Analyzing the Impact of Events Through Surveys: Formalizing Biases and Introducing the Dual Randomized Survey Design

Andrew Bertoli[†], Laura Jakli[‡], and Henry Pascoe[††]

March 2024

Abstract: Social scientists and public opinion analysts often use survey data to test how important events impact respondent beliefs, attitudes, and preferences. This paper offers a formal analysis of the pre-event/post-event survey approach, including designs that seek to reduce bias using quota sampling, rolling cross-sections, and panels. Our analysis distinguishes between various sources of bias and clarifies the comparative strengths and weaknesses of each approach. We then propose a modified panel design that can reduce bias in cases where asking respondents to complete the same survey twice could impact their responses in Wave 2. This issue is acute when fielding conventional pre-event/post-event panels due to the short time horizon between Waves 1 and 2. Our analysis elucidates important insights that can improve scholars' ability to study the causal impact of events through surveys.

[†] Assistant Professor, School of Politics, Economics, and Global Affairs, IE University, Spain, abertoli@faculty.ie.edu, *Corresponding Author*

[‡] Assistant Professor, Business, Government, and the International Economy Unit, Harvard Business School, ljakli@hbs.edu

[††] Assistant Professor, School of Politics, Economics, and Global Affairs, IE University, Spain, hpascoe@faculty.ie.edu

1. Introduction

Many social science studies use surveys to estimate the causal effect of major events on beliefs, emotions, attitudes, or preferences. These events include political debates (Brady and Johnston 2006), voter outreach campaigns (Kalla and Broockman 2022), elections (Robbins and Tessler 2012), sports competitions (Rosenzweig and Zhou 2021), announcements of pandemic restrictions (Eggers and Harding 2022), high-level state visits (Goldsmith, Horiuchi, and Matush 2021), and the imposition of economic sanctions (Seitz and Zazzaro 2020). Since 2015, the *APSR*, *AJPS*, and *JOP* alone have published more than 50 such studies.[1] Despite the seemingly straightforward nature of these pre-event/post-event comparisons, various types of bias can weaken causal inference (Muñoz, Falcó-Gimeno, and Hernández 2020). For instance, even if respondents are randomly sampled from the population of interest before and after the event, different non-response dynamics in each period can lead to systematic differences between these groups (Brady and Johnston 2006; Muñoz, Falcó-Gimeno, and Hernández 2020).

We formalize the pre-event/post-event survey approach to help distinguish between different biases that can arise in event-based designs. We develop our analysis using the potential outcomes framework (Rubin 2005; Hainmueller, Hopkins, and Yamamoto 2014; Keele 2015; Broockman, Kalla, and Sekhon 2017; Caughey et al. 2023). To establish a baseline, we begin by considering cases where the pre-event and post-event samples are constructed using simple convenience sampling. Our analysis shows that bias in this baseline model can arise from four main sources: demographic differences between Wave 1 and Wave 2 respondents, temporal factors, anticipatory factors, and differential misreporting. Researchers may be able to take steps to mitigate concerns about bias from temporal factors, anticipatory factors, and differential misreporting. However, bias from demographic differences between Wave 1 and Wave 2 respondents poses a more difficult challenge, particularly when such bias is due to unobservables.

Next, to address bias from demographic differences between Wave 1 and Wave 2 respondents, we carry out formal analyses of three survey approaches: quota sampling, rolling cross-sections, and panel designs. Our analysis sheds light on the strengths and weaknesses of each approach. For instance, we derive the conditions under which quota sampling reduces bias, which are not always

[1] For a list of these articles, see the Online Appendix.

guaranteed to hold. Similarly, we show that while rolling cross-sections should reduce imbalances between Wave 1 and Wave 2 respondents, they are unlikely to eliminate these imbalances completely, and they complicate the overall bias term in other ways. Panel designs keep respondents constant across survey waves but run the risk of conditioning effects whereby the act of completing the survey in Wave 1 affects respondents' answers in Wave 2 (Brady and Johnston 2006). This could result in attenuation bias if respondents tend to repeat their Wave 1 answers in Wave 2. To address this risk, we introduce a modified panel design that combines the main advantages of the panel and rolling cross-sections.

Our paper complements recent advances in survey methods (e.g., Ansolabehere and Schaffner 2014; Frankel and Hillygus 2014; Atkeson, Adams, and Alvarez 2014; Homola, Jackson, and Gill 2016; Broockman, Kalla, and Sekhon 2017; Atkeson and Alvarez 2018; Dafoe, Zhang, and Caughey 2018; Bansak et al. 2018; Miratrix et al. 2018; Alvarez et al. 2019; Caughey et al. 2020; Schaffner 2022; Ben-Michael, Feller, and Hartman 2024). In particular, we build on the work of Muñoz, Falcó-Gimeno, and Hernández (2020) on unexpected events during surveys. Our analysis formalizes many of their results while also providing an analytical framework that can be applied to both expected and unexpected events.

2. Potential Sources of Bias in the Baseline Model

We model the theoretical framework behind the simple convenience sampling design as follows. Each individual who a researcher (or survey firm) attempts to contact in Wave 2 has some truthful response to question k of the survey, which we denote as y_{ikt}. This y_{ikt} value represents individual i's true and honest response to question k in the world where the event happened. Researchers do not observe this value directly, but they instead measure it as individual i's observed response to question k on the survey, which we denote by y_{ikto}. The values y_{ikt} and y_{ikto} might be the same, but they also might differ due to social desirability bias, self-esteem bias, individual i not reading the survey question carefully, or some other reason that causes misreporting.

Let us denote the number of people in the population of interest by N and the number who complete the survey in Wave 2 by n_a (where "a" stands for "after"). Further, for each individual we denote whether they completed the survey in Wave 2 by the variable $r_{ia} \in \{0, 1\}$. Thus, for each of the n_a Wave 2 respondents, we observe y_{ikto}. Additionally, in theory, all individuals have a truthful response to question k in the counterfactual world where the event did not happen. We will discuss what we mean by this "counterfactual world" in more detail shortly, as there may be multiple possibilities that have different implications for analysis. For now, we assume that some well-defined counterfactual exists and denote individual i's hypothetical truthful response in this counterfactual world as y_{ikc}.

