Addressing Survey Nonresponse and Attrition in Probability-based Online Panels and Online Longitudinal Data Collections

by

Saw Htay Wah

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Survey and Data Science)
in the University of Michigan
2025

Doctoral Committee:

Professor Brady T. West, Chair
Assistant Professor Christopher Antoun, University of Maryland
Professor Frederick Conrad
Professor Megan E Patrick
Professor James Wagner

Saw Htay Wah

htaywah@umich.edu

ORCID iD: 0000-0001-7735-9837

© Saw Htay Wah 2025

Acknowledgements

First and foremost, I wish to express my immense gratitude to my entire dissertation committee, chaired by Brady T. West and comprising Christopher Antoun, Frederick Conrad, Megan E. Patrick, and James Wagner, for their unwavering support and invaluable advice throughout my dissertation journey. Brady T. West has been instrumental at every stage of my doctoral research—from conceptualization and securing grant funding to implementation, production, and completion.

I would also like to thank my collaborators who made profound contributions to various aspects of my doctoral research. I extend my gratitude to John Gonzalez, Director of the Institutional Research Office at the Rackham Graduate School, and Mike Solomon, Dean of the Rackham Graduate School, for their collaboration on the incentive experimental study. I am deeply thankful to Arie Kapteyn for his steadfast intellectual leadership, as well as to the entire Understanding America Study (UAS) Research team for their dedication and professionalism in implementing the survey burden experiment.

My heartfelt gratitude also goes to the American Family Health Study (AFHS) team, led by Brady T. West, of which I am a member. Although this dissertation does not utilize AFHS data, it greatly benefited from the invaluable experiences I gained while working on the project.

The original idea for the survey burden experiment emerged from the proposal I developed for my doctoral seminar class. I am grateful to the instructors, Trivellore Raghunathan

and Katharine Abraham, as well as the seminar participants, for their valuable insights and feedback on the proposal.

I sincerely thank my coauthors, Brady T. West, Arie Kapteyn, and Marco Angrisani, for their significant scholarly contributions to my dissertation research. Brady T. West coauthored the incentive study paper, Arie Kapteyn coauthored the papers on nonresponse patterns and survey burden analyses, and Marco Angrisani coauthored the paper on survey burden analysis.

The COVID-19 pandemic in 2020 and the significant political turmoil back home in February 2021 presented considerable challenges and disruptions. I am deeply grateful to the program's administrative team, particularly the former director, Frederick Conrad, and Assistant Director, Jill Esau, for their steadfast support during this difficult period. I also wish to extend my heartfelt thanks to my personal lawyer, Melissa Indish, whose extensive legal support spanning several months was instrumental in helping me overcome these challenges.

Finally, I gratefully acknowledge funding support from the National Institute on Aging and the Social Security Administration through a grant to Arie Kapteyn (Grant Number: U01AG077280), the National Science Foundation (Award Number: 2344816), and the Graduate School at the University of Michigan.

Table of Contents

Acknowledgements ii

List of Tables viii

List of Figures xi

List of Appendices xii

Abstract xiii

Chapter 1 Introduction 1

1.1 Background and motivation 1

1.2 An overview of the dissertation 7

1.2.1 Chapter 2 overview 7

1.2.2 Chapter 3 overview 9

1.2.3 Chapter 4 overview 11

Chapter 2 Personality Traits, Panel Tenure, Survey Topic, and Context as Predictors of Survey Nonresponse Patterns in High-Frequency Online Longitudinal Surveys 14

2.1 Introduction 14

2.2 Background 17

2.2.1 Differences between Offline and Online Panel Studies 17

2.2.2 Different Correlates of Attrition and Response Mechanisms in Offline vs. Online Panel Studies 21

2.2.3 Respondent Personality Traits as Consistent and Strong Predictors of Nonresponse Patterns in High-frequency Online Longitudinal Data Collections 23

2.2.4 The Relevance of Panel Tenure in Predicting Nonresponse Patterns in High-frequency Online Longitudinal Studies 29

2.2.5 Analyzing the Moderating Effects of Survey Topics and Context in High-Frequency Online Longitudinal Studies 31

2.3 Methods 33
2.3.1 Data and Study Population 33
2.3.2 Understanding Coronavirus in America Study (UCA) 34
2.3.3 The UAS Monthly Events Panel Surveys 35
2.3.4 Analytical Approaches 36
2.4 Results 42
2.4.1 Sample Summary Statistics 42
2.4.2 Latent Class Model Selection 44
2.4.3 Class Membership Prediction 47
2.4.4 Response Patterns in the Monthly Events Study 48
2.4.5 Response Patterns in the UCA Study 49
2.4.6 Multinomial Logistic Regression Results 51
2.5 Discussion and Conclusion 61
2.5.1 Significance of Respondent Personality Traits in Predicting Nonresponse Patterns in High-Frequency Online Longitudinal Data Collections 62
2.5.2 Practical implications related to personality traits 63
2.5.3 Relevance of Panel Tenure 65
2.5.4 Demographic and Health Effects Moderated by the COVID-19 Situations 65
2.5.5 Future Research Directions 68
2.5.6 Strength and Limitations 69
Chapter 3 Evaluating the Impact of Three Incentive Schemes on Survey Nonresponse and Attrition in Online Longitudinal Panels 71
3.1 Introduction 71
3.2 Background 74
3.2.1 Incentive vs. no incentive 74
3.2.2 Cost considerations in prepaid vs. promised incentives 77

3.2.3 Incentive size 77
3.2.4 Theoretical Frameworks 77
3.2.5 Study Hypotheses 80
3.2.6 Research Gap 81
3.3 Study Objectives 84
3.4 Methods 87
3.4.1 Study Participants 87
3.4.2 Experimental Design 87
3.4.3 Longitudinal Surveys 90
3.4.4 Outcome Variables 90
3.4.5 Statistical Analyses 91
3.5 Results 91
3.5.1 Study signup rate 91
3.5.2 Sample summary statistics 93
3.5.3 Overall treatment effects (Hypothesis 1) 93
3.5.4 Heterogeneous treatment effects (Hypothesis 2) 96
3.5.5 Cumulative response rates 102
3.5.6 Cost per completed survey 104
3.5.7 Secondary analysis 104
3.6 Discussion and Conclusion 105
3.6.1 Future Research Directions 112
3.6.2 Strength and Limitations 114
Chapter 4 Analyzing the Causal Effect of Survey Frequency on Nonresponse in Probability-Based Online Panels Among New Panel Respondents 116
4.1 Introduction 116
4.2 Background 118

4.2.1 Survey Frequency as a Measure of Respondent Burden 118
4.2.2 Use of Fresh Panel Respondents 122
4.3 Methods 124
4.3.1 Understanding American Study (UAS) 124
4.3.2 Experimental Design 125
4.3.3 Outcome Measure 126
4.3.4 Independent Variables 127
4.3.5 Statistical Analyses 128
4.4 Results 130
4.4.1 Descriptive statistics 130
4.4.2 Multivariable logit regression analyses 134
4.4.3 Heterogeneous Treatment Effects 137
4.4.4 Spillover Effects on Responses to Non-core Surveys 137
4.4.5 Similarities between Weighted and Un-weighted Model Estimates 138
4.4.6 Covariate Effects 139
4.4.7 Other Effects 139
4.5 Discussion and Conclusion 140
4.5.1 Strength and Limitations 146
Chapter 5 Conclusion 149
5.1 Summary of Key Findings 149
5.2 Broader Implications and Scholarly Contributions 150
5.3 Limitations and Future Research Directions 152
5.4 Conclusion 153
Appendices 154
Bibliography 178

List of Tables

Table 1-1: Annual attrition rate across probability-based online panels 2

Table 1-2: Attrition by panel membership year and recruitment batch in the UAS panel 3

Table 2-1: Key differences between probability-based online panel vs. offline panel studies..... 18

Table 2-2: Attrition by panel membership year by recruitment batch in the UAS panel 30

Table 2-3: Sample summary statistics by study sample 43

Table 2-4: Model fit statistics by number of response pattern classes 46

Table 2-5: Classification matrix..... 47

Table 2-6: Multinomial logistic regressions predicting class membership in the *monthly events panel study*. The reference group is "Stayers". Reported coefficients are adjusted relative-risk ratios (ARRR), which measure the proportionate change in the relative risk of belonging to a certain class, compared to the "stayers" class, when the predictor variable changes by one unit, while controlling for other predictor variables. 53

Table 2-7: Multinomial logistic regressions predicting class membership in the *COVID-19 panel study*. The reference group is "Stayers". Reported coefficients are adjusted relative-risk ratios (ARRR), which measure the proportionate change in the relative risk of belonging to a certain class, compared to the "stayers" class, when the predictor variable changes by one unit, while controlling for other predictor variables. 55

Table 3-1: Probability-based online panel vs. offline panels..... 82

Table 3-2: Study Signup Rates 92

Table 3-3: Overall distribution of baseline characteristics and by treatment status 95

Table 3-4: Response rates by treatment status..... 95

Table 3-5: Response rates by race and ethnicity..... 98

Table 3-6: Response rates by citizenship status..... 98

Table 3-7: Response rates by admission year..... 98

Table 3-8: Response rates by academic major..... 99

Table 3-9: Response rates by gender 99

Table 4-1: Unweighted sample demographic composition by survey frequency condition. Categorical variables are reported as percentages and Big-5 personality traits are reported as mean values..... 131

Table 4-2: Weighted and unadjusted overall response rates by frequency condition for core surveys, non-core surveys, and the combined core and non-core surveys. p-values are from design-adjusted Rao-Scott Chi-squared tests..... 133

Table 4-3: Response rates by date of invitation by frequency condition..... 134

Table 4-4: Weighted estimates of multivariable logit regression models predicting responses to UAS core surveys, non-core surveys, and combined core and non-core surveys. All standard errors (SEs) are cluster-robust treating individuals with repeated measures as clusters. 135

Appendix Table A-1: Weighted multinomial logistic regressions predicting class membership in the *monthly events panel study* using *probabilities predicting class membership for the chosen classes as weights*. The reference group is "Stayers". Reported coefficients are adjusted relative-risk ratios (ARRR), which measure the proportionate change in the relative risk of belonging to a certain class, compared to the "stayers" class, when the predictor variable changes by one unit, while controlling for other predictor variables. 158

Appendix Table A-2: Weighted multinomial logistic regressions predicting class membership in the *COVID-19 panel study* using *probabilities predicting class membership for the chosen classes as weights*. The reference group is "Stayers". Reported coefficients are adjusted relative-risk ratios (ARRR), which measure the proportionate change in the relative risk of belonging to a certain class, compared to the "stayers" class, when the predictor variable changes by one unit, while controlling for other predictor variables. 160

Appendix Table A-3: Multinomial logistic regressions predicting class membership in the m*onthly events panel study* for the latent class model that is *one class smaller* than the chosen number of latent classes. The reference group is "Stayers". Reported coefficients are adjusted relative-risk ratios (ARRR), which measure the proportionate change in the relative risk of belonging to a certain class, compared to the "stayers" class, when the predictor variable changes by one unit, while controlling for other predictor variables. 162

Appendix Table A-4: Multinomial logistic regressions predicting class membership in the *monthly events panel study* for the latent class model that is *one class larger* than the chosen number of latent classes. The reference group is "Stayers". Reported coefficients are adjusted relative-risk ratios (ARRR), which measure the proportionate change in the relative risk of belonging to a certain class, compared to the "stayers" class, when the predictor variable changes by one unit, while controlling for other predictor variables. 164

Appendix Table A-5: Multinomial logistic regressions predicting class membership in the *COVID-19 panel study* for the latent class model that is *one class smaller* than the chosen number of latent classes. The reference group is "Stayers". Reported coefficients are adjusted relative-risk ratios (ARRR), which measure the proportionate change in the relative risk of

belonging to a certain class, compared to the “stayers” class, when the predictor variable changes by one unit, while controlling for other predictor variables. 166

Appendix Table A-6: Multinomial logistic regressions predicting class membership in the *COVID-19 panel study* for the latent class model that is *one class larger* than the chosen number of latent classes. The reference group is “Stayers”. Reported coefficients are adjusted relative-risk ratios (ARRR), which measure the proportionate change in the relative risk of belonging to a certain class, compared to the “stayers” class, when the predictor variable changes by one unit, while controlling for other predictor variables. 168

Appendix Table C-1: Weighted sample demographic composition by survey frequency condition. Categorical variables are reported as percentages and Big-5 personality traits are reported as mean values. 174

Appendix Table C-2: Un-Weighted estimates of multivariable logit regression models predicting responses to UAS core surveys, non-core surveys, and combined core and non-core surveys. All standard errors (SEs) are cluster-robust treating individuals with repeated measures as clusters. 176

List of Figures

Figure 1-1: Response rates in online longitudinal data collections 5

Figure 2-1: Posterior response probabilities for the 5-class LCA model. Monthly events study (n=3,597)..... 48

Figure 2-2: Posterior response probabilities for the 6-class LCA model. COVID-19 study (n=5,891)..... 49

Figure 3-1: Incentive Experimental Design..... 89

Figure 3-2: Cumulative response rates by survey wave 103

Figure 4-1: Cumulative overall response rates by successive core survey waves..... 132

Appendix Figure A-1: Posterior response probabilities of the latent class model that is *one class smaller* than the chosen number of latent classes for the *monthly events panel study* 154

Appendix Figure A-2: Posterior response probabilities of the latent class model that is *one class larger* than the chosen number of latent classes for the *monthly events panel study* 155

Appendix Figure A-3: Posterior response probabilities of the latent class model that is *one class smaller* than the chosen number of latent classes for the *COVID-19 panel study*..... 156

Appendix Figure A-4: Posterior response probabilities of the latent class model that is *one class larger* than the chosen number of latent classes for the *COVID-19 panel study*..... 157

Appendix Figure B-1: Cumulative means of self-reported physical health (1 - Poor; 2 - Fair; 3 - Good; 4 - Very Good; 5 – Excellent). Standard errors are cluster robust..... 170

Appendix Figure B-2: Cumulative means of self-reported mental health (1 - Poor; 2 - Fair; 3 - Good; 4 - Very Good; 5 – Excellent). Standard errors are cluster robust..... 171

Appendix Figure B-3: Selective attrition analysis based on administrative data 172

List of Appendices

Appendix A: Chapter 2 Figures and Tables.....154

Appendix B: Chapter 3 Figures 170

Appendix C: Chapter 4 Tables.....174

Abstract

Survey nonresponse and attrition pose major challenges in probability-based online panels and online longitudinal data collections. Understanding the mechanisms and factors driving nonresponse and attrition is critical for improving data quality in online surveys. This dissertation addresses these challenges through three studies.

The first study analyzes the independent effects of non-demographic factors—respondent personality traits, panel tenure, and survey topics and context—important for predicting nonresponse in online panels, using datasets from two high-frequency online longitudinal studies. Using latent class analysis, we grouped respondents into classes based on similarities in their nonresponse patterns and estimated multinomial logistic regression models to predict these patterns. Findings indicate that personality traits—particularly conscientiousness—are the strongest predictors of nonresponse. Conscientious respondents exhibit lower wave nonresponse, while more open, extraverted, neurotic, and agreeable individuals are more prone to higher wave nonresponse, though with smaller effect sizes than conscientiousness. No demographic effects were found in the study focused on aging and well-being. However, in the study focused on COVID-19-related topics during the pandemic, some demographic effects emerged, likely due to the pandemic. These findings highlight the importance of personality traits over demographic variables in predicting nonresponse in online panel surveys.

The second study evaluates the impact of three theoretically-motivated incentive plans on wave nonresponse in online longitudinal surveys using a randomized controlled trial (RCT).

These incentives varied in timing of payments. The conditional incentive group received $5 per completed survey. The unconditional pre-per-questionnaire group received a $5 prepaid incentive before each survey invitation, regardless of participation. The unconditional pre-all-questionnaire lump sum group received a $15 prepaid incentive at Wave 1, regardless of participation. We recruited an Internet panel of 500 respondents, randomly selected from currently enrolled master's students in non-professional master's programs at the University of Michigan, Ann Arbor, who received three survey invitations over six months. Results show the two unconditional incentive groups had response rates 6 percentage points higher than the conditional group ($p < 0.05$), with no significant difference between the $5 and $15 unconditional groups. We identified variations in the effects of unconditional incentives across socioeconomic subgroups and found suggestive evidence that reminders were more effective in increasing response rates among unconditional incentive groups.

The third study presents findings from an RCT analyzing the causal effect of survey frequency on nonresponse in a probability-based online panel among new panel respondents. The experiment was implemented within the Understanding America Study (UAS), a probability-based online panel representative of the U.S. adult population. We recruited 2,057 participants aged 18+ using address-based sampling, randomly assigning half to a low-frequency condition (one survey every four weeks) and half to a high-frequency condition (one survey every two weeks). The only difference was survey invitation frequency. Overall, we found no appreciable differences in response rates between the two treatment groups. This effect remains after adjusting for baseline characteristics. However, we find significant heterogeneity in treatment effects when considering socioeconomic subgroups, particularly those currently employed, with an increase in response rate as large as 8 percentage points associated with

receiving one survey invitation every two weeks. In contrast, we found no discernible effects among lower or other SES groups.

Taken together, the results from these studies have practical implications for recruitment strategies, panel management and retention, nonresponse and attrition mitigation, enhancing longitudinal data collection in online panels, and informing future adaptive survey designs.

Chapter 1 Introduction

1.1 Background and motivation

Survey data are a major source of policy information in the United States. However, traditional methods of data collection, such as mail questionnaires, face-to-face interviews, and telephone interviews, face challenges in declining response rates and rising costs (de Leeuw et al., 2018; Dutwin & Buskirk, 2021; Luiten et al., 2020; Williams & Brick, 2018). Probability-based online panels have emerged as an attractive option for social science research and policy decisions due to their flexibility, convenience, speed, cost-effectiveness, and statistical efficiency. Baker et al. (2013) described the use of Internet panels to collect survey data as "one of the most compelling stories of the last decade" (p. 175).

Despite these inherent methodological and practical advantages, survey nonresponse and attrition pose significant challenges in online panels and online longitudinal data collections (for other challenges faced by Internet panels, see Baker et al. (2013) and Hays et al. (2015)). Over time, panel respondents may discontinue their participation, resulting in a cumulative loss of sample members. This attrition can have negative implications for subsequent interviews as it leads to a reduced sample size and decreased precision of survey estimates. Additionally, if the characteristics of panelists who drop out differ from those who continue participating in terms of key survey statistics, attrition can introduce bias into the data, which can undermine the validity and generalizability of research findings.

To provide initial context, Table 1-1 presents the average annual attrition rates observed in various online panels, ranging from 10-12% in the LISS panel to 25% in the Pew American

Trends Panel. In comparison, traditional non-Internet longitudinal studies such as the Health and Retirement Study, the Panel Study of Income Dynamics, and the National Longitudinal Survey of Youth Survey (NLSY79) report wave-to-wave attrition rates of 3-7% (Schoeni et al., 2013). These findings indicate that attrition rates in online panels tend to be higher when compared to their non-Internet counterparts.

Table 1-1: Annual attrition rate across probability-based online panels

Panel	Average Annual attrition
UAS[1]	13%
LISS[2]	10-12%
Pew American Trends Panel[3]	15-25%
Amerispeak[4]	15%
Ipsos/Knowledge Panel[5]	12-24%

However, examining average annual attrition rates across membership years can be misleading because it fails to capture an important underlying pattern: online panels experience higher attrition rates in the first membership year compared to subsequent years. Table 1-2 provides data on annual attrition in the first and second membership years for each nationally representative recruitment batch in the UAS panel from 2014 to 2018. While the average attrition rate across membership years and recruitment batches in the UAS panel is typically 13%, Table 1-2 reveals that the largest attrition occurs during the first membership year, ranging from 14% to 25% across all recruitment batches. Attrition then significantly decreases in the second and subsequent membership years. This pattern suggests that focusing efforts to reduce attrition

[1]Average across all recruitment batches. Average annual attrition rates differ by recruitment batch. Author's own calculation based on UAS attrition data. https://uasdata.usc.edu/index.php?r=eNpLtDK0qi62MrFSKkhMT1WyLrYytFwwskuTcjKT9YpSk4tKM0tyU_NKlKxrASjBDmQ,

[2] https://www.lissdata.nl/about-panel/composition-and-response

[3] https://www.pewresearch.org/methods/u-s-survey-research/american-trends-panel/

[4] https://www.amerispeak.org/

[5] Author's email communication with PI for Ipsos/Knowledge Panel, on Jan 13, 2022.

among new panel members can be an effective strategy for mitigating overall attrition in online panels.

Table 1-2: Attrition by panel membership year and recruitment batch in the UAS panel

		Attrition by panel membership year (%)	
Recruitment batch	Year of recruitment	Year 1	Year 2
1	2014	17.20	8.99
2	2015	16.47	7.39
3	2016	13.99	7.68
4	2016	15.93	8.82
5	2016	18.36	7.50
6	2016	17.31	7.29
7	2016	18.22	10.13
8	2016	17.38	9.28
9	2016	22.99	11.48
10	2018	24.84	14.24
Average annual attrition across batches		**18.27**	**9.28**

Data source: https://uasdata.usc.edu/index.php

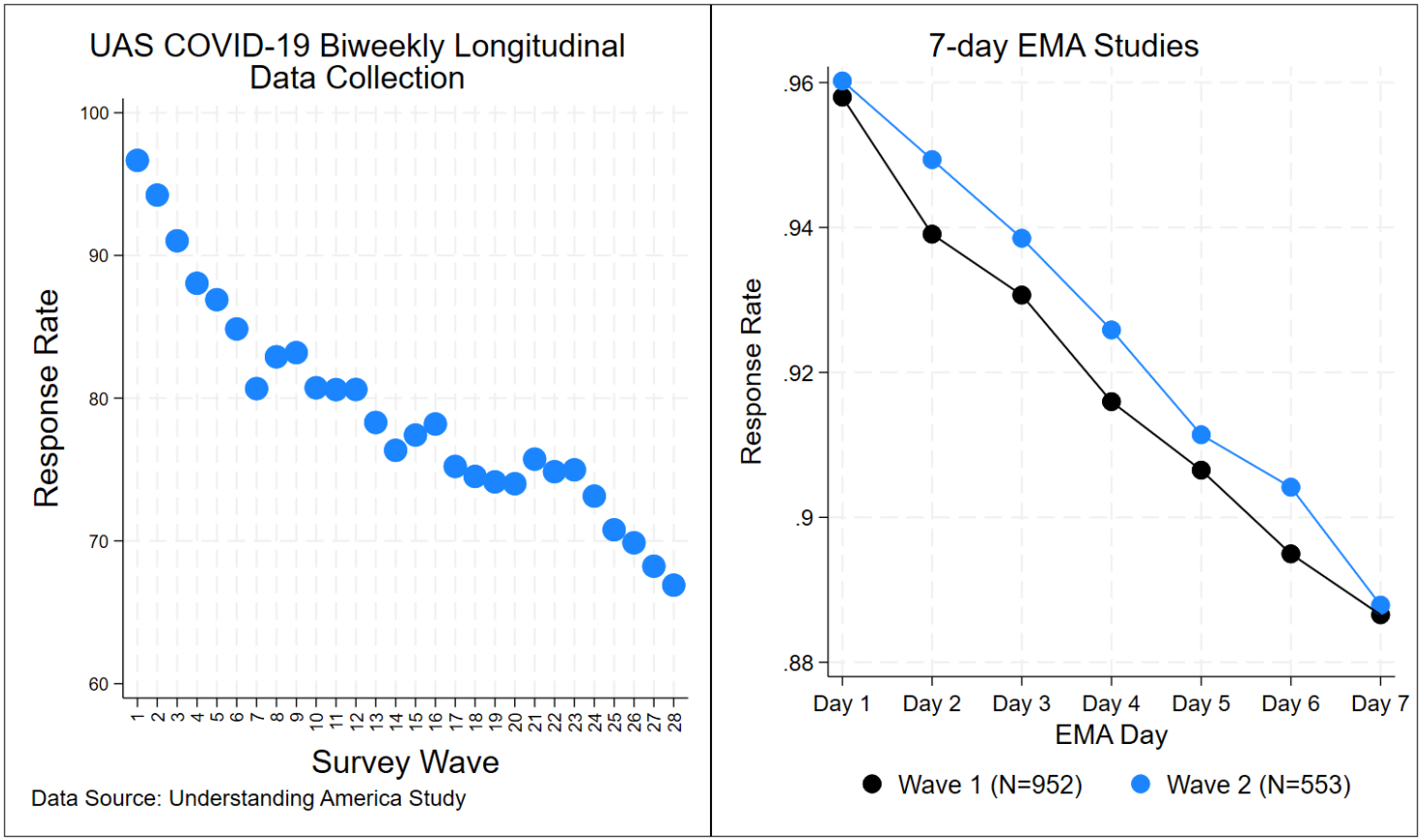
As researchers and policy makers increasingly rely on Internet panels for longitudinal data collections (Alattar et al., 2019), a similar issue emerges in online longitudinal studies, where survey participation rates tend to decline rapidly over time. To investigate this phenomenon, we examine the Understanding Coronavirus in America Study—a comprehensive online longitudinal data collection effort that aimed to capture real-time data and the dynamic shifts occurring during the COVID-19 pandemic (Kapteyn et al., 2020). In this study, approximately 7,000 participants from the UAS were recruited to complete 28 biweekly surveys using online platforms. The data collection spanned from March 2020 to July 2021, covering a wide range of critical periods during the pandemic.

The left panel of Figure 1-1 provides insights into the response rates observed throughout the 28 waves of data collection, with an average response rate of 79.04%. Notably, the study experienced a high response rate of 96.66% in the first wave, likely driven by the heightened saliency of the study topics during the early stages of the COVID-19 pandemic. However, as

data collection progressed, the response rates gradually declined, reaching a final response rate of 66.90% in wave 28.

While the decline in response rates can partially be attributed to the phenomenon of regression to the mean, where initial high response rates tend to normalize over time, it is crucial to recognize that this overall trend aligns with the commonly observed pattern of declining response rates as data collection progresses. To further support this observation, the right panel of Figure 1-1 presents daily response rates for 7 end-of-day (EOD) surveys from two population-representative Ecological Momentary Assessment (EMA) studies conducted by Stone, Schneider, Smyth, et al. (2023). These EMA studies involved participants completing 7 EOD surveys along with other tasks. Each EOD survey consisted of questions eliciting information about respondents' mood, pain, fatigue, physical activity, diet, sleep, and social and family interactions during the day. The right panel of Figure 1-1 displayed a similar trend of rapidly declining response patterns over time.

Figure 1-1: Response rates in online longitudinal data collections



Similarly, in a study by West et al. (2023), where respondents completed a long survey split into three short modules with a two-week interval between survey modules, it was found that of the respondents who completed module 1, 82.6% completed module 2, and of those respondents, 79.2% went on to complete module 3. This represents a module-to-module attrition rate of 20%. The response rate patterns observed in these three studies provide valuable insights into participant engagement in intensive online longitudinal data collection processes and have important implications for data validity and generalizability.

In the United States, online data has emerged as an indispensable resource for guiding evidence-based policy decisions. For example, the Social Security Administration (SSA) has been collecting longitudinal data using an online panel to track the U.S. population's knowledge about Social Security programs and to inform policy decisions (Alattar et al., 2018; Alattar et al., 2019). Similarly, the National Center for Health Statistics (NCHS) has been experimenting with using internet panels to collect rapid-turnaround population health data in response to emerging events, such as COVID-19 (Blumberg et al., 2021). Given the increasing use of online data for policy decision-making, it is essential to collect valid and accurate online population data. A relevant case study by Bradley et al. (2021a) exemplifies the dangers of using unrepresentative online samples. Two major studies overestimated national COVID-19 vaccination rates by 11-20 percentage points, due to the use of non-representative samples. Such biased estimates jeopardize sound policy making.

In summary, probability-based online panels offer numerous methodological and practical advantages for data collection, including flexibility, speed, convenience, and cost-effectiveness. However, higher attrition rates during the first year of membership and declining response rates over time in online longitudinal data collections present significant challenges.

These issues can impact sample size, survey estimates, and the validity of research findings. This dissertation seeks to address these challenges through three independent but interrelated studies.

1.2 An overview of the dissertation

This dissertation, including the Introduction, is organized into five chapters. Chapters 2–4 form the core of the dissertation, consisting of three independent yet interrelated studies that collectively address attrition and nonresponse issues in probability-based online panels and high-frequency online longitudinal data collections. Two of these core studies involve randomized controlled trials (RCTs), while the third analyzes two existing, novel datasets. By providing new evidence, these studies also enhance our understanding of effective strategies for improving survey engagement and data quality in online panels. Each core chapter is presented in its entirety as a standalone study, enabling readers to engage with its content independently without needing to refer to other chapters.

Below, we provide summaries of each of the three core studies. Chapter 5 summarizes the key contributions and findings of the individual studies and concludes with recommendations for future research.

1.2.1 Chapter 2 overview

Design differences between offline and online longitudinal studies can differentially impact nonresponse and attrition outcomes. In offline longitudinal studies, where interviewers must first locate respondents and then establish contact with them, household-, individual-, and interviewer-level variables primarily affect nonresponse and attrition through contactability. In contrast, in high-frequency online panel studies, surveys are delivered via email and self-administered, rendering conventional demographic variables that predict contactability less relevant for predicting nonresponse patterns. Consequently, the correlates influencing

nonresponse and attrition outcomes, as well as the underlying mechanisms driving these outcomes, may differ between the two approaches.

In this chapter, titled "Personality traits, panel tenure, survey topic, and context as predictors of survey nonresponse patterns in high-frequency online longitudinal surveys," we analyze the independent effects of non-demographic variables—respondent personality traits, panel tenure, and survey topics and context—that are likely to be important for predicting nonresponse patterns in the high-frequency online longitudinal data collection context. Additionally, we provide theoretical justification for using these non-demographic variables as predictors, drawing parallels from the medical and economic fields to further support their relevance.

Data for this study are from two recently completed high-frequency online longitudinal studies conducted in the U.S--the Understanding Coronavirus in America Study and the UAS Monthly Events Surveys. Both studies are part of the Understanding America Study (UAS) – a probability-based Internet panel that is representative of US adults and is used for conducting social science, policy, and health-related research. The two studies mainly differ in terms of survey topics (COVID-19 vs. Non COVID-19) and timing of implementation (one during COVID-19 and the other partly during a non COVID-19 period). Additionally, the two studies collected data at various frequencies (2 weeks vs. one month) with varying observation periods (14 months vs. 52 months).

Using latent class analysis (LCA), we grouped respondents into various response classes based on similarities in their nonresponse patterns across all survey waves, which revealed substantial variations in patterns of nonresponse. To predict these nonresponse patterns, we estimate multinomial logistic regression models utilizing rich background information available

for all respondents as regressors. The background information used in the regression analyses includes respondent personality traits, demographics, health outcomes, and years in the panel (referred to as panel tenure). Finally, we compare and contrast the regression coefficients between the two studies and examine whether COVID-19 related topics and the general societal environment during the pandemic moderate the relationship between individual predictor variables and nonresponse patterns.

This chapter provides robust empirical evidence on the mechanisms driving survey nonresponse patterns and the non-demographic correlates predicting them in high-frequency online longitudinal data collections in the U.S, which differ from the offline panel context. We conclude the chapter by discussing practical and theoretical implications of the findings for online panel management, nonresponse adjustments, and the design of surveys targeting specific respondent subgroups, and by offering recommendations for future research.

1.2.2 Chapter 3 overview

Most studies examining the effects of various incentive strategies in longitudinal data collections have primarily focused on traditional offline longitudinal studies, where respondents are interviewed annually or biannually face-to-face or using sequential mixed modes. Comparable incentive studies analyzing nonresponse and attrition outcomes in the online context have been limited. Additionally, due to fundamental differences between online and offline panel studies, effective incentive strategies used in offline contexts can have limited applications in online contexts.

In this chapter, titled “Evaluating the impact of three incentive schemes on survey nonresponse and attrition in online longitudinal panels,” we report findings from a randomized controlled trial (RCT) that evaluated the causal effects of three theoretically-motivated and

practically-relevant cash incentive payment strategies on nonresponse in an online longitudinal data collection effort. These incentives, tailored to the unique context of online data collection, primarily varied in the timing of the incentive payments and were informed by prior studies and existing survey response theories.

1. Conditional Per-Questionnaire Incentive (Control group): Respondents received a $5 incentive for each completed survey wave. This conditional incentive is a common practice used by most online longitudinal studies
2. Unconditional Pre-Per-Questionnaire Incentive: Respondents received a $5 unconditional incentive upfront one week prior to receiving individual survey invitations, regardless of participation.
3. Unconditional Pre-All-Questionnaire Lump Sum Incentive: Respondents received a lump sum unconditional payment of $15 upfront at Wave 1 (equivalent to the expected earnings for a given survey wave multiplied by the number of survey waves), regardless of participation.

We recruited an Internet panel of approximately 500 respondents, randomly selected from a complete list of all currently enrolled master's students in non-professional master's programs at the University of Michigan, Ann Arbor campus. Over the course of about 6 months, respondents received an online survey invitation every 2.5 months, for a total of three invitations. Each survey had an approximate completion time of 10 minutes. The survey questionnaires were identical across all waves and covered various aspects of students' well-being, including their overall mental and physical health, time management and use, perceptions of stress, life stressors, sources of stress, and work-related stress.

The frame data included background information on all respondents, which we used to explore the heterogeneity of treatment effects and study participation rates based on these characteristics. We utilized existing theories as a framework to interpret and illuminate both the overall and heterogeneous treatment effects and participation rates.

This study aims to shed new light on effective interventions aimed at enhancing online data quality, better inform the design and implementation of future online longitudinal studies, and motivate further research. We conclude the chapter by discussing the practical implications and significance of the findings with respect to recruitment strategies, panel management, nonresponse, attrition, the efficacy of longitudinal data collection in online panels and future adaptive survey design, and by offering suggestions for future research.

1.2.3 Chapter 4 overview

In his seminal work, Bradburn (1978) identified survey frequency as a factor contributing to respondent burden, which negatively affects survey response and data quality. As survey response rates for mailed questionnaires and face-to-face interviews, or telephone interviews continue to decline, the issue of respondent burden has gained significant attention in the survey research literature, especially since the publication of the *Paperwork Reduction Act Guide* (2011), which aims to address this challenge (Kaplan et al., 2022; Paperwork Reduction Act Guide, 2011). However, findings from past research have been mixed, with some studies showing positive associations between survey frequency and response rates, while others have found no significant effect. This inconsistency stems largely from the limited exogenous variation in survey frequency available in cross-sectional and conventional longitudinal surveys, as well as the reliance on observational data, which is prone to selection bias.

Most importantly, prior findings from cross-sectional and conventional longitudinal studies have limited applicability to probability-based online panels because of fundamental differences in design features and context. Given the increasing use of online data for policy decision-making and social science research, more research is needed to enhance our understanding of response mechanisms related to survey frequency in online panel settings.

