Analyzing the Impact of Events Through Surveys:

Formalizing Biases and Introducing the Dual Randomized Survey Design

Andrew Bertoli, Laura Jakli, and Henry Pascoe

Online Appendix

Contents

1 Glossary of Terms

Notation for Baseline Model

y_{ikb}: Individual i's truthful Wave 1 response

y_{ikbo}: Individual i's observed Wave 1 response

y_{ikt}: Individual i's truthful Wave 2 response

y_{ikto}: Individual i's observed Wave 2 response

y_{ikc}: Individual i's truthful Wave 2 response in the counterfactual world where the event "did not happen"

y_{ikbc}: Individual i's truthful Wave 1 response in the counterfactual world where the event "did not happen"

n_b: Number of Wave 1 respondents

n_a: Number of Wave 2 respondents

N: Number of subjects in the target population

r_{ib}: Indicator variable denoting whether individual i completed the survey in Wave 1

r_{ia}: Indicator variable denoting whether individual i completed the survey in Wave 2

$\bar{\epsilon}_{kb|r_b=1}$: Average measurement error in Wave 1

$\bar{\epsilon}_{kt|r_a=1}$: Average measurement error in Wave 2

Additional Notation for Quota Sampling

g_b: The number of people in the Wave 1 quota group

g_a: The number of people in the Wave 2 quota group

e_b: The number of people in Wave 1 who would have completed the survey but who were excluded due to quota constraints

e_a: The number of people in Wave 2 who would have completed the survey but who were excluded due to quota constraints

n_b: The number of potential Wave 1 respondents when quotas are ignored ($g_b + e_b$)

n_a: The number of potential Wave 2 respondents when quotas are ignored ($g_a + e_a$)

q_i: Indicator variable denoting whether individual i is in the quota sample

$\bar{\epsilon}_{kb|r_b=1,q=1}$: Average measurement error in the Wave 1 quota group

$\bar{\epsilon}_{kt|r_a=1,q=1}$: Average measurement error in the Wave 2 quota group

Additional Notation for Rolling Cross-Sections

n_w: The total number of people who might complete the survey in either wave (this number includes the always-responders and the sometimes-responders for both waves)

n^*: The number of always-responders

m_b: The number of sometimes-responders who would only complete the survey in Wave 1

m_a: The number of sometimes-responders who would only complete the survey in Wave 2

w_b: The number of individuals who would complete the survey if assigned to Wave 1

w_a: The number of individuals who would complete the survey if assigned to Wave 2

n: The total number of Wave 1 and Wave 2 respondents (random variable)

n_b: The number of Wave 1 respondents (random variable)

n_a: The number of Wave 2 respondents (random variable)

u_i: Indicator variable denoting whether individual i is an always-responder

s_{i1}: Indicator variable denoting whether individual i would only complete the survey if assigned to Wave 1

s_{i2}: Indicator variable denoting whether individual i would only complete the survey if assigned to Wave 2

w_i: Indicator variable denoting whether individual i would complete the survey in at least one of the two waves

w_{i1}: Indicator variable denoting whether individual i would complete the survey if assigned to Wave 1

w_{i2}: Indicator variable denoting whether individual i would complete the survey if assigned to Wave 2

$\bar{\epsilon}_{ikb|w_1=1}$: Average measurement error for the respondents who would complete the survey in Wave 1

$\bar{\epsilon}_{ikt|w_2=1}$: Average measurement error for the respondents who would complete the survey in Wave 2

α: The proportion of Wave 1 and Wave 2 respondents who are sometimes-responders (assumed to be the same in both waves for the comparison to the baseline model)

Additional Notation for the Panel Design

y_{ika}: Individual i's truthful Wave 2 response after completing the same survey in Wave 1

y_{ikao}: Individual i's observed Wave 2 response after completing the same survey in Wave 1

n^*: The number of always-responders

u_i: Indicator variable denoting whether individual i is an always-responder

$\bar{\epsilon}_{kb|u=1}$: Average Wave 1 measurement error for the always-responders

$\bar{\epsilon}_{ka|u=1}$: Average Wave 2 measurement error for the always-responders

Additional Notation for the Dual Randomized Survey Design

y_{ika}: Individual i's truthful Wave 2 response after completing Survey B in Wave 1

y_{ikao}: Individual i's observed Wave 2 response after completing Survey B in Wave 1

n^*: The number of always-responders

m_b: The number of sometimes-responders who would only complete both surveys if randomized to take Survey A in Wave 1

m_a: The number of sometimes-responders who would only complete both surveys if randomized to take Survey A in Wave 2

w_b: The number of possible respondents who might be in our sample as someone who completed Survey A in Wave 1

w_a: The number of possible respondents who might be in our sample as someone who completed Survey A in Wave 2

n: The total number of Wave 1 and Wave 2 respondents (random variable)

n_b: The number of Wave 1 respondents (random variable)

n_a: The number of Wave 2 respondents (random variable)

u_i: Indicator variable denoting whether individual i is an always-responder

s_{i1}: Indicator variable denoting whether individual i would only complete both surveys if randomized to take Survey A in Wave 1

s_{i2}: Indicator variable denoting whether individual i would only complete both surveys if randomized to take Survey A in Wave 2

w_{i1}: Indicator variable denoting whether individual i would complete both surveys if randomized to take Survey A in Wave 1

w_{i2}: Indicator variable denoting whether individual i would complete both surveys if randomized to take Survey A in Wave 2

$\bar{\epsilon}_{ikb|w_1=1}$: Average Wave 1 measurement error for the potential respondents who would complete both surveys if randomized to take Survey A in Wave 1

$\bar{\epsilon}_{ika|w_2=1}$: Average Wave 2 measurement error for the potential respondents who would complete both surveys if randomized to take Survey A in Wave 2

2 Baseline Model

2.1 Proof of Proposition 1

Proposition 1. *Bias in the baseline model can be written as*

$$Bias\left(\hat{\tau}_{k|r_a=1}\right) = Bias_{\mathbf{X}}\left(\hat{\tau}_{k|r_a=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|r_a=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|r_a=1}\right) + Bias_{\mathbf{M}}\left(\hat{\tau}_{k|r_a=1}\right) \tag{A1}$$

where

$$Bias_{\mathbf{X}}\left(\hat{\tau}_{k|r_a=1}\right) = \bar{y}_{ikb|r_a=1} - \bar{y}_{ikb|r_b=1} \quad \textit{(Demographic Bias)}$$
$$Bias_{\mathbf{T}}\left(\hat{\tau}_{k|r_a=1}\right) = \bar{y}_{ikc|r_a=1} - \bar{y}_{ikbc|r_a=1} \quad \textit{(Temporal Bias)}$$
$$Bias_{\mathbf{A}}\left(\hat{\tau}_{k|r_a=1}\right) = \bar{y}_{ikbc|r_a=1} - \bar{y}_{ikb|r_a=1} \quad \textit{(Anticipation Bias)}$$
$$Bias_{\mathbf{M}}\left(\hat{\tau}_{k|r_a=1}\right) = \bar{\epsilon}_{kt|r_a=1} - \bar{\epsilon}_{kb|r_b=1} \quad \textit{(Differential Misreporting)}$$

Proof. The bias in $\hat{\tau}_{k|r_a=1}$ can be written as

$$\begin{aligned}
Bias\left(\hat{\tau}_{k|r_a=1}\right) &= E\left[\hat{\tau}_{k|r_a=1}\right] - \bar{\tau}_{k|r_a=1} \qquad \text{(A2)}\\
&= E\left[\frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikto|r_a=1} - \frac{1}{n_b}\sum_{i=1}^{n_b} y_{ikbo|r_b=1}\right] - \frac{1}{n_a}\sum_{i=1}^{n_a}\left(y_{ikt|r_a=1} - y_{ikc|r_a=1}\right)\\
&= \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikto|r_a=1} - \frac{1}{n_b}\sum_{i=1}^{n_b} y_{ikbo|r_b=1} - \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikt|r_a=1} + \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikc|r_a=1}\\
&= \left(\frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikto|r_a=1} - \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikt|r_a=1}\right) + \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikc|r_a=1} - \frac{1}{n_b}\sum_{i=1}^{n_b} y_{ikbo|r_b=1}
\end{aligned}$$

The first of the two expressions in the line above is just the average measurement error in the Wave 2 respondents' answers. We can denote this average measurement error as

$$\bar{\epsilon}_{kt|r_a=1} = \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikto|r_a=1} - \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikt|r_a=1}$$

Further, we can define the average measurement error in the Wave 1 respondents' answers as

$$\bar{\epsilon}_{kb|r_b=1} = \frac{1}{n_b}\sum_{i=1}^{n_b} y_{ikbo|r_b=1} - \frac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}$$

We can now rewrite Equation A2 as

$$Bias\left(\hat{\tau}_{k|r_a=1}\right) = \bar{\epsilon}_{kt|r_a=1} + \tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikc|r_a=1} - \left(\tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1} + \bar{\epsilon}_{kb|r_b=1}\right) \tag{A3}$$
$$= \tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikc|r_a=1} - \tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1} + \bar{\epsilon}_{kt|r_a=1} - \bar{\epsilon}_{kb|r_b=1}$$

The first of the two differences in Equation A3 is the average difference between Wave 2 respondents' truthful counterfactual answers and Wave 1 respondents' pre-event truthful answers. Since this expression indexes over two distinct groups of respondents surveyed in two different time periods, it is challenging to interpret. We can gain traction by modifying Equation A3 slightly. First, we imagine the truthful answers of the Wave 2 respondents had they instead been surveyed in Wave 1. In other words, we imagine the y_{ikb} values for Wave 2 respondents. We can then add and subtract the average of these y_{ikb} values to Equation A3:

$$Bias\left(\hat{\tau}_{k|r_a=1}\right) = \tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikc|r_a=1} - \tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1} + \bar{\epsilon}_{kt|r_a=1} - \bar{\epsilon}_{kb|r_b=1} + \tag{A4}$$
$$\left(\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikb|r_a=1} - \tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikb|r_a=1}\right)$$

By reordering the terms, we get:

$$Bias\left(\hat{\tau}_{k|r_a=1}\right) = \left(\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikb|r_a=1} - \tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}\right) + \left(\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikc|r_a=1} - \tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikb|r_a=1}\right) + \tag{A5}$$
$$\bar{\epsilon}_{kt|r_a=1} - \bar{\epsilon}_{kb|r_b=1}$$

The first expression in Equation A5 is just the average difference in truthful responses caused by baseline demographic differences between Wave 1 and Wave 2 respondents. We can label this source of bias "demographic bias" and write it formally as $Bias_{\mathbf{X}}(\hat{\tau}_{k|r_a=1})$:

Definition 1 (Demographic Bias)**.**

$$Bias_{\mathbf{X}}\left(\hat{\tau}_{k|r_a=1}\right) \equiv \tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikb|r_a=1} - \tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1} = \bar{y}_{ikb|r_a=1} - \bar{y}_{ikb|r_b=1}$$

We can then rewrite Equation A5 as

$$Bias\left(\hat{\tau}_{k|r_a=1}\right) = Bias_{\mathbf{X}}\left(\hat{\tau}_{k|r_a=1}\right) + \left(\frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikc|r_a=1} - \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikb|r_a=1}\right) + \left(\bar{\epsilon}_{kt|r_a=1} - \bar{\epsilon}_{kb|r_b=1}\right) \tag{A6}$$

The middle expression is now limited to Wave 2 respondents only. It represents the average difference between their truthful Wave 2 answers in the counterfactual world where the event did not happen and their truthful Wave 1 answers had they completed the survey in Wave 1. Interpretation of this term now depends on what we mean by "the counterfactual world where the event did not happen." There are multiple plausible versions of this counterfactual world and which counterfactual we choose impacts how we think about this expression.