Therefore, each individual in Wave 2 has an (unobservable) individual-level treatment effect for question k:

$$\tau_{ik} = y_{ikt} - y_{ikc}$$

The first causal parameter that we might be interested in is the average treatment effect for the population of interest:

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

We cannot directly compute this quantity, since we only observe the y_{ikto} values for Wave 2 respondents. Therefore, a second causal parameter we have more information on is the average causal effect of the event for Wave 2 respondents.

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

Given the lack of information about $y_{ikt|r_a=0}$ values, researchers typically focus on estimating $\bar{\tau}_{k|r_a=1}$. Since researchers observe all $y_{ikto|r_a=1}$ values, they can use these values as a measure of the $y_{ikt|r_a=1}$ values. Without a similar measure for the counterfactual $y_{ikc|r_a=1}$ values, they must estimate the average of $y_{ikc|r_a=1}$. With the exception of panel designs, the standard estimator is the average response to the same question from a survey conducted prior to the event (Wave 1) on a different group of n_b individuals (where "b" denotes "before"). We can denote whether individual i completed the survey in Wave 1 by the variable $r_{ib} \in \{0,1\}$. Further, we can denote individual i's Wave 1 observed response to question k as y_{ikbo}. Thus, we have

$$\hat{\tau}_{k|r_a=1} = \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikto|r_a=1} - \hat{y}_{ikc|r_a=1}$$

where

$$\hat{y}_{ikc|r_a=1} = \frac{1}{n_b}\sum_{i=1}^{n_b} y_{ikbo|r_b=1}$$

In this second estimator, note that i indexes the set of individuals who completed the survey prior to the event (in Wave 1).

In sum, the estimator $\hat{\tau}_{k|r_a=1}$ can be written as

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

and the bias in $\hat{\tau}_{k|r_a=1}$ can be written as

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

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 1 as

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

The first of the two differences in Equation 2 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 2 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 2:

$$Bias\left(\hat{\tau}_{k|r_a=1}\right) = \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikc|r_a=1} - \frac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1} + \bar{\epsilon}_{kt|r_a=1} - \bar{\epsilon}_{kb|r_b=1} + \left(\frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikb|r_a=1} - \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikb|r_a=1}\right) \tag{3}$$

By reordering the terms, we get:

$$Bias\left(\hat{\tau}_{k|r_a=1}\right) = \left(\frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikb|r_a=1} - \frac{1}{n_b}\sum_{i=1}^{n_b} y_{ikb|r_b=1}\right) + \left(\frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikc|r_a=1} - \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikb|r_a=1}\right) + \bar{\epsilon}_{kt|r_a=1} - \bar{\epsilon}_{kb|r_b=1} \tag{4}$$

The first expression in Equation 4 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)**.**

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

We can then rewrite Equation 4 as

$$Bias\left(\hat{\tau}_{k|r_a=1}\right) = Bias_{\mathbf{X}}\left(\hat{\tau}_{k|r_a=1}\right) + \left(\frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikc|r_a=1} - \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikb|r_a=1}\right) + \left(\bar{\epsilon}_{kt|r_a=1} - \bar{\epsilon}_{kb|r_b=1}\right) \tag{5}$$

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 respondents' 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 people's moods, or if Wave 1 happened on a weekday and Wave 2 happened 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 anticipatory 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 5 and add and subtract the average of this potential outcome for Wave 2 respondents.

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

By reordering the terms, we obtain

$$Bias\left(\hat{\tau}_{k|r_a=1}\right) = Bias_{\mathbf{X}}\left(\hat{\tau}_{k|r_a=1}\right) + \left(\frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikc|r_a=1} - \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikbc|r_a=1}\right) + \left(\frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikbc|r_a=1} - \frac{1}{n_a}\sum_{i=1}^{n_a} y_{ikb|r_a=1}\right) + \left(\bar{\epsilon}_{kt|r_a=1} - \bar{\epsilon}_{kb|r_b=1}\right) \tag{6}$$

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 6 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 anticipatory 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 6 as

$$Bias\left(\hat{\tau}_{k|r_a=1}\right) = Bias_{\mathbf{X}}\left(\hat{\tau}_{k|r_a=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|r_a=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|r_a=1}\right) + \left(\bar{\epsilon}_{kt|r_a=1} - \bar{\epsilon}_{kb|r_b=1}\right) \tag{7}$$

The final difference in Equation 7 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.

$$Bias\left(\hat{\tau}_{k|r_a=1}\right) = Bias_{\mathbf{X}}\left(\hat{\tau}_{k|r_a=1}\right) + Bias_{\mathbf{T}}\left(\hat{\tau}_{k|r_a=1}\right) + Bias_{\mathbf{A}}\left(\hat{\tau}_{k|r_a=1}\right) + Bias_{\mathbf{M}}\left(\hat{\tau}_{k|r_a=1}\right) \tag{8}$$

where

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

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

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

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

Given that $\bar{\tau}_{k|r_a=1}$ is the average treatment effect for Wave 2 respondents, we can think of it as comparable to the average treatment effect for the treated (ATT). We might instead be tempted to estimate the average treatment effect for all Wave 1 and Wave 2 respondents (similar to the ATE). This parameter requires us to imagine Wave 1 respondents' counterfactual outcomes had they instead been surveyed in Wave 2. However, the set-up presented in this section is better suited to estimate the average treatment effect for Wave 2 respondents. This is because for Wave 2 respondents we observe y_{ikto}, which is a measure of y_{ikt}. For Wave 1 respondents, we have no comparable measure of either y_{ikt} or y_{ikc}, since we only observe y_{ikbo}. It is still possible to derive a similar but more complex expression for the bias when estimating the average treatment effect for both Wave 1 and Wave 2 respondents. In fact, this bias expression can be viewed as a special case of the bias expression for the rolling cross-section design that we analyze in the next section (see Proposition 4 and the Online Appendix for further discussion).

Of course, researchers might be most interested in estimating the average treatment effect for the broader population ($\bar{\tau}_k$). For this target parameter, we would add one more bias term to Equation 8. This term would represent the difference between the average treatment effect for Wave 2 respondents and the average treatment effect for the population: $\bar{\tau}_{k|r_a=1} - \bar{\tau}_k$. Substantively, this expression would capture the potential bias caused by the average treatment effect for Wave 2 respondents differing from the average treatment effect for the overall population. We will not include this additional term in our baseline model to avoid conflating internal and external validity. However, in the remainder of the paper, readers can draw the link to external validity by adding a simple term to the bias equations that accounts for such a potential heterogeneous treatment effect. For a formalization of this link to external validity, see the Online Appendix.