In Chapter 4, titled "Analyzing the causal effect of survey frequency on nonresponse in probability-based online panels among new panel respondents," we report findings from a randomized controlled trial (RCT) in which we recruited about 2,000 new online panelists from the US adult population aged 18 and older using address-based sampling approach, and randomly assigned them in equal numbers to either a low- or high-frequency group. The low-frequency group received one core survey invitation every four weeks, while the high-frequency group received one core survey invitation every two weeks. Both groups received the same series of surveys. Importantly, the survey length and content were the same for both experiment groups to maximize comparability. The experiment was conducted over a one-year duration, allowing sufficient time to observe and analyze the impact of survey frequency on nonresponse. We focus on new respondents because existing data indicate that the majority of attrition occurs during the first year of panel membership. The study's use of experimental data addresses selectivity issues affecting previous studies that used observational data. The experiment is embedded within the Understanding America Study (UAS), a probability-based online panel representative of U.S. adults aged 18 and older.

The availability of detailed background information on all respondents allows us to control for other factors that predict survey response, thereby enhancing statistical inferences. It also enables the exploration of whether the frequency effect differs across subgroups based on

background information, which can inform targeted interventions and enhance panel management and retention strategies. In summary, this study combines experimental data with detailed background information and longitudinal survey data to answer two research questions: (1) Is there a causal relationship between survey frequency and nonresponse among new respondents in online panel surveys? If so, what is the direction of this causal effect? (2) Does the causal effect of survey frequency on nonresponse vary across socioeconomic and racial subgroups?

We discuss the theoretical implications of the study's findings within the framework of existing theories, compare them with prior studies, examine their practical implications for adaptive design approaches in future online surveys, explore potential underlying mechanisms that influence observed nonresponse outcomes, propose directions for future research, and highlight the strengths and limitations of the study's approach.

Chapter 2 Personality Traits, Panel Tenure, Survey Topic, and Context as Predictors of Survey Nonresponse Patterns in High-Frequency Online Longitudinal Surveys

2.1 Introduction

Probability-based Internet panels offer researchers the flexibility to launch surveys at any time and collect data from a large number of respondents within a short timeframe, while also offering convenience to respondents who can complete surveys at their own pace and in a private setting. This approach is cost-effective as it eliminates or minimizes the need for intensive fieldwork and logistical operations commonly associated with in-person and telephone interviews, and mail questionnaires. Furthermore, online data enhances statistical efficiency by avoiding the clustering effect often encountered in face-to-face interviews. These advantages have been well-documented in previous studies (Alattar et al., 2018; Angrisani et al., 2019a; Callegaro et al., 2014; Cornesse et al., 2022; Das et al., 2018; Hays et al., 2015; Kapteyn et al., 2020; Kapteyn et al., 2024; MacInnis et al., 2018; Maslovskaya et al., 2022; West et al., 2023).

The Internet technology also allows for quick responses to new and emerging developments. For instance, during the COVID-19 pandemic, online panels have proven valuable in monitoring daily mood changes, observing changes in infection rates, job loss, food insecurity, and vaccination rates over time, and exploring the evolving correlation between COVID-19 vaccines and mental distress in the U.S. adult population (de la Haye et al., 2023; Kapteyn et al., 2020; Perez-Arce et al., 2021). Similarly, during the H1N1 pandemic, online panels played a crucial role in monitoring population trends related to risk behaviors and vaccine intentions in the U.S. (Gidengil et al., 2012). Additionally, internet panels offer researchers the

ability to collect high frequency longitudinal data. This also has a methodological advantage because collecting information closer in time to the occurrence of important events, and collecting it repeatedly reduces the role of measurement error and omitted variables, enabling sharper statistical inference. This is particularly important for minimizing memory errors, a key source of measurement error in survey data (Tourangeau, 1999).

Despite the inherent strengths, as with any longitudinal study, nonresponse and attrition pose a significant challenge in online data collections. Reasons for respondents discontinuing their participation may include, among others, a lack of sustained commitment, panel fatigue or cumulative burden, and external shocks such as illness or loss of employment (Hu, Melipillán, et al., 2020; Lugtig, 2014; Yan & Williams, 2022).

Both nonresponse and attrition lead to a reduced sample size and decreased precision of survey estimates. Given that data are often collected more frequently in an online panel context, even a modest attrition rate in each wave can lead to a significant loss of sample in a short period of time. Additionally, if the characteristics of panelists who drop out differ from those who continue participating, attrition may introduce bias, which can undermine the validity and generalizability of research findings (Chin et al., 2021; Hansson et al., 2018; Lugtig, 2014; Lugtig et al., 2014; Lynn, 2018; R. Olsen, 2018; Ward et al., 2017).

The implications of these challenges are underscored by several studies. Research has shown examples where statistical adjustments did not completely eliminate bias resulting from respondents not completing all longitudinal surveys (Hu, Kirlin, et al., 2020) and that nonresponse adjustments based on standard demographic variables were not sufficient to fully address attrition bias in panel surveys (Cheng et al., 2020; Frankel & Hillygus, 2014).

Additionally, some statistical estimates can be sensitive to the type of adjustment method employed (Ayala et al., 2011; Basic & Rendtel, 2007; Kristman et al., 2005).

In this study, we analyze survey nonresponse data from two recently completed studies that collected high frequency online longitudinal data. These two studies were implemented in the Understanding American Study (UAS) – a probability-based Internet panel representative of US adults and used for conducting social science, policy, and health-related research. The two studies mainly differ in terms of survey topics (COVID-19 vs. Non COVID-19) and timing of implementation (one during COVID-19 and the other partly during a non COVID-19 period). Additionally, the two studies collected data at various frequencies (every 2 weeks vs. one month) with varying observation periods (14 months vs. 52 months). While there is no universally accepted definition of what qualifies as "high" frequency, we define high-frequency online longitudinal studies as those that field at least one survey per month. We utilize a latent class analysis (LCA) approach to group respondents based on their similarities in nonresponse patterns. We then estimate multinomial logistic regression models utilizing background information available for all respondents as regressors to predict these nonresponse patterns as identified by the LCA model. The background information we used in the regression analyses include psychological traits, demographics, and health outcomes. In addition, we analyze the independent effect of respondents' time in the panel (referred to as panel tenure) in predicting nonresponse patterns due to its importance for panel management and relevance in predicting nonresponse outcomes in online panel surveys (Kapteyn et al., 2024). Finally, we compare and contrast the regression coefficients between the two studies and examine whether COVID-19 related topics and the general societal environment during the pandemic moderate the relationship between individual regression coefficients and nonresponse patterns. Findings from

the study shed new light on survey nonresponse patterns, along with their correlates and underlying response mechanisms, specific to high-frequency online longitudinal data collections in the US context.

In summary, this study makes several contributions to the emerging literature on nonresponse and attrition in high-frequency online longitudinal data collections. First, it contributes to our understanding of how design differences between offline and online panels differentially influence nonresponse, attrition outcomes, and underlying mechanisms. Second, it offers novel insights into the critical role of respondent personality traits, particularly conscientiousness, in predicting nonresponse patterns in these contexts, even when accounting for other predictors, which could have implications for the construction of nonresponse adjustment weights. Third, by analyzing response data from longitudinal studies where the same respondents were interviewed repeatedly and frequently using essentially the same questionnaire, this study better isolates the effects of personality traits from those of changing demographic characteristics and survey experiences, addressing limitations in prior studies where sample composition and survey topics varied across waves. Finally, by leveraging data from two recent and unique datasets, the study provides new evidence on the independent effect of panel tenure and the potential moderating effect of survey topics on nonresponse patterns.

2.2 Background

2.2.1 Differences between Offline and Online Panel Studies

While researchers have been collecting offline panel data for decades, it is only in recent years that the collection of high-frequency online longitudinal data has begun, made possible primarily by the high penetration of the Internet and recent advances in Internet technology. As a consequence, the majority of prior studies analyzing nonresponse and attrition outcomes have

focused on offline longitudinal studies (Fitzgerald et al., 1998; Lepkowski & Couper, 2002; Zabel, 1998), with limited comparable research using nonresponse data from probability-based online longitudinal studies (Cheng et al., 2020; Jin & Kapteyn, 2022; Kapteyn et al., 2024). Importantly, both online and offline panels typically rely on similar offline recruitment strategies, including address-based sampling and initial outreach through postal mail, telephone, or in-person contact. The differences discussed in this chapter—particularly those summarized in Table 2-1 refer primarily to the longitudinal data collection phase that follows initial recruitment and sample selection. Wherever applicable, we also highlight shared features of online and offline panels during this follow-up phase, particularly in relation to interviewing modes and contact protocols.

These differences highlighted in Table 2-1 can have important implications for nonresponse and attrition outcomes and the underlying mechanisms influencing them. In the following discussion, we examine these differences in detail. We then explore how the underlying mechanisms influencing nonresponse patterns, as well as the individual- and household-level variables that predict them, may differ between the two approaches. We then analyze the independent effects of respondent personality traits, panel tenure, and survey topic and context that are more likely to be important in predicting nonresponse patterns in high-frequency online longitudinal data collections than conventional demographics.

Table 2-1: Key differences between probability-based online panel vs. offline panel studies

	Online panel	**Offline panel**
Primary mode of data collection	• Online	• Face-to-face • Mail • Telephone • Sequential mixed modes

Data collection frequency	Very frequent: • Daily • Weekly • Monthly • Bimonthly	Less frequent: • Annually • Biannually • Or longer
Interview length	• 15-30 minutes	• 80-120 minutes
Questionnaires and survey topics across waves	• Mostly varying	• Mostly fixed
Common reason(s) for nonresponse and attrition	• Mainly refusal • Panel fatigue	• Non-location • Noncontact • Refusal

The mode of data collection is the first design aspect that distinguishes online panels from offline panel studies. In most online panel studies, the sole mode of data collection used is the Internet, although some Internet panels offer alternative modes in addition to the online option. For instance, the AmeriSpeak Panel in the U.S. uses a mixed-mode design by providing a telephone option to panelists without Internet access and to panelists with Internet access who are unwilling to provide an email address (NORC at the University of Chicago, 2022). Similarly, the GESIS panel in Germany provides paper-and-pencil questionnaires sent via postal mail to participants unable or unwilling to complete surveys online (Bosnjak et al., 2018). Providing a second mode option to online panel respondents undermines the cost and efficiency advantage of the online mode but has the potential to improve sample representation.

A few offline panel studies have switched to or begun experimenting with push-to-web sequential mixed modes (Bianchi et al., 2016; Booth et al., 2024; Cabrera-Álvarez & Lynn, 2024; Gaia, 2017; Jäckle et al., 2015; Patrick et al., 2018; Peycheva et al., 2023; Wong, 2020; Wood & Kunz, 2014). This shift from a single mode to a mixed-mode design is mainly necessitated by the need to address the rising costs associated with face-to-face and telephone interviews and to take advantage of increasing Internet penetration. In this mixed-mode design, offline panel respondents are encouraged to complete interviews online first, followed by face-

to-face or telephone interviews if necessary. This approach substantially reduces costs through decreased fieldwork and fewer field interviews. Couper and McGonagle (2019) provide a comprehensive review of recent developments in the use of web surveys in mixed-mode data collection for conventional offline longitudinal studies.

The two study approaches also differ in terms of data collection frequency. In online panel studies, data collection can be very frequent. Participants might be asked to provide data daily (as in 7-day wellbeing and diary surveys), weekly, monthly, or bimonthly. Participants in Ecological Momentary Assessment (EMA) studies may be interviewed multiple times a day over several days (Stone, Schneider, & Smyth, 2023; Stone et al., 2024). The high frequency is facilitated by the convenience and accessibility of online surveys. In EMA studies, surveys may also be delivered via push notifications through smartphone apps, where timely response is critical and email alone may not suffice. In contrast, offline panel studies generally have less frequent data collection schedules. Data might be collected annually, biannually, or even less frequently. The differences in data collection frequency have significant implications for the type of data gathered and the responsiveness of a study to new developments such as pandemics and financial crises.

Online panel interviews are typically shorter, ranging from 15 to 30 minutes. This brevity is designed to maintain participant engagement and reduce the likelihood of break-off and survey fatigue. In contrast, aided by the presence of interviewers and long between-wave gaps, offline panel interviews tend to be much longer, lasting between 80 and 120 minutes.

Online and offline studies use different questionnaire designs across survey waves. Most online panels field both longitudinal and cross-sectional surveys, with survey topics and questionnaires varying significantly across different waves. This flexibility allows online studies

to adapt to changing research needs and emerging topics of interest such as the COVID-19 pandemic. In contrast, offline studies primarily field longitudinal questionnaires with mostly fixed topics across waves, providing consistent data over time. Occasionally, offline panel studies may field supplemental questionnaires between waves to address specific research questions or capture additional data points.

Having examined the key differences and similarities between offline and online panels, several design features are especially important for interpreting the findings of this study, as outlined below. First, the study relies on a probability-based online panel recruited through an address-based sampling approach using a tailored design method. Second, respondents are surveyed frequently—typically every two weeks. Third, the questionnaire design incorporates both fixed and varying survey topics across waves. Fourth, each survey is relatively short, not exceeding 30 minutes in length. Fifth, respondents are compensated based on the estimated time required to complete each survey. Finally, all surveys are self-administered online, with invitations delivered exclusively via email.

2.2.2 Different Correlates of Attrition and Response Mechanisms in Offline vs. Online Panel Studies

The aforementioned design differences can differentially impact nonresponse and attrition outcomes, and the correlates influencing these outcomes may vary between the two study approaches.

In any follow-up survey in an offline study, interviewers must first locate respondents and then establish contact with them, conditional on being located. Once contact is made, interviewers attempt to gain cooperation from respondents. Failure at any of these three steps could result in nonresponse. As such, most offline studies have modelled nonresponse and

attrition outcomes as a function of whether: sample respondents moved between survey waves, contacts with sample respondents were established conditional on respondents being located, and sample respondents cooperated when approached by interviewers, conditional on contact being made. Non-location occurs when participants move and cannot be reached, while noncontact refers to the difficulty in reaching participants physically despite repeated attempts. These three processes can be modeled separately (Couper & Ofstedal, 2009; Lepkowski & Couper, 2002; Uhrig, 2008; Watson & Wooden, 2009) or cumulatively without making any distinction among the three processes (Frankel & Hillygus, 2014; Olson & Witt, 2011).

Given the three aforementioned processes through which nonresponse could occur in offline studies, many prior studies have focused on identifying correlates that predict non-location, non-contact, or the cumulative final outcome (Couper & Ofstedal, 2009; Fitzgerald et al., 1998; Lepkowski & Couper, 2002; McGonagle et al., 2011; Uhrig, 2008; Watson & Wooden, 2009; Zabel, 1998). Household-, individual-, and interviewer-level variables that have been found to predict these outcomes include, among others, housing arrangement (e.g., renting vs. owning), housing characteristics (e.g., gated entry, single home, communal housing), housing tenure (i.e., years at current residence as proxy for stability), household size, number of children, employment status (e.g., retired, currently working), marital status (as a proxy for geographical stability), changes in events between waves such as changes in employment and household composition (Trappmann et al., 2015), level of interviewer experience, interviewer training, and use of the same interviewers across survey waves. However, these household-, individual-, and interviewer-level variables primarily affect attrition through contactability and securing initial cooperation with respondents, making them less relevant for predicting nonresponse patterns in online panel contexts, where surveys are delivered via email and self-administered.

The use of different data collection modes between the two study approaches can impose varying levels of survey burden, which can, in turn, lead to different nonresponse patterns. For instance, since most offline panels use face-to-face and/or telephone modes administered by interviewers, the cognitive effort required to answer difficult or unclear questions may be lower for offline panel respondents as interviewers can clarify questions. Conversely, when answering sensitive questions, the presence of interviewers can create psychological stress for offline panel respondents (Newman et al., 2002; Perlis et al., 2004).

In this study, we analyze the independent effects of non-demographic variables – respondent personality traits, panel tenure, and survey topics and context – that are likely to be important for predicting nonresponse patterns in high-frequency online longitudinal studies, as detailed below.

2.2.3 Respondent Personality Traits as Consistent and Strong Predictors of Nonresponse Patterns in High-frequency Online Longitudinal Data Collections

Online surveys are typically delivered via email and self-administered, which reduces the impact of traditional predictors such as contactability and interviewer interactions. We may hypothesize that participants might decline to continue if the frequency of surveys is considered too high, or due to a lack of interest, or disruption in their survey-taking routine. Based on this expectation, we hypothesize that personality traits are important predictors of a respondent's likelihood to participate consistently in online panel surveys and may have more explanatory power than conventional demographic and household variables commonly found in offline panel studies.

Building on existing theoretical foundations such as the Social Isolation Hypothesis (Groves & Couper, 2012), which posits that individuals who are socially isolated or less

integrated into society are less likely to participate in surveys due to weaker social norms and lower perceived obligation to contribute to collective endeavors, Saßenroth (2013) emphasizes the predictive power of personality traits over traditional socio-demographic factors, particularly highlighting how subjective experiences influence survey participation decisions. This perspective underscores the role of personality traits as critical determinants of nonresponse patterns in high-frequency online longitudinal studies.

The Big Five personality traits, conscientiousness, openness, agreeableness, neuroticism, and extraversion, are widely used in psychological research and have been found to be stable throughout a person's lifespan and across cultures and languages (Costa & McCrae, 1988; McCrae & Costa Jr, 1994; McCrae et al., 2004; McCrae & John, 1992). Respondents with certain personality profiles may be more inclined to engage regularly and provide complete data, while others may be prone to drop out or respond sporadically. As such, there is a need to look beyond conventional demographic information in analyzing nonresponse patterns in high-frequency online longitudinal surveys.

This argument is supported by drawing a parallel to the literature on medication adherence, where certain personality traits, particularly conscientiousness, have been shown to play a significant role. The conscientiousness trait is defined as "socially prescribed impulse control that facilitates task- and goal-directed behavior, such as thinking before acting, delaying gratification, following norms and rules, and planning, organizing, and prioritizing tasks" (John & Srivastava, 1999). Individuals high in conscientiousness exhibit qualities such as diligence, perseverance, tenacity, orderliness, and industriousness. These traits are crucial for adhering to medication schedules, as they involve remembering to take medications, planning around dosage times, and maintaining a consistent routine. The effect of conscientiousness in predicting

medication adherence has been found to be robust across various disease conditions, population subgroups, and cultural settings (Adachi et al., 2022; Axelsson, 2013; Axelsson et al., 2011; Ko et al., 2020; Lima et al., 2018; Molloy et al., 2014; Wheeler et al., 2012).

Similarly, this characteristic is likely relevant for survey completion in high-frequency online longitudinal studies. Completing online surveys regularly requires respondents to pay attention to survey invitations sent via emails, remember due dates, and respond to survey reminders. Conscientious individuals are likely to excel in these areas due to their propensity for planning, organizing, and prioritizing tasks. Therefore, just as conscientiousness is a strong predictor of medication adherence, it should also serve as a strong predictor of nonresponse patterns characterized by low wave nonresponses in high-frequency online longitudinal studies.

In contrast to conscientiousness, a higher score in openness may reflect individuals' tendency to embrace new ideas and experiences, but this quality may not translate into a strong commitment to complete high-frequency and longitudinal surveys. More agreeable individuals are trusting, altruistic, sympathetic, and cooperative, which may predict participation in a new one-time study but may not be as strong a predictor of sustained participation in a long-term, continuous study, such as a high-frequency longitudinal survey. Neurotic individuals, being susceptible to anxiety and depression, could be more prone to panel fatigue, or study dropout. Extraversion is mainly associated with being talkative, sociable, energetic, and friendly, but high extraversion may not translate to the long-term commitment needed for participation in an online longitudinal study. Consequently, we expect openness, agreeableness, neuroticism, and extraversion to be positively associated with higher wave nonresponses in a high-frequency online longitudinal study.

Several studies in the economic and psychological literature have extensively documented personality traits as independent predictors (net of other important factors) and causes of significant life outcomes, including educational attainment, employment outcomes, wealth, mental and physical health status, well-being, marriage, divorce, criminal activity, volunteering, and productivity, among others (Almlund et al., 2011; Cubel et al., 2016; Friedman & Kern, 2014; Heckman et al., 2006; Roberts et al., 2007; Soto, 2019; Wright & Jackson, 2023). If personality traits are found to predict nonresponse patterns in high-frequency online longitudinal studies, even after accounting for conventional respondent- and household-level demographic variables, this could have practical implications for how nonresponse adjustment weights are calculated.

In offline panel studies, a few investigations have found that personality traits independently predicted survey completion. Satherley et al. (2015) analyzed the first four waves of response data from the New Zealand Attitudes and Values Study, an offline longitudinal national panel where respondents were interviewed annually using self-administered paper questionnaires, with a web option provided in waves 3 and 4. They found that conscientiousness was positively associated with consistent participation in the panel study. Similarly, Hansson et al. (2018) analyzed response data from the first three waves of the Swedish longitudinal population-based Health, Ageing, and Retirement Transitions in Sweden (HEARTS) study, an offline panel study with annual follow-ups using both paper-pencil questionnaires and online modes. They found that higher extraversion and neuroticism, and lower agreeableness, independently predicted attrition. However, given that online and offline panel studies differ in many design aspects (see Table 2-1), it is unclear to what extent these findings can be generalized to the context of high-frequency online longitudinal data collection.

Within online panel contexts, Cheng et al. (2020) analyzed wave-to-wave response data from panel surveys fielded in the UAS and found that conscientiousness and openness to experience predicted the incidence of unit nonresponse in subsequent survey waves, even after controlling for cognitive ability and demographic characteristics. Additionally, Lugtig (2014) analyzed response data from surveys conducted in the LISS panel in the Netherlands and found that respondents who dropped out were less likely to be conscientious and more likely to be extraverted and agreeable.

It should be highlighted that both Cheng et al. (2020) and Lugtig (2014) used response data generated from online panel surveys, comprising both cross-sectional and longitudinal surveys, with to some extent varying sample compositions and survey topics across survey waves. In such settings, where both respondent profiles and survey topics change, disentangling the effects of respondent personality traits from those of changing survey topics and respondent profiles over time is challenging. This challenge is underscored by prior studies revealing that prior-wave survey experiences, survey-taking experience levels, and variations in survey topics across waves predict cooperation in current and future surveys (Frankel & Hillygus, 2014; Lugtig, 2014; McLauchlan & Schonlau, 2016). Consequently, it remains unclear whether the findings found in a mix of cross-sectional and longitudinal online Internet panel surveys can be replicated in an online longitudinal study setting, where the sample mix and survey topics remain constant across waves and where the level of survey commitment required and survey burden involved are significantly higher.

In addition to influencing survey participation, personality traits have also been shown to affect satisficing behavior. Using data from two large UK panel studies—wave 3 of the UK Household Longitudinal Study (UKHLS) and wave 1 of the British Election Study Online Panel

(BESOP)—Sturgis and Brunton-Smith (2023) found that conscientious and agreeable individuals were significantly less likely to engage in satisficing behaviors, such as straightlining, selecting non-substantive response options, or skipping items, across both face-to-face and self-administered modes and in both probability and nonprobability samples. They also found that Conscientiousness and Agreeableness were stronger predictors of satisficing than Need for Cognition and Need to Evaluate, and that the effect of Agreeableness was more pronounced in face-to-face than in self-administered settings. The study did not examine the effects of Extraversion, Openness, or Neuroticism—other important domains of personality—on satisficing behavior. Nonetheless, their findings provide further support for the importance of accounting for personality traits in the design and implementation of high-frequency online longitudinal data collections.

In this current study, we analyze the independent effects of respondent personality traits using response data from two recently completed online longitudinal studies, where the same respondents were interviewed repeatedly using essentially the same questionnaire. This approach allows us to better isolate the effects of personality traits from those of changes in demographic characteristics and survey experience across survey waves.

Recent studies in the public opinion literature have begun to explore the role of personality in survey participation. Hibbing et al. (2019) show that traits like openness and conscientiousness are linked to response styles that may bias survey estimates. Valentino et al. (2020) find that openness and extraversion are negatively associated with participation in online surveys compared to face-to-face modes, suggesting potential personality-driven selection biases. While these studies rely on cross-sectional or dual-mode data, our study extends this work by using a richer 44-item personality battery and panel data to examine how personality

traits predict actual nonresponse patterns over time in high-frequency online longitudinal surveys.

2.2.4 The Relevance of Panel Tenure in Predicting Nonresponse Patterns in High-frequency Online Longitudinal Studies

Panel tenure, defined as the length of time a respondent has been part of an online panel, is a crucial factor in understanding and predicting nonresponse patterns in online data collection contexts. Over time, respondents who have been part of a panel for longer periods may develop a stronger sense of commitment and familiarity with the survey process, which can lead to higher response rates and lower attrition rates. Likely, attrition is selective, whereby individuals who are more inclined to participate may be more likely to remain in the panel. By including personality traits—particularly conscientiousness—as predictors in our analysis, we are able to partially account for the heterogeneity in the propensity to respond to surveys, acknowledging that we cannot cleanly separate the roles of selectivity and panel tenure. In what follows, panel tenure therefore is taken to reflect both possible habituation in panel response and the difference in unobserved characteristics between respondents with different tenure lengths.

New panel respondents may not have yet developed the routine of regularly taking surveys in an online environment. Previous studies have shown that prior survey experiences are predictive of cooperation in subsequent surveys, indicating that the early survey experiences of fresh respondents are crucial for their continued participation (Frankel & Hillygus, 2014; McLauchlan & Schonlau, 2016). Positive initial experiences can foster a sense of engagement and commitment, leading to longer panel tenure and reducing the likelihood of nonresponse and attrition.

Table 2-2 shows annual attrition in the first and second membership years in the UAS panel for each nationally representative recruitment batch from 2014-2018. Although the average attrition rate across membership years and recruitment batches in the UAS is typically between 7-9%, Table 2-2 reveals that the largest attrition occurs in the first membership year across all recruitment batches, ranging between 14% and 25%. Attrition then declines precipitously in the second and subsequent membership years. For example, in the 2014 recruitment batch, the attrition rate was 17.20% in the first year but dropped to 8.99% in the second year. Similarly, the 2018 recruitment batch experienced an attrition rate of 24.84% in the first year, which decreased to 14.24% in the second year. This trend indicates that respondents who persist beyond the initial year are more likely to continue participating in subsequent years, highlighting the critical nature of the early phase in a panelist's tenure.

Table 2-2: Attrition by panel membership year by recruitment batch in the UAS panel

Recruitment batch	Year of recruitment	Attrition by panel membership year (%)	
		Year 1	Year 2
1	2014	17.20	8.99
2	2015	16.47	7.39
3	2016	13.99	7.68
4	2016	15.93	8.82
5	2016	18.36	7.50
6	2016	17.31	7.29
7	2016	18.22	10.13
8	2016	17.38	9.28
9	2016	22.99	11.48
10	2018	24.84	14.24
Average annual attrition across batches		**18.27**	**9.28**

Data source: https://uasdata.usc.edu/index.php

Prior research indicates that individuals who lack a strong commitment to survey participation are more likely to drop out quickly from online panels, often after participating in only a few initial survey waves or for a short duration (Lugtig, 2014). These fast attritors can

make up a substantial proportion of the panel sample, typically ranging from 20% to 30%. Most importantly, studies have found that individuals who drop out during the first membership year have distinct demographic, psychological, and psychosocial characteristics compared to those who leave in subsequent years (Hansson et al., 2018; Ward et al., 2017).

The pattern suggests that the most productive effort in maximizing retention may be early on in a respondent's tenure. By analyzing panel tenure effects, researchers can identify factors that contribute to quick dropout and apply these insights to newer panel members. Understanding the impact of panel tenure also can help inform the design of nonresponse adjustment weights. In our panel, respondents are invited to participate in a sequence of 22 core surveys, administered biennially, and are also invited to topical surveys such as the COVID-19 series. Although there may be slight variation in the number of ad hoc survey invitations, the overall exposure to survey requests is broadly similar across respondents. For this reason, we consider panel tenure (i.e., time in the panel) to be a reasonable proxy for survey experience. Adjustment based on panel tenure is particularly important in high-frequency online longitudinal studies because modest and gradual attrition among fresh respondents can lead to a biased sample over time if fresh and experienced respondents are fundamentally different in terms of substantive variables.

Despite the importance and relevance of panel tenure in predicting nonresponse patterns in online panel contexts, to the best of our knowledge, no prior studies have systematically analyzed its independent effects in the U.S. context. This study addresses that gap.

2.2.5 Analyzing the Moderating Effects of Survey Topics and Context in High-Frequency Online Longitudinal Studies

Leverage-salience theory, which incorporates both survey features and respondent characteristics, suggests that different survey features (e.g., survey topics) have varying effects

("leverages") on participation decisions based on respondents' socioeconomic characteristics, lifestyles, and preferences (Groves et al., 2000). The relevance and interest of survey topics in influencing survey participation decisions, as predicted by leverage-salience theory, have mainly been examined in cross-sectional studies (Groves et al., 2004; Ryu et al., 2006).

Understanding the role of survey topics and context in online longitudinal studies is important for several reasons. First, longitudinal data collection requires sustained participation over time, and varying survey topics may differentially impact respondents' willingness to continue participating. Second, as online panels become an increasingly popular method for data collection, identifying factors that influence long-term engagement and reducing nonresponse is essential for maintaining data quality and representativeness. The COVID-19 pandemic has provided a unique opportunity to examine how a highly topical and relevant issue may influence nonresponse patterns.

In this study, we analyze two novel datasets generated from two recently completed online longitudinal studies. These studies primarily differ in their survey topics and contextual factors. Specifically, one study focused on well-being topics both before and during the COVID-19 pandemic, while the other solely addressed COVID-19-related topics and was fielded when COVID-19 was declared a public health emergency in March 2020. The importance and relevance of the COVID-19 study were clearly emphasized to potential participants in the consent survey by including the following text:

> "We are launching a new survey series to help us track how U.S. residents are coping with the coronavirus pandemic. Participants will be asked to fill out a survey every two weeks. We are interested in learning how life changes or stays the same over time for U.S. residents and their family and friends. For that reason, we hope you will join the project so that we have input from every part of the country and from every walk of life.

> The survey will focus on what is happening in your life, and what you may be doing the same or differently now. We want to learn about the impacts of the pandemic on your family, your job or income, and how you are coping with the changes that have been and continue to be made to help contain the virus."

We leverage the timing and relevance of the COVID-19 study, along with the disproportionate effects the pandemic had, to conduct a comparative analysis of survey nonresponse patterns and their correlates between the two studies. On the one hand, we expect the COVID-19 survey to attract a large number of participants, as it was very topical during the pandemic. On the other hand, nonresponse rates in the COVID-19 survey may be higher among subgroups of respondents whose lives and livelihoods were disproportionately affected by the pandemic.

While we did not begin with strong directional hypotheses regarding potential moderating effects of panel topic, the COVID-19 context provided a unique opportunity to explore whether broader societal conditions and topic salience could moderate the relationship between individual characteristics and survey participation. Our analyses of topic effects are primarily exploratory, but the observed differences between the two studies suggest that contextual factors may meaningfully influence nonresponse patterns.

2.3 Methods

2.3.1 Data and Study Population

Data are from two recently completed online longitudinal studies conducted in the Understanding America Study (UAS). Established at the University of Southern California (USC) in 2014, the UAS is a nationally representative, probability-based internet panel of US-households of approximately 15,000 respondents aged 18 and above. Respondents answer surveys on a computer, tablet, or smart phone, wherever they are and whenever they wish to

participate. Surveys are conducted in English and Spanish. Panel members are recruited through address-based sampling. Prior access to internet is not a pre-requisite to be in the panel; respondents without prior internet access are provided with a computer tablet and broadband internet subscription. Respondents answer surveys once or twice a month covering various substantive topics. Partly as a result of this, the UAS comprises a vast amount of background information on its respondents, including extensive measures of physical and mental health, income, labor force participation, cognitive functioning, and demographics. For more information about the UAS cohort profile, see Kapteyn et al. (2024).

UAS recruitment procedures follow the Tailored Design Method (Dillman et al., 2014), starting with a postcard, next a short recruiting survey with $5 included and $15 promised for return of the recruiting questionnaire. At the end of the questionnaire, respondents are asked for their interest in participating in a longer run study. If interested, but without Internet, they are offered an Internet-enabled tablet. The process includes several reminders and an offer of additional incentives for consenting respondents to start the first Internet survey, if they have not done so within a given time period. Panel members are paid $20 for a thirty-minute survey. The UAS has instituted a "Sleeper Protocol", which involves several follow-ups with panel members who have not responded to a survey invitation during a specified number of months. The UAS website[6] provides complete information about UAS methodologies, including response and retention rates for every recruitment batch.

2.3.2 Understanding Coronavirus in America Study (UCA)

[6] https://uasdata.usc.edu/index.php?r=eNpLtDK0qi62MrFSKkhMT1WyLrYytFwwskuTcjKT9XJTSzLyU_Jz8tMrlaxrAShuDl0,

The first data source is from the Understanding Coronavirus in America (UCA) Study (Kapteyn et al., 2020). UCA commenced on March 10, 2020, immediately following the World Health Organization's declaration of COVID-19 as a global pandemic. All active members of the UAS panel were invited to participate in the UCA. About 7,000 UAS participants signed up for the study[7]. UCA participants received a survey every two weeks. The bi-weekly surveys elicited information on a broad array of socio-economic, behavioral, and health outcomes associated with the COVID-19 pandemic. A total of 28 biweekly survey waves were conducted, spanning the period March 2020 to July 2021. While the UCA continuously added new participants to the study after the first wave, our analysis specifically focuses on the 5,891 participants recruited in the first wave. This approach ensures that we possess complete response data for the entire duration of the study, from the initial wave to the study's conclusion across all 28 waves. The UCA website[8] provides additional information about the study, including download information on available micro data files and documentation.

2.3.3 The UAS Monthly Events Panel Surveys

The second data source is from the UAS monthly events surveys conducted among UAS respondents aged 50 and above. The data collection began in July 2019 and concluded in September 2023, spanning 52 waves. Each individual survey was administered at the beginning of the month, with respondents reflecting on their experiences in the preceding calendar month. These surveys collected information about various life events, including personal and health-related experiences such as health shocks, financial events, as well as personal encounters with

[7] The number of active UAS panel members was approximately 10,000 when the UCA study was launched in early 2020. The number has since grown to 15,000.

[8] https://uasdata.usc.edu/index.php?r=eNpLtDK0qi62MrFSKkhMT1WyLrYyNAayS5NyMpP1kvOL8vMSlaxrAeJXDC8,

sickness or the death of close relatives or friends during the previous month, and solicited self-assessments of their health, life satisfaction, pain levels, hours worked, earnings, employment changes, and medical expenditures during that period. Although the study continuously added new participants after the first wave, our analysis focuses on 3,597 participants recruited in Wave 1. This is to ensure that we have complete response data for the full period of the study across 52 survey waves. More details about the monthly events study including download information on available micro data files and documentation, are presented at the study website[9].

2.3.4 Analytical Approaches

2.3.4.1 Latent Class Analysis (LCA)

We employ LCA to analyze the response data. Specifically, LCA identifies subgroups of individuals (latent classes) by identifying patterns based on categorical latent variables. In survey research, LCA has been successfully used in various applications, including evaluation of survey questions (Kreuter et al., 2008), analysis of survey error (Biemer, 2011), investigation of satisficing behaviors (Kaminska et al., 2010), and nonresponse analysis in both cross-sectional (Montagni et al., 2019) and longitudinal data collections (Hu, Melipillán, et al., 2020; James, 2023; Lugtig, 2014). We chose LCA for this study because our goal was to classify respondents into meaningful subgroups based on their longitudinal response behavior—such as consistent respondents, early dropouts, or gradual attriters—an approach adopted by prior studies by Lugtig (2014), Hu et al. (2020), and others. Identifying these subgroups allows us to inform targeted retention strategies and explore mechanisms underlying panel attrition. LCA models can also be evaluated for model fit.