One way we might conceive of this counterfactual is in a manner that we would not expect to have an impact on respondent beliefs or attitudes about issues related to the survey: for example, a scenario wherein the event was unexpectedly postponed the day prior. Such a counterfactual might be that the day before a political debate, the event is postponed for two weeks due to a water leak in the scheduled event host facility. With this counterfactual in mind, the difference between Wave 2 respondents' y_{ikb} and y_{ikc} values should merely be a short-term temporal difference. Its size would depend on whether any other salient events happened between Waves 1 and 2. It might also be affected by other temporal factors like the weather, which could impact respondents' moods, or if Wave 1 was fielded on a weekday whereas Wave 2 was fielded on a weekend.

However, we could imagine an alternative counterfactual wherein the event was never scheduled. In the debate example, this counterfactual might be that political parties had agreed a year prior to not hold any debates before the next election. With this counterfactual in mind, the difference between y_{ikb} and y_{ikc} may not just be determined by short-term temporal factors. Rather, y_{ikb} could be influenced by anticipation of the event in a way that y_{ikc} would not. For example, the lead-up to the debate might feature increased media attention to the electoral race that would not have occurred in the world where the event was never scheduled.

To distinguish between bias from temporal and anticipation factors, we first consider another potential outcome— the Wave 2 respondents' truthful answers had they been surveyed in Wave 1 and if the event "had never happened." We can denote this counterfactual outcome by y_{ikbc}. We can then take Equation A6 and add and subtract the average of this potential outcome for Wave 2 respondents.

$$\begin{aligned} Bias\left(\hat{\tau}_{k|r_a=1}\right) = Bias_{\mathbf{X}}\left(\hat{\tau}_{k|r_a=1}\right) + \left(\frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikc|r_a=1} - \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikb|r_a=1}\right) + \\ \left(\frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikbc|r_a=1} - \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikbc|r_a=1}\right) + \left(\bar{\epsilon}_{kt|r_a=1} - \bar{\epsilon}_{kb|r_b=1}\right) \end{aligned}$$

By reordering the terms, we obtain

$$Bias\left(\hat{\tau}_{k|r_a=1}\right) = Bias_{\mathbf{X}}\left(\hat{\tau}_{k|r_a=1}\right) + \left(\frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikc|r_a=1} - \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikbc|r_a=1}\right) + \left(\frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikbc|r_a=1} - \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikb|r_a=1}\right) + \left(\bar{\epsilon}_{kt|r_a=1} - \bar{\epsilon}_{kb|r_b=1}\right) \tag{A7}$$

The first of these two expressions now represents the average difference between the hypothetical post-event and pre-event truthful answers of Wave 2 respondents in the world where the event did not happen. Thus, it purely captures bias caused by temporal differences between Waves 1 and 2.

Definition 2 (Temporal Bias)**.**

$$Bias_{\mathbf{T}}\left(\hat{\tau}_{k|r_a=1}\right) \equiv \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikc|r_a=1} - \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikbc|r_a=1} = \bar{y}_{ikc|r_a=1} - \bar{y}_{ikbc|r_a=1}$$

The second expression in Equation A7 represents the average difference in the hypothetical truthful Wave 1 answers of the Wave 2 respondents in the worlds where the event did and did not happen. It thereby captures bias caused by anticipation factors.

Definition 3 (Anticipation Bias)**.**

$$Bias_{\mathbf{A}}\left(\hat{\tau}_{k|r_a=1}\right) \equiv \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikbc|r_a=1} - \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikb|r_a=1} = \bar{y}_{ikbc|r_a=1} - \bar{y}_{ikb|r_a=1}$$

We can now rewrite Equation A7 as

$$Bias\left(\hat{\tau}_{k|r_a=1}\right) = Bias_{\mathbf{X}}\left(\hat{\tau}_{k|r_a=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|r_a=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|r_a=1}\right) + \left(\bar{\epsilon}_{kt|r_a=1} - \bar{\epsilon}_{kb|r_b=1}\right) \tag{A8}$$

The final difference in Equation A8 is simply the average difference in measurement error in the Wave 1 and Wave 2 respondents' answers.

Definition 4 (Differential Misreporting Bias)**.**

$$Bias_{\mathbf{M}}\left(\hat{\tau}_{k|r_a=1}\right) \equiv \bar{\epsilon}_{kt|r_a=1} - \bar{\epsilon}_{kb|r_b=1}$$

We can therefore rewrite the overall bias term as the sum of the demographic, temporal, anticipation, and differential misreporting biases given by Definitions 1-4.

$$Bias\left(\hat{\tau}_{k|r_a=1}\right) = Bias_{\mathbf{X}}\left(\hat{\tau}_{k|r_a=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|r_a=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|r_a=1}\right) + Bias_{\mathbf{M}}\left(\hat{\tau}_{k|r_a=1}\right) \tag{A9}$$

where

$$Bias_{\mathbf{X}}\left(\hat{\tau}_{k|r_a=1}\right) = \bar{y}_{ikb|r_a=1} - \bar{y}_{ikb|r_b=1}$$
$$Bias_{\mathbf{T}}\left(\hat{\tau}_{k|r_a=1}\right) = \bar{y}_{ikc|r_a=1} - \bar{y}_{ikbc|r_a=1}$$
$$Bias_{\mathbf{A}}\left(\hat{\tau}_{k|r_a=1}\right) = \bar{y}_{ikbc|r_a=1} - \bar{y}_{ikb|r_a=1}$$
$$Bias_{\mathbf{M}}\left(\hat{\tau}_{k|r_a=1}\right) = \bar{\epsilon}_{kt|r_a=1} - \bar{\epsilon}_{kb|r_b=1}$$

□

3 External Validity

Consider the context where the target parameter that we want to estimate is the average causal effect for the population of interest:

$$\bar{\tau}_k = \frac{1}{N}\sum_{i=1}^{N}\left(y_{ikt} - y_{ikc}\right)$$

Like when we estimated the average treatment effect for the Wave 2 respondents in our baseline model, the estimator that we will use to estimate $\bar{\tau}_k$ is the average difference between the Wave 2 and Wave 1 respondents' answers to question k of the survey:

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

The bias in $\hat{\tau}_k$ can then be written as

$$Bias\left(\hat{\tau}_k\right) = Bias_{\mathbf{X}}\left(\hat{\tau}_k\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_k\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_k\right) + Bias_{\mathbf{M}}\left(\hat{\tau}_k\right) + Bias_{\mathbf{H}}\left(\hat{\tau}_k\right) \qquad \text{(A10)}$$

where

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

This expression for the overall bias is the same as in our baseline model, except for the $Bias_{\mathbf{H}}\left(\hat{\tau}_k\right)$ term that accounts for potential bias caused by the Wave 2 respondents having a heterogeneous treatment effect compared to the treatment effect in the overall population.

Deriving the bias in $\hat{\tau}_k$ is trivial. Begin by noting that $Bias\left(\hat{\tau}_{k|r_a=1}\right) = E\left[\hat{\tau}_{k|r_a=1}\right] - \bar{\tau}_{k|r_a=1}$, which can be rewritten as $E\left[\hat{\tau}_{k|r_a=1}\right] = Bias\left(\hat{\tau}_{k|r_a=1}\right) + \bar{\tau}_{k|r_a=1}$. Since $\hat{\tau}_k = \hat{\tau}_{k|r_a=1}$, we have $E[\hat{\tau}_k] = E[\hat{\tau}_{k|r_a=1}]$, so we can change the expression to $E\left[\hat{\tau}_k\right] = Bias\left(\hat{\tau}_{k|r_a=1}\right) + \bar{\tau}_{k|r_a=1}$. The bias in $\hat{\tau}_k$ is then just

$$
\begin{aligned}
Bias\left(\hat{\tau}_k\right) &= E\left[\hat{\tau}_k\right] - \bar{\tau}_k \\
&= Bias\left(\hat{\tau}_{k|r_a=1}\right) + \bar{\tau}_{k|r_a=1} - \bar{\tau}_k \\
&= Bias\left(\hat{\tau}_{k|r_a=1}\right) + Bias_{\mathbf{H}}\left(\hat{\tau}_k\right) \\
&= Bias_{\mathbf{X}}\left(\hat{\tau}_{k|r_a=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|r_a=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|r_a=1}\right) + Bias_{\mathbf{M}}\left(\hat{\tau}_{k|r_a=1}\right) + Bias_{\mathbf{H}}\left(\hat{\tau}_k\right) \\
&= Bias_{\mathbf{X}}\left(\hat{\tau}_k\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_k\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_k\right) + Bias_{\mathbf{M}}\left(\hat{\tau}_k\right) + Bias_{\mathbf{H}}\left(\hat{\tau}_k\right)
\end{aligned}
$$

where

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

4 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 article, 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.

4.1 Proof of Proposition 2

Proposition 2. *Bias in the quota sampling design is given by*

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

where

$$Bias_{\mathbf{X}}\left(\hat{\tau}_{k|r_a=1,q=1}\right) = \bar{y}_{ikb|r_a=1,q=1} - \bar{y}_{ikb|r_b=1,q=1}$$
$$Bias_{\mathbf{T}}\left(\hat{\tau}_{k|r_a=1,q=1}\right) = \bar{y}_{ikc|r_a=1,q=1} - \bar{y}_{ikbc|r_a=1,q=1}$$
$$Bias_{\mathbf{A}}\left(\hat{\tau}_{k|r_a=1,q=1}\right) = \bar{y}_{ikbc|r_a=1,q=1} - \bar{y}_{ikb|r_a=1,q=1}$$
$$Bias_{\mathbf{M}}\left(\hat{\tau}_{k|r_a=1,q=1}\right) = \bar{\epsilon}_{kt|r_a=1,q=1} - \bar{\epsilon}_{kb|r_b=1,q=1}$$

Proof. The causal parameter 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:

$$\bar{\tau}_{k|r_a=1,q=1} = \tfrac{1}{g_a}\sum_{i=1}^{g_a}\left(y_{ikt|r_a=1,q=1} - y_{ikc|r_a=1,q=1}\right)$$

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.

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

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

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

where

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

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

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

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

□

The difference between this overall bias term and the $Bias\left(\hat{\tau}_{k|r_a=1}\right)$ expression that we derived in Section 2 of the article is that this term restricts the focus to our quota sample.