Returning to internal validity, when estimating $\bar{\tau}_{k|r_a=1}$, Equation 7 breaks the bias term into four additively separable components: the bias caused by demographic differences between Wave 1 and Wave 2 respondents, the bias caused by temporal differences between Waves 1 and 2, the bias caused by anticipatory factors, and the bias caused by differential measurement error in the two waves. With unexpected events such as earthquakes or political scandals, bias from anticipatory factors should be negligible (Muñoz, Falcó-Gimeno, and Hernández 2020). With expected events, whether or not anticipation bias is concerning depends on how researchers define the counterfactual. When researchers define it as the world where the event was rescheduled at the last minute in a manner unlikely to impact respondents' relevant beliefs and attitudes, the anticipation bias term should drop out because the event lead-up would be the same in both the real and counterfactual worlds. However, when the counterfactual is the world where the event was never scheduled, the lead-up in the counterfactual world could be different, making anticipation bias a larger concern.

Ultimately, it is up to researchers to choose which counterfactual makes the most sense in the context of their studies. Even though the "last-minute rescheduling" counterfactual minimizes concerns about anticipation bias, researchers may have strong substantive reasons for choosing a less neutral counterfactual, such as the event never being scheduled. In such cases, researchers might set the Wave 1 period to end a week or more before the event, which risks bias from temporal factors but decreases possible bias arising from anticipatory factors. Regardless of which counterfactual researchers choose, they can clarify their analyses by specifying their exact counterfactual, since this affects the sources of bias they need to consider. Moreover, specifying the counterfactual helps ensure that study framing, theoretical discussion, and conclusions are consistent with research design and empirical analysis.

Researchers can take steps to minimize anticipation bias, but what about temporal bias? Muñoz, Falcó-Gimeno, and Hernández (2020) suggest multiple strategies to address temporal bias. These include checking online and print media for other salient events that might have co-occurred with the survey, considering whether the event was strategically timed, and inspecting the data for temporal trends. In particular, researchers might check if there were weather differences between Waves 1 and 2 that might have influenced respondents' moods when they took the survey. Temporal bias may be further mitigated by design choices, including fielding Waves 1 and 2 at the same time of day and the same day(s) of the week.

When it comes to bias from differential misreporting, researchers can consider whether the event itself might have changed the impact of social desirability bias or some other form of misreporting in Wave 2. For instance, Singh and Tir (2023) find that threat-inducing violent events tend to increase social desirability bias in survey responses across a wide range of contexts. Researchers can also look for other events that occurred around the time of the survey that might have lead to differential misreporting between Waves 1 and 2. Further, they might consider whether individuals might have been in more of a rush to complete the survey in one of the two waves, either because of demographic differences between Wave 1 and Wave 2 respondents or because of temporal differences between the two waves.

Can regression discontinuity help reduce some of the sources of bias discussed in this section by adjusting for temporal trends before and after the event? Unfortunately, regression discontinuity faces several notable limitations in this setting. First, regression discontinuity estimators require potential outcomes to be sufficiently smooth on both sides of the cut-point so that they can be modeled accurately with local linear regression or other smoothing techniques (Calonico, Cattaneo, and Titiunik 2014; Sekhon and Titiunik 2017). However, demographic differences between individuals are likely to lack smoothness over time. For instance, the types of individuals who complete surveys during the final hours of the workday likely differ in systematic ways from those who answer surveys an hour or two after the workday ends. If researchers aggregate responses at the level of days, we should expect systematic differences between individuals who complete surveys on weekdays and weekends. Simultaneous events could also pose a threat to the smoothness requirement, causing jumps or drops in the potential outcomes across time.

Further, standard regression discontinuity designs typically focus on scoring systems in which actors cannot precisely manipulate their scores. In fact, evidence that precise manipulation did occur is considered strong evidence against a regression discontinuity design (Caughey and Sekhon

2011). In contrast, the timing of when individuals complete surveys is typically precisely chosen, either by the individuals themselves or by those administering the survey. Even in contexts where sorting around the cut-point might be very unlikely, precise manipulation can still pose a problem for estimation by threatening smoothness in the potential outcomes on each side of the cut-point. Muñoz, Falcó-Gimeno, and Hernández (2020) highlight other concerns with using regression discontinuity in this setting. In particular, they point out that the effects of events can take time to unfold, in which case there would not be a well-defined cut-point as in standard RDs.

Rather than turning to RD estimation, the strategies discussed in this section can mitigate concerns regarding bias in pre-event/post-event survey designs. In particular, temporal, anticipation, and differential misreporting biases can often be managed with appropriate field strategies, design choices, and robustness checks. Demographic bias, on the other hand, presents a distinct set of challenges. In the next section, we analyze three approaches to reduce demographic bias. Our results clarify the extent to which these three methods can reliably reduce bias for pre-event/post-event survey designs.

3. Comparing Bias in the Baseline Model to Bias in Other Designs

To what extent do quota sampling, rolling cross-sections, and panel designs reduce bias compared to our simple baseline model? We begin with the quota sampling method. This approach is similar to the baseline model except that we keep or drop potential respondents based on whether they help us meet our demographic quotas in Wave 1 and Wave 2. Let g_b denote the number of people in our Wave 1 quota group and g_a denote the number of people in our Wave 2 quota group. Likewise, we exclude some people in Waves 1 and 2 based on quota constraints. Let e_b and e_a denote the number of people in Waves 1 and 2, respectively, who would have completed our survey but who we exclude due to quota constraints. In total, we then have $n_b = g_b + e_b$ individuals in Wave 1 and $n_a = g_a + e_a$ individuals in Wave 2 who would complete our survey if asked. We use $q_i \in \{0,1\}$ to denote whether individual i is in our quota sample or in the excluded group.

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

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

The statistic we use to estimate this parameter is the average difference between the reported answers of the g_2 respondents who completed our survey in Wave 2 and the reported answers of the g_1 respondents who completed the survey in Wave 1.

