathbf{M}}(\hat{\tau}_{k|u=1})$$

with

$$\begin{aligned}
Bias_{\mathbf{D}}(\hat{\tau}_{k|u=1}) &= \tfrac{m_a^*}{N_a^*}(\bar{y}_{ika|u=0,w_2=1} - \bar{y}_{ika|u=1}) + \tfrac{m_b^*}{N_b^*}(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|u=0,w_1=1}) \\
Bias_{\mathbf{P}}(\hat{\tau}_{k|u=1}) &= \bar{y}_{ika|u=1} - \bar{y}_{ikt|u=1} \\
Bias_{\mathbf{T}}(\hat{\tau}_{k|u=1}) &= \bar{y}_{ikc|u=1} - \bar{y}_{ikbc|u=1} \\
Bias_{\mathbf{A}}(\hat{\tau}_{k|u=1}) &= \bar{y}_{ikbc|u=1} - \bar{y}_{ikb|u=1} \\
Bias_{\mathbf{M}}(\hat{\tau}_{k|u=1}) &= \bar{\epsilon}_{ika|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}
\end{aligned}$$

□