[9] https://uasdata.usc.edu/page/UAS+Monthly+Events+Panel

For each data set, we estimated a series of latent class models, varying the number of classes from 2 to 8. Lugtig (2014) shows that LCA models perform better than other latent class analysis methods, such as latent class growth analysis (LCGA) and growth mixture models (GMM), in fitting longitudinal response data. The categorical indicators used in the LCA model are equal to 1 for respondents to a survey wave and 0 otherwise. In the monthly events study data, there are 52 response indicators, and in the UCA study data, there are 28 response indicators, each indicating response/nonresponse to individual survey waves. Using these response indicators across panel survey waves, the latent class model identifies and groups respondents into several classes based on their similarity in nonresponse patterns.

For model selection, the conventional likelihood-ratio $\chi 2$ test with H1 as the unrestricted model is invalid and cannot be used because of sparse cells resulting in a distribution that is not Chi-2 distributed. Instead, we employ other model fit criteria to select the final latent class model as recommended by prior studies (Lugtig, 2014; Nylund et al., 2007). These criteria include:

1. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) (Schwarz, 1978), where lower values of AIC and BIC indicate better-fitting models.
2. Vuong-Lo-Mendell-Rubin adjusted Likelihood Ratio Test (LMRT) (Lo et al., 2001), which assesses whether the latent class model with C classes significantly improves the fit compared to the model with (C-1) classes, with a significant LMRT p-value suggesting a better fit for the model with C classes.
3. Bootstrapped Likelihood Ratio Test (BLRT) (Nylund et al., 2007), which indicates whether the model with (C-1) classes fits significantly worse than the model with C classes, using bootstrapping to estimate the difference in Log Likelihood values between the models, with a significant BLRT p-value indicating a worse fit for the model with (C-1) classes.

4. Entropy (Celeux & Soromenho, 1996), which measures the accuracy of class separation, with entropy values above 0.8 indicating accurate participant assignment to one class.

Finally, we consider both statistical fit and substantive results in selecting a final model. As a first step, we examine the aforementioned statistical fit measures. We then evaluate the substantive interpretability and usefulness of individual classes, as well as the practicality for designing interventions targeting specific groups of respondents based on the chosen classes. For instance, a model with many classes may fail to identify a critical group in need of targeted intervention or include a class that is too small for meaningful intervention. Moreover, with the addition of more classes, the maximum-likelihood (ML) model becomes increasingly unstable, leading to convergence issues in the estimation and making parsimonious models more attractive options. Additional information about model selection in the LCA framework is provided by Nylund et al. (2007) and https://www.statmodel.com/topic5.shtml.

2.3.4.2 Predicted Posterior Probabilities and Latent Class Membership

After fitting the LCA model with the chosen number of classes, we proceed to generate predicted posterior probabilities for each class for every respondent and assign class membership to the class with the highest predicted probability for each respondent. We present summary statistics of predicted posterior probabilities grouped by predicted class membership to initially assess the accuracy of class classification. Subsequently, we plot observed response probabilities by class membership across survey waves. Alternatively, we can also plot the predicted posterior probabilities. Both approaches however yield similar graphical representations of nonresponse patterns across survey waves.

2.3.4.3 Predictor Variables

Big Five Personality Traits. The Big Five personality traits—conscientiousness, openness, extraversion, agreeableness, and neuroticism—were measured using a 44-item standardized self-report questionnaire administered to respondents (Costa & McCrae, 1999; John et al., 1991; McCrae et al., 2004). The questionnaire for measuring Big Five personality traits consisted of 44 statements related to each of the five traits, with respondents asked to rate their agreement or disagreement on a Likert scale, typically ranging from "strongly disagree" to "strongly agree." For example, statements such as "I see myself as someone who is organized and detail-oriented" were used to assess conscientiousness, while items like "I see myself as someone who is open to new experiences" were used to measure openness. Some items were reverse-coded to ensure consistency in the direction of the scoring. The responses to these statements were then aggregated to create composite scores for each trait, with higher scores indicating stronger expressions of the corresponding trait. The total score ranges for each of the Big Five personality traits were as follows: agreeableness (0-45), conscientiousness (0-45), extraversion (0-40), neuroticism (0-40), and openness (0-50). The personality data was first collected in the initial substantive survey completed by respondents immediately after they completed the baseline demographic survey that is prerequisite for becoming formal panel members. In our analysis, we used data on personality traits from the latest available UAS wave before the start of the analysis period, as this information is collected every two years from the same respondents.

Panel Tenure Variable. Official membership in the UAS begins upon completion of the first demographic survey, also referred to as the My Household survey. To calculate panel tenure, we measured the duration in days between the date the Wave 1 survey was conducted in either of the two longitudinal studies and the date a respondent became an official member of the

UAS. For the monthly events and UCA studies, the first waves were fielded on June 3, 2019, and April 1, 2020, respectively. We then grouped respondents into two categories: those with panel tenure less than 1 year (fresh or new panelists) and those with panel tenure of 1 year or more (mature or experienced panelists). This classification enables a comparative analysis of respondent engagement based on their tenure within the panel. The one-year cutoff is meaningful and practical in terms of panel management practices and implementation of targeted interventions

Other Covariates. Other covariates used in the regression analyses include the following: (1) Hispanic ethnicity (yes, no); (2) Race (white only, black only, others); (3) Gender (male, female); (4) Age groups (18-44, 45-64, 65 and above in the UCA study; 50-64, 65 and above in the monthly event study); (5) Highest educational attainment (up to GED or high school, some college, college and above); (6) Household income (below $50k, $50-$75K, $75K and above); (7) Employment status (currently working, currently not working); (8) Household size (1, 2, 3 and above); (9) Self-reported health (rated on a scale from 1=poor to 5=excellent).

Several existing theories justify the use of demographic factors as predictors of survey participation. For example, the Opportunity Cost Hypothesis (Ajzen & Fishbein, 1980; Singer, 2011) explains that individuals weigh the costs and benefits of survey participation, making socio-demographic factors important in shaping their participation decisions. The Social Exchange Theory (Dillman, 1978; Dillman et al., 2014) highlights that the norm of reciprocity and trust in exchanges—often influenced by demographic characteristics such as education or socio-economic status—enhances willingness to participate. Additionally, demographic variables, such as education, can serve as proxies for cognitive burden, where individuals with higher educational attainment or higher socioeconomic status (SES) may find survey tasks less

burdensome and are thus more likely to complete them (Holbrook et al., 2007; Krosnick, 1991; Tourangeau, 2000). A recent study by Earp et al. (2022) provides additional empirical evidence showing that demographic variables serve as good proxies for respondents' subjective burden levels that predict survey responses in a national survey.

UAS updates respondents' demographic and household information every three months. All demographic and household information used in this analysis are from the first wave of data collection in the two respective studies. Data on health outcomes are from the latest available UAS wave as this information is collected every two years on the same respondents.

2.3.4.4 Multinomial Logistic Regression

We estimate Multinomial Logistic Regression models to predict class membership for each of the two studies. For each multinomial logistic regression estimated, we report adjusted relative-risk ratios (ARRR), 95% CIs, and the level of statistical significance. All reported p values were adjusted for multiple hypothesis tests using Holm's method (Holm, 1979). Additionally, in each regression, the class with the highest response propensities across all survey waves is used as the reference group. To account for uncertainty in class membership classification in the latent class models, one could use probabilities predicting class membership for the chosen class as weights in the regression analysis, which effectively assigns higher weights to chosen classes predicted with higher probabilities (Lugtig, 2014). We compared both unweighted and weighted regression results and they are similar. This is because all chosen classes in the latent class models were predicted with high probabilities and accuracy, as discussed in the Results section. We present and discuss unweighted regression results in the main text but include weighted results in Appendix Table A-1 and Appendix Table A-2 for completeness.

Additionally, for each survey wave, the UAS provides survey weights that correct for unequal probabilities of sampling UAS members and align the sample with the U.S. adult population in terms of gender, race/ethnicity, age, education, and geographic location (Census regions) (Understanding America Study, 2023). We decided not to use the UAS survey weights in our main analyses because the focus of this study is to understand attrition dynamics within the panel—specifically, how nonresponse evolves from the original sample over time. This would appear to be most relevant for other similar longitudinal surveys. Using survey weights would make the results most relevant for the unlikely case where one is able to recruit a sample with identical inclusion probabilities, i.e. where the likelihood that an invited panel member agrees to join the panel is identical across individuals..

2.3.4.5 Statistical Packages

All analyses were conducted using Mplus for Windows Version 8.10. Mplus was mainly utilized for estimating latent class models, generating posterior response probabilities, and classifying class membership (Muthén & Muthén, 1998-2017). Additionally, Stata for Windows Version 18.0 was used primarily for data management, estimating Multinomial Logistic Regression models for predicting class membership, and generating graphs (StataCorp, 2023).

2.4 Results

2.4.1 Sample Summary Statistics

Table 2-3 presents summary statistics for the two samples. We provide mean values for the Big-5 personality traits and self-reported health, and percentages for categorical variables.

With respect to Big-5 personality traits, there are no discernible differences between the two studies. Relative to the monthly events study, the UCA study attracted a significantly higher percentage of fresh panelists (24.26% in the UCA study vs. 13.23% in the monthly events study)

and Hispanics (15.99% in the UCA study vs. 6.14% in the monthly event sample). These differences can be explained by the growth of the UAS and the different time periods covered by the two studies. In the period before March 2020, the sample size of the UAS grew considerably, resulting in more panel members with shorter tenure at the start of the UCA study.

Racial and ethnic compositions are comparable between the two studies. Participation among females was higher in the UCA study compared to the monthly events survey (57.82% in the UCA study vs. 52.32% in the monthly events study). Both studies are comparable in terms of highest educational attainment and household income. The UCA sample had a higher percentage of working individuals (56.13% in the UCA study vs. 41.90% in the monthly events sample), reflecting the age differences between the two studies. We observe differences in household composition between the two samples, with 2-person households accounting for half of the monthly events sample and 3-person households and above making up approximately half of the UCA sample; this difference mainly reflects age differences between the two samples. The mean values for self-reported health are similar between the two samples.

Table 2-3: Sample summary statistics by study sample

	Monthly events study sample (n=3,597)	COVID-19 sample (n=5,891)
Big-5 personality traits		
Conscientiousness	36.25	35.35
Openness	35.82	35.39
Extroversion	25.64	25.17
Neuroticism	20.52	21.84
Agreeableness	36.27	35.30
Panel tenure		
Less than 1 year	13.23%	24.26%
1 year and above	86.77%	75.74%
Hispanic		
Yes	6.14%	15.99%
Race & Ethnicity		
White only	83.37%	78.83%
Black only	7.47%	8.15%
Others	9.16%	13.02%

Gender		
Female	52.32%	57.82%
Age group		
18-44	NA	39.02%
45-64	NA	38.02%
50-64	59.55%	NA
65 and above	40.45%	22.95%
Education		
GED or High School	23.52%	21.53%
Some College	38.89%	37.11%
College and above	37.59%	41.36%
HH income		
Below $50K	42.44%	40.68%
$50-$75K	20.38%	19.84%
$75K and above	37.18%	39.48%
Employment status		
Currently working	41.90%	56.13%
Currently not working	58.10%	43.87%
HH size		
1	22.85%	17.70%
2	51.32%	39.90%
3+	25.83%	42.40%
Health status		
Self-reported health	2.82	2.71

2.4.2 Latent Class Model Selection

In Table 2-4, model fit statistics are presented by the number of latent classes separately for the two study samples. Across both samples, the log-likelihood values, AIC, and BIC consistently improve with an increase in the number of classes. However, in the monthly events study sample, the LMRT and BLRT statistics yield conflicting results. The LMRT suggests that a model with 4 classes best fits the data, while the BLRT indicates that a model with 8 classes performs better. Entropy values exceed 0.8 for all latent class models in both samples, indicating high accuracy in class membership assignment. Taken together, these fit statistics suggest that the models with more classes fit the data better. However, considering the substantive interpretability and practicality of implementing targeted interventions, we select latent class models with 5 classes in the monthly events sample and 6 classes in the UCA sample as our final models, highlighted in bold text. Additionally, for sensitivity analysis and completeness, we

provide in Appendix Figure A-1 through Appendix Figure A-4 showing observed attrition patterns and in Appendix Table A-3 through Appendix Table A-6 presenting multinomial logistic regression results predicting class membership: for the latent class models that are one class smaller and one class larger than the chosen number of latent classes for each of the two study samples, leading to similar substantive conclusions.

Table 2-4: Model fit statistics by number of response pattern classes

Monthly events panel study (n=3,597)

# of classes	DF [a]	LL [b]	AIC [c]	BIC [d]	Entropy [e]	LMRT p value [f]	BLRT p value [g]	Class size (%)	
								Min	Max
2	105	-55856	111923	112573	0.99	0.333	0.000	30%	70%
3	158	-46916	94148	95126	0.99	0.000	0.000	20%	60%
4	211	-43370	87162	88468	0.98	0.000	0.000	12%	55%
5✓	**264**	**-41811**	**84150**	**85784**	**0.98**	**0.637**	**0.000**	**11%**	**54%**
6	317	-40684	82002	83963	0.98	0.204	0.000	6%	54%
7	370	-39806	80352	82641	0.97	0.567	0.000	5%	52%
8	423	-39217	79281	81899	0.97	0.206	0.000	5%	51%

COVID-19 panel study (n=5,891)

# of classes	DF	LL	AIC	BIC	Entropy	LMRT p value	BLRT p value	Class size (%)	
								Min	Max
2	57	-43924	87962	88343	0.98	0.000	0.000	22%	78%
3	86	-38933	78039	78614	0.96	0.000	0.000	11%	68%
4	115	-37156	74542	75311	0.95	0.000	0.000	6%	64%
5	144	-36573	73434	74396	0.94	0.007	0.000	6%	65%
6✓	**173**	**-35988**	**72323**	**73479**	**0.93**	**0.000**	**0.000**	**4%**	**62%**
7	202	-35604	71613	72962	0.93	0.000	0.000	5%	61%
8	231	-35365	71193	72736	0.92	0.008	0.000	3%	61%

[a] DF=Degrees of freedom; [b] LL=Log-likelihood; [c] Akaike Information Criterion (AIC)=−2*LL + 2*DF; [d] Bayesian Information Criterion (BIC)=−2*LL + DF ∗ Ln(N). For LL, larger values indicate better fitting models. In contrast, for both AIC and BIC, lower values suggest better fitting models. [e] Entropy indicates the extent to which classes can be separated accurately and any entropy values above 0.8 indicates that participants can be accurately assigned to one class; Entropy=$1 - \frac{\sum_i \sum_k (-\hat{p}_{ik} \ln \hat{p}_{ik})}{n \ln k}$, where $\hat{p}_{ik}$ = posterior class probabilities; k = number of classes; n = sample size; an entropy value that is close to 1 suggests good classification because the posterior class probabilities are either close to 1 or 0. [f] LMRT= Lo-Mendell-Rubin adjusted LRT test, which indicates whether the LCA model with C classes is a significant improvement when compared with the previous LCA model with (C-1) classes. [g] BLRT= Bootstrapped LRT test, similar to LMRT, which indicates whether the LCA model with (C-1) classes fit significantly worse, but now uses bootstrapping to estimate the size of the difference in Log likelihood values between the models with C classes and (C-1) classes. LCA models in bold are selected for predicting class membership based on a combination of their model fit statistics as well as their meaningful substantive interpretations.

2.4.3 Class Membership Prediction

Table 2-5 presents the mean posterior probabilities categorized by predicted class membership. Rows represent class membership, and columns represent the corresponding predicted probabilities. For instance, for the first row in the monthly events data, conditional on being in class 1, the predicted probability of being in classes 1,2,3,4, and 5 are 0.958, 0.000, 0.002, 0.013, and 0.027, respectively. Diagonal values close to 1, and off-diagonal values close to 0, indicate a high level of accuracy in classification. For instance, the mean predicted probabilities for being in classes 1, 2, 3, 4, and 5 are 0.958, 0.988, 0.979, 0.957, and 0.991, respectively, in the monthly events study sample. The mean predicted posterior probabilities across all assigned classes are 0.983 (SD=0.071, Min=0.48, Max=1) for the monthly events study sample and 0.954 (SD=0.107, Min=0.352, Max=1) for the UCA study sample. Overall, the results indicate that classes are predicted with high accuracy, which is consistent with the high entropy values presented in Table 2-4**.**

Table 2-5: Classification matrix

5-class LCA model (Monthly events study)

	Predicted posterior probabilities				
Latent Class	Class 1	Class 2	Class 3	Class 4	Class 5
1	**0.958**	0.000	0.002	0.013	0.027
2	0.000	**0.988**	0.012	0.000	0.000
3	0.002	0.010	**0.979**	0.010	0.000
4	0.031	0.000	0.011	**0.957**	0.000
5	0.009	0.000	0.000	0.000	**0.991**

6-class LCA model (COVID-19 study)

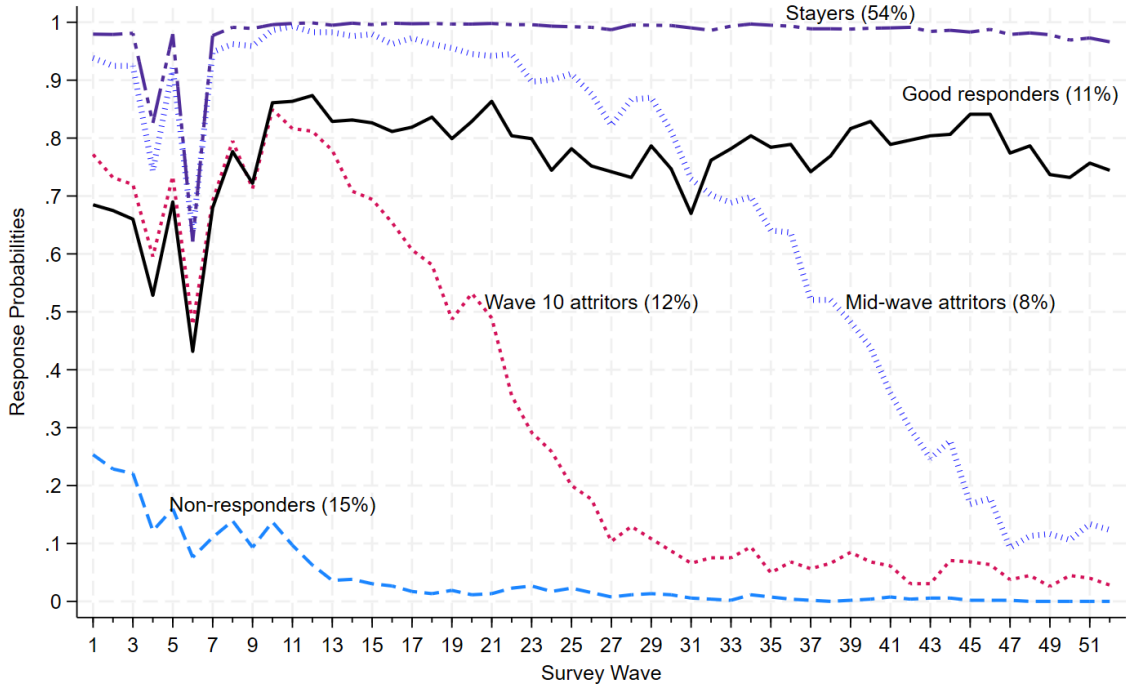
	Predicted posterior probabilities					
Latent Class	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
1	**0.950**	0.000	0.039	0.006	0.000	0.005
2	0.000	**0.876**	0.019	0.000	0.043	0.062
3	0.047	0.015	**0.886**	0.000	0.000	0.052
4	0.008	0.000	0.000	**0.992**	0.000	0.000

5	0.000	0.007	0.000	0.000	**0.982**	0.011
6	0.002	0.041	0.038	0.000	0.037	**0.881**

2.4.4 Response Patterns in the Monthly Events Study

Figure 2-1 shows observed posterior response probabilities by class membership across survey waves for the monthly events study. Notably, there were sudden drops in response rates in waves 4 and 6 that were mainly due to a glitch in the case management system, which resulted in failure to send out reminders. This issue was addressed in subsequent waves.

Figure 2-1: Posterior response probabilities for the 5-class LCA model. Monthly events study (n=3,597)



The first group of respondents consists of individuals with response propensities close to 1 across all 52 monthly survey waves. We refer to this group as the "Stayers" group, which comprises approximately 54% of the sample. Participants in this group completed almost all of the 52 monthly survey waves. Similar to the "Stayers" group, the "Good Responders" group, constituting about 11% of the sample, remained until the end of the study but missed a few surveys, with an average response rate of approximately 0.8 across all survey waves.

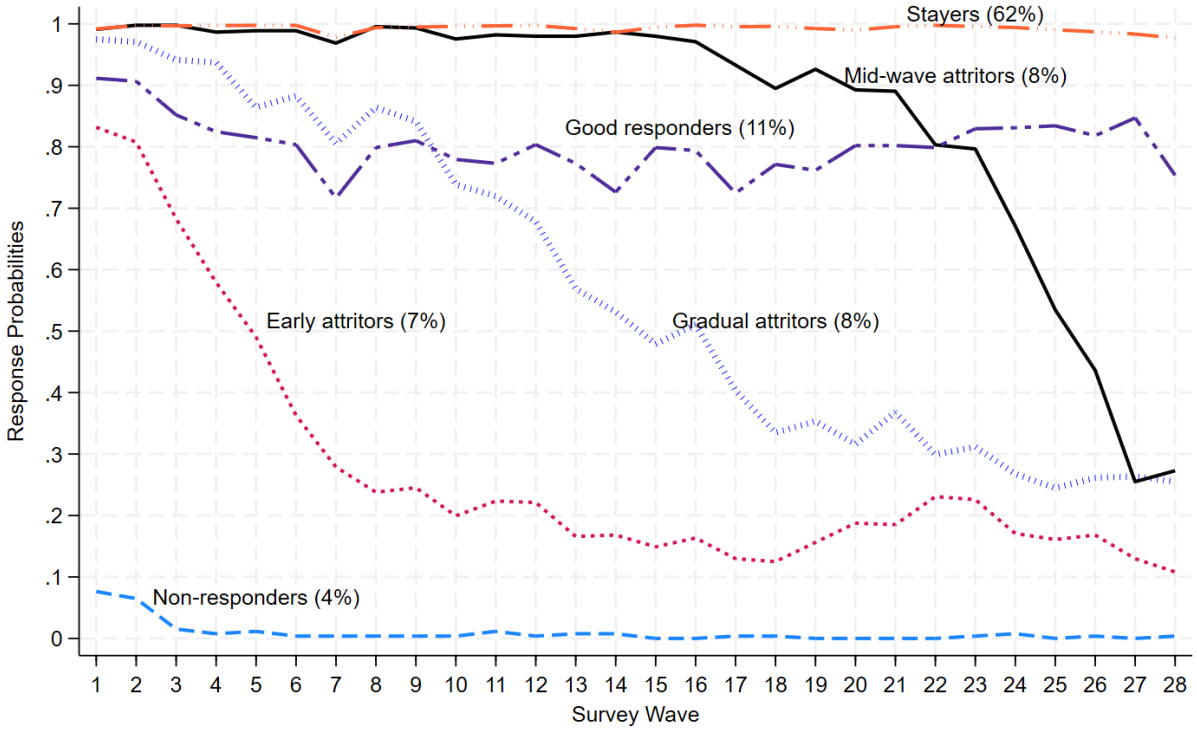
The “Mid-wave Attritors” group, constituting 8% of the sample, exhibits a response pattern similar to the “Stayers” group until Wave 15, after which nonresponse gradually increased in the Mid-waves, and they did not return. Similar to the “Mid-wave Attritors” group are the “Wave-10 Attritors” group, who make up a significant 12% of the sample and began to leave earlier than the “Mid-wave Attritors” in Wave 10, dropping linearly from an average response rate of about 85% in Wave 8 to about 5% in Wave 31 and then taper off until the end of the study period.

At the lower end of the nonresponse spectrum are a group of respondents with a response probability close to zero right from the beginning. We refer to this group as the “Non-responders” group, making up about 15% of the sample. In Appendix Figure A-1 and Appendix Figure A-2, we present nonresponse patterns for the two latent class models: one with a reduced class (i.e., 4 classes) and one with an additional class (i.e., 6 classes) compared to the selected class for comparison and completeness. These nonresponse patterns demonstrate similarities with the patterns in the chosen class, particularly among the “Stayers” and “Non-responders” groups.

2.4.5 Response Patterns in the UCA Study

Figure 2-2 presents observed posterior response probabilities across survey waves by class membership for the UCA study.

Figure 2-2: Posterior response probabilities for the 6-class LCA model. COVID-19 study (n=5,891)



There are both similarities and differences in nonresponse patterns and their sample compositions between the two studies. For instance, both studies have a "Stayers" group but their share in the UCA sample is larger than they are in the monthly events sample (62% in the UCA study vs. 54% in the monthly events study). The "Mid-wave Attritors" group in the UCA sample differs from the "Mid-wave Attritors" group in the monthly events sample. Mid-wave leaving respondents in the UCA study responded to all surveys until mid-wave (i.e., Wave 14) and then began to drop out slowly until wave 23, reaching a response rate of about 0.8, with a steep decline afterward until the end of the study period. Their sample share, however, is the same in both samples, accounting for 8% of respondents. The response pattern for "Good Responders" and their sample share (11%) is similar to a large extent in both studies.

The "Gradual Attritors" group, making up 8% of the UCA sample, completed the first survey with a response rate close to 1 and then began to leave gradually and linearly after Wave 1 until the end of the study, eventually dropping to a response rate of 25%. Notably, the response

patterns for the "Gradual Attritors" group in the UCA study is similar to that for the "Mid-wave Attritors" group in the monthly events study but they differ in the timing where the gradual attrition begins (Wave 2 in the UCA study vs. Wave 23 in the monthly events study).

The "Early Attritors" group, making up 7% of the UCA study respondents, completed the first two surveys with a response rate of 0.8. Then, their response rate began to decline steeply until Wave 8, where it reached 0.2, tapering off afterward until the end of the study period. Their response pattern is similar to the Wave-10 attrition pattern found in the monthly events study but differs mainly in the timing of the steep decline (Wave 2 in the UCA study vs. Wave 8 in the monthly events study).

Looking at the lowest end of the nonresponse spectrum, the "Non-responders" group with a response rate close to 0 across all waves make up a smaller share of the sample in the UCA study compared to the monthly events study (4% in the UCA study vs. 15% in the monthly events study). For comparison, in Appendix Figure A-3 and Appendix Figure A-4, we present response patterns for the latent class model with a reduced class (i.e., 5 classes) and the latent class model with an additional class (i.e., 7 classes), compared to the selected model. These patterns demonstrate similarities with the chosen latent class model.

2.4.6 Multinomial Logistic Regression Results

Table 2-6 and Table 2-7 present the results of multinomial regression models predicting membership classes characterized by various nonresponse patterns in the monthly events and UCA datasets, respectively. As previously mentioned, the reported coefficients are adjusted relative-risk ratios (ARRR), which measures the proportionate change in the relative risk of belonging to a specific class, compared to the base or reference class, when the predictor variable changes by one unit, while controlling for other predictor variables. For example, an ARRR of

0.964 for Conscientiousness in predicting membership in the Non-respondent class (relative to the Stayer class) indicates that, controlling for other variables, each one-unit increase in the Conscientiousness score is associated with an approximately 3.6% decrease in the relative risk of being a Non-respondent (1 − 0.964 = 0.036). The adjective "adjusted" simply means that the estimates are from a multivariate analysis. While there are several options for selecting the reference group and pairwise comparisons, we chose the class with the highest response probabilities across all survey waves—referred to as the 'Stayers'—as the reference category and focused on comparisons relative to the Stayers. This choice facilitates comparison with recent work by Lugtig (2014), Nicole (2023), and Hu et al. (2020), which also used the most stable or fully responsive class as the reference group when modeling class membership.

Below, we highlight key differences in the estimated coefficients between the two studies. Significance levels have been adjusted for multiple hypotheses testing for each latent class using Holm's method. By considering every estimated coefficient as representing a separate hypothesis we are conservative. However, as we are not just interested in the overall effect of for example personality or demographics, but rathe in their individual effects, this seems appropriate. The 95% confidence intervals have not been adjusted. The correction for multiple hypotheses testing renders very few of the estimated coefficients significantly different from one in Table 2-6 except for personality traits (the monthly events survey). In Table 2-7 (the UCA study), more coefficients, including some demographic variables, differ significantly from one. In the Discussion and Conclusion section, we provide additional contexts and detailed explanations about these differences.

Table 2-6: Multinomial logistic regressions predicting class membership in the *monthly events panel study*. The reference group is "Stayers". Reported coefficients are adjusted relative-risk ratios (ARRR), which measure the proportionate change in the relative risk of belonging to a certain class, compared to the "stayers" class, when the predictor variable changes by one unit, while controlling for other predictor variables.

	Non-responders	Wave 10 attritors	Mid-wave attritors	Good responders	Stayers
Big-5 Personality Traits					
Conscientiousness Score	0.964**	0.978	0.961**	0.975	
	[0.943,0.986]	[0.956,1.000]	[0.936,0.986]	[0.952,0.997]	
Openness Score	1.019	1.019	1.002	1.013	
	[1.000,1.039]	[1.000,1.038]	[0.981,1.024]	[0.994,1.032]	
Extroversion Score	1.042***	1.010	1.012	1.010	Ref.
	[1.022,1.062]	[0.991,1.029]	[0.990,1.035]	[0.991,1.029]	
Neuroticism Score	1.029	1.034**	1.000	1.016	
	[1.008,1.050]	[1.013,1.055]	[0.976,1.024]	[0.994,1.037]	
Agreeableness Score	1.021	1.037**	1.018	1.032	
	[0.997,1.044]	[1.013,1.061]	[0.991,1.046]	[1.008,1.057]	
Panel Tenure					
(Ref: Less than 1 year)					
1 year and above	0.687	0.933	0.710	0.834	Ref.
	[0.502,0.941]	[0.661,1.316]	[0.488,1.031]	[0.594,1.173]	
Hispanic					
(Ref: No)					
Yes	1.554	1.014	0.475	1.096	Ref.
	[1.037,2.328]	[0.640,1.609]	[0.226,0.995]	[0.703,1.710]	
Race & Ethnicity					
(Ref: White only)					
Black only	1.138	1.155	1.267	1.403	
	[0.741,1.745]	[0.764,1.745]	[0.801,2.004]	[0.941,2.092]	Ref.
Others	1.167	0.94	0.871	1.387	
	[0.812,1.677]	[0.637,1.387]	[0.543,1.396]	[0.974,1.974]	
Gender					
(Ref: Female)					
Male	1.227	1.187	1.068	0.933	Ref.
	[0.977,1.542]	[0.945,1.491]	[0.819,1.394]	[0.739,1.179]	
Age Group					
(Ref: 50-64)					
65 and above	1.136	0.913	1.291	0.769	Ref.
	[0.879,1.468]	[0.706,1.181]	[0.960,1.736]	[0.586,1.008]	
Education					
(Ref: GED or high school)					
Some College	0.941	0.913	0.951	0.908	
	[0.712,1.244]	[0.691,1.207]	[0.686,1.316]	[0.683,1.207]	Ref.
College and above	0.684	0.816	0.843	0.766	
	[0.496,0.942]	[0.596,1.119]	[0.583,1.218]	[0.554,1.058]	
HH Income					
(Ref: Below $50K)					
$50-$75K	0.968	1.001	1.045	0.751	
	[0.715,1.311]	[0.745,1.346]	[0.742,1.473]	[0.544,1.037]	Ref.
$75K and above	0.943	0.824	0.852	0.92	
	[0.705,1.262]	[0.615,1.106]	[0.602,1.205]	[0.688,1.232]	
Employment Status					
(Ref: Currently working)					

Currently not working	0.927	0.919	0.871	0.681*	Ref.
	[0.716,1.199]	[0.712,1.187]	[0.642,1.182]	[0.526,0.882]	
Household Size (Ref: 1)					
2	0.826	0.972	0.86	1.049	Ref.
	[0.624,1.094]	[0.729,1.294]	[0.620,1.194]	[0.784,1.403]	
3 and above	1.037	1.248	1.168	1.045	
	[0.753,1.430]	[0.906,1.721]	[0.806,1.691]	[0.750,1.455]	
Health Status					
	0.98	1.135	1.327***	1.122	Ref.
	[0.864,1.111]	[1.003,1.284]	[1.150,1.530]	[0.987,1.275]	
Constant	0.061***	0.022***	0.151**	0.080***	
	[0.013,0.284]	[0.005,0.102]	[0.026,0.879]	[0.017,0.380]	
n	3446				
Pseudo R^2	0.022				

Exponentiated coefficients (i.e., adjusted relative-risk ratios); 95% confidence intervals in brackets; * p < 0.10, ** p < 0.05, *** p < 0.01. All p values were adjusted for multiple hypothesis tests using Holm's method (Holm, 1979). Note the 95% CIs were not adjusted for multiple hypothesis tests.

Table 2-7: Multinomial logistic regressions predicting class membership in the *COVID-19 panel study*. The reference group is "Stayers". Reported coefficients are adjusted relative-risk ratios (ARRR), which measure the proportionate change in the relative risk of belonging to a certain class, compared to the "stayers" class, when the predictor variable changes by one unit, while controlling for other predictor variables.

	Non-responders	Early attritors	Gradual attritors	Mid-wave attritors	Good responders	Stayers
Big-5 Personality Traits						
Conscientiousness Score	0.977	0.962**	0.953***	0.982	0.956***	
	[0.951,1.004]	[0.941,0.984]	[0.934,0.972]	[0.962,1.003]	[0.939,0.973]	
Openness Score	1.030	1.043***	1.027*	1.029**	1.008	
	[1.006,1.055]	[1.023,1.064]	[1.009,1.045]	[1.011,1.047]	[0.992,1.024]	
Extroversion Score	0.998	1.035***	1.014	1.007	1.012	Ref.
	[0.975,1.021]	[1.017,1.054]	[0.997,1.032]	[0.990,1.025]	[0.997,1.027]	
Neuroticism Score	1.013	1.028*	1.018	1.015	1.020	
	[0.987,1.039]	[1.008,1.049]	[0.998,1.037]	[0.996,1.035]	[1.003,1.038]	
Agreeableness Score	1.000	1.024	1.006	0.995	1.032**	
	[0.973,1.028]	[1.002,1.048]	[0.985,1.027]	[0.975,1.016]	[1.013,1.051]	
Panel Tenure						
(Ref: Less than 1 year)						
1 year and above	4.848***	0.888	0.999	1.400	0.770	Ref.
	[3.071,7.653]	[0.694,1.135]	[0.789,1.266]	[1.075,1.824]	[0.629,0.944]	
Hispanic						
(Ref: No)						
Yes	4.718***	1.306	1.229	0.99	1.697***	Ref.
	[3.412,6.523]	[0.978,1.745]	[0.936,1.614]	[0.726,1.349]	[1.341,2.148]	
Race & Ethnicity						
(Ref: White only)						
Black only	1.670	1.639*	0.918	0.925	1.495	
	[1.021,2.730]	[1.151,2.333]	[0.619,1.361]	[0.625,1.370]	[1.094,2.044]	Ref.
Others	1.377	0.947	0.845	0.83	1.091	
	[0.952,1.992]	[0.688,1.303]	[0.624,1.143]	[0.601,1.147]	[0.845,1.408]	
Gender						
(Ref: Female)						
Male	0.696	1.203	1.014	0.946	1.064	Ref.
	[0.517,0.937]	[0.955,1.515]	[0.817,1.260]	[0.760,1.176]	[0.878,1.288]	
Age Group						
(Ref: 18-44)						
45-64	0.590**	0.346***	0.359***	0.545***	0.525***	Ref.