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

We provide the full analysis of this design, including proofs of the propositions that follow, in the Online Appendix. Proposition 1 establishes that the bias of the quota sampling approach can be written as the sum of biases from the demographic differences between Wave 1 and Wave 2 respondents, temporal factors, anticipatory factors, and differential misreporting.

Proposition 1. *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

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

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

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

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

This overall bias term is similar to the $Bias(\hat{\tau}_{k|r_a=1})$ term we derived in Section 2, except that it restricts the focus to our quota sample.

Does quota sampling reduce bias compared to convenience sampling? Proposition 2 shows the conditions under which quotas reduce or amplify bias.

Proposition 2. *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.*

The first two expressions to the right of the inequality are the weighted average of the biases in the quota sample and excluded group. The intuition is that we should expect quota sampling to increase bias when the bias in the excluded group is smaller than the bias in the quota group or when the bias terms for the two groups point in opposite directions and thus partially cancel each other out.

The final expression is a residual term that arises when $Bias\left(\hat{\tau}_{k|r_a=1}\right)$ is separated into a weighted average of $Bias\left(\hat{\tau}_{k|r_a=1,q=1}\right)$ and $Bias\left(\hat{\tau}_{k|r_a=1,q=0}\right)$. This residual term can grow large when there is a notable difference between the proportion of Wave 1 and Wave 2 individuals who are in the quota groups (rather than the excluded groups). In particular, this can happen when it is easier to find individuals to meet the quotas in Wave 1 than in Wave 2. In such cases, quota sampling may increase bias if the weighted average and residual term point in opposite directions.

A brief example illustrates how quota sampling can exacerbate bias. Imagine that a researcher hires a survey firm to administer a survey in a province before and after an event. The province has one large town in its center and many smaller communities on the outskirts. The researcher sets quotas related to age, gender, education, political ideology, and income. However, the researcher does not set quotas on whether respondents live in town or in the surrounding communities. In Wave 1, the survey firm reaches out to people in town and meets its quotas staying within city limits. However, in Wave 2, the firm cannot find enough new people in town to meet its quotas and proceeds to recruit many respondents from the outskirt communities, inducing imbalance between Wave 1 and Wave 2 samples on location. Had there been no demographic quotas, the company may have been able to find a sufficient number of Wave 2 respondents without leaving town, maintaining balance on location. Whether one lives in town or in the outskirts of the province could matter substantially for baseline beliefs and attitudes related to the survey, even after conditioning on covariates that were balanced through quotas. In such cases, quota sampling could exacerbate bias.

Even if the researcher had anticipated this issue and specified location-based quotas, quota sampling might still increase bias. For instance, canvassers might still have trouble meeting their quotas in the town (or in the outskirts) in Wave 2, leading them to employ different strategies to reach respondents. Such changes in strategy could lead to certain types of imbalances between the Wave 1 and Wave 2 respondents that would not have arisen if the canvassers had not been restricted by quota constraints.

We offer additional examples of how quota sampling can amplify bias in the Online Appendix that involve phone and internet surveys. Given the potential of quota sampling to exacerbate bias, it is a sub-optimal approach to estimating the causal effects of events through surveys. This conclusion accords with the conventional view in statistics that quota sampling, despite its convenience, is unreliable for obtaining representative samples (e.g., see Bowling and Ebrahim 2005: 199; Freedman, Pisani, and Purves 2007: 334–38; Yang and Banamah 2014).

Next, we examine the rolling cross-section design. There are various ways of carrying out this design, including random-digit dialing or selecting a specific set of individuals and randomizing whether they are asked to complete the survey in Wave 1 or 2. Alternatively, researchers often take existing survey data from a lengthy rolling cross-section project lasting months and compare the responses just before and after a certain event (see Brady and Johnston 2006: 166). For such cases, we will restrict our attention to the individuals surveyed in these two periods, labeling the period just before the event as Wave 1 and the period just after the event as Wave 2. Of course, these two periods may not correspond to Waves 1 and 2 of the broader project.

We conceptualize this design as beginning with an initial sample that includes a group of "always-responders" who would complete the survey in either wave, a group of "sometimes-responders" who would complete the survey only in Wave 1 or Wave 2 but not both, and a group of "never-responders" who would not complete the survey in either wave. We focus on the always-responders and sometimes-responders since never-responders are inaccessible to us.

Let n_w denote the total number of always-responders and sometimes-responders initially asked to complete the survey in Waves 1 or 2. We again denote the number who complete the survey in each wave by n_a (Wave 2) and n_b (Wave 1). We can also denote the total number of always-responders in the sample by n^*. Further, let m_a be the number of sometimes-responders who would only complete the survey in Wave 2 and m_b be the number of sometimes-responders who would only complete the survey in Wave 1. Note that $n^* + m_a + m_b = n_w$.

We let $u_i \in \{0,1\}$ denote whether individual i is an always-responder and $s_i \in \{0,1\}$ denote whether individual i is a sometimes-responder. We will also let $s_{i1} \in \{0,1\}$ denote whether individual i would only complete the survey if assigned to Wave 1 and $s_{i2} \in \{0,1\}$ denote whether individual i would only complete the survey if assigned to 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. Similarly, let $w_{i1} \in \{0,1\}$ 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.

Lastly, we can denote the number of individuals who would complete the survey if assigned to Wave 1 as $w_b = n^* + m_b = \sum_i^N w_{i1}$ and the number who would complete the survey if assigned to Wave 2 as $w_a = n^* + m_a = \sum_i^N w_{i2}$. Note that n^*, m_a, m_b, w_a, w_b, and n_w are all parameters that do not depend on the randomization. Meanwhile, n_a and n_b are random variables, since they can change based on the numbers of always-responders and sometimes-responders randomized to Waves 1 or 2.

There are two causal parameters we might want to estimate in a rolling cross-section design. The first is the average treatment effect on the always-responders:

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

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

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

Because we cannot distinguish between the always-responders and sometimes-responders in the sample, the statistic we would use to estimate these two parameters is the same. It 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.