4.2 Proof of Proposition 3

Proposition 3. *Quota designs reduce bias if and only if*

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

When the inequality is 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

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

We can decompose $Bias(\hat{\tau}_{k|r_a=1})$ into a weighted average of the bias in the sub-sample who we would have surveyed if we had done quota sampling and the bias in the sub-sample who we would have 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

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

which we can separate into

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

We can simplify this expression as follows

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

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

$$Bias\left(\hat{\tau}_{k|r_a=1}\right) = \left(\frac{g_a}{n_a}\right) Bias\left(\hat{\tau}_{k|r_a=1,q=1}\right) + \left(\frac{e_a}{n_a}\right) Bias\left(\hat{\tau}_{k|r_a=1,q=0}\right) +$$
$$\left(\frac{g_a}{n_a}\right)\bar{y}_{ikb|r_b=1,q=1} - \left(\frac{g_b}{n_b}\right)\bar{y}_{ikb|r_b=1,q=1} + \left(\frac{e_a}{n_a}\right)\bar{y}_{ikb|r_b=1,q=0} - \left(\frac{e_b}{n_b}\right)\bar{y}_{ikb|r_b=1,q=0} +$$
$$\left(\frac{g_a}{n_a}\right)\bar{\epsilon}_{kb|r_b=1,q=1} - \left(\frac{g_b}{n_b}\right)\bar{\epsilon}_{kb|r_b=1,q=1} + \left(\frac{e_a}{n_a}\right)\bar{\epsilon}_{kb|r_b=1,q=0} - \left(\frac{e_b}{n_b}\right)\bar{\epsilon}_{kb|r_b=1,q=0}$$
$$Bias\left(\hat{\tau}_{k|r_a=1}\right) = \left(\frac{g_a}{n_a}\right) Bias\left(\hat{\tau}_{k|r_a=1,q=1}\right) + \left(\frac{e_a}{n_a}\right) Bias\left(\hat{\tau}_{k|r_a=1,q=0}\right) +$$
$$\left(\frac{g_a}{n_a}\right)\bar{y}_{ikbo|r_b=1,q=1} - \left(\frac{g_b}{n_b}\right)\bar{y}_{ikbo|r_b=1,q=1} + \left(\frac{e_a}{n_a}\right)\bar{y}_{ikbo|r_b=1,q=0} - \left(\frac{e_b}{n_b}\right)\bar{y}_{ikbo|r_b=1,q=0}$$
$$Bias\left(\hat{\tau}_{k|r_a=1}\right) = \left(\frac{g_a}{n_a}\right) Bias\left(\hat{\tau}_{k|r_a=1,q=1}\right) + \left(\frac{e_a}{n_a}\right) Bias\left(\hat{\tau}_{k|r_a=1,q=0}\right) +$$
$$\left(\frac{g_a}{n_a}\right)\bar{y}_{ikbo|r_b=1,q=1} - \left(\frac{g_b}{n_b}\right)\bar{y}_{ikbo|r_b=1,q=1} + \left(1-\frac{g_a}{n_a}\right)\bar{y}_{ikbo|r_b=1,q=0} - \left(1-\frac{g_b}{n_b}\right)\bar{y}_{ikbo|r_b=1,q=0}$$
$$Bias\left(\hat{\tau}_{k|r_a=1}\right) = \left(\frac{g_a}{n_a}\right) Bias\left(\hat{\tau}_{k|r_a=1,q=1}\right) + \left(\frac{e_a}{n_a}\right) Bias\left(\hat{\tau}_{k|r_a=1,q=0}\right) +$$
$$\left(\frac{g_a}{n_a}-\frac{g_b}{n_b}\right)\bar{y}_{ikbo|r_b=1,q=1} - \left(\frac{g_a}{n_a}-\frac{g_b}{n_b}\right)\bar{y}_{ikbo|r_b=1,q=0}$$
$$Bias\left(\hat{\tau}_{k|r_a=1}\right) = \left(\frac{g_a}{n_a}\right) Bias\left(\hat{\tau}_{k|r_a=1,q=1}\right) + \left(\frac{e_a}{n_a}\right) Bias\left(\hat{\tau}_{k|r_a=1,q=0}\right) + \tag{A11}$$
$$\left(\frac{g_a}{n_a}-\frac{g_b}{n_b}\right)\left(\bar{y}_{ikbo|r_b=1,q=1} - \bar{y}_{ikbo|r_b=1,q=0}\right) \tag{A12}$$

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

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

or (utilizing Lines A11-A12)

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

$\square$

4.3 Examples of How Quota Sampling Could Amplify Bias

We begin with a hypothetical phone survey 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 had to start 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 ones, even after conditioning on the covariates balanced through quotas. As such, quota sampling could either reduce or amplify bias, depending on the relationship between these covariates and the potential outcomes.

We next consider a hypothetical involving an online survey. In Wave 1, the survey firm is able to meet its quotas without an issue. However, in Wave 2, the quota constraints make it difficult for the firm to obtain a sufficiently large sample. For this reason, the firm has to work harder, either by advertising the survey more broadly or by offering potential respondents further incentives. This change in sampling procedures could lead to large demographic differences between Wave 1 and Wave 2 respondents on unobservables. Whether quota sampling would increase bias would depend on the relationship between the imbalanced factors and the potential outcomes.

5 Rolling Cross-Sections

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 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_w denote the total number of always-responders and sometimes-responders. 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:

$$n^* \leq n_w$$

$$n_b + n_a = n$$

$$n_b^* + m_b^* = n_b$$

$$n_a^* + m_a^* = n_a$$

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

Also note that n_w 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 continue to 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, we will let $s_i \in \{0, 1\}$ denote whether individual i is a sometimes-responder, $s_{i1} \in \{0, 1\}$ denote whether individual i is a sometimes-responder who would only complete the survey in Wave 1, and $s_{i2} \in \{0, 1\}$ denote whether individual i is a sometimes-responder who would only complete the survey in Wave 2. We will also let m_b denote the number of sometimes-responders who would complete the survey if they were assigned to take it in Wave 1 and m_a denote the number of sometimes-responders who would complete the survey if they were assigned to take it in Wave 2. Further, let $w_i \in \{0, 1\}$ denote whether individual i would complete the survey in at least one of the two waves, $w_{i1} \in \{0, 1\}$ denote whether individual i would complete the survey if assigned to Wave 1, and $w_{i2} \in \{0, 1\}$ denote whether individual i would complete the survey if assigned to Wave 2. Then the number of individuals who would complete the survey if assigned to Wave 1 can be written as $w_b = n^* + m_b = \sum_i^N w_{i1}$ and the number who would complete the survey if assigned to Wave 2 can be written as $w_a = n^* + m_a = \sum_i^N w_{i2}$. For clarity, note that w_a, w_b, m_a, and m_b are parameters and that $n^* + m_a + m_b = n_w$.

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

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

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:

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

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:

$$\bar{\tau}_{k|w=1} = \frac{1}{n_w} \sum_{i=1}^{n_w} \left(y_{ikt|w=1} - y_{ikc|w=1} \right)$$

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

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

5.1 Proof of Proposition 4

Proposition 4. *When estimating* $\bar{\tau}_{k|u=1}$*, the bias in* $\hat{\tau}_{k|u=1}$ *can be written as*

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

where

$$\begin{aligned}
Bias_{\mathbf{S}}\left(\hat{\tau}_{k|u=1}\right) &= \frac{m_a}{w_a}\left(\bar{y}_{ikt|s_2=1} - \bar{y}_{ikt|u=1}\right) + \frac{m_b}{w_b}\left(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|s_1=1}\right) \\
Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right) &= \bar{y}_{ikc|u=1} - \bar{y}_{ikbc|u=1} \\
Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right) &= \bar{y}_{ikbc|u=1} - \bar{y}_{ikb|u=1} \\
Bias_{\mathbf{M}}\left(\hat{\tau}_{k|u=1}\right) &= \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}
\end{aligned}$$

Proof. When estimating $\bar{\tau}_{k|u=1}$, the bias in $\hat{\tau}_{k|u=1}$ is

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

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

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

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

We can further rewrite the overall bias term as

$$
\begin{aligned}
Bias\left(\hat{\tau}_{k|u=1}\right) &= E\left[\frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikt|r_a=1}\right] - E\left[\frac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}\right] - \frac{1}{n^*}\sum_{i=1}^{n^*}\left(y_{ikt|u=1} - y_{ikc|u=1}\right) + \\
&\quad \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
&= E\left[\frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikt|r_a=1}\right] - E\left[\frac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}\right] - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} + \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} + \\
&\quad \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
&= \left(E\left[\frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikt|r_a=1}\right] - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1}\right) + \left(\frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - E\left[\frac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}\right]\right) + \qquad \text{(A13)} \\
&\quad \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \qquad \text{(A14)}
\end{aligned}
$$

We can begin by focusing on the first of the two differences in Line A13. We can break this expression down as follows:

$$
\begin{aligned}
E\left[\frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikt|r_a=1}\right] - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} &= E\left[\frac{1}{n_a}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1} + \frac{1}{n_a}\sum_{i=1}^{m_a^*} y_{ikt|r_a=1,s_2=1}\right] - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} \\
&= E\left[\frac{1}{n_a}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}\right] + E\left[\frac{1}{n_a}\sum_{i=1}^{m_a^*} y_{ikt|r_a=1,s_2=1}\right] - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} \\
&= E\left[\frac{1}{n_a}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}\right] - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} + E\left[\frac{1}{n_a}\sum_{i=1}^{m_a^*} y_{ikt|r_a=1,s_2=1}\right]
\end{aligned}
\qquad \text{(A15)}
$$

Focusing on the first term in Line A15, note that

$$
\begin{aligned}
E\left[\frac{1}{n_a}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}\right] &= E\left[\frac{n_a^*}{n_a n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}\right] \\
&= E\left[\frac{n_a^*+(n_a-n_a)}{n_a n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}\right] \\
&= E\left[\frac{n_a^*-n_a}{n_a n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}\right] + E\left[\frac{n_a}{n_a n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}\right] \\
&= E\left[\frac{-m_a^*}{n_a n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}\right] + E\left[\frac{1}{n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}\right] \qquad \text{(A16)}
\end{aligned}
$$

The second term in Line A16 is the expected value of a random 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.

$$
E\left[\frac{1}{n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}\right] = \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} \qquad \text{(A17)}
$$

Combining Lines A13-A14, A15, A16, and A17, we get

$$
\begin{aligned}
Bias\left(\hat{\tau}_{k|u=1}\right) = &\left(E\left[\frac{-m_a^*}{n_a n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}\right] + \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} + E\left[\frac{1}{n_a}\sum_{i=1}^{m_a^*} y_{ikt|r_a=1,s_2=1}\right]\right) + \\
&\left(\frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - E\left[\frac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}\right]\right) + \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
= &\left(E\left[\frac{1}{n_a}\sum_{i=1}^{m_a^*} y_{ikt|r_a=1,s_2=1}\right] - E\left[\frac{m_a^*}{n_a n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}\right]\right) + \qquad \text{(A18)} \\
&\left(\frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - E\left[\frac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}\right]\right) + \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \qquad \text{(A19)}
\end{aligned}
$$

We can now focus on the first difference in Line A19. We can rewrite this expression as

$$
\begin{aligned}
\frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - E\left[\frac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}\right] &= \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - \left(E\left[\frac{1}{n_b}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}\right] + E\left[\frac{1}{n_b}\sum_{i=1}^{m_b^*} y_{ikb|r_b=1,s_1=1}\right]\right) \\
&= \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - E\left[\frac{1}{n_b}\sum_{i=1}^{m_b^*} y_{ikb|r_b=1,s_1=1}\right] - E\left[\frac{1}{n_b}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}\right]
\end{aligned}
\qquad \text{(A20)}
$$

The last term in Line A20 can be written as

$$
\begin{aligned}
E\left[\frac{1}{n_b}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}\right] &= E\left[\frac{n_b^*}{n_b n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}\right] \\
&= E\left[\frac{n_b^*+(n_b-n_b)}{n_b n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}\right] \\
&= E\left[\frac{n_b^*-n_b}{n_b n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}\right] + E\left[\frac{n_b}{n_b n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}\right] \\
&= E\left[\frac{-m_b^*}{n_b n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}\right] + E\left[\frac{1}{n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}\right] \qquad \text{(A21)}
\end{aligned}
$$

Similar to what we did in Line A17, we can note that the second term in Line A21 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.

$$
E\left[\frac{1}{n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}\right] = \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} \qquad \text{(A22)}
$$

Combining Lines A18-A19, A20, A21, and A22, we get

$$
\begin{aligned}
Bias\left(\hat{\tau}_{k|u=1}\right) = &\left(E\left[\frac{1}{n_a}\sum_{i=1}^{m_a^*} y_{ikt|r_a=1,s_2=1}\right] - E\left[\frac{m_a^*}{n_a n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}\right]\right) + \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - \\
&E\left[\frac{1}{n_b}\sum_{i=1}^{m_b^*} y_{ikb|r_b=1,s_1=1}\right] - \left(E\left[\frac{-m_b^*}{n_b n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}\right] + \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1}\right) + \\
&\bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
= &\left(E\left[\frac{1}{n_a}\sum_{i=1}^{m_a^*} y_{ikt|r_a=1,s_2=1}\right] - E\left[\frac{m_a^*}{n_a n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}\right]\right) + \qquad \text{(A23)} \\
&\frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} + \qquad \text{(A24)} \\
&\left(E\left[\frac{m_b^*}{n_b n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}\right] - E\left[\frac{1}{n_b}\sum_{i=1}^{m_b^*} y_{ikb|r_b=1,s_1=1}\right]\right) + \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \qquad \text{(A25)}
\end{aligned}
$$