	[0.424,0.821]	[0.266,0.450]	[0.282,0.458]	[0.428,0.693]	[0.424,0.649]	
65+	0.75	0.260***	0.224***	0.378***	0.436***	
	[0.482,1.167]	[0.173,0.390]	[0.151,0.330]	[0.266,0.537]	[0.319,0.596]	
Education						
(Ref: GED or high school)						
Some College	0.736	1.063	0.691*	0.972	1.00	
	[0.516,1.050]	[0.794,1.423]	[0.531,0.898]	[0.743,1.272]	[0.788,1.268]	Ref.
College and above	0.702	0.724	0.567***	0.662	0.759	
	[0.474,1.040]	[0.522,1.004]	[0.424,0.759]	[0.489,0.897]	[0.582,0.990]	
HH Income						
(Ref: Below $50K)						
$50-$75K	0.896	0.887	0.826	1.058	1.08	
	[0.600,1.338]	[0.655,1.200]	[0.616,1.106]	[0.796,1.406]	[0.844,1.381]	Ref.
$75K and above	1.158	0.705	0.839	0.93	0.879	
	[0.807,1.661]	[0.527,0.943]	[0.640,1.098]	[0.706,1.223]	[0.691,1.118]	
Employment Status						
(Ref: Currently working)						
Currently not working	1.176	0.680*	0.785	0.967	0.788	Ref.
	[0.856,1.613]	[0.523,0.883]	[0.617,0.999]	[0.760,1.228]	[0.637,0.975]	
Household Size						
(Ref: 1)						
2	0.82	0.973	1.152	1.023	1.279	
	[0.552,1.219]	[0.700,1.352]	[0.832,1.595]	[0.757,1.383]	[0.967,1.691]	Ref.
3 and above	1.073	1.271	1.504	1.187	1.481*	
	[0.718,1.604]	[0.919,1.759]	[1.091,2.073]	[0.871,1.618]	[1.115,1.965]	
Health Status						
Self-report of health	0.908	0.931	0.978	1.059	1.001	Ref.
	[0.777,1.061]	[0.821,1.056]	[0.870,1.100]	[0.942,1.191]	[0.903,1.110]	
Constant	0.017***	0.025***	0.293*	0.073***	0.147***	
	[0.003,0.110]	[0.006,0.110]	[0.075,1.150]	[0.018,0.293]	[0.044,0.496]	
n			5743			
Pseudo R^2			0.062			

Exponentiated coefficients (i.e., adjusted relative-risk ratios); 95% confidence intervals in brackets; * p < 0.10, ** p < 0.05, *** p < 0.01. All p values were adjusted for multiple hypothesis tests using Holm's method (Holm, 1979). Note the 95% CIs were not adjusted for multiple hypothesis tests.

2.4.6.1 Differences by Respondent Personality Traits

Our results in both Table 2-6 and Table 2-7 show consistent effects of conscientiousness in predicting nonresponse patterns. In the monthly events study, the regression results in Table 2-6 reveal that respondents with higher conscientiousness scores are less likely to be in the ”Non-responders” and “Mid-wave Attritors” classes and are more likely to be in the “Stayers” class. These differences are statistically significant even after adjusting for multiple hypotheses testing. Although the estimated coefficients for conscientiousness are less than one in the regression models predicting membership in the “Wave 10 Attritors” and “Good Responders” attrition classes in the monthly events study, they are not statistically significant after applying adjustments for multiple hypotheses tests.

Similarly, in the UCA study, results from Table 2-7 show that respondents with higher conscientiousness scores are more likely to be associated with the “Stayers” class than with the “Early Attritors,” “Gradual Attritors,” and “Good Responders” classes. These differences remain significant even after adjustments for multiple hypothesis tests. For the remaining attrition classes in the UCA study, the estimated coefficients for the conscientiousness trait are less than one but are not statistically significant.

It is worth mentioning that without adjustments for multiple hypotheses tests, as done in prior studies such as Lugtig (2014) and Cheng et al. (2020), all pairwise comparisons between each of the attrition classes and the “Stayers” class would have reached statistical significance levels in both study samples. Despite the more stringent and conservative approach we used in handling the estimated regression coefficients, our results still emphasize the importance, robustness, and significance of conscientiousness in predicting nonresponse patterns in high-frequency online longitudinal studies, while minimizing the risk of false conclusions.

In contrast to conscientiousness, individuals with high openness scores are more likely to be associated with the attrition classes than with the "Stayers" class in both studies. This difference is more pronounced in the UCA sample than in the monthly events data, reaching significance levels in the multinomial logistic regression models predicting membership in the "Early Attritors", "Gradual Attritions", and "Mid-wave Attritors" classes. Without adjustments for multiple hypothesis tests, the effects of openness would have been statistically significant for the contrasts between each of the attrition classes and the "Stayers" class in both studies. However, as previously mentioned, the adjusted results provide a more robust conclusion.

Similar to openness, more extroverted, neurotic, and agreeable respondents are more likely to be associated with each of the attrition classes than with the "Stayers" class, although some of the differences are not statistically significant. In particular, more extroverted respondents are significantly more likely to be in the "Non-responders" class in the monthly events study and the "Early Attritors" class in the UCA study than in the "Stayers" class. Similarly, respondents with high neuroticism scores are significantly more likely to belong to the "Wave 10 Attritors" class in the monthly events study and the "Early Attritors" class in the UCA study than to the "Stayers" class. More agreeable respondents are significantly more likely to be associated with the "Wave 10 Attritors" class in the monthly events study and the "Good Responders" class in the UCA study than with the "Stayers" class.

Taken together, our results underscore the robustness, importance, and significance of respondent personality traits, particularly conscientiousness, in predicting nonresponse patterns in high-frequency online longitudinal studies. The relationships we found are consistent with our previously stated hypotheses, as well as with the general findings in the medication adherence literature.

2.4.6.2 Differences by Panel Tenure Status

The results in Table 2-6 and Table 2-7 show striking differences in the panel tenure effect between the two studies. The monthly events attrition regression results in Table 2-6 show no significant differences between each of the attrition classes and the "Stayers" class by panel tenure status. In contrast, the results in Table 2-7 reveals that, compared to fresh respondents, experienced panelists are significantly more likely to be in the "Non-responders" class than in the "Stayers" class (ARRR= 4.848, p-value<0.01). There are no significant differences between each of the remaining attrition classes and the "Stayer" class in the UCA study, indicating that experienced respondents who decided to leave the UCA study did so from the beginning.

2.4.6.3 Differences Among Demographic Subgroups

Hispanics. The regression results in Table 2-6 indicate no significant difference between Hispanic and non-Hispanic individuals across all attrition groups in the monthly events study. However, the results from Table 2-7 reveal a different pattern, indicating that Hispanics are more likely than non-Hispanic individuals to be associated with either the "Non-Responders" class (ARRP=4.718, p-value<0.01) or the "Good Responders" class (ARRR=1.697, p-value<0.001), rather than the "Stayers" class.

Race. In the monthly events study, we did not find significant differences in nonresponse patterns by respondent race. However, in the UCA study, compared to Whites, Blacks had a higher relative risk of being categorized as "Early Attritors" (ARRR=1.639, p-value<0.1) than as "Stayers".

Gender. No significance differences by respondent gender are found in either of the two studies.

Age Groups. In the monthly events study, age has no significant impact. Note however that the age range is restricted to 50+, a plausible explanation for the absence of age effects. In

the UCA study, we observe a large, significant, and negative association between age and nonresponse patterns, with younger participants displaying significantly higher dropout rates compared to their older counterparts across all attrition classes and these differences widen with age.

Education. In the monthly events study, we find no association between education and membership of any of the attrition classes. In contrast, in the UCA study, we found that higher education is associated with a lower likelihood of belonging to the gradual attritors class.

Household Income, Household Size, and Employment Status. Household income has no effect on nonresponse patterns in either study after controlling for other factors. Household size and employment status seem to better predict nonresponse patterns in the UCA study than in the monthly events study. In the UCA study, respondents from households with at least three people (relative to single-person households) and respondents who were employed (relative to those not currently working) are more likely to be associated with the attrition classes rather than the "Stayers" class. However, some of the estimated coefficients do not reach the significance level after applying adjustments for multiple hypothesis tests.

2.4.6.4 Differences by Health Status

In the monthly events study, health is a significant predictor among "Mid-wave Attritors" group (ARRR=1.327, p-value<0.01) (Note that the rating scale for the self-reported health variable ranges from 1=Excellent to 5=Poor). Our analysis of the interaction effects between age and health status show that the health effects are largest among older respondents aged 65 and above (results not shown). This finding aligns with existing literature, which suggests that individuals, particularly in the older age group, with poor health often prioritize healthcare needs

over survey participation, leading to higher dropout rates (Gerry & Papadopoulos, 2015; Radler & Ryff, 2010; Uhrig, 2008; Watson & Wooden, 2009).

However, the UCA study presents a contrasting picture, as no discernible association between health status and any of the attrition classes is observed. Additionally, we did not find any interaction effects between age and health status in the UCA study (results not shown). In our Discussion and Conclusion section, we suggest a substantive explanation for the lack of the health effect in the UCA study.

2.5 Discussion and Conclusion

In this study, we examined nonresponse patterns and their correlates in two recently completed high-frequency online longitudinal studies that differed primarily in terms of survey topics, frequency, relevance, and target age groups (50+ or 18+). Our findings provide suggestive evidence that design differences between offline and online panels may differentially influence nonresponse and attrition outcomes. Specifically, we found that respondent personality traits play a more significant role, while conventional demographic and household variables, which are commonly used in offline panel studies, were often insignificant. By analyzing nonresponse data from longitudinal studies where the same respondents were repeatedly interviewed using essentially the same questionnaire, we better isolated the effects of personality traits from those of changing demographic characteristics and survey experiences. Leveraging unique datasets from two studies that primarily differ in topic relevance and interest, as an exploratory analysis, we assessed the COVID-19-related context as a potential factor moderating the relationship between individual regression coefficients and nonresponse patterns. Below, we discuss the significance and implications of our findings, provide suggestions for future research directions, and highlight the strengths and weaknesses of our study approach.

2.5.1 Significance of Respondent Personality Traits in Predicting Nonresponse Patterns in High-Frequency Online Longitudinal Data Collections

Respondents high in conscientiousness exhibit qualities such as diligence, perseverance, tenacity, orderliness, and industriousness, making them more likely to sustain the level of commitment required to remain engaged in intensive online longitudinal surveys. Indeed, we found in both studies that more conscientious respondents are most likely to be "Stayers" and less likely to belong to any of the other attrition classes, which is consistent with prior research by Lugtig (2014) and Cheng et al. (2020). However, our effect sizes for conscientiousness are larger and stronger, particularly given that our results account for multiple hypothesis tests, which previous studies did not. This finding aligns with our previously stated expectations and further supports Lugtig (2014)'s survey commitment explanation as an underlying mechanism. Our finding is also consistent with the literature on medication adherence, which shows the consistent and robust effect of conscientiousness in predicting medication adherence across various disease conditions, population subgroups, and cultural settings.

In contrast to conscientiousness, we found that individuals with high openness scores are more likely to be associated with the attrition classes than with the "Stayers" class, with this difference being more pronounced in the UCA sample than in the monthly events data. This finding is consistent with Cheng et al. (2020) but less so with Lugtig (2014), where the coefficients for the openness reached significance in only one of the nine attrition classes and had inconsistent signs. In contrast, the signs for the estimated coefficients for openness are consistent (all greater than one) across all attrition classes in both studies we examined. Our results may suggest that as online data collection becomes more frequent and longitudinal, the role of openness as a predictor of attrition becomes more pronounced and consistent.

Similar to openness, we found that more extroverted, neurotic, and agreeable respondents are more likely to be associated with each of the attrition classes than with the "Stayers" class. Although some of these differences are not statistically significant, all have the expected signs, with estimated coefficients greater than one across all attrition classes in both studies. Our finding that more open, extroverted, neurotic, and agreeable respondents are more likely to drop out of the two studies aligns with prior research, is consistent with our previously stated theoretical predictions and expectations, and supports the panel fatigue explanation as an underlying mechanism proposed by Lugtig (2014).

Collectively, these findings underscore the growing importance of personality traits, beyond conventional demographic variables, in predicting nonresponse patterns as online data collection becomes more frequent, intense, and burdensome. Among all Big Five personality traits, conscientiousness exhibits the most consistent and strongest effect size, as expected. This has practical implications for panel management, retention strategies, and nonresponse adjustment strategies, given that personality traits are found to independently predict and causally affect significant life outcomes.

Our findings indicate that panel commitment and fatigue are the most likely mechanisms of dropout and suggest the need for more tailored approaches to panel management and retention strategies.

2.5.2 Practical implications related to personality traits

Our findings have several important implications for the design, implementation, and management of online panel surveys. For instance, survey organizations could implement targeted interventions to keep respondents with lower conscientiousness engaged, such as offering frequent reminders via text messages or additional monetary incentives, or giving them

less frequent surveys in order to reduce their burden while they are still acclimating to the online survey environment. Additionally, personality traits could be utilized as predictive tools to identify participants at higher risk of attrition, allowing for early intervention before dropout occurs. Providing additional support to individuals high in neuroticism or agreeableness, who may be more prone to attrit, can also enhance their commitment to the study.

In terms of nonresponse adjustment strategies, online survey researchers should consider incorporating personality measures into weighting schemes to improve the accuracy of bias adjustments, given that personality traits like conscientiousness are not only strong predictors of survey participation but also causally linked to significant life outcomes. While population-level benchmarks for personality traits do not exist—making traditional post-stratification infeasible—researchers can still leverage internal benchmarks by collecting personality data from all respondents at baseline. By incorporating personality measures into longitudinal weighting adjustments based on internal benchmarks, online longitudinal studies can achieve more accurate and representative estimates, even in the context of high-frequency and intensive online data collection, where modest early-stage attrition can lead to substantial sample loss and bias over time.

To facilitate this, since personality data are not available in the frame data, online longitudinal studies should collect personality data at the beginning of the study, preferably immediately after respondents become formal panel members or immediately after respondents complete the baseline mandatory demographic survey that is prerequisite for becoming panel members, to ensure that baseline data are available for all respondents (Sturgis and Brunton-Smith (2023) suggest collecting this information in the first wave of an offline panel study, such as the UK Household Longitudinal Study). Having personality data for the full initial sample

enables researchers to incorporate these variables into longitudinal weights, track potential attrition-related bias over time, implement targeted intervention, and inform future adaptive design based on personality profiles.

2.5.3 Relevance of Panel Tenure

We found no difference in attrition classes by panel tenure status in the monthly events study. In contrast, our results from the UCA data showed that fresh respondents were less likely to drop out early than experienced respondents. Consistent with the prediction by the Leverage-salience theory (Groves et al., 2000), we attribute this effect to the societal importance of the COVID-19 study, particularly to the fresh respondents, which was unequivocally made salient to them during the consent survey. Our results contrast with the findings of Biddle and Sollis (2023), who reported that recently recruited respondents were less likely to participate in COVID-19 online panel surveys in Australia.

The contrasting results on attrition by panel tenure status between the UCA data and the monthly events study highlight the role that panel tenure can play in optimizing engagement strategies. Understanding that fresh respondents may be more motivated by the societal relevance of a study suggests that survey organizations should emphasize the importance and impact of the study during the recruitment and consent process, particularly for new panelists. This approach could help maintain engagement and reduce early attrition among fresh respondents, especially in studies with high societal relevance. Additionally, in studies where panel tenure does not significantly influence attrition and topic saliency is low, focusing on other factors such as survey frequency and incentives may be more effective in reducing dropout rates.

2.5.4 Demographic and Health Effects Moderated by the COVID-19 Situations

We found no significant differences in nonresponse patterns across demographic subgroups in the monthly events study. However, in the UCA study, differences emerged in certain demographic subgroups, likely due to the unique circumstances of the COVID-19 pandemic. Specifically, we observed lower participation rates among younger individuals, Hispanics, Blacks, and those with lower levels of education—groups whose lives and livelihoods were disproportionately affected by the pandemic. Additionally, we found varying health effects between the two studies, which may be attributable to the pandemic.

In contrast to prior studies that found mixed age effects, the age effects in the UCA study are the largest and strongest among all demographic effects. Specifically, in the UCA study, we found large and negative effects among younger age groups across all attrition classes. This could be due to several factors, including the resilience, positivity, and better coping mechanisms often associated with older individuals during crises and challenging situations (Bruine de Bruin, 2021; Chappell et al., 2022; Na et al., 2022). Moreover, older individuals are more likely to be retired, which meant that they were less affected by factors such as job and earnings loss. This made older participants less vulnerable to dropping out of the study compared to younger respondents.

Regarding racial ethnic differences, minorities, particularly Blacks and Hispanics, typically exhibit lower survey response rates compared to Whites during non COVID-19 periods in the US (Jang & Vorderstrasse, 2019). However, the further decline in their survey participation during COVID-19 aligns with the substantive literature demonstrating the disproportionate impact of COVID-19 on Hispanics and Blacks in terms of mortality, hospitalization, infection, and job loss (Alcendor, 2020; Raifman & Raifman, 2020; Rossen et al., 2020). Additionally, they faced heightened challenges such as housing insecurity,

overcrowding, food insecurity, and poor neighborhood quality during the pandemic (Chun et al., 2023; Kim-Mozeleski et al., 2023; Lauren et al., 2021; Morales et al., 2021; Trabilsy et al., 2023; Versey, 2021). Notably, during the pandemic, opportunities for remote work varied significantly across racial/ethnic groups, with Blacks and Hispanics often employed in sectors of the US economy where remote work arrangements were impractical (Angelucci et al., 2020). These challenging factors collectively may have contributed to the further decreased survey participation among Hispanics and Blacks during the pandemic.

Traditionally, survey response theories have suggested that women exhibit higher response rates than men due to their presumed greater availability at home for interviews (Lepkowski & Couper, 2002; Nicoletti & Buck, 2003; Watson & Wooden, 2009). Obviously, the dynamics shifted during the COVID-19 pandemic, as women grappled with increased household and childcare duties, leading to a reversal of this nonresponse pattern. Indeed, we found suggestive evidence indicating that the decline in participation among females during COVID-19 was strongest among women with children under age 18 (results not shown). The net effect appears to be that we don't find differences in attrition across gender.

In the monthly events study, consistent with prior studies, we found that poor health predicts nonresponse patterns, particularly in the "Mid-wave attrition" class. In contrast, in the UCA study, we found no significant health effect across all attrition classes. While several explanations are possible, one possible explanation is that COVID-19 caused widespread closures of health centers, which may have hindered individuals' ability to seek medical care regardless of their health situations and healthcare needs. According to the UAS COVID-tracking data, compared to March 2020, medical visits declined by 37% in April and by 20% in May of 2020. Compared to 2019, routine physical exams declined by 39%, and well-woman

exams fell by 46% in 2020. Several other studies analyzing administrative record data and other national survey data in the US found similar trends (Birkmeyer et al., 2020; Jeffery et al., 2020; Nourazari et al., 2021; Rennert-May et al., 2021).

These findings suggest potential moderation effects by context surrounding COVID-19 and may have important implications for survey design and retention strategies, particularly during societal disruptions like the COVID-19 pandemic. For instance, lower participation rates among vulnerable subgroups such as younger individuals, Hispanics, Blacks, and those with lower education levels suggest a need for more targeted outreach and engagement strategies. The shift in gender dynamics during COVID-19 suggests the importance of flexible survey designs that accommodate women balancing childcare and household responsibilities. The absence of significant health-related attrition in the UCA study also suggests that health barriers may be reduced during crises, though poor health remains a strong predictor of attrition under normal conditions. Adapting survey methods and procedures, such as use of differential incentive strategies and tailored messaging, to address these demographic factors, especially during disruptive periods, is crucial for minimizing attrition and maintaining panel engagement.

2.5.5 Future Research Directions

Future research should explore the extent to which incorporating personality traits into the adjustment process can address attrition biases, in addition to conventional demographic variables such as age, gender, education, and race/ethnicity. This analysis will help determine the independent contribution of personality traits in reducing attrition biases. Another promising direction is to investigate the effects of significant events, such as illness or job loss, occurring between survey waves on respondent behavior. Testing the hypothesis that these events act as "shocks" to participation patterns could provide valuable insights into the drivers of mid-wave

attrition and improve the understanding of nonresponse triggers, as suggested by Lugtig (2014). Additionally, designing and evaluating the effectiveness of interventions targeting individuals with low conscientiousness scores based on baseline data or prior surveys represents a valuable area for future inquiry. Tailored strategies aimed at this group may lead to more effective retention efforts, and understanding which interventions work best could enhance overall panel engagement. Finally, future studies should replicate our research approach in other online data collection settings and contexts different from the one we examined. For example, it is unclear whether the same nonresponse patterns related to personality traits would emerge in non-probability panels. Such panels often differ in recruitment methods, participant motivations, and respondent expectations, which may influence how personality traits affect participation decisions differently. Future research examining whether the predictive power of personality traits generalizes to non-probability panel contexts would be a valuable extension of this work.

2.5.6 Strength and Limitations

A key strength of this study is the use of data from two recently completed studies, both conducted within the same probability-based online panel dedicated to social science and health-related research. The inclusion of rich background information for all respondents enhances the precision of statistical inferences, and the availability of Big Five personality traits data allowed us to emphasize the significance of personality factors, particularly in relation to attrition. However, a limitation is that the study does not assess the extent to which selective attrition may result in biases in some key substantive variables. Another limitation is that our results may have limited generalizability beyond the context we examined. We also acknowledge that the age restriction of 50+ in the Monthly Event Panel Surveys overlaps with panel topic, which limits our ability to disentangle topic effects from age-related sample composition. As such,

comparisons between the two studies should be interpreted with caution. However, both studies were implemented within the same probability-based online panel, which enhances comparability and internal consistency across the analyses. Despite this, our findings provide valuable insights into the complexities and unique patterns of nonresponse and attrition, as well as their correlates and underlying mechanisms, in the context of high-frequency online longitudinal data collection—a topic that has received limited attention thus far.

Chapter 3 Evaluating the Impact of Three Incentive Schemes on Survey Nonresponse and Attrition in Online Longitudinal Panels

3.1 Introduction

Recent advancements in online surveys have allowed researchers to collect data over time from the same respondents more frequently than is possible with conventional face-to-face longitudinal studies that typically collect data annually or biannually. Frequent data collection has a methodological advantage because collecting information closer in time to the occurrence of important events, and collecting it repeatedly reduces the role of measurement error and omitted variables, enabling sharper statistical inference. Online surveys also offer researchers the flexibility to launch surveys at any time and collect data from a large number of respondents within a short timeframe, while also offering convenience to respondents who can complete surveys at their own pace and in a private setting. This approach is also cost-effective as it eliminates or minimizes the need for intensive fieldwork and logistical operations commonly associated with in-person and telephone interviews, and mail questionnaires. Furthermore, online data enhances statistical efficiency by avoiding the clustering effect often encountered in face-to-face interviews associated with multi-stage cluster sampling designs.

Unlike offline data collection methods, online surveys allow for quick response to emerging developments. For instance, during the COVID-19 pandemic, online panels have proven valuable in monitoring daily mood changes, observing changes in infection rates, job loss, food insecurity, and vaccination rates over time, and exploring the evolving correlation between COVID-19 vaccines and mental distress in the U.S. adult population (de la Haye et al.,

2023; Kapteyn et al., 2020; Perez-Arce et al., 2021). Similarly, during the H1N1 pandemic, online panels played a crucial role in monitoring population trends related to risk behaviors and vaccine intentions in the U.S. (Gidengil et al., 2012).

Despite their inherent strengths, as with any longitudinal study, wave nonresponse and attrition pose a significant challenge in online data collections. Wave nonresponse occurs when respondents fail to participate in some survey waves. One particular case of wave nonresponse is attrition in which respondents miss a few survey waves initially and then drop out from the study completely. The focus of this study is the former situation: analyzing responses to individual survey waves.

Both nonresponse and attrition lead to a reduced sample size and decreased precision of survey estimates. Given that data are collected more frequently in online panel context, even a modest attrition rate in each wave can lead to a significant loss of sample members over time in a short period of time. Additionally, if the characteristics of panelists who drop out differ from those who continue participating in terms of key survey statistics, attrition may introduce bias, which can undermine the validity and generalizability of research findings (Chin et al., 2021; Hansson et al., 2018; Lugtig, 2014; Lugtig et al., 2014; Lynn, 2018; R. Olsen, 2018; R. J. Olsen, 2018; Ward et al., 2017).

The implications of these challenges are underscored by several studies. Research has shown that statistical adjustments may not completely eliminate bias resulting from respondents not completing all longitudinal surveys (Hu, Kirlin, et al., 2020) and that nonresponse adjustments based on standard demographic variables were not sufficient to fully address attrition bias in panel surveys (Cheng et al., 2020; Frankel & Hillygus, 2014). Additionally, some statistical estimates can be sensitive to the type of adjustment method employed (Ayala et al.,

2011; Basic & Rendtel, 2007; Kristman et al., 2005). As such, minimizing nonresponse and attrition through effective panel management and interventions should be part of a comprehensive mitigating strategy rather than solely relying on post data-collection statistical adjustments. As probability-based online panels have emerged as an attractive option for social science research and policy decisions, it is essential to identify and rigorously evaluate the effectiveness of strategies aimed at enhancing sample representativeness across survey waves.

Despite extensive research on survey incentives, little is known about how different payment structures—particularly their timing and frequency—affect nonresponse and attrition in high-frequency online panels. Most prior studies have focused on cross-sectional or offline panel surveys, leaving key questions about the effects of incentive timing in online panels unanswered.

In this study, we report findings from a randomized controlled trial (RCT) that evaluated the causal effects of three theoretically-motivated cash incentive payment strategies on nonresponse in an online longitudinal data collection effort. These incentives, tailored to the unique context of online data collection, primarily varied in the timing of the incentive payments and were informed by prior studies and existing survey response theories. For this study, we recruited an online panel of 500 respondents, randomly selected from a complete list of all currently enrolled master's students in non-professional master's programs at the University of Michigan, Ann Arbor campus. Over the course of about 6 months, respondents received an online survey invitation every 2.5 months, totaling three invitations. Our frame data included background information on all respondents, which we used to explore the heterogeneity of treatment effects and study participation rates based on these characteristics. This research aims to shed new light on effective interventions aimed at enhancing online data quality, better inform

the design and implementation of future online longitudinal studies, and motivate further research.

3.2 Background

Numerous studies have investigated the effectiveness of incentives on response rates in both cross-sectional surveys and offline panel surveys. These studies generally fall into one of the following categories, or a combination of them: (1) examining whether to provide an incentive or not, (2) analyzing the effect of various forms of incentives (cash, in-kind, or gifts), (3) exploring the effect of incentives by survey mode (mail, telephone, face-to-face, Internet, or mixed modes), (4) evaluating the timing of incentives (prepaid [unconditional] vs. promised [conditional]), and (5) assessing the effect of increasing incentive size. Notably, findings from cross-sectional and panel studies sometimes differ, highlighting the complexity and variability in the impact of incentives on survey participation across different contexts and methodologies (Church, 1993; Edwards et al., 2002; Laurie & Lynn, 2009; Singer, 2002; Singer et al., 1999; Singer & Ye, 2013; Toepoel, 2012; Wong, 2020).

3.2.1 Incentive vs. no incentive

While the majority of prior studies have found incentives to be effective (Church, 1993), their overall effectiveness varies based on several factors. Below, we highlight some of the key factors.

3.2.1.1 Form of incentive

All else being equal, cash incentives are found to be more effective than in-kind incentives, gifts, lotteries, or donations to charitable organizations (Church, 1993; Felderer et al., 2018; Göritz & Neumann, 2016; Henderson et al., 2010; Laurie & Lynn, 2009; Ryu et al., 2006; Singer et al., 1999; Singer & Ye, 2013; Tzamourani, 2000). This is largely because cash has a

universally accepted and understood value, whereas non-monetary incentives can have limited practical use. This finding applies to both cross-sectional and offline longitudinal surveys to a large extent.

3.2.1.2 Incentive effects by mode

Numerous prior studies have found that incentives are effective in all survey modes, but their effects tend to be larger in self-administered surveys (e.g., mail) than in interviewer-administered face-to-face surveys (Church, 1993; Edwards et al., 2023; Gelman et al., 2003; Göritz, 2006; Jäckle & Lynn, 2008; Mercer et al., 2015; Ryu et al., 2006; Singer et al., 1999). This is because, in surveys where there is direct interaction between interviewers and respondents, interviewers can crowd out the effects of incentives through the use of various persuasion and doorstep strategies to gain respondents' cooperation. This is especially the case in offline longitudinal studies where interactions and communications between respondents and interviewers are needed between waves.

Between online and offline surveys, prior studies have found that incentives are more effective in online surveys. This is true for both cross-sectional and longitudinal surveys where respondents have the option of completing the surveys over the Internet (Göritz, 2006; McMaster et al., 2017).

To address the rising costs associated with face-to-face interviews, several studies have switched to push-to-web sequential mixed modes (McMaster et al., 2017). In this mixed-mode design, respondents are encouraged to complete interviews online first, followed by face-to-face or telephone interviews if necessary. With this approach, costs can be substantially reduced through decreased fieldwork and fewer field interviews. Despite the clear cost-saving advantage, response rates are typically lower in mixed-mode designs compared to single modes (Manfreda

et al., 2008; Patrick et al., 2018). However, offering additional incentives has been found to offset this disadvantage and can result in achieving a response rate comparable to single-mode face-to-face designs while still realizing the cost-saving goal (Bianchi et al., 2016; Gaia, 2017; Jäckle et al., 2015; Messer & Dillman, 2011; Wood & Kunz, 2014).

With recent advancements in mobile technologies, a growing number of surveys have collected data via mobile devices, sometimes requiring respondents to download and install mobile apps. The key advantages of mobile approaches include respondent convenience and quick turnaround time. However, response rates are typically lower in mobile data collections compared to those conducted on PCs (Toepoel & Funke, 2018). Adding incentives to mobile data collection has been found to help address the low response rates in mobile surveys (Keusch et al., 2019; Mavletova & Couper, 2016).

3.2.1.3 Prepaid (unconditional) vs. promised (conditional) incentives

The majority of evidence from existing literature suggests that prepaid or unconditional incentives are more effective in increasing response rates compared to promised or conditional incentives, non-monetary incentives, or no-incentive conditions (Church, 1993; Messer & Dillman, 2011; Petrolia & Bhattacharjee, 2009; Singer et al., 1999; Singer & Ye, 2013; Toepoel, 2012). However, as previously mentioned, this effect can be moderated by interviewers in face-to-face interviews. Additionally, it is worth highlighting that the evidence for the effectiveness of prepaid incentives has remained inconclusive in some specific contexts, mainly outside the US. For instance, while Castiglioni et al. (2008) found that conditional incentives were more effective than unconditional incentives in a three-wave panel survey in Germany, Jäckle and Lynn (2008) found the opposite effect, and Gaia (2017) found an insignificant unconditional incentive effect in two panel studies in the UK.

3.2.2 Cost considerations in prepaid vs. promised incentives

Prepaid incentives typically involve higher upfront costs because they are distributed to all sampled respondents regardless of actual response. However, when evaluated on a cost-per-completed-response basis, prepaid incentives can be more cost-effective. A meta-review study by Mercer et al. (2015), for example, found that prepaid incentives in mail surveys yielded the largest increase in response rates per dollar spent, followed by prepaid incentives in telephone surveys. These cost advantages are attributed to the lower baseline response rates in mail and telephone surveys, which provide more opportunities for improvement from each dollar spent. In contrast, postpaid incentives are given only to respondents who complete the survey. This approach incurs lower costs initially but can lead to higher recruitment and follow-up costs, especially when overall response rates are low to begin with.

3.2.3 Incentive size

In addition to the form of the incentive, survey mode, and timing, the size of the incentive is also found to influence response rates. Prior studies have largely found a positive, albeit non-linear, relationship between incentive size and response rate. That is, increasing the incentive amount leads to higher response rates, but the increase occurs at a declining rate (Cantor et al., 2008; Church, 1993; Mercer et al., 2015; Singer & Ye, 2013). This finding is not limited to cross-sectional surveys but also extends to longitudinal surveys to a large extent (Booker et al., 2011; Börsch-Supan et al., 2013; Gaia, 2017; James, 1997; Laurie, 2007; Martin & Winters, 2001; Rodgers, 2011; Westra et al., 2015).

3.2.4 Theoretical Frameworks

This study draws on several theoretical frameworks to inform the development of incentive strategies suitable for implementation in online panel contexts and to generate

hypotheses regarding both the overall and subgroup-specific effects of these strategies in high-frequency online longitudinal surveys. The following subsections describe the relevant frameworks, Economic Exchange Theory, Leverage-Salience Theory, and the Norm of Reciprocity and Social Exchange Theories, and explain how each informs our hypotheses and design choices.

3.2.4.1 Economic Exchange Theory

According to the economic exchange theory, when approached by survey organizations or researchers, potential respondents make a rational economic decision by weighing the potential benefits against the costs (Ajzen & Fishbein, 1980; Biner & Kidd, 1994). In such circumstances, the theory suggests that incentives should at least compensate for respondents' time and effort associated with the survey request for the economic exchange (i.e., responding to a survey request) to occur. This cost-benefit mechanism suggests that, all else being equal, a survey request will be more attractive to low-income individuals and that there is a positive, albeit nonlinear, relationship between incentive size and response propensity. Additionally, the fact that cash incentives are more effective than non-monetary incentives is consistent with the cost-benefit explanation. Singer (2011) provides additional insights on the cost-benefit theoretical framework, which takes into consideration non-monetary aspects of costs and benefits, along with additional empirical evidence supporting the cost-benefit argument.

3.2.4.2 Social Exchange Theory

Unlike the economic exchange explanation, the social exchange theory (Dillman, 1978; Dillman et al., 2014) postulates that the trust researchers place in potential respondents motivates them to respond to the survey. The unconditional incentive embodies this trust. In return, respondents reciprocate by completing the survey. Within this social exchange framework, any

amount of unconditional incentive can be effective in improving overall response rates. Greenberg and Dillman (2023) recently tested the theory in a mail survey conducted in West Virginia and found empirical support for it.