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

Proposition 3 gives the bias in estimating the first parameter:

Proposition 3. *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

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

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

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

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

The first term in $Bias_{\mathbf{S}}\left(\hat{\tau}_{k|u=1}\right)$ is the proportion of the possible Wave 2 respondents who are sometimes-responders ($s_{i2} = 1$) multiplied by the average difference between these sometimes-responders' y_{ikt} values and the always-responders' y_{ikt} values. The second expression is the proportion of the possible Wave 1 respondents who are sometimes-responders ($s_{i1} = 1$) multiplied by the average difference between the always-responders' y_{ikb} values and these 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 when our parameter of interest is the average treatment effect for always-responders. The $Bias_{\mathbf{S}}\left(\hat{\tau}_{k|u=1}\right)$ term disappears if the number of sometimes-responders is zero or if the average differences in y_{ikt} and y_{ikb} values between the always-responders and sometimes-responders is zero.

Proposition 4 gives the bias if we are interested in the second parameter $\bar{\tau}_{k|w=1}$.

Proposition 4. *When estimating $\bar{\tau}_{k|w=1}$, the bias in $\hat{\tau}_{k|w=1}$ is*

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

where,

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

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

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

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

The $Bias_{\mathbf{S}}\left(\hat{\tau}_{k|w=1}\right)$ term now arises from the fact that we do not see any y_{ikto} values for Wave 1 sometimes-responders, nor do we see any y_{ikbo} values for Wave 2 sometimes-responders. The size of this bias term is determined by the prevalence of each type of sometimes-responder and how much they differ from each other and the always-responders in terms of potential outcomes.

Which of the two parameters are we better positioned to estimate, $\bar{\tau}_{k|u=1}$ or $\bar{\tau}_{k|w=1}$? The answer depends on whether sometimes-responders in Wave 1 and sometimes-responders in Wave 2 are more similar to each other than they are to always-responders, as well as whether sometimes-responders are less influenced by temporal and anticipatory factors than always-responders. If these conditions hold, $\hat{\tau}_{k|w=1}$ should be less biased than $\hat{\tau}_{k|u=1}$. However, in cases where researchers expect Wave 1 and Wave 2 sometimes-responders to differ from each other drastically, or if Wave 1 sometimes-responders seem much more likely than always-responders to be influenced by temporal and anticipatory factors, then we should expect $\hat{\tau}_{k|u=1}$ to be less biased than $\hat{\tau}_{k|w=1}$.

How do the bias terms for these two estimators compare to bias in the baseline model, formalized in Equation 8? The rolling cross-section design trades the bias in demographic differences between Wave 1 and Wave 2 respondents for bias caused by sometimes-responders. Focusing first on $Bias_{\mathbf{S}}\left(\hat{\tau}_{k|u=1}\right)$, we can better understand this term 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, $\frac{m_a}{w_a} = \frac{m_b}{w_b}$. We will denote these proportions as $\alpha \leq 1$. As we show in the Online Appendix, this symmetry allows us to write the overall bias in $\hat{\tau}_{k|u=1}$ as

$$\begin{aligned}(9)\qquad Bias\left(\hat{\tau}_{k|u=1}|m_a = m_b\right) = {} & \alpha Bias_{\mathbf{X}}\left(\hat{\tau}_{k|s=1}\right) + \left(\alpha Bias_{\mathbf{T}}\left(\hat{\tau}_{k|s=1}\right) + (1-\alpha)\, Bias_{\mathbf{T}}\left(\hat{\tau}_{k|u=1}\right)\right) + \\ & \left(\alpha Bias_{\mathbf{A}}\left(\hat{\tau}_{k|s=1}\right) + (1-\alpha)\, Bias_{\mathbf{A}}\left(\hat{\tau}_{k|u=1}\right)\right) + Bias_{\mathbf{M}}\left(\hat{\tau}_{k|u=1}\right) + \\ & \alpha\left(\bar{\tau}_{k|s=1} - \bar{\tau}_{k|u=1}\right)\end{aligned}$$

The expression involving demographic differences between Wave 1 and Wave 2 respondents is now limited to sometimes-responders. The bias from temporal factors is just a weighted average of the temporal bias for sometimes-responders and the temporal bias for always-responders. The same logic holds for bias in anticipatory factors. Meanwhile, the term for bias caused by differential misreporting does not change. We also add a new bias term involving the difference in average treatment effect between sometimes-responders and always-responders.

Turning to the overall bias term in Proposition 4, $Bias_{\mathbf{S}}\left(\hat{\tau}_{k|w=1}\right)$, again consider the special case where the initial numbers of Wave 1 and Wave 2 sometimes-responders are equivalent ($m_a = m_b$). In this case, the proportions $\frac{m_a}{n_w}$ and $\frac{m_b}{n_w}$ will be equal. We can denote these proportions as $\lambda \leq 1$. We can then write the overall bias in the estimator $\hat{\tau}_{k|w=1}$ as

$$
\begin{aligned}
(10) \qquad Bias\left(\hat{\tau}_{k|w=1}\right) = {} & \lambda Bias_{\mathbf{X}}\left(\hat{\tau}_{k|w_2=1}\right) + \Big(\lambda Bias_{\mathbf{T}}\left(\hat{\tau}_{k|w_2=1}\right) + (1-\lambda)\, Bias_{\mathbf{T}}\left(\hat{\tau}_{k|w=1}\right)\Big) + \\
& \Big(\lambda Bias_{\mathbf{A}}\left(\hat{\tau}_{k|w_2=1}\right) + (1-\lambda)\, Bias_{\mathbf{A}}\left(\hat{\tau}_{k|w=1}\right)\Big) + Bias_{\mathbf{M}}\left(\hat{\tau}_{k|w=1}\right) + \\
& \lambda\left(\bar{\tau}_{k|w_2=1} - \bar{\tau}_{k|w=1}\right)
\end{aligned}
$$

which is similar to the bias in $\hat{\tau}_{k|u=1}$ formalized in Equation 9.

In sum, the rolling cross-section estimator may reduce bias caused by demographic differences between Wave 1 and Wave 2 respondents, but it is unlikely to entirely eliminate this bias because of the sometimes-responders. This result accords with the empirical finding by Muñoz, Falcó-Gimeno, and Hernández (2020: 193) of detectable relationships between demographic characteristics and the timing of interviews in two widely-used rolling cross-section datasets. Moreover, our analysis shows that using the rolling cross-section design complicates the rest of the overall bias term in ways that could either decrease or amplify total bias compared to our baseline model.