The expression in Line A24 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.

$$\begin{aligned}\frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} =& \bar{y}_{ikc|u=1} - \bar{y}_{ikb|u=1} && \text{(A26)}\\ =& Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right) && \text{(A27)}\end{aligned}$$

where

$$\begin{aligned}Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right) &= \bar{y}_{ikc|u=1} - \bar{y}_{ikbc|u=1} && \text{(A28)}\\ Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right) &= \bar{y}_{ikbc|u=1} - \bar{y}_{ikb|u=1} && \text{(A29)}\end{aligned}$$

Therefore, we can write the overall bias term as

$$\begin{aligned}Bias\left(\hat{\tau}_{k|u=1}\right) = &\left(E\left[\frac{1}{n_a}\sum_{i=1}^{m_a^*} y_{ikt|r_a=1,s_2=1}\right] - E\left[\frac{m_a^*}{n_a n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}\right]\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right) + && \text{(A30)}\\ &Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right) + \left(E\left[\frac{m_b^*}{n_b n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}\right] - E\left[\frac{1}{n_b}\sum_{i=1}^{m_b^*} y_{ikb|r_b=1,s_1=1}\right]\right) + && \text{(A31)}\\ &\bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} && \text{(A32)}\end{aligned}$$

We will start with the first difference in Line A30. Note that

$$\begin{aligned}E\left[\frac{1}{n_a}\sum_{i=1}^{m_a^*} y_{ikt|r_a=1,s_2=1}\right] - E\left[\frac{m_a^*}{n_a n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}\right] =& E\left[\frac{m_a^*}{n_a}\left(\frac{1}{m_a^*}\sum_{i=1}^{m_a^*} y_{ikt|r_a=1,s_2=1}\right) - \frac{m_a^*}{n_a}\left(\frac{1}{n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}\right)\right]\\ =& E\left[\frac{m_a^*}{n_a}\left(\frac{1}{m_a^*}\sum_{i=1}^{m_a^*} y_{ikt|r_a=1,s_2=1} - \frac{1}{n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}\right)\right]\end{aligned} \quad \text{(A33)}$$

So the expression inside the parentheses in Line A33 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.

Therefore, the first average inside the parentheses in Line A33 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

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

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

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

Also, the expected proportions of sometimes-responders in Waves 1 and 2 are

$$
\begin{aligned}
E\left[\frac{m_b^*}{n_b}\right] &= \frac{m_b}{w_b} \\
E\left[\frac{m_a^*}{n_a}\right] &= \frac{m_a}{w_a}
\end{aligned}
$$

Returning to the expression in Line A33, 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

$$
\begin{aligned}
E\left[\frac{m_a^*}{n_a}\left(\frac{1}{m_a^*}\sum_{i=1}^{m_a^*} y_{ikt|r_a=1,s_2=1} - \frac{1}{n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}\right)\right] =& E\left[\frac{m_a^*}{n_a}\right] E\left[\frac{1}{m_a^*}\sum_{i=1}^{m_a^*} y_{ikt|r_a=1,s_2=1} - \frac{1}{n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}\right] \\
=& \frac{m_a}{w_a}\left(E\left[\frac{1}{m_a^*}\sum_{i=1}^{m_a^*} y_{ikt|r_a=1,s_2=1}\right] - E\left[\frac{1}{n_a^*}\sum_{i=1}^{n_a^*} y_{ikt|r_a=1,u=1}\right]\right)
\end{aligned}
\tag{A34}
$$

Inside the parentheses of Line A34, 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

$$
E\left[\frac{1}{m_a^*}\sum_{i=1}^{m_a^*} y_{ikt|r_a=1,s_2=1}\right] = \frac{1}{m_a}\sum_{i=1}^{m_a} y_{ikt|s_2=1}
$$

and

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

We can then write

$$\frac{m_a}{w_a}\left(E\left[\frac{1}{m_a^*}\sum_{i=1}^{m_a^*} y_{ikt|r_a=1,s_2=1}\right] - E\left[\frac{1}{n_a^*}\sum_{i=1}^{n_a^*} = y_{ikt|r_a=1,u=1}\right]\right) = \frac{m_a}{w_a}\left(\frac{1}{m_a}\sum_{i=1}^{m_a} y_{ikt|s_2=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1}\right)$$

Substituting this expression into the overall bias term, we get

$$Bias\left(\hat{\tau}_{k|u=1}\right) = \frac{m_a}{w_a}\left(\frac{1}{m_a}\sum_{i=1}^{m_a} y_{ikt|s_2=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right) + \tag{A35}$$

$$Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right) + \left(E\left[\frac{m_b^*}{n_b n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}\right] - E\left[\frac{1}{n_b}\sum_{i=1}^{m_b^*} y_{ikb|r_b=1,s_1=1}\right]\right) + \tag{A36}$$

$$\bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \tag{A37}$$

We can now turn to the difference in Line A36. Similar to before, we can begin by noting that

$$\begin{aligned} E\left[\frac{m_b^*}{n_b n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}\right] - E\left[\frac{1}{n_b}\sum_{i=1}^{m_b^*} y_{ikb|r_b=1,s_1=1}\right] &= E\left[\frac{m_b^*}{n_b}\left(\frac{1}{n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}\right) - \frac{m_b^*}{n_b}\left(\frac{1}{m_b^*}\sum_{i=1}^{m_b^*} y_{ikb|r_b=1,s_1=1}\right)\right] \\ &= E\left[\frac{m_b^*}{n_b}\left(\frac{1}{n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1} - \frac{1}{m_b^*}\sum_{i=1}^{m_b^*} y_{ikb|r_b=1,s_1=1}\right)\right] \end{aligned} \tag{A38}$$

As in Line A33, the expression inside the parentheses in Line A38 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_{ikb} value of the Wave 1 respondents who are always-responders, and the second average is the mean y_{ikb} 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

$$\begin{aligned} E\left[\frac{m_b^*}{n_b}\left(\frac{1}{n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1} - \frac{1}{m_b^*}\sum_{i=1}^{m_b^*} y_{ikb|r_b=1,s_1=1}\right)\right] &= E\left[\frac{m_b^*}{n_b}\right] E\left[\frac{1}{n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1} - \frac{1}{m_b^*}\sum_{i=1}^{m_b^*} y_{ikb|r_b=1,s_1=1}\right] \\ &= \frac{m_b}{w_b}\left(E\left[\frac{1}{n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}\right] - E\left[\frac{1}{m_b^*}\sum_{i=1}^{m_b^*} y_{ikb|r_b=1,s_1=1}\right]\right) \end{aligned} \tag{A39}$$

Inside the parentheses of Line A39, the first term is the average of n_b^* random draws from the y_{ikb} values of the n^* always-responders. Likewise, the second term is the average of m_b^* random draws from the y_{ikb} values of the m_b sometimes-responders who would only complete the survey in Wave 1. Therefore, we have

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

and

$$E\left[\frac{1}{m_b^*}\sum_{i=1}^{m_b^*} y_{ikb|r_b=1,s_1=1}\right] = \frac{1}{m_b}\sum_{i=1}^{m_b} y_{ikb|s_1=1}$$

We can then write

$$\frac{m_b}{w_b}\left(E\left[\frac{1}{n_b^*}\sum_{i=1}^{n_b^*} y_{ikb|r_b=1,u=1}\right] - E\left[\frac{1}{m_b^*}\sum_{i=1}^{m_b^*} y_{ikb|r_b=1,s_1=1}\right]\right) = \frac{m_b}{w_b}\left(\frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} - \frac{1}{m_b}\sum_{i=1}^{m_b} y_{ikb|s_1=1}\right)$$

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

$$\begin{aligned}
Bias\left(\hat{\tau}_{k|u=1}\right) &= \frac{m_a}{w_a}\left(\frac{1}{m_a}\sum_{i=1}^{m_a} y_{ikt|s_2=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right) + \\
&\frac{m_b}{w_b}\left(\frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} - \frac{1}{m_b}\sum_{i=1}^{m_b} y_{ikb|s_1=1}\right) + \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
&= \frac{m_a}{w_a}\left(\bar{y}_{ikt|s_2=1} - \bar{y}_{ikt|u=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right) + \\
&\frac{m_b}{w_b}\left(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|s_1=1}\right) + \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
&= \frac{m_a}{w_a}\left(\bar{y}_{ikt|s_2=1} - \bar{y}_{ikt|u=1}\right) + \frac{m_b}{w_b}\left(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|s_1=1}\right) + \qquad \text{(A40)} \\
&Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right) + \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \qquad \text{(A41)}
\end{aligned}$$

The first expression in Line A40 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 A40 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.

$$Bias_{\mathbf{S}}\left(\hat{\tau}_{k|u=1}\right) = \frac{m_a}{w_a}\left(\bar{y}_{ikt|s_2=1} - \bar{y}_{ikt|u=1}\right) + \frac{m_b}{w_b}\left(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|s_1=1}\right)$$

We now have

$$Bias\left(\hat{\tau}_{k|u=1}\right) = Bias_{\mathbf{S}}\left(\hat{\tau}_{k|u=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right) + \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \tag{A42}$$

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

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

In sum, we can write the overall bias as

$$Bias\left(\hat{\tau}_{k|u=1}\right) = Bias_{\mathbf{S}}\left(\hat{\tau}_{k|u=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right) + Bias_{\mathbf{M}}\left(\hat{\tau}_{k|u=1}\right) \tag{A43}$$

where

$$\begin{aligned}
Bias_{\mathbf{S}}\left(\hat{\tau}_{k|u=1}\right) &= \tfrac{m_a}{w_a}\left(\bar{y}_{ikt|s_2=1} - \bar{y}_{ikt|u=1}\right) + \tfrac{m_b}{w_b}\left(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|s_1=1}\right) \\
Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right) &= \bar{y}_{ikc|u=1} - \bar{y}_{ikbc|u=1} \\
Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right) &= \bar{y}_{ikbc|u=1} - \bar{y}_{ikb|u=1} \\
Bias_{\mathbf{M}}\left(\hat{\tau}_{k|u=1}\right) &= \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}
\end{aligned}$$

□

5.2 Deriving the Bias When Estimating $\bar{\tau}_{k|w=1}$

Begin by recalling that 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 is written as

$$\bar{\tau}_{k|w=1} = \tfrac{1}{n_w}\sum_{i=1}^{n_w}\left(y_{ikt|w=1} - y_{ikc|w=1}\right) \tag{A44}$$

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 the survey in Wave 2 and the n_b respondents who complete it in Wave 1.