3.2.4.3 Norm of Reciprocity

Similar to the social exchange theory is the norm of reciprocity theory (Gouldner, 1960). However, unlike the social exchange theory, where trust is the key component, the norm of reciprocity theory proposes that it is the goodwill or kindness from a stranger symbolized by the unconditional incentive that motivates participation. The receipt of this unconditional incentive from a researcher or survey organization that the respondent does not know creates an obligation for the respondent to reciprocate by completing the survey request. In this framework, similar to the social exchange theory, any type and amount of unconditional incentive will be effective. A recent study by Becker (2023) provides empirical evidence consistent with this explanation.

3.2.4.4 Leverage-Saliency Theory

Leverage-salience theory (Groves et al., 2000) incorporates both survey features and respondent characteristics, suggesting that different survey features (e.g., study topics, incentive, survey sponsor) have varying effects ("leverages") on participation decisions based on respondents' socioeconomic characteristics, lifestyles, and preferences. According to this theory, the likelihood of survey participation depends on how saliently the positive features of the survey are presented to potential respondents in the initial survey request. One implication of this theory is that unconditional incentives will be more effective among subgroups with low response propensities and low-income groups if the incentive feature is made salient to them in the survey request. In this framework, incentives can serve as a substitute for respondents' lack of motivation or interest in participating in a survey. We will be utilizing insights from these

aforementioned theories to guide our hypotheses development and justify our intervention design as explained in the next sections.

3.2.5 Study Hypotheses

3.2.5.1 Overall Effect

According to Economic Exchange Theory, individuals weigh the perceived costs of participation (e.g., time and effort) against the anticipated benefits (e.g., monetary incentives). From this perspective, incentives that are guaranteed and delivered upfront should be more effective than conditional rewards in motivating participation in online panel surveys. Similarly, the Norm of Reciprocity and Social Exchange Theories suggest that unconditional incentives can foster a sense of obligation or goodwill, motivating individuals to reciprocate by participating.

> **Hypothesis 1**: Unconditional incentives will be more effective than conditional incentives in motivating participation in high-frequency online panel surveys, as the upfront nature of the incentive may foster a stronger sense of commitment and motivation.

3.2.5.2 Treatment Heterogeneity

Economic Exchange Theory also implies that the perceived value of incentives varies across subgroups, with individuals from lower socioeconomic backgrounds likely to find incentives more attractive and thus more responsive to them. Similarly, the Leverage-Salience Theory posits that survey features—such as incentives, topic relevance, and sponsor—exert varying levels of influence ("leverage") on participation decisions depending on respondents' characteristics. For instance, unconditional incentives may be especially more effective among individuals with lower intrinsic motivation, weaker topic relevance, or fewer economic resources.

Hypothesis 2: Unconditional incentives will be particularly effective among subgroups with lower baseline response propensities.

As an exploratory analysis, we will also examine whether repeated unconditional payments (delivered per wave) sustain engagement more effectively than a single large lump sum provided at the beginning of the study, the salience of which may diminish over time.

3.2.6 Research Gap

Most studies examining the effects of various incentive strategies in longitudinal data collection settings have primarily focused on traditional offline longitudinal studies, where respondents are interviewed annually or biannually face-to-face or using sequential mixed modes. Since respondents are interviewed repeatedly in panel studies, various combinations of incentives can be applied across survey waves (Laurie & Lynn, 2009; Wong, 2020). One common strategy employed is the use of differential incentives, where a significantly larger incentive is offered to specific groups particularly those with low response propensity during some data collection waves. For instance, the "early bird" strategy targets non-responders from the previous wave or respondents with a low response propensity by offering them larger incentives at the beginning of the data collection period (McGonagle et al., 2023). Similarly, the "time-limited" strategy offers larger incentives to all remaining respondents during specified time periods determined by the researchers (Fomby et al., 2017). Additionally, the "end-game" strategy involves providing higher incentives to remaining sample cases during the final weeks of the data collection period (McGonagle, 2020). Several other cohort studies in the UK also experimented with similar incentive strategies (Booth et al., 2024; Cabrera-Álvarez & Lynn, 2023; Gaia, 2017; Peycheva et al., 2023; Wong, 2020).

These differential incentive approaches have generally been found to boost the response rate in offline panel studies. However, implementing these strategies in an online longitudinal context presents challenges due to fundamental differences in design features between offline and online panel studies. Table 3-1 highlights the key differences. One significant difference is the frequency of data collection. In online data collection, survey waves typically have shorter durations, with the gap between two successive waves often being less than a month, compared to one year or longer in most offline studies. This shorter time frame limits the applicability of strategies like the "time-limited" or "end-game" approaches, which are less suitable for online longitudinal data collections.

The design features summarized in Table 3-1 are useful for better understanding the implications of panel structure for optimal incentive strategy. For example, the higher frequency of online panel surveys may increase the cumulative burden of participation, making unconditional incentives more salient and potentially more effective at sustaining respondent engagement over time. Similarly, the shorter time between survey waves in online panels makes it impractical to implement strategies such as the “end-game” and “time-limited” approaches. In contrast, prepaid incentives, particularly those delivered per wave or as a lump sum, may be more practical and effective in the online panel context.

Table 3-1: Probability-based online panel vs. offline panels

	Online panel	**Offline panel**
Survey frequency	• High frequency: 1 survey/two months – 4 surveys/month	• Low frequency: 1 survey every year – one survey every five years
Interview length	• 15-30 minutes	• 80-120 minutes
Observation period	• Mostly not life-long	• Mostly life-long
Mode of interview	• Online	• Face-to-face • Mail • Telephone

		• Mixed modes
Common reason(s) for attrition	• Mainly refusal, panel fatigue • Noncontact to a lesser extent	• Non-location • Noncontact conditional on being located • Refusal, panel fatigue
Common strategies for reducing attrition	• Incentives • Panel engagement and management • Frequent update of email and telephone number	• Incentives • Panel engagement and management • In-between wave tracking • In-between wave address updates • Tailored messaging • Interviewer training • Use of the same interviewers across waves • Longer field period
Questionnaires and survey topics across waves	• Mostly varying	• Mostly fixed

Other strategies employed in offline panel studies to increase the response rates include offering one-time incentives in a specific data collection wave, introducing incentives for the first time after a few more waves of data collection have been carried out, withdrawing existing incentives, providing a one-time bonus on top of the baseline incentives, increasing the level of the incentive offered periodically, offering a large bonus if respondents call in to do the interview by telephone, and using pre-paid (unconditional) incentives for a few waves and then switching back to post-paid (conditional) incentives (Booth et al., 2024; Cabrera-Álvarez & Lynn, 2023; Jäckle & Lynn, 2008; Laurie, 2007; Laurie & Lynn, 2009; Singer & Ye, 2013; Westra et al., 2015; Wong, 2020; Zagorsky & Rhoton, 2008).

Although these strategies have been found to be effective in reducing nonresponse and attrition in some specific contexts in offline panel studies, it is important to recognize the fundamental differences in study characteristics and design features between online and offline panel studies. The most significant differences, in addition to data collection frequency, include

the mode of data collection, questionnaire length, study period, questionnaire design, and common reasons for attrition (see Table 3-1). Therefore, the applicability of these strategies to online longitudinal data collections remains uncertain. Thus, there is a critical need to explore the effectiveness of other incentive payment structures suitable for implementation in an online longitudinal study context to maximize survey response rates and enhance data quality.

3.3 Study Objectives

In this study, using an experimental design, we evaluate the effectiveness of three incentive payment schemes tailored to the context of high-frequency online longitudinal surveys in the U.S and test our stated hypotheses. In all the incentive plans evaluated, respondents are compensated for their time completing individual survey waves but receive their incentives at different time points, as explained below. This contrasts with some offline longitudinal studies that either never use incentives or provide incentives only in one or a few survey waves without repeating the incentives in future waves. Additionally, the total amount respondents can earn is the same across all incentive plans, contrasting with some prior studies that utilized differential incentive payments for refusal conversion (Laurie & Lynn, 2009; Singer & Ye, 2013; Wong, 2020).

In the first incentive condition, called the conditional per questionnaire incentive group, compensation is based on the completion of individual longitudinal surveys, which is a common approach used by existing studies. As such, this incentive group will serve as the control group in the evaluation.

In the second incentive condition, called the unconditional pre-per questionnaire incentive group, respondents receive a cash payment in advance equivalent to the expected earnings for a given survey wave, regardless of their response to the survey invitation for that

wave. This plan is based on previous evidence from cross-sectional and offline longitudinal studies that have shown the effectiveness of unconditional incentives compared to conditional incentives in increasing response rates. The aim is to establish trust or goodwill with potential respondents and enhance the saliency of survey requests among respondents with low response propensity by offering them an incentive payment in advance for each survey wave. By providing the unconditional incentive repeatedly across survey waves, this plan also aims to strengthen the sense of obligation and reciprocity among all respondents, regardless of their initial survey commitment level. This approach is grounded in the concept that unconditional incentives have a greater impact when the salience of the research is low and can help improve overall response rates and sample representation.

In the third treatment condition, called the unconditional pre-all questionnaire lump sum incentive group, participants receive an upfront lump sum payment at wave 1 equivalent to the expected earnings for a given survey wave multiplied by the number of survey waves, regardless of their response to the survey invitations in all waves. Previous evidence from offline panel studies indicates that providing a one-time substantial unconditional incentive has an immediate impact not only on the current wave but also on subsequent waves (Singer & Ye, 2013; Wong, 2020). It is reasonable to expect this carryover effect in an online longitudinal study context because the time intervals between survey waves are significantly shorter in the online context than in the offline context. This incentive condition aims to establish trust with potential respondents, as well as enhance the saliency of the survey in wave 1 as predicted by the Leverage-salience theory. However, in this condition, the saliency gradually diminishes from wave 2 onwards as the large unconditional incentive is not repeated after wave 1. This treatment approach acknowledges the initial boost in engagement and participation by offering a

significant incentive upfront while recognizing that the saliency effect may decrease over time without the continued provision of repeated incentives, while still compensating respondents for all future survey time. Additionally, this incentive plan requires less administration compared to the previous two conditions, thus potentially reducing administrative costs.

No other incentive conditions were considered in this study to maintain a focused evaluation of the three incentive schedules that are theoretically grounded and practically relevant. These three interventions were selected because they represent the most commonly used and are feasible for implementation in online panel context. The conditional per questionnaire incentive serves as a baseline for comparison, reflecting standard practices in many online longitudinal surveys. The two unconditional incentive conditions were chosen to test the well-documented efficacy of upfront payments in enhancing response rates while exploring the effects of payment timing (per wave vs. upfront lump sum) and sufficiently compensating respondents for their time completing individual surveys. Additionally, introducing additional conditions would have increased complexity and required a larger sample size to ensure sufficient statistical power, which was not feasible given the study's constraints. The current design allows for a rigorous examination of the three interventions while providing actionable insights for practitioners and researchers alike.

The remainder of the paper is organized as follows. In the Methods section, we present our methodological approach, followed by a presentation of results in the Results section. In the Discussion & Conclusion section, we interpret our results, discuss their implications and significance with respect to recruitment strategies, panel management, nonresponse, attrition, the efficacy of longitudinal data collection in online panels and future adaptive survey design,

highlight the strengths and limitations of the study, make suggestions for future research, and conclude with final remarks.

3.4 Methods

3.4.1 Study Participants

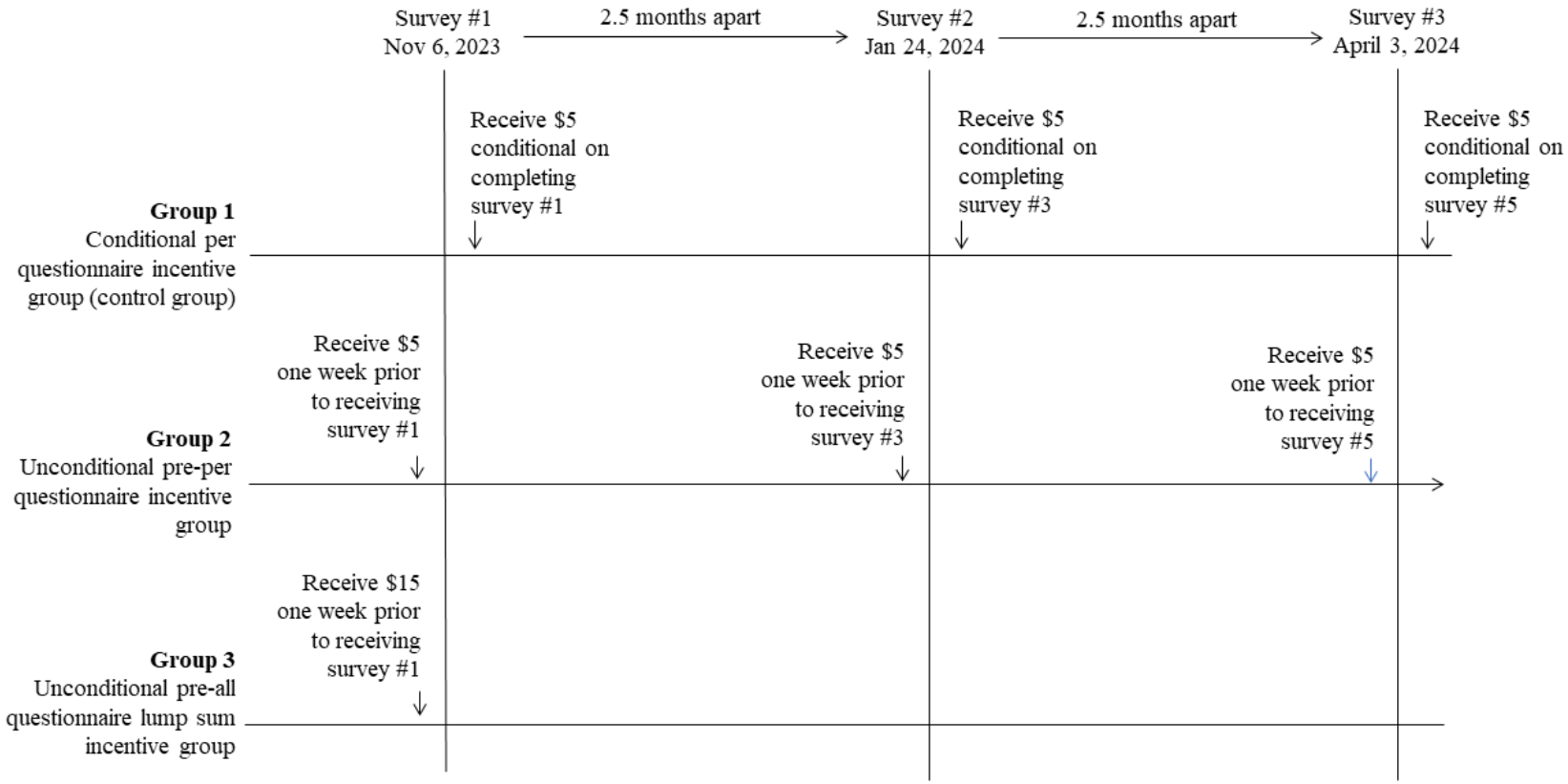
Participants for this study are master's students at the University of Michigan. The study received a complete list of all 4,435 master's students currently enrolled in non-professional master's programs in the Ann Arbor campus. The list contained information about students' first and last names, emails, gender, age, race & ethnicity, citizenship status, underrepresented minority (URM) status, academic major, year of admission, undergraduate GPA, Pell grant status, and first-generation college student status. Participant recruitment, which involved sending out initial emails to all 4,435 students and following up with up to 4 additional reminders to those who had not signed up for the study yet, began on October 10, 2023 and ended on October 25, 2023. The recruitment email provided brief background information about the study and the organization conducting it, various tasks participants would be asked to complete in the study, and the incentive amounts participants could expect to earn for completing individual tasks. Additionally, the email contained an embedded URL, which directed potential participants to the Qualtrics online initial intake survey, where they provided consent. The intake survey also collected participants' contact information and mailing addresses for the sole purpose of mailing their cash incentives. The contact information was updated throughout the study period.

3.4.2 Experimental Design

Participants who provided consent were randomly assigned to one of three incentive groups, each representing distinct payment plans with equal group sizes, as mentioned earlier

and illustrated in Figure 3-1. Specifically, participants randomized into the conditional per questionnaire incentive group (the control group) received a $5 incentive for each completed survey wave, and these cash incentives were mailed to them within a week of survey completion. As previously mentioned, this conditional incentive is a common practice used by most online longitudinal studies. Participants in the unconditional pre-per questionnaire incentive group received a $5 unconditional incentive one week prior to receiving individual survey invitations. Participants in the unconditional pre-all-questionnaire lump sum incentive group received a lump sum unconditional payment of $15 (equivalent to $5 for each of the three surveys they would complete throughout the study) one week before receiving their first survey invitation.

Figure 3-1: Incentive Experimental Design



All cash incentives were dispatched via USPS First Class mailing, and participants typically received them within 2-3 days. However, a small number of incentive mailings (less than 2% of the mailings) were returned as undeliverable, primarily due to incorrect or incomplete addresses or a lack of proof of participant residency (e.g., absence of names of current residents on the mailboxes). To address this issue, we implemented two approaches: either asking participants to collect their incentives in person at our campus office or arranging for one of our graduate research assistants to hand-deliver the incentives to participants' residences. For the latter option, we verified addresses with participants beforehand to ensure smooth delivery.

3.4.3 Longitudinal Surveys

Participants received three longitudinal surveys throughout the 2023-2024 academic year. The first, second, and third surveys were distributed on November 6, 2023, January 24, 2024, and April 3, 2024, respectively (Figure 3-1). Each survey had an approximate completion time of 10 minutes. Participants securely filled out individual surveys using U-M licensed Qualtrics software. The survey questionnaires were identical across all waves and covered various aspects of students' well-being, including their overall mental and physical health, time management and use, perceptions of stress, life stressors, sources of stress, and work-related stress. The longitudinal design allowed the U-M Graduate School to monitor students' mental and emotional well-being over the course of one academic year. In each survey wave, non-responders received up to two reminder emails, with the first reminder sent one week after the initial invitation and the second reminder one week later.

3.4.4 Outcome Variables

The primary outcome variable for this analysis is the response to individual survey waves, measured as a binary indicator with a value of 1 if participants complete a survey wave and 0 otherwise. As a secondary outcome, we also analyzed item nonresponse.

3.4.5 Statistical Analyses

We first provide summary statistics of recruitment data, including the overall study signup rate and signup rates broken down by baseline characteristics available from the administrative data. We also present summary statistics of the baseline characteristics of our final analytical sample. Next, we assess the balance of these baseline characteristics across incentive groups by conducting t-tests for continuous variables and Chi-squared tests for categorical variables, reporting their p-values. Since the treatment was randomly assigned, simply taking differences in average outcome values across incentive groups will provide an unbiased estimate of the treatment effects. Additionally, we combined the two unconditional incentive groups to compare their pooled response rates with the control group. This pooling decision was made post hoc, after data collection was completed. Consequently, we report p-values from Chi-squared tests assessing bivariate associations between our main outcome variable and the treatment variable. To explore treatment heterogeneity, we conduct a series of bivariate association tests (and report their p-values) between the main outcome variable and treatment variable for each of the subgroups based on the information available in the administrative data. We also provide cumulative response rates for each survey wave to analyze the dynamics of response rates over time and how the response rates changed after reminders were sent out. All analyses were conducted using Stata Version 18.0.

3.5 Results

3.5.1 Study signup rate

Table 3-2 presents summary statistics for the recruitment data. Overall, of the 4,435 master's students contacted, 492 agreed to participate, yielding an overall study signup rate of approximately 11.09%[10]. The results presented in Table 3-2 indicate that signup rates differ across subgroups. In particular, women (compared to men), Whites (compared to non-Whites), domestic students (compared to international students), Pell Grant recipients (compared to non-Pell Grant recipients), and students admitted in Fall 2023 (compared to those admitted before 2023) are significantly more likely to sign up for the study (p-value < 0.001). Additionally, signup rates tend to differ across academic units, with students majoring in Architecture, Urban Planning, Environment and Sustainability, Education, Medicine, and Public Policy being significantly more likely to participate compared to those in Dentistry, Engineering, Kinesiology, Public Health, Science, Social Science, Language, and Art majors (p-value < 0.001). In contrast, participation rates do not seem to differ by undergraduate GPA or first-generation college student status.

Table 3-2: Study Signup Rates

	Population Counts	Signup Rate	p-value
Overall	4435	11.09%	NA
Gender			
Female	1969	16.15%	0.000
Male	2466	7.06%	
Race & Ethnicity			
White	1294	15.69%	0.000
Non-White	3141	9.20%	
Citizenship Status			
Domestic Students	2232	13.80%	0.000
International Students	2203	8.35%	
If First Generation Student			
No	3668	11.07%	0.908
Yes	767	11.21%	
Pell Grant			
No	3815	10.46%	0.001
Yes	620	15.00%	
Admission Year			
Before Fall 2023	2513	7.48%	0.000

[10] Defined as the number of respondents who became panel members divided by the total number of respondents initially contacted via email.

Fall 2023	1922	15.82%	
Undergraduate GPA			
Below 3.5	1204	11.13%	0.963
3.5 and above	3231	11.08%	
Academic Discipline			
Architecture, Urban Planning, Environment and Sustainability	615	19.35%	
Dentistry	119	5.88%	
Education	210	20.95%	
Engineering	1860	6.77%	
Kinesiology	122	9.84%	0.000
Medicine	263	17.11%	
Public Health	213	14.55%	
Public Policy	224	18.30%	
Science	517	8.90%	
Social Science, Language, and Art	292	7.19%	

3.5.2 Sample summary statistics

Table 3-3 presents the sample summary statistics for respondent baseline characteristics. The first column provides the summary statistics for the overall sample. The second through fourth columns present the summary statistics for each of the three incentive groups. The last column presents p-values from Chi-squared tests assessing the balance of the baseline characteristics across the incentive groups. The overall sample compositions by baseline characteristics are as follows: female students (64.63%), White students (41.26%), domestic students (62.60%), first-generation college students (17.48%), Pell Grant recipients (18.90%), fall 2023 admitted students (61.79%), students with an undergraduate GPA of 3.5 and above (72.76%), and students majoring in Architecture, Urban Planning, Environment and Sustainability; Education; Medicine; and Public Policy (50.61%). The p-values presented in the last column reveal that these baseline characteristics are balanced across the incentive groups.

Of the 492 sample participants invited to all three panel surveys, 28 (about 6%) did not complete any surveys, 40 (about 8%) completed one survey, 65 (about 13%) completed two surveys, and 359 (about 73%) completed all three surveys.

3.5.3 Overall treatment effects (Hypothesis 1)

Table 3-4 presents the response rates for each of the three incentive groups for each survey wave and for all waves combined, along with p-values from Chi-squared tests assessing the associations between the incentive treatment and outcome variables. The last two rows in Table 3-4 provide the response rates for the control group and the two unconditional incentive groups combined, along with their associated p-values. Since the difference in response rates between the two unconditional incentive groups did not reach a significance level across all survey waves, combining them helps to increase the power of the analysis and provides an overall estimate of the unconditional incentive effect. Standard errors (SEs) are adjusted for clustering when data are pooled across survey waves.

Table 3-3: Overall distribution of baseline characteristics and by treatment status

	Overall (n=492)	Conditional (control group) (n=164)	Unconditional (per questionnaire) (n=164)	Unconditional (all questionnaire) (n=164)	p-value
Female (%)	64.63	67.07	64.02	62.80	0.707
White (%)	41.26	42.07	40.24	41.46	0.943
Domestic students (%)	62.60	63.41	65.24	59.15	0.504
First generation student (%)	17.48	17.68	17.07	17.68	0.986
Received Pell grant (%)	18.90	16.46	20.12	20.12	0.620
Admitted in fall 2023 (%)	61.79	60.37	57.93	67.07	0.211
GPA 3.5 and above (%)	72.76	71.95	71.95	74.39	0.849
High signup academic majors [a] (%)	50.61	51.22	49.39	47.56	0.803

[a] High signup majors include: Architecture, Urban Planning, Environment and Sustainability; Education; Medicine; Public Policy (See Table 3-2). The remaining majors are classified as low signup group in this analysis.

Table 3-4: Response rates by treatment status

	Wave 1	p-value	Wave 2	p-value	Wave 3	p-value	All waves	p-value [b]
Treat1	86.59		79.27		75.00		80.28	
Treat2	92.68	0.045	87.80	0.081	80.49	0.474	86.99	0.103
Treat3	93.90		85.98		78.66		86.18	
Conditional	86.59	0.014	79.27	0.028	75.00	0.248	80.28	0.035
Unconditional	93.29		86.89		79.57		86.59	

Treat1=Conditional; Treat2= Unconditional (per questionnaire); Treat3=Unconditional (all-questionnaire). [b] SEs are cluster-robust.

The last column and the last two rows in Table 3-4 reveal that, on average, unconditional incentives increased the response rate by about 6 percentage points across all survey waves, and this difference is statistically significant (p-value = 0.035). Among individual survey waves, response rates in the unconditional incentive groups are also higher than in the conditional group for wave 1 (p-value = 0.014), wave 2 (p-value = 0.028), but the difference failed to reach significance in wave 3 (p-value = 0.248), likely due to small sample size.

The results in Table 3-4 also indicate that, across all survey waves, the difference in response rates between the two unconditional incentive groups and the control group is smallest in wave 3. This is largely due to the timing of wave 3 launch (April 3, 2024) and fielding, which coincided with the last weeks of the winter semester when students were busy preparing for finals and graduation. This timing seems to have weakened the effectiveness of the unconditional payments in wave 3, as evidenced by the larger decline in response rates from wave 2 to wave 3 among the two unconditional incentive groups compared to the conditional incentive group.

Comparing the two unconditional incentive groups, the response rate among the group that received a $15 unconditional pre-all-questionnaire lump sum is slightly higher than the group that received a $5 unconditional pre-per-questionnaire incentive in the first wave (93.90% vs. 92.68%), but the difference is not significant. In the second and third waves, however, response rates are consistently higher among the group that received a $5 unconditional pre-per-questionnaire incentive compared to the group that received a $15 lump sum payment, though these differences are not significant.

3.5.4 Heterogeneous treatment effects (Hypothesis 2)

Our recruitment results, as presented in Table 3-2, show that signup rates differ significantly by race and ethnicity, citizenship status, admission year, academic major, and

gender. In addition to boosting the overall response rate, unconditional incentives are designed to improve sample representation by attracting respondents with low response propensities. Therefore, we analyze whether the unconditional incentives have differential effects among subgroups with varying signup rates. Table 3-5 through Table 3-9 provide response rates for these subgroups. For each subgroup analysis, we present cross-tabulations and the associated statistical test results.

Table 3-5: Response rates by race and ethnicity

	Wave 1				Wave 2				Wave 3				All waves [b]			
	Non-White	p-value	White	p-value	Non-White	p-value	White	p-value	Non-White	p-value	White	p-value	Non-White	p-value	White	p-value
Treat1	81.05		94.20		73.68		86.96		67.37		85.51		74.04		88.89	
Treat2	91.84	0.011	93.94	0.998	89.80	0.007	84.85	0.934	80.61	0.089	80.30	0.684	87.41	0.009	86.36	0.843
Treat3	93.75		94.12		86.46		85.29		77.08		80.88		85.76		86.76	
Conditional	81.05	0.003	94.20	0.961	73.68	0.002	86.96	0.717	67.37	0.034	85.51	0.386	74.04	0.002	88.89	0.562
Unconditional	92.78		94.03		88.14		85.07		78.87		80.60		86.60		86.57	

Treat1=Conditional; Treat2= Unconditional (per questionnaire); Treat3=Unconditional (all-questionnaire). [b] SEs are cluster-robust.

Table 3-6: Response rates by citizenship status

	Wave 1				Wave 2				Wave 3				All waves [b]			
	Inter.	p-value	Dom.	p-value	Inter.	p-value	Dom.	p-value	Inter.	p-value	Dom.	p-value	Inter.	p-value	Dom.	p-value
Treat1	80.00		90.38		71.67		83.65		65.00		80.77		72.22		84.94	
Treat2	96.49	0.005	90.65	0.626	84.21	0.113	89.72	0.431	78.95	0.237	81.31	0.951	86.55	0.046	87.23	0.769
Treat3	94.03		93.81		85.07		86.60		73.13		82.47		84.08		87.63	
Conditional	80.00	0.001	90.38	0.597	71.67	0.037	83.65	0.263	65.00	0.125	80.77	0.815	72.22	0.016	84.94	0.474
Unconditional	90.22		92.16		84.68		88.24		75.81		81.86		85.22		87.42	

Dom.=Domestic; Treat1=Conditional; Treat2= Unconditional (per questionnaire); Treat3=Unconditional (all-questionnaire). [b] SEs are cluster-robust.

Table 3-7: Response rates by admission year

	Wave 1				Wave 2				Wave 3				All waves [b]			
	Before FA 23	p-value	FA 23	p-value	Before FA 23	p-value	FA 23	p-value	Before FA 23	p-value	FA 23	p-value	Before FA 23	p-value	FA 23	p-value
Treat1	81.54		89.90		78.46		79.80		75.38		74.75		78.46		81.48	
Treat2	92.75	0.006	92.63	0.781	89.86	0.083	86.32	0.473	85.51	0.334	76.84	0.841	89.37	0.073	85.26	0.657
Treat3	98.15		91.82		90.74		83.64		79.63		78.18		89.51		84.55	
Conditional	81.54	0.003	89.90	0.502	78.46	0.026	79.80	0.267	75.38	0.216	74.75	0.587	78.46	0.022	81.48	0.373
Unconditional	95.12		92.20		90.24		84.88		82.93		77.56		89.43		84.88	

Treat1=Conditional; Treat2= Unconditional (per questionnaire); Treat3=Unconditional (all-questionnaire). [b] SEs are cluster-robust.

Table 3-8: Response rates by academic major

	Wave 1				Wave 2				Wave 3				All waves [b]			
	Low signup	p-value	High signup	p-value	Low signup	p-value	High signup	p-value	Low signup	p-value	High signup	p-value	Low signup	p-value	High signup	p-value
Treat1	80.95		92.50		72.62		86.25		67.86		82.50		73.81		87.08	
Treat2	91.36	0.045	93.98	0.743	82.72	0.120	92.77	0.359	75.31	0.375	85.54	0.658	83.13	0.092	90.76	0.569
Treat3	92.31		95.35		84.62		87.21		76.92		80.23		84.62		87.60	
Conditional	80.95	0.013	92.50	0.501	72.62	0.042	86.25	0.389	67.86	0.167	82.50	0.947	73.81	0.031	87.08	0.561
Unconditional	91.82		94.67		83.65		89.94		76.10		82.84		83.86		89.15	

Treat1=Conditional; Treat2= Unconditional (per questionnaire); Treat3=Unconditional (all-questionnaire).[b] SEs are cluster-robust.

Table 3-9: Response rates by gender

	Wave 1				Wave 2				Wave 3				All waves [b]			
	Male	p-value	Female	p-value	Male	p-value	Female	p-value	Male	p-value	Female	p-value	Male	p-value	Female	p-value
Treat1	77.78		90.91		75.93		80.91		74.07		75.45		75.93		82.42	
Treat2	86.44	0.166	96.19	0.157	81.36	0.680	91.43	0.064	71.19	0.494	85.71	0.149	79.66	0.419	91.11	0.046
Treat3	90.16		96.12		81.97		88.35		80.33		77.67		84.15		87.38	
Conditional	77.78	0.071	90.91	0.054	75.93	0.382	80.91	0.024	74.07	0.803	75.45	0.187	75.93	0.303	82.42	0.041
Unconditional	88.33		96.15		81.67		89.90		75.83		81.73		81.94		89.26	

Treat1=Conditional; Treat2= Unconditional (per questionnaire); Treat3=Unconditional (all-questionnaire). [b] SEs are cluster-robust.

Beginning with differences by race and ethnicity, the overall response rate among White participants is found to be 5 percentage points higher than among Non-White participants (87.36% vs. 82.47%), and the difference is significant (p-value = 0.061). However, the results in Table 3-5 reveal that the two unconditional incentives are most effective among Non-White participants and have no discernible effects among White participants. Specifically, on average across all waves, the two unconditional incentives boosted the response rate among Non-White participants by 12 percentage points (p-value = 0.002). This difference is also significant across individual survey waves. As observed in the overall patterns, among Non-White participants, a $15 unconditional lump sum payment is more effective than a $5 unconditional per-questionnaire payment in the first wave. However, in the second and third waves, the $5 unconditional payment is more effective than the $15 lump sum payment in increasing response rates, but the differences between the two unconditional incentive groups did not reach a significant level.

Table 3-6 presents response rates by citizenship status. On average across all waves, the response rate among domestic students is found to be 5.6 percentage points higher than that of international students (86.58% vs. 80.98%; p-value = 0.005). Notably, the two unconditional incentives are most effective in boosting response rates among international students and did not produce any significant effects among domestic students. Across all waves, the two unconditional incentives increased the response rate among international students by an average of 13 percentage points (p-value = 0.016), and this positive effect among international students persists across individual survey waves.

With respect to differences by admission year (Table 3-7), overall response rates are not significantly different between those admitted before Fall 2023 (85.64%) and those admitted in

Fall 2023 (83.77%). However, the two unconditional incentives have a stronger and significant impact among those admitted before Fall 2023 compared to those admitted in Fall 2023. Specifically, among those admitted before Fall 2023, the two unconditional incentives increased the overall response rate by an average of 11 percentage points compared to the conditional incentive group (p-value = 0.022), while no significant effect was found for those admitted in Fall 2023. The unconditional incentive effect among those admitted before Fall 2023 is also observed in individual survey waves. It is worth highlighting that, among those admitted before Fall 2023, a $15 unconditional lump sum incentive had a larger effect in the first wave and a smaller effect in the third wave than a $5 unconditional per-questionnaire incentive, with both unconditional incentives having an equal effect in the second wave.