We next examine the panel design. Panels start with a group of individuals who have the opportunity to take the same survey before and after the event. Researchers might conduct their own two-wave panel around the event. However, they sometimes instead examine existing survey data from a lengthy panel study with many waves, focusing on the waves before and after the event. For this discussion, we label these Waves 1 and 2. We will call the individuals who complete the survey in both waves "always-responders." We denote whether individual i is an always-responder by $u_i \in \{0,1\}$ and the total number of always-responders by n^*. The causal parameter we estimate is the average treatment effect of the event on the n^* always-responders' truthful answers to question k of the survey:

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

In the above line, we consider $y_{ikt|u=1}$ to be individual i's truthful answer in the world where individual i did not complete the survey in Wave 1.

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

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

In this 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.

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

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

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

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

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

The $Bias_{\mathbf{C}}\left(\hat{\tau}_{k|u=1}\right)$ term can be thought of as the average difference between the n^* respondents' 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 respondents' truthful answers in Wave 2, which can be thought of as a type of conditioning effect (Brady and Johnston 2006: 164). Panel conditioning might also impact measurement error in Wave 2 ($\bar{\epsilon}_{ka|u=1}$), as we will discuss shortly.

Therefore, the panel design resembles the baseline model except that it exchanges bias in demographic differences between Wave 1 and Wave 2 respondents for potential conditioning effects. If the two waves are spaced weeks or months apart, conditioning effects may be minimal, but temporal bias may be large. However, if the two waves are spaced closer together, temporal bias may be minimal, but the risk of conditioning effects increases. In part, this is because the desire for internal consistency in a panel response is likely stronger on a condensed survey timeline than on quarterly, annual, or biannual panels. As such, panel respondents might be inclined to repeat their Wave 1 answers in Wave 2, which could result in attenuation bias.

Additionally, panel participation might lead respondents to update issue positions or attitudes when the topic is initially low salience and thus less crystallized (Bach 2021). In this case, the panel design may initiate a cognitive process that crystallizes or updates the attitude or issue position of interest by Wave 2. In turn, some portion of the observed treatment effect is a function of the survey stimulus rather than the event stimulus (Amaya, Hatley, and Lau 2021). Panel studies measuring engagement in illicit behavior confirm that the aforementioned conditioning effects are exacerbated by short duration between survey waves (Halpern-Manners, Warren, and Torche 2014).

Another concern arises from speeding and inattentiveness. On a short time horizon, respondents may be particularly likely to recall survey structure and resort to shortcuts (Bach 2021). Illustratively, Schonlau and Toepoel (2015) find that straightlining increases with repeated exposure to survey elements through panels. In this case, the panel structure could lead to higher levels of misreporting in Wave 2 than in Wave 1, which could amplify the overall bias through the $Bias_{\mathbf{M}}\left(\hat{\tau}_{k|u=1}\right)$ term.

The panel design may also be impractical in instances where researchers wish to include an experimental component in their survey. For ethics and transparency, researchers typically debrief survey participants about the experimental manipulation and full study objectives at the end of each survey wave. However, the debrief that accompanies experimental treatments could bias measures of treatment effect across panels, especially when there is a short amount of time between two waves.

In the next section, we propose a modified survey design that avoids asking respondents to complete the same survey twice. Our approach is particularly advantageous in situations where researchers suspect the $Bias_{\mathbf{C}}\left(\hat{\tau}_{k|u=1}\right)$ term to be large, which may often be likely in pre-event/post-event survey designs.

4. The Dual Randomized Survey (DRS) Design

The goal of the DRS design is to identify respondents who would be willing to complete the survey in both waves and in turn randomize whether they take the survey in the pre-event or post-event period. DRS ensures that the pre-event group is similar to the post-event group on baseline characteristics while eliminating potential bias caused by asking people to take the same survey twice. Additionally, this design allows researchers to run two different surveys on separate topics within a single research design, as we explain below.

The DRS design implementation strategy is as follows. The first step is to identify a group of people from the overall population who seem likely to complete a survey in both the pre-event and post-event periods. We call these people "likely-responders." An attempt should be made to contact each of these likely-responders in the pre-event period (Wave 1). Those successfully contacted should be asked to fill out one of two surveys—either Survey A or Survey B. We will consider Survey A to be the main survey of interest. Survey B could be a different survey that forms the basis of a separate study. Which of the two surveys respondents are asked to complete in Wave 1 should be randomized, either beforehand or at time of contact. Those who complete the survey they are randomized to receive in Wave 1 comprise the Wave 1 respondents. Likely-responders who do not complete a survey in Wave 1 are dropped from the sample. In the post-event period (Wave 2), each Wave 1 respondent is recontacted, this time to complete whichever survey they did not take in Wave 1. Those who complete both surveys will constitute the main sample. Wave 1 respondents who do not complete their allocated Wave 2 survey are dropped from the analysis.

We formalize this design as follows. We denote the number of individuals who complete a survey in both Wave 1 and Wave 2 by n. These n individuals constitute our sample. We focus on estimating the causal effect of the event on question k of Survey A. Our conclusions apply to Survey B without loss of generality.

We conceptualize each individual as having a truthful Wave 1 response (y_{ikb}), a truthful Wave 2 response after having completed Survey B in Wave 1 (y_{ika}), a truthful Wave 2 response in the counterfactual world where individual i did not complete Survey B in Wave 1 (y_{ikt}), and a truthful Wave 2 response in the world where the event did not occur (y_{ikc}). The difference between y_{ika} and y_{ikt} helps formalize the causal impact of completing Survey B in Wave 1 on individual i's response to question k of Survey A in Wave 2. Each individual also has a reported Wave 1 response (y_{ikbo}) and a reported Wave 2 response (y_{ikao}), only one of which we observe. Consistent with our previous analyses, we 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.

We cannot observe any individual's y_{ikt} or y_{ikc}, as we only observe y_{ikao} or y_{ikbo}. Therefore, we use the average response of individuals in our sample who completed Survey A in Wave 1 to estimate $\bar{y}_{ikc}$ and the average response of individuals in our sample who completed Survey A in Wave 2 to estimate $\bar{y}_{ikt}$. We denote the number of individuals in our sample who completed Survey A in Wave 1 by n_b and the number who completed Survey A in Wave 2 by n_a. Note that $n_a + n_b = n$, as we drop anyone from our sample who did not complete both surveys.