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

The bias in this estimator is therefore

$$
\begin{aligned}
Bias\left(\hat{\tau}_{k|w=1}\right) &= E\left[\hat{\tau}_{k|w=1}\right] - \bar{\tau}_{k|w=1} \\
&= E\left[\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikto|r_a=1} - \tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikbo|r_b=1}\right] - \tfrac{1}{n_w}\sum_{i=1}^{n_w}\left(y_{ikt|w=1} - y_{ikc|w=1}\right) \\
&= E\left[\tfrac{1}{n_a}\sum_{i=1}^{n_a}\left(y_{ikt|r_a=1} + \epsilon_{ikt|r_a=1}\right) - \tfrac{1}{n_b}\sum_{i=1}^{n_b}\left(y_{ikb|r_b=1} + \epsilon_{ikb|r_b=1}\right)\right] - \tfrac{1}{n_w}\sum_{i=1}^{n_w}\left(y_{ikt} - y_{ikc}\right) \\
&= E\left[\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikt|r_a=1}\right] + \bar{\epsilon}_{ikt|w_2=1} - E\left[\tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}\right] - \bar{\epsilon}_{ikb|w_1=1} - \tfrac{1}{n_w}\sum_{i=1}^{n_w}\left(y_{ikt|w=1} - y_{ikc|w=1}\right) \\
&= E\left[\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikt|r_a=1}\right] - E\left[\tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}\right] - \tfrac{1}{n_w}\sum_{i=1}^{n_w} y_{ikt|w=1} + \tfrac{1}{n_w}\sum_{i=1}^{n_w} y_{ikc|w=1} + \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
&= \left(E\left[\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikt|r_a=1}\right] - \tfrac{1}{n_w}\sum_{i=1}^{n_w} y_{ikt|w=1}\right) + \left(\tfrac{1}{n_w}\sum_{i=1}^{n_w} y_{ikc|w=1} - E\left[\tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}\right]\right) + \quad \text{(A45)} \\
&\bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \quad \text{(A46)}
\end{aligned}
$$

Note that, as before, $w_{i1} \in \{0,1\}$ denotes whether an individual would complete the survey if assigned to Wave 1 and $w_{i2} \in \{0,1\}$ denotes whether they would complete the survey if assigned to Wave 2. Similarly, $w_i \in \{0,1\}$ denotes whether individual i would complete the survey in at least one of the two waves. Likewise, $w_b = n^* + m_b$ denotes the number of individuals who would complete the survey if asked to do it in Wave 1, and $w_a = n^* + m_a$ denotes the number of individuals who would complete the survey if asked to do it in Wave 2. Focusing on the first difference in Line A45, we can rewrite the expression as

$$
\begin{aligned}
E\left[\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikt|r_a=1}\right] - \tfrac{1}{n_w}\sum_{i=1}^{n_w} y_{ikt|w=1} &= \tfrac{1}{w_a}\sum_{i=1}^{w_a} y_{ikt|w_2=1} - \tfrac{1}{n_w}\sum_{i=1}^{n_w} y_{ikt|w=1} \\
&= \bar{y}_{ikt|w_2=1} - \bar{y}_{ikt|w=1} \quad \text{(A47)}
\end{aligned}
$$

Similarly, the second difference in Line A45 can be rewritten as

$$
\begin{aligned}
\tfrac{1}{n_w}\sum_{i=1}^{n_w} y_{ikc|w=1} - E\left[\tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}\right] &= \tfrac{1}{n_w}\sum_{i=1}^{n_w} y_{ikc|w=1} - \tfrac{1}{w_b}\sum_{i=1}^{w_b} y_{ikb|w_1=1} \\
&= \tfrac{1}{n_w}\sum_{i=1}^{n_w} y_{ikc|w=1} - \tfrac{1}{w_b}\sum_{i=1}^{w_b} y_{ikb|w_1=1} + \tfrac{1}{n_w}\sum_{i=1}^{m_a} y_{ikb|s_2=1} - \tfrac{1}{n_w}\sum_{i=1}^{m_a} y_{ikb|s_2=1} \\
&= \tfrac{1}{n_w}\sum_{i=1}^{n_w} y_{ikc|w=1} - \left(\tfrac{1}{w_b}\sum_{i=1}^{w_b} y_{ikb|w_1=1} + \tfrac{1}{n_w}\sum_{i=1}^{m_a} y_{ikb|s_2=1}\right) + \quad \text{(A48)} \\
&\tfrac{1}{n_w}\sum_{i=1}^{m_a} y_{ikb|s_2=1} \quad \text{(A49)}
\end{aligned}
$$

We can rewrite the expression inside the parentheses in Line A48 as

$$\begin{aligned}
\tfrac{1}{w_b}\sum_{i=1}^{w_b} y_{ikb|w_1=1} + \tfrac{1}{n_w}\sum_{i=1}^{m_a} y_{ikb|s_2=1} &= \tfrac{n_w}{w_b n_w}\sum_{i=1}^{w_b} y_{ikb|w_1=1} + \tfrac{1}{n_w}\sum_{i=1}^{m_a} y_{ikb|s_2=1} \\
&= \tfrac{n_w+(w_b-w_b)}{w_b n_w}\sum_{i=1}^{w_b} y_{ikb|w_1=1} + \tfrac{1}{n_w}\sum_{i=1}^{m_a} y_{ikb|s_2=1} \\
&= \tfrac{n_w-w_b}{w_b n_w}\sum_{i=1}^{w_b} y_{ikb|w_1=1} + \tfrac{w_b}{w_b n_w}\sum_{i=1}^{w_b} y_{ikb|w_1=1} + \tfrac{1}{n_w}\sum_{i=1}^{m_a} y_{ikb|s_2=1} \\
&= \tfrac{n_w-w_b}{w_b n_w}\sum_{i=1}^{w_b} y_{ikb|w_1=1} + \tfrac{1}{n_w}\sum_{i=1}^{w_b} y_{ikb|w_1=1} + \tfrac{1}{n_w}\sum_{i=1}^{m_a} y_{ikb|s_2=1} \\
&= \tfrac{n_w-w_b}{w_b n_w}\sum_{i=1}^{w_b} y_{ikb|w_1=1} + \tfrac{1}{n_w}\sum_{i=1}^{n_w} y_{ikb|w=1}
\end{aligned}$$

Combining Lines A47, A48-A49, and A50, we get

$$\begin{aligned}
Bias\left(\hat{\tau}_{k|w=1}\right) &= \bar{y}_{ikt|w_2=1} - \bar{y}_{ikt|w=1} + \tfrac{1}{n_w}\sum_{i=1}^{n_w} y_{ikc|w=1} - \left(\tfrac{n_w-w_b}{w_b n_w}\sum_{i=1}^{w_b} y_{ikb|w_1=1} + \tfrac{1}{n_w}\sum_{i=1}^{n_w} y_{ikb|w=1}\right) + \tfrac{1}{n_w}\sum_{i=1}^{m_a} y_{ikb|s_2=1} + \\
&\quad \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
&= \left(\bar{y}_{ikt|w_2=1} - \bar{y}_{ikt|w=1}\right) + \left(\tfrac{1}{n_w}\sum_{i=1}^{n_w} y_{ikc|w=1} - \tfrac{1}{n_w}\sum_{i=1}^{n_w} y_{ikb|w=1}\right) + \tfrac{w_b-n_w}{w_b n_w}\sum_{i=1}^{w_b} y_{ikb|w_1=1} + \tfrac{1}{n_w}\sum_{i=1}^{m_a} y_{ikb|s_2=1} + \\
&\quad \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
&= \left(\bar{y}_{ikt|w_2=1} - \bar{y}_{ikt|w=1}\right) + \left(\bar{y}_{ikc|w=1} - \bar{y}_{ikb|w=1}\right) + \tfrac{w_b-n_w}{w_b n_w}\sum_{i=1}^{w_b} y_{ikb|w_1=1} + \tfrac{1}{n_w}\sum_{i=1}^{m_a} y_{ikb|s_2=1} + \\
&\quad \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}
\end{aligned}$$

Similar to our discussion in Section 2 of the article, this difference $\bar{y}_{ikc|w=1} - \bar{y}_{ikb|w=1}$ can be decomposed into the bias from temporal factors and the bias from anticipatory factors. Thus, we can write

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

where

$$\begin{aligned}
Bias_{\mathbf{T}}\left(\hat{\tau}_{k|w=1}\right) &= \bar{y}_{ikc|w=1} - \bar{y}_{ikbc|w=1} \\
Bias_{\mathbf{A}}\left(\hat{\tau}_{k|w=1}\right) &= \bar{y}_{ikbc|w=1} - \bar{y}_{ikb|w=1}
\end{aligned}$$

We can now write the overall bias term as:

$$Bias\left(\hat{\tau}_{k|w=1}\right) = \left(\bar{y}_{ikt|w_2=1} - \bar{y}_{ikt|w=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|w=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|w=1}\right) + \tag{A50}$$

$$\left(\tfrac{w_b - n_w}{w_b n_w}\sum_{i=1}^{w_b} y_{ikb|w_1=1} + \tfrac{1}{n_w}\sum_{i=1}^{m_a} y_{ikb|s_2=1}\right) + \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \tag{A51}$$

The first difference in Line A51 can be written as

$$\begin{aligned}
\tfrac{w_b - n_w}{w_b n_w}\sum_{i=1}^{w_b} y_{ikb|w_1=1} + \tfrac{1}{n_w}\sum_{i=1}^{m_a} y_{ikb|s_2=1} &= \tfrac{1}{n_w}\sum_{i=1}^{w_b} y_{ikb|w_1=1} - \tfrac{1}{w_b}\sum_{i=1}^{w_b} y_{ikb|w_1=1} + \tfrac{1}{n_w}\sum_{i=1}^{m_a} y_{ikb|s_2=1} \\
&= \tfrac{1}{n_w}\sum_{i=1}^{w_b} y_{ikb|w_1=1} + \tfrac{1}{n_w}\sum_{i=1}^{m_a} y_{ikb|s_2=1} - \tfrac{1}{w_b}\sum_{i=1}^{w_b} y_{ikb|w_1=1} \\
&= \tfrac{1}{n_w}\sum_{i=1}^{n_w} y_{ikb|w=1} - \tfrac{1}{w_b}\sum_{i=1}^{w_b} y_{ikb|w_1=1} \\
&= \bar{y}_{ikb|w=1} - \bar{y}_{ikb|w_1=1}
\end{aligned}$$

We can now write the overall bias term as

$$\begin{aligned}
Bias\left(\hat{\tau}_{k|w=1}\right) &= \left(\bar{y}_{ikt|w_2=1} - \bar{y}_{ikt|w=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|w=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|w=1}\right) + \left(\bar{y}_{ikb|w=1} - \bar{y}_{ikb|w_1=1}\right) + \\
&\quad \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
&= \left(\bar{y}_{ikt|w_2=1} - \bar{y}_{ikt|w=1}\right) + \left(\bar{y}_{ikb|w=1} - \bar{y}_{ikb|w_1=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|w=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|w=1}\right) + \quad \text{(A52)} \\
&\quad \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \quad \text{(A53)}
\end{aligned}$$

The differences within the two sets of parentheses in Line A52 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

$$\left(\bar{y}_{ikt|w_2=1} - \bar{y}_{ikb|w_1=1}\right) - \left(\bar{y}_{ikt|w=1} - \bar{y}_{ikb|w=1}\right) = \left(\bar{y}_{ikt|w_2=1} - \bar{y}_{ikb|w_1=1}\right) - \left(\tfrac{w_a}{n_w}\right)\bar{y}_{ikt|w_2=1} - \left(\tfrac{m_b}{n_w}\right)\bar{y}_{ikt|s_1=1} + \tag{A54}$$

$$\left(\tfrac{w_b}{n_w}\right)\bar{y}_{ikb|w_1=1} + \left(\tfrac{m_a}{n_w}\right)\bar{y}_{ikb|s_2=1} \tag{A55}$$

$$= \tfrac{m_b}{n_w}\left(\bar{y}_{ikt|w_2=1} - \bar{y}_{ikt|s_1=1}\right) + \tfrac{m_a}{n_w}\left(\bar{y}_{ikb|s_2=1} - \bar{y}_{ikb|w_1=1}\right) \tag{A56}$$

We can label this bias

$$Bias_{\mathbf{S}}\left(\hat{\tau}_{k|w=1}\right) = \tfrac{m_b}{n_w}\left(\bar{y}_{ikt|w_2=1} - \bar{y}_{ikt|s_1=1}\right) + \tfrac{m_a}{n_w}\left(\bar{y}_{ikb|s_2=1} - \bar{y}_{ikb|w_1=1}\right)$$

We can now write the overall bias term as

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

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

$$Bias_{\mathbf{M}}\left(\hat{\tau}_{k|w=1}\right) = \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}$$

In sum, we can write the overall bias as

$$Bias\left(\hat{\tau}_{k|w=1}\right) = Bias_{\mathbf{S}}\left(\hat{\tau}_{k|w=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|w=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|w=1}\right) + Bias_{\mathbf{M}}\left(\hat{\tau}_{k|w=1}\right) \tag{A57}$$

where

$$\begin{aligned}
Bias_{\mathbf{S}}\left(\hat{\tau}_{k|w=1}\right) &= \tfrac{m_b}{n_w}\left(\bar{y}_{ikt|w_2=1} - \bar{y}_{ikt|s_1=1}\right) + \tfrac{m_a}{n_w}\left(\bar{y}_{ikb|s_2=1} - \bar{y}_{ikb|w_1=1}\right) \\
Bias_{\mathbf{T}}\left(\hat{\tau}_{k|w=1}\right) &= \bar{y}_{ikc|w=1} - \bar{y}_{ikbc|w=1} \\
Bias_{\mathbf{A}}\left(\hat{\tau}_{k|w=1}\right) &= \bar{y}_{ikbc|w=1} - \bar{y}_{ikb|w=1} \\
Bias_{\mathbf{M}}\left(\hat{\tau}_{k|w=1}\right) &= \bar{\epsilon}_{ikt|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}
\end{aligned}$$

5.3 Comparing Bias in the Rolling Cross-Section Design to Bias in the Baseline Model

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 average treatment effect for all Wave 1 and Wave 2 respondents. The $Bias_{\mathbf{S}}\left(\hat{\tau}_{k|w=1}\right)$ term becomes a weighted average of the pre-event and post-event differences between the Wave 1 and Wave 2 respondents' truthful answers. Clearly, bias when estimating the average treatment effect for Wave 2 respondents in the baseline model is more straightforward to comprehend than bias when estimating the average treatment effect for both Wave 1 and Wave 2 respondents.