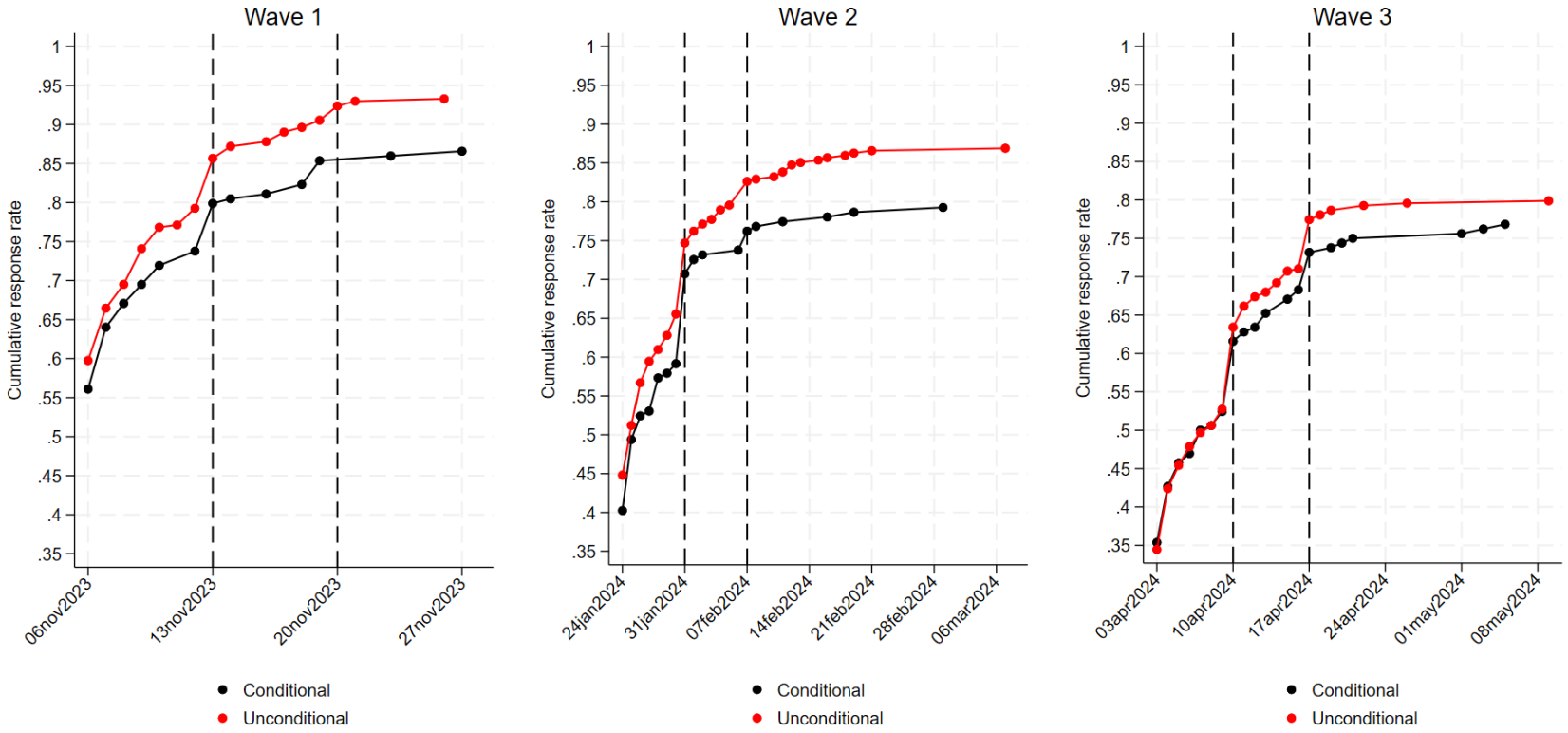
Table 3-8 presents response rates by academic major, revealing that participants from high-signup academic units (i.e., Architecture, Urban Planning, Environment and Sustainability, Education, Medicine, and Public Policy) completed the longitudinal surveys at a significantly higher rate than participants from low-signup academic units (i.e., Dentistry, Engineering, Kinesiology, Public Health, Science, Social Science, Language, and Art). The overall response rate among high-signup academic units is 88.49%, compared to 80.38% among low-signup units, and this difference is significant (p-value = 0.001). However, our results also indicate that the two unconditional incentives are most effective among participants from low-signup academic units. Among low-signup academic units, the overall response rate is about 10 percentage points higher in the two unconditional incentive groups than in the conditional group (83.86% vs. 73.81%; p-value = 0.031). In contrast, among high-signup academic units, the unconditional incentives have no significant effect. Similar to previous observations, we found no significant difference in response rates between the two unconditional incentive groups.

Table 3-9 presents response rates by gender. The average response rate across all waves is about 7 percentage points higher for females than for males (86.90% vs. 80.08%; p-value = 0.020). Unconditional incentives appear to be effective for both gender groups, but their effect does not reach a significance level for males, largely due to their relatively small sample size. However, the effect is significant for females (p-value = 0.041).

3.5.5 Cumulative response rates

To analyze how the response rates changed over time in each of the survey waves, we provide cumulative response rates by survey wave in Figure 3-2. The dotted vertical lines represent the dates reminders were sent out. Figure 3-2 reveals two key observations. First, across all survey waves, the difference in response rates between the unconditional and conditional incentives tended to widen immediately after the first reminder was sent out. Second, in wave 3, the effect of the unconditional incentives was realized only after the first reminder was sent out. As previously mentioned, the timing of the wave 3 launch and fielding coincided with the last weeks of the winter semester when students were busy preparing for finals and graduation. This could potentially explain why the unconditional incentives were effective only after the reminder was sent out and why the unconditional incentive effect was weaker in wave 3 than in waves 1 and 2.

Figure 3-2: Cumulative response rates by survey wave



3.5.6 Cost per completed survey

To complement the response rate analysis, we also examined the overall incentive costs and cost-efficiency of each treatment group. The total incentive cost for the Conditional group was $1,975, compared to $2,460 each for the two Unconditional groups. The Conditional group, where respondents received payment only after completing each survey, yielded 395 completed person-surveys across all three waves, resulting in a cost of $5.00 per completion. The $5 Unconditional (per questionnaire) group generated 428 completed person-surveys, translating to $5.75 per completion. The $15 Unconditional (all-questionnaire) group produced 424 completions, or $5.80 per completion.

3.5.7 Secondary analysis

In addition to analyzing survey nonresponse as our primary outcome variable, we also examined item nonresponse across the different incentive conditions as our secondary outcome variable. However, we did not find significant or systematic differences in item nonresponse by incentive groups (results not shown).

Additionally, we conducted supplementary analyses comparing cumulative estimates of key estimates among the three treatment conditions. Appendix Figure B-1 and Appendix Figure B-2 present the cumulative means of two key variables—self-reported physical and mental health—separately for each treatment condition. The graphs indicate that differential response rates across treatment conditions do not result in differing distributions for these two key variables. To explore whether selective attrition occurred based on background information available in the frame data, Appendix Figure B-3 presents the sample composition of international students, female students, students majoring in fields with low study take-up rates, and students admitted before Fall 2023, by treatment condition over time. The results show no

discernible evidence of selective attrition based on these background characteristics from the administrative data.

3.6 Discussion and Conclusion

Longitudinal data collection utilizing probability-based online panels offers several methodological and practical advantages over conventional offline panel studies. Despite this inherent strength, nonresponse and attrition pose significant challenges. Given the increasing use of online longitudinal data for social science research and policy decision-making, it is essential to collect valid and accurate online population data. Most studies examining the effects of various incentive strategies in longitudinal data collections have primarily focused on traditional offline longitudinal studies, where respondents are interviewed annually or biannually face-to-face or using sequential mixed modes. Comparable incentive studies analyzing nonresponse and attrition outcomes in the online context have been limited. Additionally, due to fundamental differences between online and offline panel studies, effective incentive strategies used in offline contexts can have limited applications in online contexts.

In this study, we report findings from a randomized controlled trial that evaluated the causal impact of three incentive payment plans specifically tailored for high-frequency online longitudinal studies in the U.S. While the incentive plans we evaluated were informed by prior studies and existing theories, they are distinct from those implemented in offline panel studies. Our findings contribute to the existing literature on the use of incentives in survey research, providing valuable insights for improving response rates in online longitudinal studies and informing future research.

Our study found that unconditional incentives are more effective than conditional incentives in increasing response rates (Hypothesis 1). The effects we found are consistent with

the previously stated theoretical expectations. We found no significant difference in response rates between the $5 and $15 unconditional incentives. Additionally, we identified substantial variations in study participation rates and unconditional incentive effects across subgroups based on race and ethnicity, citizenship status, admission year, academic major, and gender (Hypothesis 2). These findings have practical implications for future online longitudinal data collection efforts, highlighting the need for tailored incentive strategies to improve response rates and sample representation.

Below, we discuss the theoretical implications of our findings using existing theories as a framework for interpreting the various outcomes observed in our data. We also discuss the practical implications of our results for future studies in online longitudinal data collection, make suggestions for future research directions, and highlight the strengths and weaknesses of our study approach.

Our study's participant recruitment rate of 11.09% is consistent with those found in other probability-based online panels representative of U.S. adults aged 18 and above [11]. Our results also reveal substantial variations in participation rates across subgroups. In particular, we found that non-Whites, international students, those admitted before Fall 2023, those majoring in Dentistry, Engineering, Kinesiology, Public Health, Science, Social Science, Language, and Art, and male students participated at significantly lower rates. Additionally, our study found that unconditional incentives are most effective among these subgroups with low survey motivation and response propensities. This is consistent with the predictions of the Leverage-Salience Theory (Groves et al., 2000), which posits that survey features such as topics and incentives have

[11] https://uasdata.usc.edu/index.php?r=eNpLtDK0qi62MrFSKkhMT1WyLrYytFwwskuTcjKT9YpSk4tKM0tyU_NKlKxrASjBDmQ,

differential leverage on the participation decisions of participants based on their background characteristics. In this framework, unconditional incentives are expected to be more effective when made salient to subgroups with lower response propensities or lower intrinsic motivation to participate, as the incentive can act as a substitute for weak intrinsic interest or perceived relevance of the survey.

Furthermore, the heterogeneous findings align with the Economic Exchange Theory (Ajzen & Fishbein, 1980; Biner & Kidd, 1994), which suggests that potential respondents weigh the costs and benefits of participation when deciding whether to complete a survey. From this perspective, incentives are most effective when they sufficiently compensate for the time and effort required to respond. The theory implies that survey participation is more attractive to individuals with constrained economic resources, particularly international students, as observed in this study.

In addition to incentives, the relevance of (and interest in) survey topics also play crucial roles in influencing survey participation decisions as implied by the Leverage-Salience Theory (Groves et al., 2004; Ryu et al., 2006). Our study's focus on the mental well-being of students might explain our finding that the study attracted a larger number of participants majoring in Architecture, Urban Planning, Environment and Sustainability, Education, Medicine, and Public Policy. These fields often emphasize the societal importance of mental health and general well-being in their curricula and professional practices, making the survey topic particularly relevant and engaging for these participants. Consequently, the alignment between the survey topic and the interests of these academic majors likely contributed to the higher participation rates observed among these subgroups.

Taken together, these theoretical frameworks help explain the heterogeneous treatment effects observed in our study and underscore the importance of tailoring incentive strategies to target specific subgroups based on their level of motivation.

Our results show that the $5 and $15 unconditional incentives yielded nearly identical response rates in the first wave of data collection, with the response rate for the $15 unconditional incentive only slightly higher than that for the $5 incentive. However, this difference was not statistically significant. One explanation for this outcome is the principle of diminishing returns from economic exchange theory, which suggests that after a certain point, increasing the incentive amount has a progressively smaller effect on response rates. Essentially, once an incentive sufficiently compensates participants for their time and effort, further increases yield minimal additional impact. Another explanation is that for the $5 unconditional incentives, respondents only had to do one survey, whereas for the $15, they had to do three surveys, potentially reducing the impact of the $15 unconditional incentives in the first wave.

In contrast to economic exchange theory, social exchange theory offers an alternative explanation. This theory suggests that even small incentives can be highly effective because they serve as a symbol of trust from the researcher. By providing a prepaid incentive, the researcher conveys a gesture of appreciation and trust, expecting that the respondent will reciprocate by completing the survey. Thus, both the $5 and $15 unconditional incentives effectively communicate this trust, making the additional $10 in the larger incentive less influential. This insight highlights that the perception of trust and reciprocity can be as crucial as the monetary value of the incentive in influencing response rates, underscoring the need to balance economic and social considerations in formulating survey incentive strategies.

In the second and third waves of data collection, we found that the response rates for the $5 unconditional incentive were higher than those for the $15 unconditional incentive, which had been provided as a lump sum in the first wave. Since the $15 incentive covered all three waves and was not repeated in the second and third waves, its initial saliency diminished over time. However, the difference in response rates between the $5 and $15 unconditional incentives in the second and third waves was not statistically significant.

This pattern can be explained by both economic exchange theory and social exchange theory. Economic exchange theory suggests that the upfront $15 unconditional incentive, while initially impactful, loses its effectiveness when not reinforced in subsequent waves, as participants no longer perceive a direct benefit for continued participation. On the other hand, social exchange theory highlights the importance of ongoing gestures of trust and appreciation. The $5 unconditional incentive, given before each wave, continually reinforces the researcher's trust and expectation of reciprocity, maintaining a higher level of engagement over time. Another possible explanation is that the repeated delivery of the $5 incentive also served as a timely reminder of the study over time, in addition to invoking the expected reciprocity. Although the periodic reminder and trust components are conceptually distinct, their independent effects are difficult to disentangle empirically in the context of this study design.

Therefore, the consistent application of the $5 unconditional incentive in each wave sustains its saliency and effectiveness, while the one-time $15 incentive's impact wanes without repetition. Additionally, the finding that response rates for the $15 unconditional incentive group were higher than those for the conditional incentive group in the second and third waves suggests a carryover effect of the unconditional incentive well beyond its application in the first wave.

This finding underscores the importance of both immediate and sustained incentive strategies to maintain high response rates in high-frequency online longitudinal data collections.

Consistent with prior studies (Crawford et al., 2001), our results indicate that the timing of a survey is an important factor in predicting survey completion among student populations. Our results also indicate that providing unconditional incentives has no negative effects on data quality in online longitudinal data collections, as measured by item nonresponse. This finding is consistent with prior research in cross-sectional and offline longitudinal studies (Booth et al., 2024; Cabrera-Álvarez & Lynn, 2023; Jäckle & Lynn, 2008; Ryu et al., 2006; Singer & Ye, 2013; Wong, 2020).

Our findings have practical implications for panel management, nonresponse, and attrition in future online longitudinal data collection efforts. Researchers must weigh the trade-off between cost and data quality. Given the effectiveness of unconditional incentives demonstrated in this study, researchers need to balance the benefits of increased response rates against the higher costs associated with providing these incentives. Additionally, assuming unconditional incentives are cost-effective, researchers must decide whether to use a lump-sum payment for all questionnaires or a per-questionnaire-based prepaid incentive. As found in this study, both methods are equally effective in boosting response rates across survey waves; however, the per-questionnaire approach can be more effective than the lump-sum payment in later waves if the data collection period exceeds a certain duration (e.g., one year), as the effectiveness of a lump-sum payment can decline over time. On the other hand, a per-questionnaire-based approach can be difficult to administer for studies with frequent data collection, such as daily or weekly surveys, and will likely incur additional costs associated with processing incentive payments in waves beyond the first. Thus, the choice between a lump-sum

and a per-questionnaire-based unconditional incentive will depend on the frequency and duration of data collection, as well as logistical and budgetary considerations.

Our study findings have implications for adaptive design approaches in future online longitudinal surveys. Our observational finding indicates that sending email reminders is effective in maintaining respondent engagement and increasing response rates, especially during busy periods and among respondents who received unconditional incentives. In the context of online data collection, respondents who consistently skip surveys can be easily and immediately identified. For those who do not respond consistently, additional reminders such as text messages or telephone calls can be employed to encourage participation. Moreover, our results show that the study attracted a larger number of participants for whom the survey topics were motivating and relevant. To engage respondents for whom the survey topics are less relevant, future studies should consider designing tailored communication strategies during recruitment, assuming these individuals can be easily and sufficiently identified using available frame data, as in our study. These strategies could include personalized messages that emphasize the broader significance of their participation and the potential impact of the survey results to enhance engagement among these respondents.

Our findings that unconditional incentives effectively increase response rates can also inform future adaptive designs. In particular, future studies could consider using differential incentives to target groups with low socioeconomic characteristics, such as low-income individuals, ethnic minorities, and those with lower education levels, to reduce participation barriers among these subgroups. Tailoring incentives in this way can help address specific challenges faced by these groups and improve overall response rates. Additionally, the availability of information from prior survey waves—a key strength of longitudinal studies—

should be leveraged in the adaptive design process of future studies. By analyzing respondents' behavior in previous waves, researchers can identify those who are less likely to participate consistently and implement targeted interventions, such as enhanced reminders or additional incentives, to encourage their continued participation. This adaptive design approach not only maximizes response rates but also enhances sample representation and potentially reduces sample selection bias.

3.6.1 Future Research Directions

Building on the methodology and findings from this study, future research should implement the same incentive design in a population-representative sample with a sufficiently large sample size. However, most panel members in probability-based panels, such as those in the Understanding America Study (UAS), are accustomed to receiving postpaid incentives. Additionally, online panels based on student populations may differ from commercial or professional panels regarding participant motivation, response behavior, and incentive sensitivity. As a result, the relative effectiveness of unconditional prepaid incentives observed in this student sample may not fully generalize to these settings. Future research should examine whether prior panel experience and expectations about incentive timing in probability-based panels moderate the effectiveness of unconditional incentive found in this study.

Since various online studies collect longitudinal data at different frequencies, future research should extend the analysis to settings that collect data at different intervals (e.g., daily, weekly, biweekly, monthly) and analyze if the relationship between unconditional incentives and nonresponse outcomes is moderated by data collection frequency. Future studies can also experimentally analyze the efficacy of a hybrid incentive plan where prepaid incentives are used in some survey waves and post-paid incentives in others. This hybrid approach has the potential

to deliver similar results to those found in this study but at a lower cost. A special case of the hybrid design involves alternating between prepaid and postpaid incentives across waves. Additionally, future research should collect more auxiliary and para data similar to Becker (2023) to uncover the underlying mechanisms influencing observed nonresponse and attrition outcomes in online longitudinal data collections. Future studies should also analyze the extent of attrition biases in online longitudinal data and assess the efficacy of unconditional incentives in mitigating potential selection biases. In addition to these directions, future work should consider improvements to the study design based on lessons learned from this experiment. For instance, extending the data collection period beyond three waves would enable a better understanding of long-term panel retention and the sustainability of participation under different incentive structures. Another promising avenue for design improvement is to provide larger incentives to specific subgroups with chronically low response propensities at targeted waves—an approach commonly used and demonstrated to be effective in offline panels—as a way to reduce attrition and improve data quality.

While unconditional incentive strategies may be more expensive in the short term, as shown in this study, they may prove more cost-effective over time by reducing the need to recruit replacement sample members and by improving sample representativeness and data quality. A promising direction for future research is to examine whether unconditional incentive payments help retain panel members over longer time horizons and ultimately offset their initial higher cost. Retaining original panelists not only limits recruitment expenses but may also reduce the risk of attrition-related bias by maintaining a more stable and representative sample. Although this study focused on short-term participation across three waves, future work should

assess the cumulative cost-benefit of unconditional incentive strategies in extended online longitudinal data collection efforts.

3.6.2 Strength and Limitations

The main strength of this study lies in its use of experimental data, which enables us to make causal inferences and provide causal estimates. Another strength is the availability of background information for all respondents, which allowed for the exploration of treatment heterogeneity. Additionally, the use of a fresh panel mitigates panel conditioning effects, where previous interviews potentially affect future survey responses, and panel maturity effects, where attrition in prior waves results in a sample of committed and loyal respondents who are less sensitive to incentives (Kraemer et al., 2024; Wong, 2020). However, a significant limitation is that the sample is restricted to master's students enrolled in non-professional degree programs at public universities in the U.S., with an overrepresentation of individuals who are younger and have higher levels of education. As such, the findings should be interpreted with caution. . Despite this limitation, the study provides valuable insights into the effectiveness of different incentive structures in improving response rates in online longitudinal panel surveys. Additionally, the directions of the overall and heterogeneous treatment effects, as well as other findings observed in this study, are consistent with broader theoretical predictions and prior research findings. This consistency suggests that the same results could be replicated in a population-representative sample, although further empirical investigation is needed. Another limitation of this study is the lack of sufficient statistical power to detect potential variations in the outcome variable among all subgroups. In addition, this study did not assess whether unconditional incentives influenced the honesty of responses to sensitive questions—such as citizenship status, undergraduate GPA, Pell Grant receipt, URM status, and first-generation

status—because this information was obtained directly from administrative records rather than through self-reports. While social exchange theory suggests that unconditional incentives might enhance honest reporting, this potential effect could not be evaluated in the current study. While the timing of wave 3, coinciding with the end of the Winter academic semester, is a plausible explanation for the decline in response rates, particularly among the unconditional incentive groups, additional factors such as panel fatigue may also have contributed to this decline. However, these independent effects cannot be disentangled within the context of the current study design. Despite these limitations, the study provides insights into the effectiveness of using unconditional incentives at various time points to reduce survey nonresponse and attrition in online longitudinal data collections, an area that has received little attention so far.

Chapter 4 Analyzing the Causal Effect of Survey Frequency on Nonresponse in Probability-Based Online Panels Among New Panel Respondents

4.1 Introduction

Probability-based Internet panels offer researchers the ability to collect high-frequency longitudinal data. This approach has a methodological advantage, as collecting information closer in time to the occurrence of important events, and collecting it repeatedly reduces the role of measurement error and omitted variables. Internet technology also allows for quick responses to new and emerging developments. For instance, during the COVID-19 pandemic, online panels have proven valuable in monitoring daily mood changes, observing changes in infection rates, job loss, food insecurity, and vaccination rates over time, and exploring the evolving correlation between COVID-19 vaccines and mental distress in the U.S. adult population (de la Haye et al., 2023; Kapteyn et al., 2020; Perez-Arce et al., 2021).

Despite these inherent strengths, frequent data collection may increase the level of burden on respondents, which can negatively impact data quality, primarily through higher nonresponse and attrition (Crawford et al., 2001; Rolstad et al., 2011; Yan et al., 2020; Yan & Williams, 2022). Both nonresponse and attrition reduce the sample size and decrease the precision of survey estimates. Given that data are collected more frequently, even a modest attrition rate in each survey wave can lead to a significant loss of sample members over time in a short period of time. Additionally, if the characteristics of panelists who drop out differ from those who continue participating in terms of key survey statistics, attrition may introduce bias (Chin et al., 2021; Hansson et al., 2018; Lugtig, 2014; Lugtig et al., 2014; Ward et al., 2017).

Survey data are a major source of policy information in the United States. However, traditional methods of data collection, such as mailed questionnaires, face-to-face and telephone interviews, face challenges in declining response rates and rising cost (de Leeuw et al., 2018; Dutwin & Buskirk, 2021; Luiten et al., 2020; Williams & Brick, 2018). Probability-based online panels have emerged as an attractive option due to their flexibility, convenience, speed, cost-effectiveness, and statistical efficiency. Given the increasing use of online data for policy decision-making and social science research (Alattar et al., 2019), understanding whether there is a causal relationship between survey frequency and nonresponse is essential for addressing nonresponse and enhancing the quality of online data.

In this study, we report findings from a randomized controlled trial (RCT) in which we recruited about 2,000 new online panelists and randomly assigned them in equal numbers to either a low- or high-frequency group. The low-frequency group received one core survey invitation every four weeks, while the high-frequency group received one core survey invitation every two weeks. Both groups received the same series of surveys. Importantly, the survey length and content were the same for both experiment groups to maximize comparability. The experiment was conducted over a one-year duration, allowing sufficient time to observe and analyze the impact of survey frequency on nonresponse. We focus on new respondents because existing data indicate that the majority of attrition occurs during the first year of panel membership. The study's use of experimental data addresses selectivity issues affecting previous studies that used observational data. Our experiment is embedded within the Understanding America Study (UAS), a probability-based online panel representative of U.S. adults aged 18 and older.

Additionally, the availability of detailed background information on all respondents allows us to control for other factors that predict survey response, thereby enhancing statistical inferences. It also enables the exploration of whether the frequency effect differs across subgroups based on background information, which can inform targeted interventions and enhance panel management and retention strategies. In summary, this study combines experimental data with detailed background information and longitudinal survey data to answer the following research questions:

RQ1: Is there a causal relationship between survey frequency and nonresponse among new respondents in online panel surveys? If so, what is the direction of this causal effect?

RQ2: Does the causal effect of survey frequency on nonresponse vary across socioeconomic and racial subgroups?

4.2 Background

4.2.1 Survey Frequency as a Measure of Respondent Burden

In his burden model, Bradburn (1978) identified survey frequency as a factor contributing to respondent burden. Survey frequency has been operationalized in various ways across different data collection contexts. For instance, McCarthy et al. (2006) examined the relationship between survey frequency and response rates in cross-sectional agricultural surveys, where different agricultural operations had varying probabilities of being sampled, with some entities more likely to be sampled due to their rare characteristics and larger size. They found only a negligible relationship between the frequency of survey invitations and response rates, partly due to insufficient variations in survey frequency.

Eggleston (2024) analyzed survey frequency effects using cross-sectional household survey data collected by the U.S. Census Bureau from 2015-2020. He found that households

sampled for the American Community Survey (ACS) or the Current Population Survey (CPS) between 2015 and 2019 exhibited lower response rates to the 2020 U.S. Census, with the effect varying by the time gap between survey requests. Specifically, the negative effect was significantly larger when the gap between the ACS/CPS and the 2020 Census was shorter. Additionally, the ACS showed a stronger negative effect than the CPS, likely due to its greater similarity to the Census.

These findings suggest that survey fatigue and confusion are primary mechanisms driving the negative survey frequency effects in cross-sectional data collections. Households sampled for multiple surveys may feel overburdened by repeated requests (survey fatigue), particularly when surveys are closely spaced. The greater similarity between the ACS and the Census--both mandatory and use web and mail modes--likely contributed to higher levels of confusion, as respondents may have mistakenly assumed they were part of the same study. In contrast, the CPS, which is voluntary and relies on in-person interviews, generated less confusion. Similar survey fatigue effects have been documented by Porter et al. (2004), while confusion effects were reported by Sinibaldi and Karlsson (2017), both within the context of cross-sectional data collection.

In longitudinal surveys, most studies have a fixed periodicity (typically annual or biannual), with no variation in frequency to exploit. In such a scenario, one must rely on cross-study comparisons to analyze the relationship between survey frequency and response rate in longitudinal surveys. For instance, Zabel (1998) compares PSID and SIPP and finds that attrition is higher in the SIPP. During the observation period, PSID respondents were interviewed once a year and SIPP respondents three times a year. However, the cross-study comparison approach

makes it challenging to isolate the independent effects of survey frequency, as studies often differ not only in survey frequency but also in other survey and study characteristics.

As some longitudinal studies occasionally field between-wave supplemental surveys for which only subsets of the full sample are eligible, another approach is to compare the response rates of respondents eligible for these supplemental surveys with those who are not. For example, Earp et al. (2022) examine variations in survey frequency within the Current Population Survey (CPS), a monthly longitudinal survey of US households, to estimate the effect of survey frequency on future response rates. The CPS periodically fields supplementary surveys, for which only subsets of the entire CPS sample are eligible. They found that increased survey frequency predicts future survey completion only when it interacts with respondents' characteristics.

Similarly, Sastry et al. (2021) exploit variations in surveying frequency generated by the introduction of the Child Development Supplement (CDS) in the PSID in 1997 to analyze the association between survey frequency and attrition in later PSID waves from 1997 to 2015. The CDS was a major inter-wave supplemental study that added additional survey burden on a subset of PSID families. The authors found no systematic association between participation in the CDS and post-1997 attrition, except for CDS families with the youngest children.

Additionally, a study by Pashazadeh et al. (2021) uses a quasi-experimental approach to analyze the effects of the additional survey burden caused by the collection of biomarker data in wave 2 of the UK Household Longitudinal Study (UKHLS), on survey cooperation in waves 3-7 (McFall, 2013). They find that the odds of cooperating in wave 3 were significantly reduced for respondents who were randomly sampled for biological data collection in wave 2 compared to those who were not sampled, but this had no effect on cooperation in waves 4-7.

Beyond conventional longitudinal studies, other types of high-burden data collection methods, such as Ecological Momentary Assessments (EMAs), provide additional insights into how survey frequency affects respondent cooperation. EMA surveys collect real-time data on participants' behaviors, emotions, or experiences in their natural environments, typically by prompting respondents with brief surveys up to six times a day. Several studies have examined the relationship between survey frequency and response rates in EMA studies and found no significant correlation between survey frequency and completion rates (Eisele et al., 2022; Ono et al., 2019; Stone et al., 2003; Walsh & Brinker, 2015). However, the cross-study comparison approach makes it challenging to disentangle the independent effects of survey frequency.

Online panel surveys provide another setting to examine how different aspects of survey burden—such as questionnaire length and survey frequency—affect response rates. Jin and Kapteyn (2022) explored the effect of survey burden, measured by either the length of the most recent questionnaire (defined as the number of screens shown to a respondent) or the number of survey invitations (survey frequency) in the year preceding a new survey, on the response probability to a new panel survey. They identified a positive relationship between questionnaire length and subsequent response probability for about one third of the sample and a negative relation for only 3% (the remainder of the sample did not show an effect of the length of the preceding questionnaire). About 11% of the sample exhibited a negative relationship between the number of previous survey invitations and the probability of responding to a survey. The relationship was positive for 9% of the sample, while no significant association was found for the remainder. The authors conclude that the positive relationship between questionnaire length and response probability may be due to the fact that respondents get paid per minute of question time, so that a longer questionnaire earns respondents more money.

While previous studies provide valuable insights into the burden effects of frequent surveying, using observational data to analyze the relationship between survey frequency and response rates can introduce significant selection bias. Moreover, the context of online panel surveys differs substantially from that of cross-sectional and traditional longitudinal surveys, making it challenging to generalize findings from non-online contexts to online panels.

The mechanisms driving survey frequency effects may also vary across different data collection contexts. For instance, confusion effects commonly reported in cross-sectional studies (Eggleston, 2024; Sinibaldi & Karlsson, 2017), where respondents mistakenly perceive multiple surveys as part of the same study, are unlikely in online panels. Respondents in these panels are typically aware that each invitation corresponds to a distinct survey. Moreover, frequent surveying in online panels may foster positive habits and enhance engagement and commitment among participants. Frequent surveys may also integrate into respondents' routine and lead to sustained participation, rather than survey fatigue as observed in non-online panel settings (Engstrom & Sinibaldi, 2024; Porter et al., 2004). Additionally, as online panel respondents are typically compensated based on their time, receiving more frequent surveys offers greater earning opportunities. These differences underscore the urgent need to design experimental studies specifically tailored to the unique context of online panels to analyze the causal effects of survey frequency.

4.2.2 Use of Fresh Panel Respondents

Our study focuses on fresh panel respondents, motivated by three key considerations. First, previous studies have shown that prior survey experiences are predictive of cooperation in subsequent surveys, indicating that the early survey experiences of fresh respondents play a

crucial role in their continued participation (Galesic, 2006; Lepkowski & Couper, 2002; Watson & Wooden, 2009).

Second, prior research indicates that individuals who lack a strong commitment to survey participation are more likely to drop out quickly from online panels, often after participating in only a few initial survey waves or for a short duration. These fast attritors can make up a substantial proportion of the panel sample, typically ranging from 20% to 30% (Kapteyn et al., 2024; Lugtig, 2014). Importantly, prior studies have found that individuals who drop out during the first membership year have distinct demographic, psychological and psychosocial characteristics compared to those who leave in subsequent years (Hansson et al., 2018; Ward et al., 2017). On the other hand, loyal panel members (late or slow attritors) are more likely to discontinue their participation as a result of cumulative burden or panel fatigue after an extended period (Lugtig, 2014).

Third, the use of a fresh panel mitigates panel conditioning effects, where previous interviews potentially affect future survey responses, and panel maturity effects, where attrition in prior waves results in a sample of committed and loyal respondents in later waves, which may limit the effectiveness of interventions (Kraemer et al., 2024; Wong, 2020).

In summary, this study utilizes experimental data to address selection bias and analyze the causal effects of survey frequency. It also utilizes rich background information available for all respondents to enhance statistical inferences and explore treatment heterogeneity. By focusing on fresh panel respondents, this study provides new insights into how frequent surveying uniquely and independently influences survey completion in online panel contexts, as well as the population subgroups that are most responsive to frequent surveying.

4.3 Methods

4.3.1 Understanding American Study (UAS)

The context for this study is the Understanding America Study (UAS). Established at the University of Southern California in 2014, the UAS is a nationally representative, probability-based internet panel of US-households of approximately 15,000 respondents aged 18 and above. Respondents answer surveys on a computer, tablet, or smart phone, wherever they are and whenever they wish to participate. Surveys are conducted in English and Spanish. Panel members are recruited through address-based sampling. Prior access to internet is not a pre-requisite to be in the panel; respondents without prior internet access are provided with a computer tablet and broadband internet subscription. Respondents answer surveys once or twice a month covering various substantive topics. Partly as a result of this, the UAS comprises a vast amount of background information on its respondents, including extensive measures of physical and mental health, income, labor force participation, cognitive functioning, and demographics. Recruitment for this study began on January 9, 2024 and ended on March 8, 2024. Recruitment materials were mailed to a total of 19,231 addresses across the U.S. Of these, 2,057 respondents became panel members, yielding an overall recruitment rate (ORR) of approximately 10.70% [12]. Additional details on overall recruitment rates by individual recruitment batches are available on

[12] Overall recruitment rate (ORR) is the product of three individual response rates: Initial Response Rate (IRR)* Recruitment Rate (RECR) * Profile Rate (PROR). All rates conform to AAPOR specifications. The initial response rate is based on AAPOR RR1. Please see the following web site for a typical calculation of the overall recruitment rate: https://uasdata.usc.edu/page/Response+And+Attrition+--+MSG+2015-07+Nationally+Representative+Batch+1

the UAS website [13]. The UAS website [14] also provides complete information about UAS methodologies, including recruitment protocols, response and retention rates.

4.3.2 Experimental Design

We randomly assigned half of the new sample to the high-frequency group (n=1,029), which receives one core survey invitation every two weeks, and the remaining half to the low-frequency group (n=1,028), which receives one core survey invitation every four weeks. The high-frequency group represents current UAS respondents who typically receive one survey invitation every two weeks. Core surveys are a series of longitudinal surveys that the UAS administers to the same respondents every two years. They form the backbone of the UAS, and are made publicly available as comprehensive files [15]. Each core survey takes no more than 30 minutes, with identical individual survey lengths in both conditions. Both the low- and high-frequency groups receive an identical sequence of core surveys covering the same topics. The observation period for this experiment was one year. By the end of that period, participants in the high-frequency group completed 22 core surveys, while those in the low-frequency group completed 12 core surveys. Both groups received an equal number of non-core surveys during that time.

In addition to the core surveys, the UAS regularly administers one-off cross-sectional non-core surveys. All respondents, regardless of their assigned treatment condition, receive

[13] https://uasdata.usc.edu/index.php?r=eNpLtDK0qi62MrFSKkhMT1WyLrYytFwwskuTcjKT9YpSk4tKM0tyU_NKlKxrASjBDmQ,

[14] https://uasdata.usc.edu/index.php?r=eNpLtDK0qi62MrFSKkhMT1WyLrYytFwwskuTcjKT9XJTSzLyU_Jz8tMrlaxrAShuDl0,

[15] https://uasdata.usc.edu/page/Comprehensive+File

cross-sectional non-core surveys as they become available. Typically, non-core surveys are shorter in length compared to core surveys. Specifically, existing UAS survey timing data indicate that respondents spend an average of 15 minutes (SD=12 mins) to complete a core survey and 8 minutes (SD=8 mins) to complete a non-core survey [16]. The combination of core and non-core surveys in our experimental setup is designed to simulate actual data collection conditions in online panels, as most field both longitudinal and cross-sectional surveys. With this experimental setup, survey frequency varies only for core surveys, whereas it is the same for non-core surveys across treatment conditions. This setup also allows us to examine whether manipulating survey frequency in core surveys has spillover effects on the completion of non-core surveys and whether these spillover effects differ across socioeconomic subgroups.