An initial complication is that whether respondents are assigned to complete Survey A or Survey B in Wave 1 might impact whether they participate in Wave 2 (Frankel and Hillygus 2014). This issue might be particularly concerning if one survey is much longer than the other or if either survey asks certain sensitive questions that might discourage participation in another survey a short time later. Therefore, we denote whether individual i would only complete both surveys if randomized to take Survey A in Wave 1 by $s_{i1} \in \{0,1\}$, and we denote the number of such individuals by m_b. Similarly, we denote whether individual i would only complete both surveys if randomized to do Survey A in Wave 2 by $s_{i2} \in \{0,1\}$ and the number of such individuals by m_a. Meanwhile, let $u_i \in \{0,1\}$ denote whether individual i would complete both surveys regardless of which survey they were assigned to do first, and denote the number of such individuals by n^*.

We can also denote whether individual i would complete both surveys if randomized to take Survey A in Wave 1 by $w_{i1} \in \{0,1\}$ and whether individual i would complete both surveys if randomized to take Survey A in Wave 2 by $w_{i2} \in \{0,1\}$. In total, we have $w_a = m_a + n^*$ possible respondents who might be in our sample as someone who completed Survey A in Wave 2 and $w_b = m_b + n^*$ possible respondents who might be in our sample as someone who completed Survey A in Wave 1.

The ideal scenario is that which survey respondents are asked to complete in Wave 1 has no impact on their decision to complete both surveys. In this case, $m_a = 0$, $m_b = 0$, and $n^* = n$. However, to derive conclusions that are as general as possible, we allow for both m_a and m_b to be greater than 0.

The causal parameter we estimate is the average treatment effect of the event on the truthful responses to question k of the n^* individuals who would complete both surveys regardless of which one they were assigned to do first:

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

The statistic we use to estimate this parameter is as follows:

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

This expression is simply the average difference in the outcomes of the respondents in our sample who completed Survey A in Wave 2 compared to those who completed Survey A in Wave 1.

Proposition 6 presents the bias decomposition for the DRS design.

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

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

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

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

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

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

The $Bias_{\mathbf{D}}(\hat{\tau}_{k|u=1})$ term captures the bias induced by differential dropout based on which survey individuals completed in Wave 1. Ideally, this bias is 0 because researchers carefully construct both surveys to be similar in factors such as length and sensitivity to avoid differential attrition. The $Bias_{\mathbf{P}}(\hat{\tau}_{k|u=1})$ term captures the average causal effect of completing Survey B in Wave 1 on respondents' answers to question k of Survey A in Wave 2, which can be thought of as a possible undesired priming effect. Here, too, this bias is ideally 0 because the researcher thoughtfully constructed Survey B so that completing it in Wave 1 would be very unlikely to impact respondents' answers to questions on Survey A in Wave 2. Even if completing Survey B might impact some individuals' answers on Survey A, we just need the average effect to be 0 (or close to it) for the $Bias_{\mathbf{P}}(\hat{\tau}_{k|u=1})$ term to be negligible.

Thus, the DRS design trades bias from demographic differences between Wave 1 and Wave 2 respondents for potential bias from differential attrition and the possibility of Survey B affecting individuals' responses on Survey A. There is also still the potential for bias from temporal factors, anticipatory factors, and differential misreporting. As we discussed in Section 2, researchers can take steps to address concerns regarding these potential sources of bias.

Researchers using the DRS design can test for differential attrition by checking the proportion of respondents who completed Survey A in Wave 1 (compared to Wave 2). If no differential attrition occurred, then this proportion should be similar to a draw from a binomial distribution with size n and probability 0.5. However, if the observed proportion differs significantly from a 50/50 split, then it would suggest that differential attrition took place. Likewise, researchers can use balance tests to verify that the group that completed Survey A in Wave 1 is similar on average to the group that completed Survey A in Wave 2. Larger-than-expected imbalances would indicate differential attrition. This test could also identify notable chance imbalances that might have arisen and could cast doubt on the results.

The DRS design also allows researchers to embed survey experiments to examine if the treatment effect measured by the experiment changed after the event. This option might not be feasible in a traditional panel design due to required debriefing after the Wave 1 survey. This test would also not be straightforward with convenience sampling, quota, or rolling cross-section designs due to demographic imbalances.

There are two weaknesses of the DRS design compared to the standard panel. First, in situations where the bias from completing the same survey twice is minimal, the panel design provides more statistical power because Wave 1 and Wave 2 responses come from the same respondents. In comparison, the DRS design should feature small demographic differences between respondents who complete Survey A in Waves 1 and 2 (simply due to chance variation), which can decrease statistical power. Second, the standard panel design allows researchers to compute changes for specific individuals, which may be of substantive interest (Brady and Johnston 2006). These individual-level differences cannot be interpreted as "individual-level treatment effects" since they can be influenced by temporal factors, anticipatory factors, and differential misreporting. However, they may still be of interest to researchers in certain contexts. Of course, along with these two limitations, researchers can only use DRS for expected events due to the planning required.

5. Conclusion

Using the potential outcomes framework, this study disaggregated four sources of bias in the pre-event/post-event survey design: bias from demographic differences between the Wave 1 and Wave 2 respondents, bias from temporal factors, bias from anticipatory factors, and bias from differential misreporting. It also clarified the strengths and weaknesses of approaches that try to reduce bias through quota sampling, rolling cross-sections, and panel designs. Using the standard panel design as a baseline framework, we propose a new approach that can increase internal validity in contexts where asking respondents to complete the same survey twice may affect their Wave 2 answers in a variety of unobserved and underspecified ways. Our new method provides a novel contribution to the social scientist's methodological toolkit.

The analytical framework that we outline in this paper can also be extended to study expected or unexpected events beyond the social sciences. For instance, it could be applied to test how Super Bowl ads affect individuals' interest in buying certain products or how the announcement of a new strain of a virus impacts individuals' mental health evaluations. Further, future research could extend our framework to settings where scholars measure public opinion before and after important events using social media data. Researchers frequently want to understand how specific events influence public beliefs, attitudes, and preferences, making the framework introduced in this paper valuable in a wide range of contexts.

Funding: This paper was partially funded by MCIN /AEI /10.13039/501100011033 / FEDER, UE Grant No PID2022-139998NB-I00.