We can now examine how the bias in the estimators $\hat{\tau}_{k|u=1}$ and $\hat{\tau}_{k|w=1}$ from the rolling cross-section design compare to the bias in the baseline model from the article. 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 A43, we can better understand $Bias_{\mathbf{S}}\left(\hat{\tau}_{k|u=1}\right)$ 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, $\frac{m_a}{w_a} = \frac{m_b}{w_b}$, which we will denote as $\alpha \leq 1$. This symmetry allows us to rewrite $Bias_{\mathbf{S}}\left(\hat{\tau}_{k|u=1}\right)$ as

$$\begin{aligned} Bias_{\mathbf{S}}\left(\hat{\tau}_{k|u=1}|m_a = m_b\right) &= \tfrac{m_a}{w_a}\left(\bar{y}_{ikt|s_2=1} - \bar{y}_{ikt|u=1}\right) + \tfrac{m_b}{w_b}\left(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|s_1=1}\right) \\ &= \alpha\left(\bar{y}_{ikt|s_2=1} - \bar{y}_{ikt|u=1}\right) + \alpha\left(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|s_1=1}\right) \\ &= \alpha\left(\bar{y}_{ikt|s_2=1} - \bar{y}_{ikb|s_1=1} - \bar{y}_{ikt|u=1} + \bar{y}_{ikb|u=1}\right) \\ &= \alpha\left[\left(\bar{y}_{ikt|s_2=1} - \bar{y}_{ikb|s_1=1}\right) - \left(\bar{y}_{ikt|u=1} - \bar{y}_{ikb|u=1}\right)\right] \end{aligned} \quad \text{(A58)}$$

Note that the terms inside both sets of parentheses in Line A58 resemble our estimator in the baseline model of Section 2 of the article. 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 $\hat{\tau}_{k|r_a=1}$ 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 the equation $Bias\left(\hat{\tau}\right) = E\left[\hat{\tau}\right] - \bar{\tau}$. We can therefore write

$$\begin{aligned} Bias_{\mathbf{S}}\left(\hat{\tau}_{k|u=1}|m_a = m_b\right) = \alpha\Big[&\left(\bar{\tau}_{k|s_2=1} + Bias_{\mathbf{X}}\left(\hat{\tau}_{k|s_2=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|s_2=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|s_2=1}\right)\right) - \\ &\left(\bar{\tau}_{k|u=1} + Bias_{\mathbf{X}}\left(\hat{\tau}_{k|u=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right)\right)\Big] \\ = \alpha\Big[&\left(\bar{\tau}_{k|s_2=1} + Bias_{\mathbf{X}}\left(\hat{\tau}_{k|s_2=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|s_2=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|s_2=1}\right)\right) - \\ &\left(\bar{\tau}_{k|u=1} + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right)\right)\Big] \end{aligned}$$

Substituting this expression into Equation A43, we obtain

$$Bias\left(\hat{\tau}_{k|u=1}|m_a = m_b\right) = \alpha Bias_{\mathbf{X}}\left(\hat{\tau}_{k|s_2=1}\right) + \left(\alpha Bias_{\mathbf{T}}\left(\hat{\tau}_{k|s_2=1}\right) + (1-\alpha)\, Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right)\right) + \quad \text{(A59)}$$

$$\left(\alpha Bias_{\mathbf{A}}\left(\hat{\tau}_{k|s_2=1}\right) + (1-\alpha)\, Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right)\right) + Bias_{\mathbf{M}}\left(\hat{\tau}_{k|u=1}\right) + \quad \text{(A60)}$$

$$\alpha\left(\bar{\tau}_{k|s_2=1} - \bar{\tau}_{k|u=1}\right) \quad \text{(A61)}$$

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, 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 A57 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 $\frac{m_a}{n_w}$ and $\frac{m_b}{n_w}$ will be equal. In this situation, we can define $\lambda = \frac{m_a}{n_w} = \frac{m_b}{n_w}$. We can then rewrite the equation for $Bias_{\mathbf{S}}\left(\hat{\tau}_{k|w=1}|m_a = m_b\right)$ as

$$Bias_{\mathbf{S}}\left(\hat{\tau}_{k|w=1}|m_a = m_b\right) = \left(\bar{y}_{ikt|w_2=1} - \bar{y}_{ikb|w_1=1}\right) - \left(\bar{y}_{ikt|w=1} - \bar{y}_{ikb|w=1}\right)$$

Here we are utilizing the way that we wrote $Bias_{\mathbf{S}}(\hat{\tau}_{k|w=1})$ at the beginning of Line A54.

This expression is very similar to what we saw in Line A58. As 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. Also as 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

$$Bias_{\mathbf{S}}\left(\hat{\tau}_{k|w=1}|m_a = m_b\right) = \left(\bar{\tau}_{k|w_2=1} + Bias_{\mathbf{X}}\left(\hat{\tau}_{k|w_2=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|w_2=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|w_2=1}\right)\right) - \tag{A62}$$

$$\left(\bar{\tau}_{k|w=1} + Bias_{\mathbf{X}}\left(\hat{\tau}_{k|w=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|w=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|w=1}\right)\right) \tag{A63}$$

Note that in Line A62, $Bias_{\mathbf{X}}\left(\hat{\tau}_{k|w_2=1}\right) = \bar{y}_{ikb|w_2=1} - \bar{y}_{ikb|w_1=1}$. Since the pre-event and post-event samples in Line A63 consist of n_w individuals with exactly the same demographic characteristics, we can drop the $Bias_{\mathbf{X}}(\hat{\tau}_{k|w=1})$ term.

$$Bias_{\mathbf{S}}\left(\hat{\tau}_{k|w=1}|m_a = m_b\right) = \left(\bar{\tau}_{k|w_2=1} + Bias_{\mathbf{X}}\left(\hat{\tau}_{k|w_2=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|w_2=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|w_2=1}\right)\right) -$$

$$\left(\bar{\tau}_{k|w=1} + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|w=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|w=1}\right)\right)$$

Substituting this expression into Line A57, we get

$$Bias\left(\hat{\tau}_{k|w=1}|m_a = m_b\right) = Bias_{\mathbf{X}}\left(\hat{\tau}_{k|w_2=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|w_2=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|w_2=1}\right) + Bias_{\mathbf{M}}\left(\hat{\tau}_{k|w=1}\right) +$$

$$\left(\bar{\tau}_{k|w_2=1} - \bar{\tau}_{k|w=1}\right)$$

Note that this equation can be rewritten as

$$Bias\left(\hat{\tau}_{k|w=1}|m_a=m_b\right)=\alpha Bias_{\mathbf{X}}\left(\hat{\tau}_{k|s_2=1}\right)+\alpha Bias_{\mathbf{T}}\left(\hat{\tau}_{k|s_2=1}\right)+(1-\alpha)Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right)+$$
$$\alpha Bias_{\mathbf{A}}\left(\hat{\tau}_{k|s_2=1}\right)+(1-\alpha)Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right)+Bias_{\mathbf{M}}\left(\hat{\tau}_{k|w=1}\right)+$$
$$\left(\bar{\tau}_{k|w_2=1}-\bar{\tau}_{k|w=1}\right)$$

which is very similar to the expression we derived for $Bias\left(\hat{\tau}_{k|u=1}|m_a=m_b\right)$.

6 Panel Designs

In a panel design, we begin with a group of individuals who have the opportunity to take the same survey in Wave 1 and Wave 2. We can denote whether individual i takes 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 we estimate is the average treatment effect of the event on these n^* respondents' truthful answers to question k of the survey:

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

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 $\bar{\tau}_{k|u=1}$ is

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

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.

6.1 Proof of Proposition 5

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

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

where

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

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

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

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

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

$$\begin{aligned}
Bias\left(\hat{\tau}_{k|u=1}\right) &= E\left[\hat{\tau}_{k|u=1}\right] - \bar{\tau}_{k|u=1} \\
&= E\left[\frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikao|u=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikbo|u=1}\right] - \frac{1}{n^*}\sum_{i=1}^{n^*}\left(y_{ikt|u=1} - y_{ikc|u=1}\right) \\
&= \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikao|u=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikbo|u=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} + \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} \\
&= \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ika|u=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} + \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} + \bar{\epsilon}_{ka|u=1} - \bar{\epsilon}_{kb|u=1} \\
&= \left(\frac{1}{n^*}\sum_{i=1}^{n^*} y_{ika|u=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1}\right) + \left(\frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1}\right) + \bar{\epsilon}_{ka|u=1} - \bar{\epsilon}_{kb|u=1}
\end{aligned} \tag{A64}$$

The first difference in Line A64 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

$$Bias_{\mathbf{C}}\left(\hat{\tau}_{k|u=1}\right) = \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ika|u=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} = \bar{y}_{ika|u=1} - \bar{y}_{ikt|u=1} \tag{A65}$$

Returning to Line A64, the second difference is similar to what we saw in Equation 5 from Section 2 of the article. Following what we did in Section 2, we can decompose this expression into bias from temporal factors and bias from anticipatory factors:

$$\frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} = Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right) \tag{A66}$$

where

$$Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right) = \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikbc|u=1} = \bar{y}_{ikc|u=1} - \bar{y}_{ikbc|u=1} \tag{A67}$$

$$Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right) = \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikbc|u=1} - \frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} = \bar{y}_{ikbc|u=1} - \bar{y}_{ikb|u=1} \tag{A68}$$

Finally, the last difference in Line A64 is just the potential difference in misreporting between Waves 1 and 2.

$$Bias_{\mathbf{M}}\left(\hat{\tau}_{k|u=1}\right) = \bar{\epsilon}_{ka|u=1} - \bar{\epsilon}_{kb|u=1} \tag{A69}$$

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

$$Bias\left(\hat{\tau}_{k|u=1}\right) = Bias_{\mathbf{C}}\left(\hat{\tau}_{k|u=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right) + Bias_{\mathbf{M}}\left(\hat{\tau}_{k|u=1}\right) \tag{A70}$$

where

$$Bias_{\mathbf{C}}\left(\hat{\tau}_{k|u=1}\right) = \bar{y}_{ika|u=1} - \bar{y}_{ikt|u=1} \tag{A71}$$

$$Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right) = \bar{y}_{ikc|u=1} - \bar{y}_{ikbc|u=1} \tag{A72}$$

$$Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right) = \bar{y}_{ikbc|u=1} - \bar{y}_{ikb|u=1} \tag{A73}$$

$$Bias_{\mathbf{M}}\left(\hat{\tau}_{k|u=1}\right) = \bar{\epsilon}_{ka|u=1} - \bar{\epsilon}_{kb|u=1} \tag{A74}$$

□

7 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 Survey A in Wave 2 and n_b denote the number of individuals who complete Survey A in Wave 1, such that $n_a + n_b = n$. Due to 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 to do first 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 sometimes-responders who would only complete both surveys if assigned to do Survey A in Wave 2 by m_a, and the number of sometimes-responders who would only complete both surveys if assigned to do Survey A in Wave 1 by m_b. We will also let w_a denote the total possible number of individuals who would complete both surveys if assigned to do Survey A in Wave 2 and w_b denote the total possible number of individuals who would complete both surveys if assigned to do Survey A in Wave 1, such that $w_a = n^* + m_a$ and $w_b = n^* + m_b$. 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. As before, we will use $u_i \in \{0, 1\}$ to denote whether individual i is an always-responder. We will also use $w_{i1} \in \{0, 1\}$ to denote whether individual i would complete both surveys if assigned to do Survey A in Wave 1 and $w_{i2} \in \{0, 1\}$ to denote whether individual i would complete both surveys if assigned to do Survey A 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}).