The recruitment letter did not mention the number of survey invitations respondents would receive within a given time period. Treatment assignment occurred after participants became panel members. Regardless of the frequency condition, respondents receive a reminder every two weeks to complete any pending core surveys. The reminder text states that surveys are awaiting completion on their dashboard without listing specific surveys to avoid discouragement from seeing a long list of incomplete tasks.

4.3.3 Outcome Measure

In our analytical data file, each respondent has multiple records corresponding to the number of survey invitations they received. Our outcome variable is an indicator variable coded as 1 if the respondent completed an individual survey invite and 0 otherwise (partial completes are extremely rare, constituting less than 0.1% of all cases, and are coded as nonresponse). Thus,

[16] https://uasdata.usc.edu/page/Panel+Recruitment+And+Participation+Data

each respondent contributes multiple (1,0) measures to the dataset, representing their responses to different surveys over time. We chose to analyze survey responses rather than attrition because the criteria for categorizing a case as attrition differ substantially across studies. Some studies consider respondents as having attrited after missing a few surveys initially, while others allow respondents to return at any time after they left. However, nonresponse and attrition outcomes are closely related, as respondents who frequently miss surveys are more likely to drop out completely.

4.3.4 Independent Variables

Experimental variable. A binary variable coded as 0 for participants randomized into the low-frequency group, receiving one core survey invitation every four weeks, and, coded as 1 for those assigned to the high-frequency group, receiving one core survey invitation every two weeks.

Covariates. Covariates used in the regression analyses include the following variables, which have been shown to serve as good proxies for respondents' subjective burden levels and cognitive burden that predict survey responses (Cheng et al., 2020; Earp et al., 2022; Holbrook et al., 2007; Tourangeau, 2000): Gender (female, male); Highest educational attainment (GED or high school and below, some college, college and above); Age groups (18-44, 45-64, 65 and above); Household income (below $50k, $50-$75K, $75K and above); Employment status (currently working, currently not working); Race (White only, Black only, Others); Hispanic ethnicity (yes, no); Household size (1, 2, 3+); Household structure (Have children below 18, have no children); Big Five personality traits, which include conscientiousness, openness, extraversion, agreeableness, and neuroticism; Survey wave indicators for core surveys; and Core vs. non-core survey indicator.

Personality traits have been shown to predict responses to Internet panel surveys (Cheng et al., 2020; Lugtig, 2014). Among these, conscientiousness, characterized by diligence and perseverance, fosters task persistence and sustained engagement, making it a strong predictor of survey completion. In contrast, neuroticism, marked by emotional instability, increases susceptibility to fatigue and dropout. Drawing on Bradburn's (1978) burden model and Leverage-Salience Theory (Bradburn, 1978; Groves et al., 2000), which posit that individual characteristics, such as personality traits, interact with survey features, such as frequency, to influence survey responses, we expect that personality traits moderate these effects. Specifically, conscientious respondents are more likely to engage under high-frequency conditions, whereas neurotic respondents are more likely to disengage.

The baseline demographic data were collected from the mandatory demographic survey (the so-called "My Household" survey), a prerequisite for all respondents to become panel members. Personality trait data were collected from the first wave of the core survey, administered immediately after respondents completed the prerequisite demographic survey. The traits were measured using a standardized self-report questionnaire consisting of 44 statements related to each of the five traits (Costa & McCrae, 1999; John et al., 1991; McCrae et al., 2004).

4.3.5 Statistical Analyses

We first provide unweighted sample summary statistics for the baseline characteristics and assess their balance across frequency conditions by conducting Chi-squared tests and reporting their p-values. For comparison and completeness, we also provide weighted sample summary statistics in the Appendix. We then present graphs that illustrate the cumulative response rates across successive core surveys by frequency condition. To estimate the frequency effect, we first present unadjusted overall response rates by frequency condition for core surveys,

non-core surveys, and the combination of core- and non-core surveys, along with the corresponding test statistics. We also consider that the low frequency group receives core surveys later than the high frequency group, which may confound the comparison. Next, we estimate three multivariable logit regression models that include the treatment variable, covariates, and significant interactions between the treatment variable and covariates (The process for selecting interactions is explained below) to explore treatment heterogeneity, restricting the analyses to core surveys, non-core surveys, and their combination. In the regression analysis of core survey data, we include core survey wave indicators as additional controls. In the regression analysis of the combined core and non-core survey data, we include an indicator variable for core vs. non-core surveys. All reported standard errors (SEs) are cluster-robust to account for clustering effects at the respondent level, treating individuals with repeated measures as clusters.

To explore treatment heterogeneity in responses to core surveys, we iteratively interact the treatment variable with individual covariates, examining p-values across all interaction terms. In each iteration, we drop the insignificant term with the largest p-value and continue the process until only significant interactions remain. We excluded treatment × wave interactions in the iterative process, as they have no practical implications for targeted interventions or informing future adaptive designs. As a result, the three multivariable logit regression models we estimate only include the significant interactions remaining after the iterative process, as well as the individual covariates in their level form. For each estimated logit regression model, we report estimated coefficients, 95% confidence intervals (95% CIs) and level of statistical significance.

All bivariable analyses and multivariable regression analyses are weighted to account for the complex survey design**.** The weights correct for unequal probabilities of sampling UAS

members and align the sample with the U.S. adult population in terms of gender, race/ethnicity, age, education, and geographic location (Census regions) (Understanding America Study, 2023). The weighting enhances the generalizability of the findings to the U.S. adult population, whereas the unweighted results are directly relevant for the UAS. For most estimated regression coefficients, we found that the weighted and unweighted results are similar both quantitatively and qualitatively. For comparison and completeness, we also present the unweighted logit regression results in the Appendix. All analyses were conducted using Stata for Windows Version 18.0 (StataCorp, 2023). Stata's "survey" module was used to analyze the complex survey data.

4.4 Results

4.4.1 Descriptive statistics

Table 4-1 presents the unweighted sample summary statistics for the baseline covariates. We provide percentages for categorical variables and mean values for the Big-5 personality traits. The first column provides the overall sample summary statistics, while the second and third columns provide the same statistics for the high and low frequency groups, respectively. The last column presents p-values from Chi-squared tests, assessing the balance of covariates across treatment groups. The p-values indicate that the baseline covariates are balanced between the two treatment groups, except for the variable indicating the percentage of households with children under 18, which is higher in the low-frequency group (5.83% for the high-frequency group vs. 9.92% for the low-frequency group). After weighting, the baseline characteristics remain balanced between the two treatment conditions (Appendix Table C-1 in the Appendix), except for the percentage of female respondents, which is higher in the low-frequency group (48.65% for the high-frequency group vs. 54.02% for the low-frequency group), and the

percentage of households with children under 18, which is also higher in the low-frequency group (5.69% for the high-frequency group vs. 10.42% for the low-frequency group). As a result, we adjust for all covariates in our multivariable logit regression, including these two variables that are imbalanced between the burden conditions.

Table 4-1: Unweighted sample demographic composition by survey frequency condition. Categorical variables are reported as percentages and Big-5 personality traits are reported as mean values.

	Overall (n=2,057)	High burden: Control group (n=1,029)	Low burden: Treated group (n=1,028)	p-value
Employment status				
Currently working	58.58	60.25	56.91	0.123
Gender				
Female	59.65	58.11	61.19	0.156
Education				
High school or GED	25.33	24.39	26.26	0.585
Some college	31.06	31.78	30.35	
College and above	43.61	43.83	43.39	
Age				
18-44	45.22	43.46	46.98	0.261
45-64	30.93	31.64	30.21	
65+	23.85	24.90	22.81	
HH Income				
Less than 50K	44.05	43.90	44.21	0.425
50K-75K	17.30	18.34	16.26	
75K and above	38.65	37.76	39.53	
Race				
White only	64.00	65.40	62.60	0.178
Black only	17.68	16.13	19.24	
Others	18.32	18.48	18.16	
Hispanic				
Yes	15.65	15.84	15.47	0.816
HH Size				
1	30.31	30.29	30.32	0.601
2	36.90	37.86	35.93	
3 and above	32.80	31.85	33.75	
Household structure				
Households with children below 18	7.88	5.83	9.92	0.001
Personality traits				
Conscientiousness	34.67	34.71	34.63	0.404
Openness	35.14	35.01	35.27	0.770
Extraversion	24.89	24.99	24.79	0.271
Agreeableness	34.94	34.82	35.06	0.761
Neuroticism	21.73	21.83	21.63	0.277

Figure 4-1 presents the cumulative response rates by the number of days since survey invitations were sent out for successive core surveys. The wave-specific graphs reveal that final response rates are similar between the two frequency conditions for the initial core survey waves but are somewhat higher for the high-frequency group in later waves, resulting in an overall higher response rate for the high-frequency group.

Figure 4-1: Cumulative overall response rates by successive core survey waves

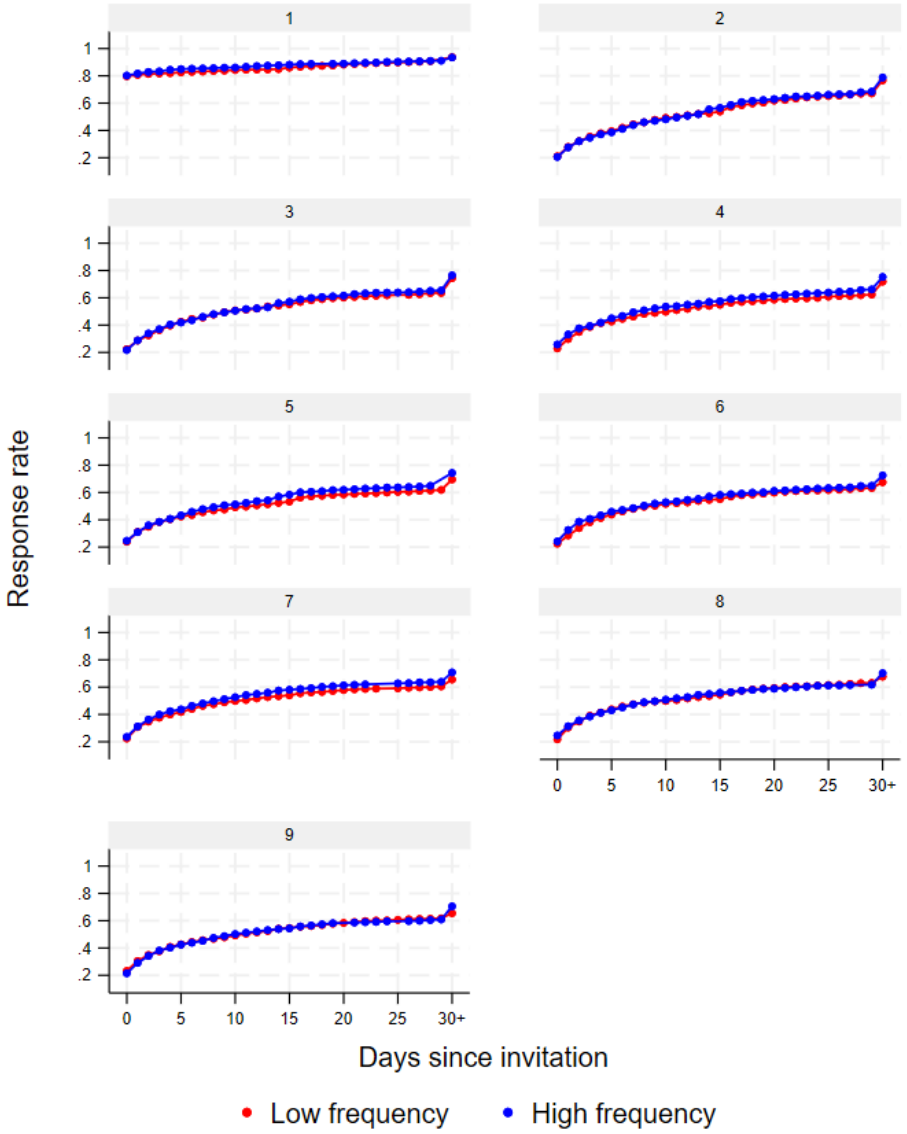


Table 4-2 presents weighted and unadjusted overall response rates by frequency condition for core surveys, non-core surveys, and their combination, along with the corresponding p-values. The results indicate that the unadjusted overall response rate for core surveys is about 3 percentage points higher in the high-frequency group compared to the low-frequency group, but the difference is not statistically significant (p-value = 0.117). The unadjusted overall response rates for non-core surveys and combined core and non-core surveys are identical across both frequency conditions. It should be noted that these are not final response rates. Core surveys are kept in the field for at least a year and experience shows that even after a substantial delay, respondents may still answer a survey.

Table 4-2: Weighted and unadjusted overall response rates by frequency condition for core surveys, non-core surveys, and the combined core and non-core surveys. p-values are from design-adjusted Rao-Scott Chi-squared tests

Overall weighted response rates (%)						
	Core surveys	p-value	Non-core surveys	p-value	Core + Non-core surveys	p-value
High burden	75.12	0.117	59.72	0.865	65.46	0.577
Low burden	72.13		60.06		64.40	

As noted, the high frequency group receives survey invitations every two weeks, while the low frequency group receives invitations every four weeks. The first survey a UAS respondent completes is the demographic “My Household” survey. After that they are immediately eligible for a core survey. Table 4-3 shows the consequences for the timing of core survey invitations. In the first two-week period both the low and high frequency group receive an invitation for the first core survey. However, the high frequency group also receives an invitation for the second core survey at the end of the first two-week period.

The table shows the response rates by wave and by period. By the time the high frequency group has received the invitation for the ninth survey, the low frequency group

receives the invitation for the fifth survey. The table shows two stylized facts: (1) response rates in later waves are lower than in earlier waves, which may partly reflect that respondents have had less time to respond to the later waves; (2) the response rates of high and low frequency groups are essentially identical when we control for the timing of a survey. However, if we were to simply compare response rates by waves, we would conclude that the low frequency group exhibits lower response rates than the high frequency group. Table 4-3 suggests that this may be a time effect rather than a treatment effect.

Table 4-3: Response rates by date of invitation by frequency condition

	Low Frequency		**High Frequency**	
Days since first My Household Survey	**Wave**	**RR**	**Wave**	**RR**
0-14 days	1	93.7	1&2	93.4/78.8
15-28 days	2	76.7	3	76.5
29-42 days			4	75.3
43-56 days	3	74.4	5	74.3
57-70 days			6	72.5
71-84 days	4	71.8	7	70.8
85-98 days			8	70.3
99-112 days	5	69.5	9	70.5

4.4.2 Multivariable logit regression analyses

In Table 4-4, we present weighted estimates from multivariable logit regression models predicting responses to core surveys (Column 1), non-core surveys (Column 2), and combined core and non-core surveys (Column 3). After iteratively selecting significant interaction terms, only the interaction between the treatment variable and employment remains significant. As such, Table 4-4 presents logit regression models that include the interaction term for employment status alongside the individual covariates in their level forms.

Table 4-4: Weighted estimates of multivariable logit regression models predicting responses to UAS core surveys, non-core surveys, and combined core and non-core surveys. All standard errors (SEs) are cluster-robust treating individuals with repeated measures as clusters.

	Coefficient (95% CI)		
	Core surveys	Non-core surveys	Core + Non-core surveys
	(1)	(2)	(3)
Frequency condition (Ref: Low frequency)			
High frequency	-0.043	-0.254*	-0.253*
	(-0.387,0.302)	(-0.536,0.028)	(-0.538,0.033)
Demographic Variables			
Employment status (Ref: Not working)			
Currently working	-0.213	-0.268*	-0.249*
	(-0.551,0.125)	(-0.540,0.005)	(-0.529,0.031)
High frequency # Currently working	0.532**	0.449**	0.469**
	(0.081,0.984)	(0.089,0.810)	(0.099,0.839)
Gender (Ref: female)			
Male	-0.169	-0.073	-0.108
	(-0.404,0.065)	(-0.262,0.116)	(-0.300,0.085)
Education (Ref: High School or GED)			
Some College	0.372**	0.263**	0.290**
	(0.078,0.665)	(0.026,0.500)	(0.048,0.532)
College and above	0.544***	0.398***	0.437***
	(0.221,0.867)	(0.137,0.659)	(0.170,0.703)
Age (Ref: 18-44)			
45-64	-0.134	0.042	-0.011
	(-0.411,0.144)	(-0.179,0.263)	(-0.238,0.215)
65+	0.014	0.113	0.081
	(-0.323,0.352)	(-0.157,0.383)	(-0.197,0.358)
Income (Ref: Less than 50K)			
50K-75K	-0.400**	-0.165	-0.235
	(-0.740,-0.060)	(-0.438,0.107)	(-0.516,0.047)
75K and above	-0.274*	-0.214*	-0.227*
	(-0.599,0.052)	(-0.469,0.040)	(-0.489,0.035)
Race (Ref: White only)			
Black only	-0.055	-0.232*	-0.173
	(-0.382,0.271)	(-0.489,0.025)	(-0.434,0.088)
Others	-0.303*	-0.240*	-0.257**
	(-0.610,0.004)	(-0.494,0.013)	(-0.514,-0.001)
Hispanic (Ref: Non-Hispanic)			
Yes	-0.328**	-0.401***	-0.371***
	(-0.648,-0.009)	(-0.646,-0.156)	(-0.623,-0.119)
HH Size (Ref: One)			
Two	0.409***	0.269**	0.304**
	(0.129,0.690)	(0.042,0.497)	(0.070,0.537)
Three and above	0.549***	0.389***	0.428***

	(0.244,0.853)	(0.141,0.637)	(0.176,0.680)
Have children below 18 (Ref: No children at home)			
Have children at home	0.383	-0.031	0.081
	(-0.108,0.874)	(-0.357,0.294)	(-0.258,0.420)
Personality Traits			
Conscientiousness	0.040***	0.042***	0.040***
	(0.017,0.063)	(0.025,0.060)	(0.022,0.058)
Openness	0.027***	0.016**	0.019**
	(0.009,0.046)	(0.001,0.032)	(0.003,0.034)
Extroversion	-0.006	-0.015*	-0.013
	(-0.027,0.014)	(-0.031,0.001)	(-0.029,0.004)
Neuroticism	0.035***	0.020***	0.022***
	(0.017,0.053)	(0.007,0.033)	(0.009,0.036)
Agreeableness	0.017	0.011	0.011
	(-0.007,0.041)	(-0.007,0.029)	(-0.007,0.030)
Core survey wave (Ref: Wave=1)			
2	-2.653***		
	(-2.999,-2.308)		
3	-2.768***		
	(-3.111,-2.426)		
4	-2.872***		
	(-3.215,-2.529)		
5	-2.959***		
	(-3.302,-2.616)		
6	-3.043***		
	(-3.385,-2.701)		
7	-3.144***		
	(-3.487,-2.802)		
8	-3.115***		
	(-3.457,-2.772)		
9	-3.155***		
	(-3.500,-2.810)		
Core vs. Noncore (Ref: Non-core)			
Core surveys			0.492***
			(0.429,0.555)
High frequency # Core surveys			0.184***
			(0.073,0.295)
Constant	0.060	-2.030***	-2.191***
	(-0.924,1.044)	(-2.692,-1.368)	(-2.907,-1.474)
Observations	15,538	26,392	41,930

* $p < 0.10$-** $p < 0.05$-*** $p < 0.01$. 95% confidence intervals in brackets.

4.4.3 Heterogeneous Treatment Effects

While survey frequency has no overall effect on response rate to core surveys (Table 4-2 and Table 4-3), we find treatment effects conditional on employment status. When introducing heterogeneous treatment effects in Table 4-4, Column 1, we-find that the strongest effects are concentrated among individuals who are currently employed. Specifically, among individuals currently employed, the odds of responding to a core survey under the high-frequency condition are 1.63 [17] times those in the low-frequency condition, with the difference being statistically significant (95% CI= 1.149-2.113, p-value < 0.01). This corresponds to an 8-percentage-point difference in response rates to core surveys (78.94% for currently employed persons receiving surveys at a high frequency vs. 70.70% for currently employed persons receiving surveys at a low frequency). This positive survey frequency effect in online panel surveys contrasts with the negative frequency effect found by Eggleston (2024) in cross-sectional surveys. In contrast to employed individuals, non-working individuals did not exhibit increased engagement when they received core surveys more frequently, as shown in Table 4-4, Column 1, suggesting that the intervention had no effect on non-working respondents. In the discussion section, we explain the practical implications of the positive frequency effect observed among working individuals, discuss why the survey frequency effect differs between cross-sectional and online panel surveys, and provide a plausible explanation for the positive frequency effect found in our study.

4.4.4 Spillover Effects on Responses to Non-core Surveys

[17] Calculated as exp[estimated coefficient for the high-frequency variable + estimated coefficient for the employment interaction term]. The correct 95% confidence interval (CI) is obtained using Stata's "nlcom" command as the following: "nlcom exp[estimated coefficient for the high-frequency variable + estimated coefficient for the employment interaction term]". The coefficients must be in the log-odds scale. Minor discrepancies may arise due to rounding the estimated coefficients to three decimal places. A similar calculation applies to other interaction models presented in Columns 2 and 3 of Table 4-4.

Additionally, the results indicate spillover treatment effects from core surveys to non-core surveys among individuals currently employed. Specifically, among employed individuals, the odds of responding to non-core surveys in the high-frequency condition are 1.22 times (95% CI=0.937-1.495) those in the low-frequency condition (Table 4-4, Column 2), with the difference being marginally significant ($p < 0.05$). This corresponds to a response rate difference of approximately 4 percentage points (64% for currently employed persons receiving core surveys at a high frequency vs. 60% for currently employed persons receiving core surveys at a low frequency). Similarly, among employed individuals, the odds of responding to both core and non-core surveys under the high-frequency condition are 1.24 times (95% CI=0.953-1.531) those under the low-frequency condition (Table 4-4, Column 3), with the difference also being marginally significant ($p < 0.05$). This corresponds to a response rate difference of 5–6 percentage points (69% for currently employed persons receiving core surveys at a high frequency vs. 64% for currently employed persons receiving core surveys at a low frequency). Notably, this is observed even though respondents in both treatment conditions received the same number of non-core survey requests.

4.4.5 Similarities between Weighted and Un-weighted Model Estimates

The unweighted results presented in Appendix Table C-2 in the Appendix are similar to their weighted counterparts in Table 4-4, both quantitatively and qualitatively. The similarity in estimated coefficients between the weighted and unweighted models is primarily due to balanced covariates across treatment conditions. This finding suggests that the treatment effects are robust regardless of whether weights are applied. However, among working individuals, the estimated high-frequency treatment effect on response to core surveys and its spillover effects on response to non-core surveys and the combined core and non-core surveys are more pronounced in the

unweighted models than in the weighted models (Table 4-4 and Appendix Table C-2), primarily because weighting inflates standard errors.

4.4.6 Covariate Effects

Regarding the effects of the covariates, the regression results in Table 4-4 reveal that males (compared to females), racial minority groups (compared to Whites), and Hispanic individuals (compared to non-Hispanic subgroup) are less likely to respond to survey requests, but the difference by gender fails to reach significance. Similarly, higher income is negatively associated with response rates across all survey types. In contrast, respondents with at least some college education (compared to those with a GED or high school education and lower) and those from households with at least two members (compared to single-person households) are significantly more likely to respond to all types of survey requests, while age has no discernible effect across all survey types. Respondents with higher scores on conscientiousness, openness, and neuroticism have significantly higher adjusted odds of responding to all three survey types, regardless of whether weights are applied. The covariate effects observed in this study are consistent across both weighted (Table 4-4) and unweighted models (Appendix Table C-2 in the Appendix) and largely align with findings from prior research (Cheng et al., 2020; Jin & Kapteyn, 2022; Lugtig, 2014).

4.4.7 Other Effects

Beyond these covariate influences, our results indicate that response rates to core surveys tend to decline initially but stabilize as the number of survey waves increases (Table 4-4, Column 1). As noted, core surveys are kept in the field for at least a year, so that the response rates for the later surveys are likely to increase with time. Core surveys exhibit significantly higher response rates compared to non-core surveys (Table 4-4, Column 3; AOR = 1.64, 95% CI

= 1.54–1.74, $p < 0.01$). This however is mainly an artefact, as non-core surveys are often kept in the field for a shorter period, and respondents consequently have less time to complete a survey. The significant interaction term between the treatment variable and the core survey indicator in Table 4-4, Column 3 suggests that respondents in the high-frequency condition have higher odds of completing core surveys compared to those in the low-frequency condition keeping other factors fixed, confirming the successful experimental manipulation of survey frequency for core surveys.

In summary, Table 4-3 suggests no substantial differences in response rates across treatment groups, once the timing of survey invitations is taken into account (RQ1). However, when interactions are introduced, we find that the more frequent schedule resulted in the largest increases in response rates for core surveys among employed individuals, with effect sizes as large as 8 percentage points (RQ2). In contrast, we find no significant effects among lower or other SES groups. Additionally, we find evidence that among employed individuals, the positive treatment effect of a more frequent survey schedule on responses to core surveys extends to non-core surveys and to the combined core and non-core surveys as well. In the following section, we discuss the theoretical implications of these findings within the framework of existing theories, compare them with prior studies, examine their practical implications, propose directions for future research and highlight the strengths and limitations of our approach.

4.5 Discussion and Conclusion

As survey response rates for mailed questionnaires and face-to-face or telephone interviews continue to decline, the issue of respondent burden has gained significant attention in the survey research literature, especially since the publication of the *Paperwork Reduction Act Guide* (2011), which aims to address this challenge (Kaplan et al., 2022; Paperwork Reduction

Act Guide, 2011). However, findings from past research have been mixed, with some studies showing positive associations between survey frequency and response rates, while others have found no significant effect. This inconsistency stems largely from the limited exogenous variation in survey frequency available in cross-sectional and conventional longitudinal surveys, as well as the reliance on observational data, which is prone to selection bias.

Most importantly, prior findings from cross-sectional and conventional longitudinal studies have limited applicability to probability-based online panels because of fundamental differences in design features and context. Given the increasing use of online data for policy decision-making and social science research, more research is needed to enhance our understanding of response mechanisms related to survey frequency in online panel settings. Our study extends the existing literature by leveraging the unique high-frequency data collection structure of probability-based online panels, which allows us to experimentally test the causal link between survey frequency and nonresponse while providing new evidence on the causal effects of survey frequency in online panel settings.

Rather than lowering response rates as found in prior cross-sectional studies (Eggleston, 2024), our results indicate no difference in overall response rate between new respondents on a 1-survey-per-2-weeks schedule and those on a 1-survey-per-4-weeks schedule.

The null-effects found in our study contrast with the negative frequency effects identified by Eggleston (2024) in his analysis of responses to the 2020 U.S. Census. That paper suggests two possible explanations: respondents may have been confused by the similarity of multiple survey requests, leading some to believe they had already provided the requested information, or they may have felt overwhelmed by repeated resurvey requests, resulting in survey fatigue. In contrast, online panel respondents are aware that they are part of an ongoing study and that each

survey is distinct. In addition, a working relationship between panel organization and respondents is established before survey requests are sent, and frequent surveys may foster engagement rather than causing panel fatigue, particularly among respondents from certain socioeconomic subgroups, as suggested by our results. Additionally, UAS respondents receive $20 per 30 minutes of survey time, so that responding to more survey requests implies higher earnings, which may also boost participation. However, this interpretation assumes that respondents are primarily motivated by cumulative earnings. It is also plausible that, for low-frequency respondents, each survey represents a rarer opportunity. As such, the opportunity cost of missing a survey may be higher for the low-frequency group than for high-frequency respondents. From this perspective, the marginal value of each incentive may be greater for the low-frequency group, potentially offsetting the advantage of higher cumulative earnings available to the high-frequency group. Future research should examine whether participation decisions are more strongly influenced by total earning potential or by the perceived value of each individual opportunity.

While we found no overall treatment effect, our subgroup analyses reveal that employed individuals experienced a positive effect from receiving more frequent core survey invitations. This group not only demonstrated higher response rates to core surveys but also exhibited positive spillover effects in their participation in non-core surveys that were not experimentally varied. This pattern underscores the potential for tailoring survey frequencies to the characteristics of different subgroups to enhance overall response and retention rates in online panels. Most importantly, the fact that the intervention is most effective among employed individuals, combined with their lower response rates in the absence of intervention, suggests that increasing survey frequency could be an effective strategy to improve sample representation

in these subgroups. Future research could explore how varying engagement strategies influence these dynamics across employment groups to optimize panel retention.

Leverage-salience theory (Groves et al., 2000) can help us further distinguish between engagement and incentive effects based on actual response behavior. The theory posits that different survey features such as incentive have varying effects ("leverages") on participation decisions based on respondents' socioeconomic characteristics, lifestyles, and preferences. One implication of this theory is that incentives will be more effective among subgroups with low response propensities and those with low socioeconomic status.

Applying the predictions of Leverage-Salience Theory to our data, if there are any incentive effects, we should expect the treatment effects to be stronger among low SES groups such as low-education and low-income groups, racial minority groups, and those currently not working. Conversely, if the engagement effects dominate the incentive effects, we should expect the treatment effects to be larger among high SES subgroups such as high-education and high-income groups, Whites, and those who are currently working. Our subgroup analyses of the treatment effects showed that the effects primarily accrued to those who are currently working, with no discernible effects found for low-education and low-income groups, racial minority groups, and individuals who are not currently working. Additionally, our results show positive spillover effects, suggesting that completing more core surveys also boosts completion rates for non-core surveys among individuals with busy schedules, such as those who are currently working. Despite the evidence in our data supporting the engagement explanation, it remains important for future research to clearly separates engagement effects from incentive effects.

In our current design, the high-frequency group not only receives surveys more frequently but also completes twice as many core surveys as the low-frequency group. As a

result, it is difficult to disentangle whether the observed effects are driven by the frequency of survey contact, the total amount of effort required, or the greater incentive earnings available to the high-frequency group. The current study's setup potentially confounds survey frequency with total respondent burden and cumulative incentive, making it challenging to isolate the engagement effects of survey frequency from the incentive effects associated with increased survey volume.

Future research should collect additional subjective data on respondent engagement, motivation, and attitudes toward receiving more frequent surveys, as well as the dose response relationship between incentives and response rates in online panel settings to disentangle the various mechanisms driving these effects.

Subjective data would help disentangle the engagement mechanism by comparing response rates among high-frequency respondents with differing levels of engagement, motivation, and attitudes toward receiving more frequent surveys. Significant differences in response behavior across these dimensions would suggest the presence of engagement effects. Alternatively, interacting the treatment variable with these subjective measures could help uncover how engagement moderates the effect of survey frequency.

While the engagement mechanism is relatively straightforward to analyze, disentangling the incentive mechanism would require a more elaborate experimental design, which may be prohibitively expensive to implement. A more viable approach would be to draw on insights from Leverage-Salience Theory by comparing response rates among high-frequency respondents across demographic subgroups for whom incentives carry different levels of leverage in their participation decisions. This strategy could help uncover how incentive effects vary across respondent profiles and serve as a more practical alternative to direct experimental manipulation.

However, this approach requires a sufficiently large sample to meaningfully detect differences across subgroups.

Our study findings have implications for adaptive design approaches in future online surveys. Given that respondents in the high-frequency group exhibited higher engagement than those in the low-frequency group among certain subgroups, future adaptive designs could adjust survey frequency based on individual response patterns. For example, new panelists demonstrating high engagement, particularly those currently employed, could be scheduled for more frequent surveys to maximize data collection. Conversely, those showing signs of fatigue or decreased responsiveness, such as non-working respondents (as well as respondents with lower education or income), could receive surveys less frequently to prevent dropout as they adapt to the experience of responding to frequent surveys. Additionally, future adaptive designs could incorporate differential incentive strategies by offering larger rewards to non-working respondents (as well as to racial minority groups, and low-income or low-education respondents) who may be less responsive to increased survey frequency, if feasible.

Our choice of a one-year observation period was based on practical considerations. For instance, the decision to target fresh respondents was informed by evidence that most attrition occurs during the first year of panel membership. Nonetheless, it is possible that certain effects may only become apparent over a longer timeframe. Our findings suggest that frequent survey requests did not reduce overall participation during this period, but we cannot rule out the possibility that different patterns might emerge beyond the first year. Future research should assess whether high-frequency survey designs produce different effects over longer durations.

Finally, although we observed no statistically significant difference in overall response rates, the results can also be interpreted from a non-inferiority perspective. That is, a higher-

frequency survey design does not necessarily lead to meaningfully worse outcomes than a lower-frequency schedule in the context of online panel surveys. Framing the findings this way has practical relevance, as it suggests that survey designers may implement more frequent data collection—at least to some extent—without compromising participation. Nevertheless, these findings should be replicated in other settings and populations to assess their robustness and generalizability.

4.5.1 Strength and Limitations

The main strength of this study is the use of probability-based online panel. First of all, probability Internet panels have to be distinguished from convenience Internet panels, where respondents are recruited only from among existing Internet users, with unknown inclusion probabilities. Several studies have shown that even with relatively low response rates, probability Internet panels provide superior information to convenience panels (Chang & Krosnick, 2009; Krosnick et al., 2013; Yeager et al., 2011). Non-probability samples generally appear to be incapable of providing unbiased estimates of population parameters. Couper et al. (2007) and Schonlau et al. (2009) show that weighting and matching do not eliminate differences in estimates between non-probability and probability online samples.

On the other hand, probability-based Internet panels mitigate selection problems facing opt-in or convenience Internet panels. Several studies have shown that widely used opt-in panels contain small but measurable shares of "bogus respondents" (about 4% to 7%, depending on the source), who tend to select positive answer choices, thus introducing systematic biases (Kennedy et al., 2020; Schneider et al., 2018). Moreover, prior research has shown that probability-based Internet panels provide superior information to convenience panels, even when response rates are low (Chang & Krosnick, 2009; Krosnick et al., 2013; MacInnis et al., 2018; Malhotra &

Krosnick, 2007; Tourangeau, 2020; Yeager et al., 2011). A recent dramatic example of how nonprobability-based or biased samples may generate wildly inaccurate estimates is provided by (Bradley et al., 2021b), who show that both the Delphi-Facebook survey (250,000 responses per week) and the Census Household Pulse (about 75,000 responses every two weeks) overestimated national COVID-19 vaccination rates by 11-20 percentage points. On the other hand, the Axios/Ipsos and UAS probability-based Internet panels provided accurate estimates. Similarly, Angrisani et al. (2019b) found that UAS estimates of population parameters were very similar to those obtained from traditional, high quality representative surveys such as the Health and Retirement Study (HRS) and the Current Population Survey (CPS).

The advantages of using probability samples have also been documented in cross-sectional surveys. In their meta-analysis of 59 studies based on probability samples, Groves and Peytcheva (2008) found that nonresponse rates are a poor predictor of nonresponse bias in probability samples. While surveys can experience large nonresponse biases, these biases vary more across individual survey estimates within a study than across studies with different response rates. The authors emphasize that bias is driven by the relationship between response propensity and the specific survey variable, not by response rate alone. They recommend focusing on the design features and salience of survey content that influence this relationship to better understand and reduce nonresponse bias.