Acknowledgements: The authors thank Vincent Pons, Chad Hazlett, Daniel Kselman, Michael Becher, Nikitas Konstantinidis, Onyebuchi Arah, and Ryan Baxter-King for helpful feedback. In addition, they thank all participants at UCLA's Practical Causal Inference Lab.

Competing interests: The authors declare none.

References

Alvarez, R. M., L. R. Atkeson, I. Levin, and Y. Li. 2019. "Paying Attention to Inattentive Survey Respondents." *Political Analysis* 27(2): 145–162.

Amaya, A., N. Hatley, and A. Lau. 2021. "Measuring the Risks of Panel Conditioning in Survey Research." *Pew Research Center*.

Ansolabehere, S., and B. F. Schaffner. 2014. "Does Survey Mode Still Matter? Findings From a 2010 Multi-Mode Comparison." *Political Analysis* 22(3): 285–303.

Atkeson, L. R., A. N. Adams, and R. M. Alvarez. 2014. "Nonresponse and Mode Effects in Self- and Interviewer-Administered Surveys." *Political Analysis* 22(3): 304–320.

Atkeson, L. R., and R. M. Alvarez. 2018. *The Oxford Handbook of Polling and Survey Methods*. New York: Oxford University Press.

Bach, R. L. 2021. "A Methodological Framework for the Analysis of Panel Conditioning Effects." *Measurement Error in Longitudinal Data*. Ed. by A. Cernat, and J. W. Sakshaug. Oxford, UK: Oxford University Press: 19–41.

Bansak, K., J. Hainmueller, D. J. Hopkins, and T. Yamamoto. 2018. "The Number of Choice Tasks and Survey Satisficing in Conjoint Experiments." *Political Analysis* 26(1): 112–119.

Ben-Michael, E., A. Feller, and E. Hartman. 2024. "Multilevel Calibration Weighting for Survey Data." *Political Analysis* 32(1): 65–83.

Bowling, A., and S. Ebrahim. 2005. "Quantitative Social Science: The Survey." *Handbook of Health Research Methods: Investigation, Measurement and Analysis*: 190–214.

Brady, H. E., and R. Johnston. 2006. "The Rolling Cross-Section and Causal Attribution." *Capturing Campaign Effects*. Ed. by H. E. Brady, and R. Johnston. Ann Arbor, MI: University of Michigan Press: 164–195.

Broockman, D. E., J. L. Kalla, and J. S. Sekhon. 2017. "The Design of Field Experiments with Survey Outcomes: A Framework for Selecting More Efficient, Robust, and Ethical Designs." *Political Analysis* 25(4): 435–464.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82(6): 2295–2326.

Caughey, D., A. J. Berinsky, S. Chatfield, E. Hartman, E. Schickler, and J. S. Sekhon. 2020. *Target Estimation and Adjustment Weighting for Survey Nonresponse and Sampling Bias.* New York: Cambridge University Press.

Caughey, D., A. Dafoe, X. Li, and L. Miratrix. 2023. "Randomisation Inference Beyond the Sharp Null: Bounded Null Hypotheses and Quantiles of Individual Treatment Effects." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 85(5): 1471–1491.

Caughey, D., and J. S. Sekhon. 2011. "Elections and the Regression Discontinuity Design: Lessons from Close US House Races, 1942–2008." *Political Analysis* 19(4): 385–408.

Dafoe, A., B. Zhang, and D. Caughey. 2018. "Information Equivalence in Survey Experiments." *Political Analysis* 26(4): 399–416.

Eggers, A. C., and R. Harding. 2022. "Rallying in Fear? Estimating the Effect of the UK COVID-19 Lockdown with a Natural Experiment." *European Journal of Political Research* 61(2): 586–600.

Frankel, L. L., and D. S. Hillygus. 2014. "Looking Beyond Demographics: Panel Attrition in the ANES and GSS." *Political Analysis* 22(3): 336–353.

Freedman, D., R. Pisani, and R. Purves. 2007. *Statistics, 4th ed.* W.W. Norton.

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

Hainmueller, J., D. J. Hopkins, and T. Yamamoto. 2014. "Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments." *Political Analysis* 22(1): 1–30.

Halpern-Manners, A., J. R. Warren, and F. Torche. 2014. "Panel Conditioning in a Longitudinal Study of Illicit Behaviors." *Public Opinion Quarterly* 78(3): 565–590.

Homola, J., N. Jackson, and J. Gill. 2016. "A Measure of Survey Mode Differences." *Electoral Studies* 44: 255–274.

Kalla, J. L., and D. 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.

Keele, L. 2015. "The Statistics of Causal Inference: A View from Political Methodology." *Political Analysis* 23(3): 313–335.

Miratrix, L. W., J. S. Sekhon, A. G. Theodoridis, and L. F. Campos. 2018. "Worth Weighting? How to Think About and Use Weights in Survey Experiments." *Political Analysis* 26(3): 275–291.

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

Robbins, M. D. H., and M. Tessler. 2012. "The Effect of Elections on Public Opinion Toward Democracy: Evidence From Longitudinal Survey Research in Algeria." *Comparative Political Studies* 45(10): 1255–1276.

Rosenzweig, L. R., and Y.-Y. Zhou. 2021. "Team and Nation: Sports, Nationalism, and Attitudes Toward Refugees." *Comparative Political Studies* 54(12): 2123–2154.

Rubin, D. B. 2005. "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions." *Journal of the American Statistical Association* 100(469): 322–331.

Schaffner, B. F. 2022. "Optimizing the Measurement of Sexism in Political Surveys." *Political Analysis* 30(3): 364–380.

Schonlau, M., and V. Toepoel. 2015. "Straightlining in Web Survey Panels Over Time." *Survey Research Methods* 9(2): 125–137.

Seitz, W., and A. Zazzaro. 2020. "Sanctions and Public Opinion: The Case of the Russia-Ukraine Gas Disputes." *The Review of International Organizations* 15: 817–843.

Sekhon, J. S., and R. Titiunik. 2017. "Understanding Regression Discontinuity Designs as Observational Studies." *Observational Studies* 3(2): 174–182.

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

Yang, K., and A. Banamah. 2014. "Quota Sampling as an Alternative to Probability Sampling? An Experimental Study." *Sociological Research Online* 19(1): 56–66.