7.1 Proof of Proposition 6

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

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

where

$$\begin{aligned}
Bias_{\mathbf{D}}\left(\hat{\tau}_{k|u=1}\right) &= \tfrac{m_a}{w_a}\left(\bar{y}_{ika|s_2=1} - \bar{y}_{ika|u=1}\right) + \tfrac{m_b}{w_b}\left(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|s_1=1}\right) \\
Bias_{\mathbf{P}}\left(\hat{\tau}_{k|u=1}\right) &= \bar{y}_{ika|u=1} - \bar{y}_{ikt|u=1} \\
Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right) &= \bar{y}_{ikc|u=1} - \bar{y}_{ikbc|u=1} \\
Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right) &= \bar{y}_{ikbc|u=1} - \bar{y}_{ikb|u=1} \\
Bias_{\mathbf{M}}\left(\hat{\tau}_{k|u=1}\right) &= \bar{\epsilon}_{ika|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}
\end{aligned}$$

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

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

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

$$\begin{aligned}
Bias\left(\hat{\tau}_{k|u=1}\right) &= E\left[\hat{\tau}_{k|u=1}\right] - \bar{\tau}_{k|u=1} \\
&= E\left[\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ikao|r_a=1} - \tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikbo|r_b=1}\right] - \tfrac{1}{n^*}\sum_{i=1}^{n^*}\left(y_{ikt|u=1} - y_{ikc|u=1}\right) \\
&= E\left[\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ika|r_a=1} - \tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}\right] - \tfrac{1}{n^*}\sum_{i=1}^{n^*}\left(y_{ikt|u=1} - y_{ikc|u=1}\right) + \bar{\epsilon}_{ika|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
&= E\left[\tfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ika|r_a=1}\right] - E\left[\tfrac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}\right] - \tfrac{1}{n^*}\sum_{i=1}^{n^*}\left(y_{ikt|u=1} - y_{ikc|u=1}\right) + \bar{\epsilon}_{ika|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
&= \tfrac{1}{w_a}\sum_{i=1}^{w_a} y_{ika|w_2=1} - \tfrac{1}{w_b}\sum_{i=1}^{w_b} y_{ikb|r_b=1} - \tfrac{1}{n^*}\sum_{i=1}^{n^*}\left(y_{ikt|u=1} - y_{ikc|u=1}\right) + \bar{\epsilon}_{ika|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
&= \tfrac{1}{w_a}\sum_{i=1}^{w_a} y_{ika|w_2=1} - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} + \left(\tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikc|u=1} - \tfrac{1}{w_b}\sum_{i=1}^{w_b} y_{ikb|w_1=1}\right) + \bar{\epsilon}_{ika|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
&= \tfrac{1}{w_a}\sum_{i=1}^{w_a} y_{ika|w_2=1} - \bar{y}_{ikt|u=1} + \left(\bar{y}_{ikc|u=1} - \tfrac{1}{w_b}\sum_{i=1}^{w_b} y_{ikb|w_1=1}\right) + \bar{\epsilon}_{ika|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \qquad \text{(A75)}
\end{aligned}$$

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

$$\tfrac{1}{w_a}\sum_{i=1}^{w_a} y_{ika|w_2=1} = \tfrac{1}{w_a}\sum_{i=1}^{m_a} y_{ika|s_2=1} + \tfrac{1}{w_a}\sum_{i=1}^{n^*} y_{ika|u=1} \qquad \text{(A76)}$$

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

$$\begin{aligned}
\tfrac{1}{w_a}\sum_{i=1}^{n^*} y_{ika|u=1} &= \tfrac{n^*}{w_a n^*}\sum_{i=1}^{n^*} y_{ika|u=1} \\
&= \tfrac{n^*+(w_a-w_a)}{w_a n^*}\sum_{i=1}^{n^*} y_{ika|u=1} \\
&= \tfrac{n^*-w_a}{w_a n^*}\sum_{i=1}^{n^*} y_{ika|u=1} + \tfrac{w_a}{w_a n^*}\sum_{i=1}^{n^*} y_{ika|u=1} \\
&= \tfrac{-m_a}{w_a n^*}\sum_{i=1}^{n^*} y_{ika|u=1} + \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ika|u=1} \\
&= -\tfrac{m_a}{w_a n^*}\sum_{i=1}^{n^*} y_{ika|u=1} + \bar{y}_{ika|u=1} \qquad \text{(A77)}
\end{aligned}$$

We can then use Lines A76 and A77 to rewrite the overall bias term in Line A75 as

$$Bias\left(\hat{\tau}_{k|u=1}\right) = \tfrac{1}{w_a}\sum_{i=1}^{m_a} y_{ika|s_2=1} - \tfrac{m_a}{w_a n^*}\sum_{i=1}^{n^*} y_{ika|u=1} + \bar{y}_{ika|u=1} - \bar{y}_{ikt|u=1} + \tag{A78}$$

$$\left(\bar{y}_{ikc|u=1} - \tfrac{1}{w_b}\sum_{i=1}^{w_b} y_{ikb|w_1=1}\right) + \bar{\epsilon}_{ika|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \tag{A79}$$

The second difference in Line A81 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:

$$Bias_{\mathbf{P}}\left(\hat{\tau}_{k|u=1}\right) = \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ika|u=1} - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikt|u=1} = \bar{y}_{ika|u=1} - \bar{y}_{ikt|u=1} \tag{A80}$$

Therefore, we can rewrite the overall bias as

$$Bias\left(\hat{\tau}_{k|u=1}\right) = \tfrac{1}{w_a}\sum_{i=1}^{m_a} y_{ika|s_2=1} - \tfrac{m_a}{w_a n^*}\sum_{i=1}^{n^*} y_{ika|u=1} + Bias_{\mathbf{P}}\left(\hat{\tau}_{k|u=1}\right) + \tag{A81}$$

$$\left(\bar{y}_{ikc|u=1} - \tfrac{1}{w_b}\sum_{i=1}^{w_b} y_{ikb|w_1=1}\right) + \bar{\epsilon}_{ika|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \tag{A82}$$

We can now turn our attention to the first difference in Line A82. We can begin by noting that

$$\tfrac{1}{w_b}\sum_{i=1}^{w_b} y_{ikb|w_1=1} = \tfrac{1}{w_b}\sum_{i=1}^{m_b} y_{ikb|s_1=1} + \tfrac{1}{w_b}\sum_{i=1}^{n^*} y_{ikb|u=1} \tag{A83}$$

We can break the second term down further as

$$\begin{aligned}
\tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} &= \tfrac{n^*}{w_b n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} \\
&= \tfrac{n^*+(w_b-w_b)}{w_b n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} \\
&= \tfrac{n^*-w_b}{w_b n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} + \tfrac{w_b}{w_b n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} \\
&= \tfrac{-m_b}{w_b n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} + \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} \\
&= -\tfrac{m_b}{w_b n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} + \bar{y}_{ikb|u=1} \\
&= \bar{y}_{ikb|u=1} - \tfrac{m_b}{w_b n^*}\sum_{i=1}^{n^*} y_{ikb|u=1}
\end{aligned} \tag{A84}$$

Using Lines A83 and A84, we can now rewrite the overall bias term from Lines A81-A82 as

$$\begin{aligned}
Bias\left(\hat{\tau}_{k|u=1}\right) &= \tfrac{1}{w_a}\sum_{i=1}^{m_a} y_{ika|s_2=1} - \tfrac{m_a}{w_a n^*}\sum_{i=1}^{n^*} y_{ika|u=1} + Bias_{\mathbf{P}}\left(\hat{\tau}_{k|u=1}\right) + \\
&\quad \bar{y}_{ikc|u=1} - \left(\tfrac{1}{w_b}\sum_{i=1}^{m_b} y_{ikb|s_1=1} + \bar{y}_{ikb|u=1} - \tfrac{m_b}{w_b n^*}\sum_{i=1}^{n^*} y_{ikb|u=1}\right) + \bar{\epsilon}_{ika|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \\
&= \tfrac{1}{w_a}\sum_{i=1}^{m_a} y_{ika|s_2=1} - \tfrac{m_a}{w_a n^*}\sum_{i=1}^{n^*} y_{ika|u=1} + Bias_{\mathbf{P}}\left(\hat{\tau}_{k|u=1}\right) + \\
&\quad \left(\bar{y}_{ikc|u=1} - \bar{y}_{ikb|u=1}\right) + \tfrac{m_b}{w_b n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} - \tfrac{1}{w_b}\sum_{i=1}^{m_b} y_{ikb|s_1=1} + \bar{\epsilon}_{ika|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}
\end{aligned}$$

Similar to what we did in Section 2 of the article, 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 by anticipatory factors:

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

where

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

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

$$Bias\left(\hat{\tau}_{k|u=1}\right) = \tfrac{1}{w_a}\sum_{i=1}^{m_a} y_{ika|s_2=1} - \tfrac{m_a}{w_a n^*}\sum_{i=1}^{n^*} y_{ika|u=1} + Bias_{\mathbf{P}}\left(\hat{\tau}_{k|u=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right) + \tag{A85}$$

$$Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right) + \tfrac{m_b}{w_b n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} - \tfrac{1}{w_b}\sum_{i=1}^{m_b} y_{ikb|s_1=1} + \bar{\epsilon}_{ika|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \tag{A86}$$

We can next note that

$$\begin{aligned}
\tfrac{1}{w_a}\sum_{i=1}^{m_a} y_{ika|s_2=1} - \tfrac{m_a}{w_a n^*}\sum_{i=1}^{n^*} y_{ika|u=1} &= \left(\tfrac{m_a}{w_a}\right)\left(\tfrac{1}{m_a}\sum_{i=1}^{m_a} y_{ika|s_2=1} - \tfrac{1}{n^*}\sum_{i=1}^{n^*} y_{ika|u=1}\right) \\
&= \left(\tfrac{m_a}{w_a}\right)\left(\bar{y}_{ika|s_2=1} - \bar{y}_{ika|u=1}\right)
\end{aligned} \tag{A87}$$

Similarly,

$$\frac{m_b}{w_b n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} - \frac{1}{w_b}\sum_{i=1}^{m_b} y_{ikb|s_1=1} = \left(\frac{m_b}{w_b}\right)\left(\frac{1}{n^*}\sum_{i=1}^{n^*} y_{ikb|u=1} - \frac{1}{m_b}\sum_{i=1}^{m_b} y_{ikb|s_1=1}\right)$$
$$= \left(\frac{m_b}{w_b}\right)\left(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|s_1=1}\right)$$

We can now rewrite the overall bias in the estimator from Lines A85-A86 as

$$Bias\left(\hat{\tau}_{k|u=1}\right) = \frac{m_a}{w_a}\left(\bar{y}_{ika|s_2=1} - \bar{y}_{ika|u=1}\right) + Bias_{\mathbf{P}}\left(\hat{\tau}_{k|u=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right) +$$
$$Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right) + \frac{m_b}{w_b}\left(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|s_1=1}\right) + \bar{\epsilon}_{ika|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}$$
$$= \frac{m_a}{w_a}\left(\bar{y}_{ika|s_2=1} - \bar{y}_{ika|u=1}\right) + \frac{m_b}{w_b}\left(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|s_1=1}\right) + \tag{A88}$$
$$Bias_{\mathbf{P}}\left(\hat{\tau}_{k|u=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right) + \bar{\epsilon}_{ika|w_2=1} - \bar{\epsilon}_{ikb|w_1=1} \tag{A89}$$

The expression in Line A88 can be labeled

$$Bias_{\mathbf{D}}\left(\hat{\tau}_{k|u=1}\right) = \frac{m_a}{w_a}\left(\bar{y}_{ika|s_2=1} - \bar{y}_{ika|u=1}\right) + \frac{m_b}{w_b}\left(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|s_1=1}\right)$$

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 A89 is just the potential bias caused by differential misreporting between the respondents in Waves 1 and 2. We can label this bias

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

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

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

where

$$Bias_{\mathbf{D}}\left(\hat{\tau}_{k|u=1}\right) = \frac{m_a}{w_a}\left(\bar{y}_{ika|s_2=1} - \bar{y}_{ika|u=1}\right) + \frac{m_b}{w_b}\left(\bar{y}_{ikb|u=1} - \bar{y}_{ikb|s_1=1}\right)$$
$$Bias_{\mathbf{P}}\left(\hat{\tau}_{k|u=1}\right) = \bar{y}_{ika|u=1} - \bar{y}_{ikt|u=1}$$
$$Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right) = \bar{y}_{ikc|u=1} - \bar{y}_{ikbc|u=1}$$
$$Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right) = \bar{y}_{ikbc|u=1} - \bar{y}_{ikb|u=1}$$
$$Bias_{\mathbf{M}}\left(\hat{\tau}_{k|u=1}\right) = \bar{\epsilon}_{ika|w_2=1} - \bar{\epsilon}_{ikb|w_1=1}$$

□

8 Pre-event/Post-event Survey Designs in APSR, AJPS, and JOP (2015-24)

Table 1 lists more than 25 studies that use the pre-event/post-event survey design. These studies examine a wide range of important events and outcomes of interest to political scientists. Muñoz, Falcó-Gimeno, and Hernández (2020: A7) also provide a table with information on 44 studies published in a variety of journals that focus on unexpected events. Their list spans political science, sociology, and economics. In short, studying the impact of important events through surveys is a key part of social science and will likely remain common well into the future.

Table 1: Selected pre-event/post-event survey design studies

Study	**Event(s)**	**Outcome(s)**
Balcells, Tellez, and Villamil 2024	Russian invasion of Ukraine	Spanish nationalism
Bartels, Horowitz, and Kramon 2023	Kenyan supreme court ruling	Judicial support; partisan backlash
Cohen et al. 2023	Bolsonaro election	Allegiance to political system
Epifanio, Giani, and Ivandic 2023	2005 London bombings	Support toward curbing freedoms
Harding and Nwokolo 2023	Boko Haram attacks	Political trust; national identification; ethnic identification
Mettler, Jacobs, and Zhu 2023	Republicans gaining control of Congress and Presidency	Support for the Affordable Care Act
Pop-Eleches and Way 2023	Repression of Moldova electoral protests	Opposition support
Singh and Tir 2023	Terrorist attacks	Reported electoral participation
Bateson and Weintraub 2022	2016 US presidential election	Trust in the United States
Berliner and Wehner 2022	Audits	Approval of Mayors
Hale 2022	Invasion of Crimea	Reported support for Putin
Holman, Merolla, and Zechmeister 2022	2017 Manchester terrorist attack	Support for Teresa May
Kalla and Broockman 2022	Personal persuasion campaigns	Affective polarization

Table 1: Selected pre-event/post-event survey design studies (continued)

Study	Event(s)	Outcome(s)
Ayoub, Page, and Whitt 2021	Pride event	Attitudes toward LGBT+ community
Croke 2021	Anti-malaria campaign	Leader approval
Goldsmith, Horiuchi, and Matush 2021	High-level state visits	Approval of visiting leader
Reny and Newman 2021	George Floyd protests	Attitudes toward the police and African-Americans
Batto and Beaulieu 2020	Legislative brawl	Evaluation of the legislature
Mikulaschek, Pant, and Tesfaye 2020	Iraqi PM resignation	Support for violent opposition; public service provision optimism
Frye and Borisova 2019	Election; protest	Trust in government
Alkon and Wang 2018	Pollution reduction intervention	Regime evaluation
Flesken 2018	Romanian campaign and election	National and ethnic salience
Baker et al. 2016	Party brand change	Party support/identification
Bishin et al. 2016	Supreme Court ruling	Attitudes toward gays and lesbians
Bisgaard 2015	Economic shock	Attitudes related to the economy
Branton et al. 2015	2006 immigration protests	Immigration policy preferences
Tesler 2015	Elite political communication	Public opinion
Hirano et al. 2015	Primary campaigns and elections	Perceptions of candidates

References

Alkon, Meir, and Erik H. Wang. 2018. "Pollution Lowers Support for China's Regime: Quasi-Experimental Evidence from Beijing." *The Journal of Politics* 80(1): 327–331.

Ayoub, Phillip M., Douglas Page, and Sam Whitt. 2021. "Pride amid Prejudice: The Influence of LGBT+ Rights Activism in a Socially Conservative Society." *American Political Science Review* 115(2): 467–485.

Baker, Andy, Barry Ames, Anand E. Sokhey, and Lucio R. Renno. 2016. "The Dynamics of Partisan Identification When Party Brands Change: The Case of the Workers Party in Brazil." *The Journal of Politics* 78(1): 197–213.

Balcells, Laia, Juan Fernando Tellez, and Francisco Villamil. 2024. "The Wars of Others: The Effect of the Russian Invasion of Ukraine on Spanish Nationalism." *The Journal of Politics* 86(1): 352–357.

Bartels, Brandon L., Jeremy Horowitz, and Eric Kramon. 2023. "Can Democratic Principles Protect High Courts from Partisan Backlash? Public Reactions to the Kenyan Supreme Court's Role in the 2017 Election Crisis." *American Journal of Political Science* 67(3): 790–807.

Bateson, Regina, and Michael Weintraub. 2022. "The 2016 Election and America's Standing Abroad: Quasi-Experimental Evidence of a Trump Effect." *The Journal of Politics* 84(4): 2300–2304.

Batto, Nathan F., and Emily Beaulieu. 2020. "Partisan Conflict and Citizens' Democratic Attitudes: How Partisanship Shapes Reactions to Legislative Brawls." *The Journal of Politics* 82(1): 315–328.

Berliner, Daniel, and Joachim Wehner. 2022. "Audits for Accountability: Evidence from Municipal By-Elections in South Africa." *The Journal of Politics* 84(3): 1581–1594.

Bisgaard, Martin. 2015. "Bias Will Find a Way: Economic Perceptions, Attributions of Blame, and Partisan-Motivated Reasoning During Crisis." *The Journal of Politics* 77(3): 849–860.

Bishin, Benjamin G., Thomas J. Hayes, Matthew B. Incantalupo, and Charles Anthony Smith. 2016. "Opinion Backlash and Public Attitudes: Are Political Advances in Gay Rights Counterproductive?" *American Journal of Political Science* 60(3): 625–648.

Branton, Regina, Valerie Martinez-Ebers, Tony E. Carey Jr., and Tetsuya Matsubayashi. 2015. "Social Protest and Policy Attitudes: The Case of the 2006 Immigrant Rallies." *American Journal of Political Science* 59(2): 390–402.

Cohen, Mollie J., Amy Erica Smith, Mason W. Moseley, and Matthew L. Layton. 2023. "Winners' Consent? Citizen Commitment to Democracy When Illiberal Candidates Win Elections." *American Journal of Political Science* 67(2): 261–276.

Croke, Kevin. 2021. "The Impact of Health Programs on Political Opinion: Evidence from Malaria Control in Tanzania." *The Journal of Politics* 83(1): 340–353.

Epifanio, Mariaelisa, Marco Giani, and Ria Ivandic. 2023. "Wait and See? Public Opinion Dynamics After Terrorist Attacks." *The Journal of Politics* 85(3): 843–859.

Flesken, Anaid. 2018. "Ethnic Parties, Ethnic Tensions? Results of an Original Election Panel Study." *American Journal of Political Science* 62(4): 967–981.

Frye, Timothy, and Ekaterina Borisova. 2019. "Elections, Protest, and Trust in Government: A Natural Experiment from Russia." *The Journal of Politics* 81(3): 820–832.

Goldsmith, Benjamin E., Yusaku Horiuchi, and Kelly Matush. 2021. "Does Public Diplomacy Sway Foreign Public Opinion? Identifying the Effect of High-Level Visits." *American Political Science Review* 115(4): 1342–1357.

Hale, Henry E. 2022. "Authoritarian Rallying as Reputational Cascade? Evidence from Putin's Popularity Surge After Crimea." *American Political Science Review* 116(2): 580–594.

Harding, Robin, and Arinze Nwokolo. 2023. "Terrorism, Trust, and Identity: Evidence from a Natural Experiment in Nigeria." *American Journal of Political Science.* Forthcoming.

Hirano, Shigeo, Gabriel S. Lenz, Maksim Pinkovskiy, and James M. Snyder Jr. 2015. "Voter Learning in State Primary Elections." *American Journal of Political Science* 59(1): 91–108.

Holman, Mirya R., Jennifer L. Merolla, and Elizabeth J. Zechmeister. 2022. "The Curious Case of Theresa May and the Public That Did Not Rally: Gendered Reactions to Terrorist Attacks Can Cause Slumps Not Bumps." *American Political Science Review* 116(1): 249–264.

Kalla, Joshua L, and David E Broockman. 2022. "Voter Outreach Campaigns Can Reduce Affective Polarization Among Implementing Political Activists: Evidence from Inside Three Campaigns." *American Political Science Review* 116(4): 1516–1522.

Mettler, Suzanne, Lawrence R. Jacobs, and Ling Zhu. 2023. "Policy Threat, Partisanship, and the Case of the Affordable Care Act." *American Political Science Review* 117(1): 296–310.

Mikulaschek, Christoph, Saurabh Pant, and Beza Tesfaye. 2020. "Winning Hearts and Minds in Civil Wars: Governance, Leadership Change, and Support for Violent Groups in Iraq." *American Journal of Political Science* 64(4): 773–790.

Muñoz, Jordi, Albert Falcó-Gimeno, and Enrique Hernández. 2020. "Unexpected Event During Survey Design: Promise and Pitfalls for Causal Inference." *Political Analysis* 28(2): 186–206.

Pop-Eleches, Grigore, and Lucan A. Way. 2023. "Censorship and the Impact of Repression on Dissent." *American Journal of Political Science* 67(2): 456–471.

Reny, Tyler T., and Benjamin J. Newman. 2021. “The Opinion-Mobilizing Effect of Social Protest Against Police Violence: Evidence from the 2020 George Floyd Protests.” *American Political Science Review* 115(4): 1499–1507.

Singh, Shane P., and Jaroslav Tir. 2023. “Threat-Inducing Violent Events Exacerbate Social Desirability Bias in Survey Responses.” *American Journal of Political Science* 67(1): 154–169.

Tesler, Michael. 2015. “Priming Predispositions and Changing Policy Positions: An Account of When Mass Opinion Is Primed or Changed.” *American Journal of Political Science* 59(4): 806–824.