Another main strength of this study lies in its use of experimental data, which effectively addresses selection bias and provides causal estimates. Additionally, the availability of rich background information and longitudinal survey data, combined with a large sample size, enhances statistical inferences and enables the exploration of treatment heterogeneity. The use of nationally representative data, combined with the application of survey weights, enhances the

generalizability of our results to the U.S. adult population, although as noted the unweighted results are most relevant to the UAS practice. Finally, the study's focus on a fresh panel mitigates panel conditioning effects, where previous interviews potentially influence future survey responses, and panel maturity effects, where attrition in earlier waves results in a sample of committed and loyal respondents in later waves, which may limit the effectiveness of interventions (Kraemer et al., 2024; Wong, 2020). However, a limitation of the study's approach is the lack of substantial variation in survey frequency, which could have revealed a non-linear relationship between survey frequency and survey responses. Another limitation is the lack of data on respondents' reasons for responding to or not responding to more frequent survey requests, which would help uncover the mechanisms driving the observed outcomes. Despite this limitation, the study generates findings that may motivate further research and better inform strategies for enhancing panel management and retention practices.

Chapter 5 Conclusion

5.1 Summary of Key Findings

This dissertation addresses critical challenges in probability-based online panels and high-frequency online longitudinal data collections by examining mechanisms driving and factors predicting nonresponse and attrition and evaluating the efficacy of theoretically-informed interventions to mitigate them. Through three independent yet interrelated studies, this research enhances our understanding of how non-demographic factors, incentive structures, and survey frequency influence respondent behavior and data quality in online panel surveys.

The first study demonstrates that personality traits—particularly conscientiousness—are the strongest, most consistent, and independent predictors of nonresponse patterns in high-frequency online longitudinal data collections. Conscientious individuals exhibit lower wave nonresponse, while more open, extraverted, neurotic, and agreeable individuals are more prone to higher wave nonresponse, though with smaller effect sizes than conscientiousness. The study also reveals that demographic effects are largely absent in an aging and well-being study conducted during a normal period but emerge in a COVID-19-focused study, likely due to the unique societal context of the pandemic. These findings highlight the increasing importance of personality traits over traditional demographic variables in predicting nonresponse as online data collection becomes more frequent and intensive.

The second study evaluates the impact of three theoretically motivated incentive plans on wave nonresponse using a randomized controlled trial. The study finds that unconditional incentives are more effective than conditional incentives in improving response rates, with the

two unconditional incentive groups exhibiting a 6-percentage point higher response rate than the conditional incentive group. However, there is no significant difference between the $5 and $15 unconditional incentive groups. Additionally, the study uncovers heterogeneity in the effects of unconditional incentives across socioeconomic subgroups, with these effects being strongest among individuals with lower motivation and study participation rates. The study also finds suggestive evidence that survey reminders are more effective in increasing response rates among unconditional incentive recipients.

The third study presents experimental evidence on the causal effect of survey frequency on nonresponse in a probability-based online panel among new panel respondents. Overall, we found no significant survey frequency effect, a finding that contrasts with the negative frequency effect found in prior cross-sectional studies. The effect remains robust even after adjusting for baseline characteristics and incorporating survey weights. However, our subgroup analyses reveal that among employed individuals, the likelihood of responding to survey requests is significantly higher under the high-frequency condition than the low-frequency condition, with an increase in response rate as large as 8 percentage points. In contrast, no discernible effects are found among lower or other SES subgroups.

5.2 Broader Implications and Scholarly Contributions

Collectively, these studies contribute to the methodological literature on online panel survey research by providing new empirical evidence into the mechanisms underlying and factors driving nonresponse and attrition. The findings have significant implications for survey practitioners and policymakers alike aiming to improve data quality, enhance panel retention, and design more effective survey protocols.

First, the identification of personality traits as key predictors of nonresponse suggests that survey practitioners should consider incorporating personality measures into weighting adjustments, given that these traits strongly and independently predict important life outcomes. While population-level benchmarks for personality traits do not exist—making traditional post-stratification infeasible—researchers can still leverage internal benchmarks by collecting personality data from all respondents at baseline. To facilitate this, since personality data are not available in the frame data, online longitudinal studies should collect personality data at the beginning of the study, preferably immediately after respondents become formal panel members or immediately after respondents complete the baseline mandatory demographic survey that is prerequisite for becoming panel members, to ensure that baseline data are available for all respondents (In offline panel study such as the UK Household Longitudinal Study, Sturgis and Brunton-Smith (2023) suggest collecting this information in the first wave). Having personality data for the full initial sample enables researchers to incorporate these variables into longitudinal weights, track potential attrition-related bias over time, implement targeted intervention, and inform future adaptive design based on personality profiles. Additionally, understanding how personality traits influence panelists' response behavior can help survey practitioners develop targeted engagement strategies, such as personalized survey invitations and adaptive incentive structures, to minimize attrition and improve the representativeness of longitudinal data over time.

Second, the findings on incentive strategies provide new evidence for optimizing incentive structures to improve participation, particularly among students. While unconditional incentives are generally more effective, as shown in prior cross-sectional and offline panel studies, variations in their impact across socioeconomic subgroups highlight the need for

targeted incentive strategies and adaptive designs that account for respondent heterogeneity. Importantly, the results of this study are limited to the population of master's students enrolled in non-professional master's degree programs at public universities in the U.S.

Third, the experimental evidence on survey frequency demonstrates that, in contrast to prior cross-sectional studies where increased survey frequency negatively affects survey response, more frequent survey invitations can lead to higher response rates in the online panel context, particularly among individuals with higher socioeconomic status (SES) such as those currently employed. This suggests that well-designed high-frequency online longitudinal studies can be viable for collecting rich, continuous data without sacrificing data quality.

5.3 Limitations and Future Research Directions

While this dissertation provides important insights, several limitations should be acknowledged. First, the generalizability of findings from the study analyzing the independent effects of non-demographic factors on nonresponse patterns may be influenced by the specific contexts and populations studied. Future research should examine whether similar patterns hold across different panel structures, study settings, population groups, geographic regions, and cultural contexts.

Second, the incentive study was conducted using a student population, which significantly limits the generalizability of the findings. Although the results align with broader theoretical predictions and prior research, future studies should replicate this approach using a population-representative sample with sufficient statistical power.

Finally, the findings on survey frequency suggest potential benefits of higher-frequency data collection. However, additional research is needed to assess the effects of frequency variation across different online panel settings with substantial variation in survey frequency,

which could reveal a non-linear relationship between survey frequency and survey responses. Future studies examining respondents' reasons for responding to or not responding to more frequent survey requests would help uncover the mechanisms driving the observed outcomes. Furthermore, studies are needed to evaluate the efficacy of adaptive survey designs that tailor invitation schedules based on individual respondent behaviors.

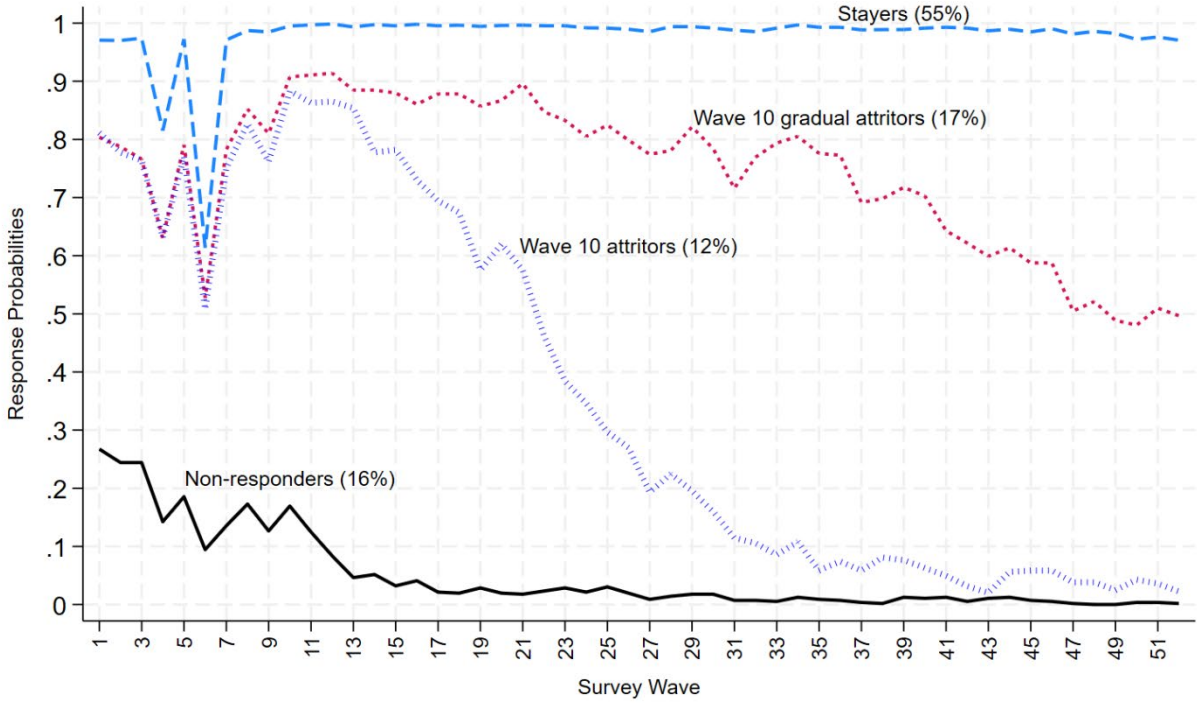
5.4 Conclusion

This dissertation advances knowledge on survey nonresponse and attrition in probability-based online panels and longitudinal data collections by identifying key predictors of nonresponse, evaluating incentive strategies tailored to the online panel context, and establishing a causal link between survey frequency and response rates. These findings provide practical recommendations for improving panel management, designing more effective engagement strategies, and ensuring the long-term viability of online longitudinal data collections. Most importantly, the dissertation generates findings that can inform and motivate future research. Future studies building on these insights will refine best practices for optimizing data quality and representativeness in the design and implementation of online panel surveys, which are becoming increasingly important in social science research and policy decision-making.

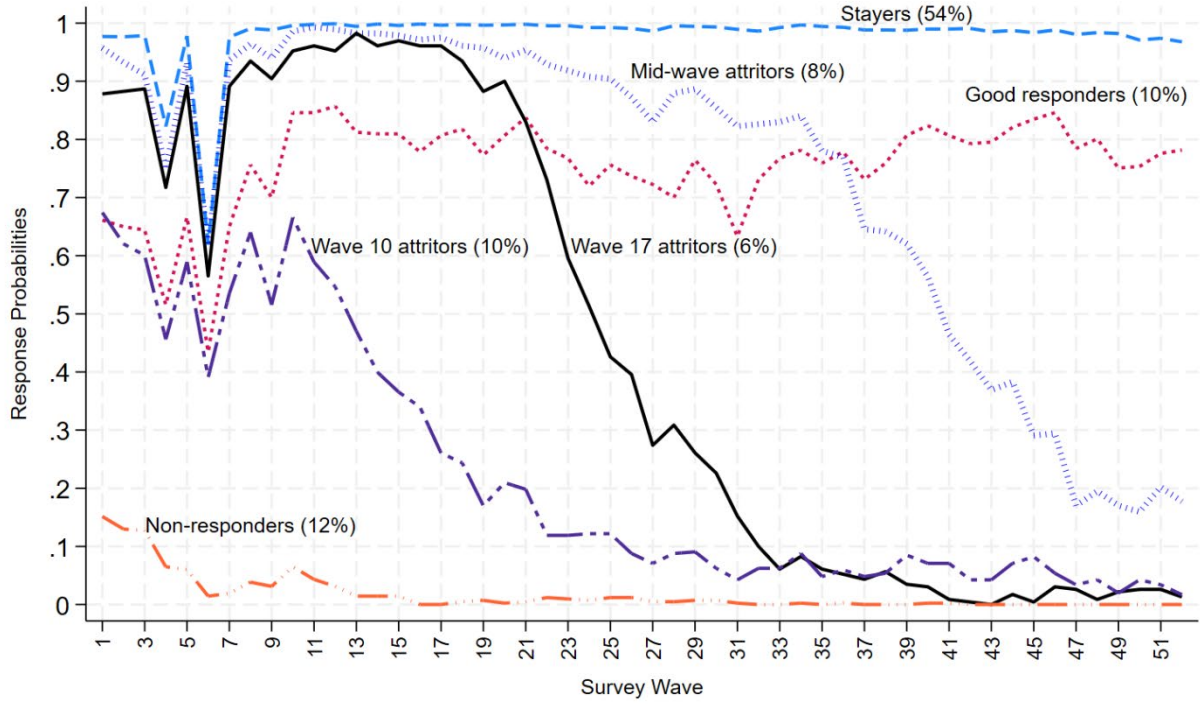
Appendices

Appendix A: Chapter 2 Figures and Tables

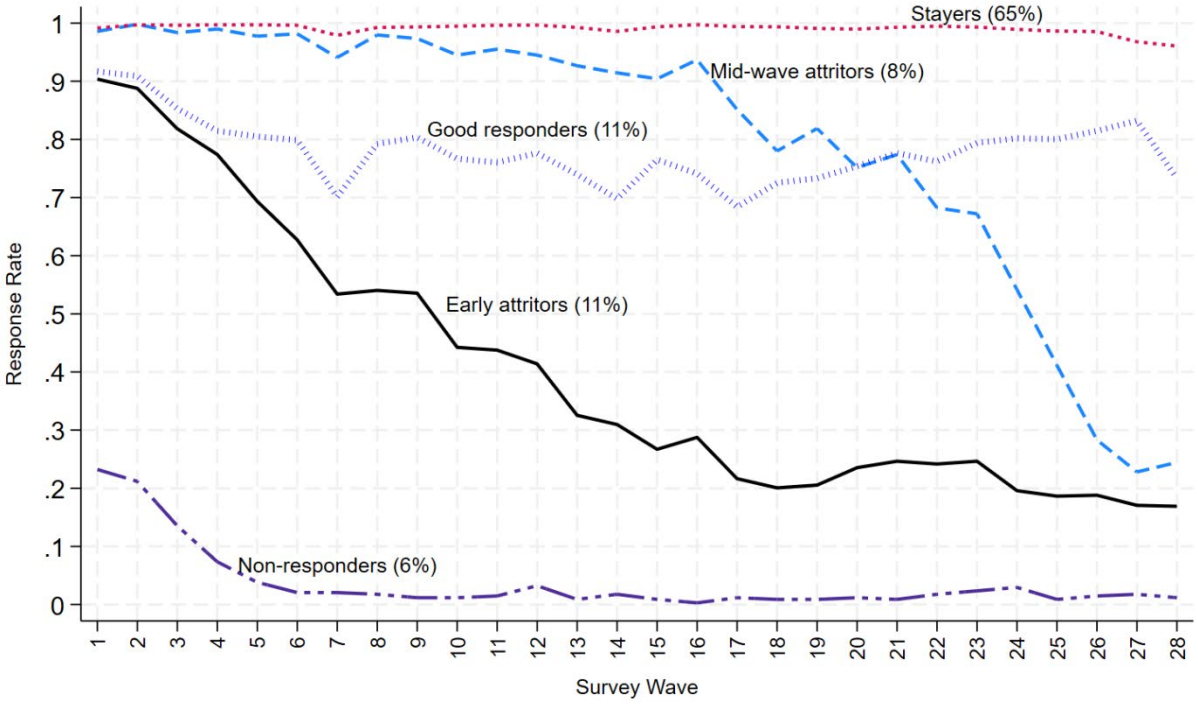
Appendix Figure A-1: Posterior response probabilities of the latent class model that is *one class smaller* than the chosen number of latent classes for the *monthly events panel study*



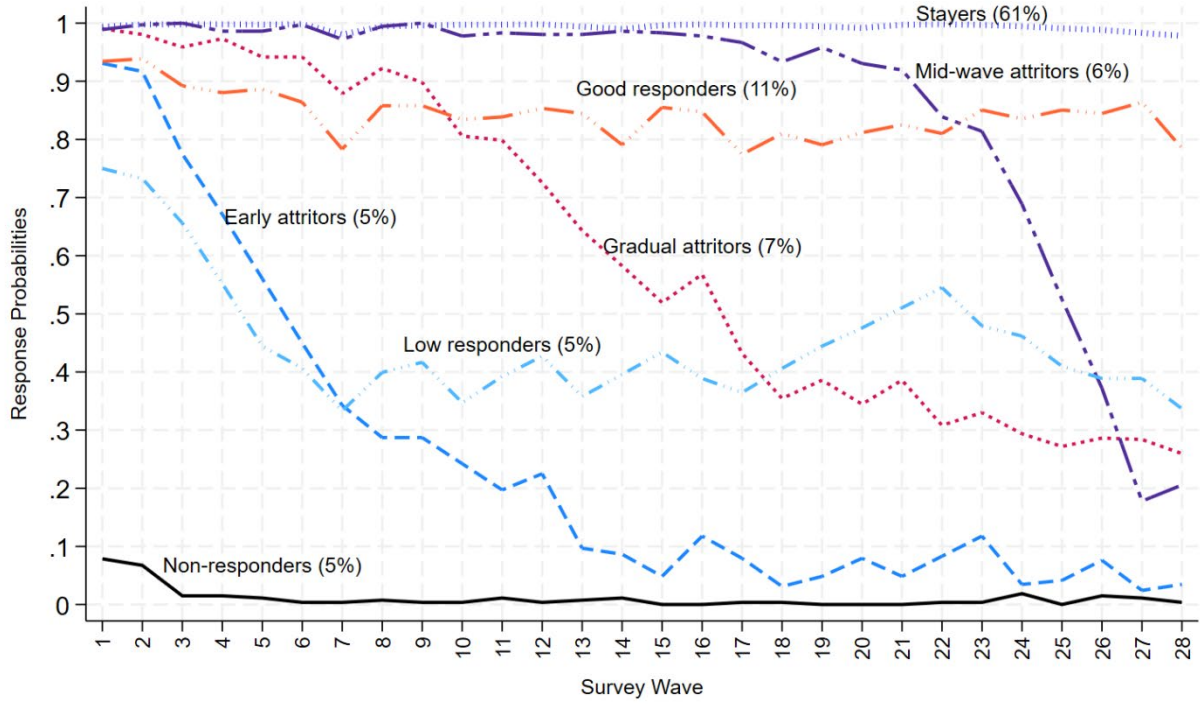
Appendix Figure A-2: Posterior response probabilities of the latent class model that is *one class larger* than the chosen number of latent classes for the *monthly events panel study*



Appendix Figure A-3: Posterior response probabilities of the latent class model that is *one class smaller* than the chosen number of latent classes for the *COVID-19 panel study*



Appendix Figure A-4: Posterior response probabilities of the latent class model that is *one class larger* than the chosen number of latent classes for the *COVID-19 panel study*



Appendix Table A-1: Weighted multinomial logistic regressions predicting class membership in the *monthly events panel study* using *probabilities predicting class membership for the chosen classes as weights*. The reference group is "Stayers". Reported coefficients are adjusted relative-risk ratios (ARRR), which measure the proportionate change in the relative risk of belonging to a certain class, compared to the "stayers" class, when the predictor variable changes by one unit, while controlling for other predictor variables.

	Non-responders	Wave 10 attritors	Mid-wave attritors	Good responders	Stayers
Big-5 Personality Traits					
Conscientiousness Score	0.964**	0.977	0.959**	0.973	
	[0.942,0.986]	[0.955,0.999]	[0.934,0.985]	[0.951,0.996]	
Openness Score	1.019	1.018	1.002	1.013	
	[1.000,1.038]	[0.999,1.037]	[0.980,1.024]	[0.994,1.033]	
Extroversion Score	1.042***	1.011	1.012	1.009	Ref.
	[1.023,1.062]	[0.992,1.030]	[0.990,1.035]	[0.990,1.029]	
Neuroticism Score	1.028	1.032**	1.002	1.014	
	[1.007,1.050]	[1.011,1.053]	[0.978,1.027]	[0.993,1.036]	
Agreeableness Score	1.021	1.035**	1.022	1.032	
	[0.998,1.045]	[1.012,1.060]	[0.995,1.050]	[1.007,1.057]	
Panel Tenure					
(Ref: Less than 1 year)					
1 year and above	0.689	0.932	0.716	0.827	Ref.
	[0.504,0.943]	[0.661,1.315]	[0.491,1.044]	[0.587,1.165]	
Hispanic					
(Ref: No)					
Yes	1.566	1.041	0.483	1.099	Ref.
	[1.045,2.346]	[0.657,1.649]	[0.227,1.024]	[0.700,1.726]	
Race & Ethnicity					
(Ref: White only)					
Black only	1.128	1.128	1.288	1.429	
	[0.736,1.728]	[0.745,1.709]	[0.812,2.045]	[0.958,2.130]	Ref.
Others	1.185	0.951	0.893	1.393	
	[0.827,1.700]	[0.646,1.401]	[0.556,1.434]	[0.975,1.989]	
Gender					
(Ref: Female)					
Male	1.218	1.191	1.076	0.930	Ref.
	[0.970,1.530]	[0.948,1.496]	[0.822,1.409]	[0.735,1.177]	
Age Group					
(Ref: 50-64)					
65 and above	1.132	0.909	1.326	0.763	Ref.
	[0.876,1.462]	[0.702,1.176]	[0.983,1.789]	[0.581,1.004]	
Education					
(Ref: GED or high school)					
Some College	0.921	0.920	0.944	0.922	
	[0.697,1.216]	[0.696,1.216]	[0.679,1.313]	[0.691,1.229]	Ref.
College and above	0.682	0.826	0.857	0.768	
	[0.496,0.938]	[0.603,1.132]	[0.590,1.243]	[0.554,1.066]	
HH Income					
(Ref: Below $50K)					
$50-$75K	0.956	0.989	1.055	0.753	
	[0.707,1.293]	[0.736,1.329]	[0.746,1.490]	[0.544,1.041]	Ref.
$75K and above	0.917	0.799	0.846	0.895	
	[0.686,1.225]	[0.596,1.072]	[0.596,1.202]	[0.667,1.201]	

Employment Status (Ref: Currently working)					
Currently not working	0.913 [0.706,1.180]	0.925 [0.716,1.196]	0.881 [0.647,1.201]	0.684* [0.527,0.888]	Ref.
Household Size (Ref: 1)					
2	0.827 [0.625,1.094]	0.986 [0.740,1.314]	0.880 [0.631,1.227]	1.061 [0.791,1.423]	Ref.
3 and above	1.042 [0.757,1.434]	1.268 [0.919,1.749]	1.205 [0.828,1.754]	1.064 [0.762,1.487]	
Health Status					
	0.985 [0.869,1.117]	1.136 [1.003,1.285]	1.335*** [1.156,1.542]	1.115 [0.979,1.270]	Ref.
Constant	0.062*** [0.013,0.284]	0.024*** [0.005,0.113]	0.121** [0.020,0.722]	0.087*** [0.018,0.421]	
n	3,446				
Pseudo R^2	0.005				

Exponentiated coefficients (i.e., adjusted relative-risk ratios); 95% confidence intervals in brackets; * p < 0.10, ** p < 0.05, *** p < 0.01. All p values were adjusted for multiple hypothesis tests using Holm's method. Note the 95% CIs were not adjusted for multiple hypothesis tests.

Appendix Table A-2: Weighted multinomial logistic regressions predicting class membership in the *COVID-19 panel study* using *probabilities predicting class membership for the chosen classes as weights*. The reference group is "Stayers". Reported coefficients are adjusted relative-risk ratios (ARRR), which measure the proportionate change in the relative risk of belonging to a certain class, compared to the "stayers" class, when the predictor variable changes by one unit, while controlling for other predictor variables.

	Non-responders	Early attritors	Gradual attritors	Mid-wave attritors	Good responders	Stayers
Big-5 Personality Traits						
Conscientiousness Score	0.976	0.963**	0.951***	0.982	0.955***	
	[0.950,1.002]	[0.942,0.985]	[0.932,0.971]	[0.961,1.003]	[0.937,0.973]	
Openness Score	1.030	1.044***	1.028*	1.027*	1.009	
	[1.007,1.055]	[1.024,1.064]	[1.009,1.047]	[1.008,1.046]	[0.993,1.026]	
Extroversion Score	0.999	1.036***	1.014	1.009	1.012	Ref.
	[0.976,1.021]	[1.017,1.055]	[0.996,1.032]	[0.991,1.027]	[0.997,1.028]	
Neuroticism Score	1.012	1.030*	1.016	1.014	1.019	
	[0.987,1.037]	[1.009,1.051]	[0.997,1.036]	[0.995,1.034]	[1.001,1.037]	
Agreeableness Score	1.001	1.026	1.007	0.996	1.031**	
	[0.974,1.028]	[1.003,1.049]	[0.986,1.029]	[0.975,1.017]	[1.012,1.051]	
Panel Tenure						
(Ref: Less than 1 year)						
1 year and above	4.889***	0.871	0.983	1.443	0.750*	Ref.
	[3.114,7.674]	[0.681,1.113]	[0.771,1.254]	[1.093,1.905]	[0.608,0.923]	
Hispanic						
(Ref: No)						
Yes	4.786***	1.325	1.205	0.954	1.705***	Ref.
	[3.480,6.581]	[0.992,1.770]	[0.910,1.597]	[0.688,1.322]	[1.338,2.172]	
Race & Ethnicity						
(Ref: White only)						
Black only	1.661	1.683*	0.891	0.917	1.456	
	[1.026,2.689]	[1.185,2.392]	[0.592,1.341]	[0.611,1.378]	[1.053,2.012]	Ref.
Others	1.375	0.956	0.839	0.825	1.084	
	[0.956,1.977]	[0.694,1.317]	[0.614,1.147]	[0.588,1.156]	[0.834,1.410]	
Gender						
(Ref: Female)						
Male	0.685	1.216	0.998	0.983	1.096	Ref.
	[0.512,0.918]	[0.965,1.532]	[0.798,1.249]	[0.785,1.232]	[0.900,1.335]	
Age Group						
(Ref: 18-44)						
45-64	0.593**	0.350***	0.350***	0.540***	0.531***	Ref.
	[0.429,0.821]	[0.269,0.455]	[0.272,0.449]	[0.420,0.695]	[0.426,0.662]	

65+	0.744	0.260***	0.203***	0.383***	0.439***	
	[0.482,1.149]	[0.173,0.391]	[0.135,0.305]	[0.267,0.550]	[0.318,0.607]	
Education (Ref: GED or high school)						
Some College	0.730	1.057	0.698	0.991	1.008	
	[0.515,1.034]	[0.788,1.416]	[0.532,0.917]	[0.750,1.309]	[0.787,1.291]	Ref.
College and above	0.697	0.735	0.574***	0.665	0.806	
	[0.474,1.026]	[0.529,1.019]	[0.425,0.776]	[0.485,0.912]	[0.613,1.060]	
HH Income (Ref: Below $50K)						
$50-$75K	0.900	0.888	0.816	0.992	1.059	
	[0.607,1.335]	[0.656,1.202]	[0.602,1.105]	[0.737,1.337]	[0.822,1.365]	Ref.
$75K and above	1.158	0.709	0.851	0.920	0.828	
	[0.812,1.650]	[0.530,0.947]	[0.644,1.123]	[0.693,1.221]	[0.646,1.062]	
Employment Status (Ref: Currently working)						
Currently not working	1.178	0.681*	0.810	0.953	0.777	
	[0.863,1.608]	[0.523,0.885]	[0.632,1.038]	[0.743,1.222]	[0.623,0.968]	Ref.
Household Size (Ref: 1)						
2	0.811	1.014	1.147	1.057	1.289	
	[0.550,1.198]	[0.727,1.414]	[0.819,1.605]	[0.773,1.444]	[0.965,1.720]	Ref.
3 and above	1.061	1.323	1.471	1.188	1.517*	
	[0.715,1.575]	[0.953,1.836]	[1.056,2.049]	[0.860,1.639]	[1.133,2.033]	
Health Status						
Self-report of health	0.906	0.931	0.977	1.073	0.988	
	[0.777,1.056]	[0.821,1.055]	[0.866,1.102]	[0.950,1.212]	[0.888,1.099]	Ref.
Constant	0.018***	0.021***	0.277*	0.062***	0.145***	
	[0.003,0.111]	[0.005,0.091]	[0.068,1.136]	[0.015,0.262]	[0.041,0.506]	
n			5,743			
Pseudo R^2			0.019			

Exponentiated coefficients (i.e., adjusted relative-risk ratios); 95% confidence intervals in brackets; * p < 0.10, ** p < 0.05, *** p < 0.01. All p values were adjusted for multiple hypothesis tests using Holm's method. Note the 95% CIs were not adjusted for multiple hypothesis tests.

Appendix Table A-3: Multinomial logistic regressions predicting class membership in the m*onthly events panel study* for the latent class model that is *one class smaller* than the chosen number of latent classes. The reference group is "Stayers". Reported coefficients are adjusted relative-risk ratios (ARRR), which measure the proportionate change in the relative risk of belonging to a certain class, compared to the "stayers" class, when the predictor variable changes by one unit, while controlling for other predictor variables.

	Non-responders	Wave 10 attritors	Wave 10 gradual attritors	Stayers
Big-5 Personality Traits				
Conscientiousness Score	0.966**	0.976	0.961***	
	[0.945,0.987]	[0.954,0.997]	[0.942,0.980]	
Openness Score	1.019	1.015	1.008	
	[1.000,1.038]	[0.997,1.034]	[0.992,1.024]	
Extroversion Score	1.038***	1.010	1.012	Ref.
	[1.019,1.057]	[0.992,1.029]	[0.996,1.028]	
Neuroticism Score	1.034**	1.023	1.009	
	[1.013,1.055]	[1.003,1.044]	[0.992,1.028]	
Agreeableness Score	1.027	1.021	1.029*	
	[1.004,1.050]	[0.998,1.044]	[1.009,1.050]	
Panel Tenure				
(Ref: Less than 1 year)				
1 year and above	0.709	0.857	0.853	Ref.
	[0.522,0.962]	[0.618,1.190]	[0.638,1.140]	
Hispanic				
(Ref: No)				
Yes	1.483	1.081	0.882	Ref.
	[0.994,2.211]	[0.691,1.691]	[0.584,1.332]	
Race & Ethnicity				
(Ref: White only)				
Black only	1.258	0.949	1.447	
	[0.843,1.877]	[0.618,1.459]	[1.033,2.026]	Ref.
Others	1.227	0.842	1.235	
	[0.866,1.738]	[0.570,1.245]	[0.905,1.687]	
Gender				
(Ref: Female)				
Male	1.302	1.112	1.013	Ref.
	[1.043,1.623]	[0.889,1.390]	[0.832,1.233]	
Age Group				
(Ref: 50-64)				
65 and above	1.124	0.970	0.905	Ref.
	[0.877,1.441]	[0.754,1.247]	[0.724,1.132]	
Education				
(Ref: GED or high school)				
Some College	0.862	0.959	0.863	
	[0.660,1.127]	[0.727,1.266]	[0.679,1.096]	Ref.
College and above	0.655	0.925	0.751	
	[0.482,0.891]	[0.679,1.262]	[0.573,0.986]	
HH Income				
(Ref: Below $50K)				
$50-$75K	0.938	1.013	0.933	
	[0.700,1.259]	[0.759,1.352]	[0.718,1.211]	Ref.
$75K and above	0.898	0.794	0.914	
	[0.678,1.190]	[0.596,1.059]	[0.712,1.174]	

Employment Status (Ref: Currently working)				
Currently not working	0.931 [0.726,1.194]	0.940 [0.730,1.209]	0.790 [0.634,0.984]	Ref.
Household Size (Ref: 1)				
2	0.829 [0.632,1.088]	0.999 [0.755,1.321]	1.033 [0.807,1.322]	Ref.
3 and above	1.087 [0.799,1.480]	1.266 [0.922,1.737]	1.215 [0.919,1.605]	
Health Status				
	0.991 [0.878,1.119]	1.179 [1.045,1.331]	1.216*** [1.092,1.353]	Ref.
Constant	0.047*** [0.011,0.208]	0.056*** [0.013,0.249]	0.184** [0.049,0.684]	
n		3,446		
Pseudo R^2		0.020		

Exponentiated coefficients (i.e., adjusted relative-risk ratios); 95% confidence intervals in brackets; * p < 0.10, ** p < 0.05, *** p < 0.01. All p values were adjusted for multiple hypothesis tests using Holm's method. Note the 95% CIs were not adjusted for multiple hypothesis tests.

Appendix Table A-4: Multinomial logistic regressions predicting class membership in the *monthly events panel study* for the latent class model that is *one class larger* than the chosen number of latent classes. The reference group is "Stayers". Reported coefficients are adjusted relative-risk ratios (ARRR), which measure the proportionate change in the relative risk of belonging to a certain class, compared to the "stayers" class, when the predictor variable changes by one unit, while controlling for other predictor variables.

	Non-responders	Wave 10 attritors	Wave 17 attritors	Mid-wave attritors	Good responders	Stayers
Big-5 Personality Traits						
Conscientiousness Score	0.967	0.962**	0.999	0.956**	0.972	
	[0.942,0.992]	[0.939,0.986]	[0.970,1.029]	[0.931,0.982]	[0.949,0.996]	
Openness Score	1.014	1.021	1.015	1.004	1.010	
	[0.993,1.036]	[1.001,1.042]	[0.991,1.040]	[0.982,1.026]	[0.990,1.030]	
Extroversion Score	1.049***	1.015	1.009	1.006	1.009	Ref.
	[1.027,1.072]	[0.994,1.036]	[0.985,1.034]	[0.984,1.029]	[0.989,1.029]	
Neuroticism Score	1.033	1.032*	1.020	1.001	1.011	
	[1.010,1.058]	[1.010,1.056]	[0.994,1.047]	[0.977,1.026]	[0.989,1.034]	
Agreeableness Score	1.025	1.035	1.005	1.021	1.036	
	[0.999,1.052]	[1.010,1.061]	[0.976,1.035]	[0.994,1.049]	[1.010,1.062]	
Panel Tenure						
(Ref: Less than 1 year)						
1 year and above	0.700	0.788	0.795	0.889	0.752	Ref.
	[0.491,0.998]	[0.552,1.124]	[0.519,1.217]	[0.592,1.335]	[0.531,1.066]	
Hispanic						
(Ref: No)						
Yes	1.424	1.384	0.994	0.557	1.107	Ref.
	[0.892,2.273]	[0.883,2.169]	[0.531,1.863]	[0.276,1.126]	[0.695,1.764]	
Race & Ethnicity						
(Ref: White only)						
Black only	1.249	1.266	0.774	1.308	1.557	
	[0.786,1.987]	[0.812,1.972]	[0.415,1.444]	[0.822,2.083]	[1.037,2.337]	Ref.
Others	1.120	1.238	0.670	0.867	1.508	
	[0.737,1.702]	[0.844,1.816]	[0.379,1.187]	[0.536,1.403]	[1.048,2.170]	
Gender						
(Ref: Female)						
Male	1.147	1.435*	1.020	1.053	0.945	Ref.
	[0.887,1.484]	[1.119,1.839]	[0.758,1.372]	[0.805,1.378]	[0.739,1.209]	
Age Group						
(Ref: 50-64)						
65 and above	1.192	0.863	1.189	1.173	0.799	Ref.
	[0.894,1.590]	[0.650,1.146]	[0.857,1.648]	[0.868,1.587]	[0.601,1.062]	

Education (Ref: GED or high school)						
Some College	0.837	1.025	0.929	0.849	0.927	
	[0.614,1.140]	[0.751,1.397]	[0.648,1.332]	[0.612,1.177]	[0.689,1.248]	Ref.
College and above	0.632**	0.917	0.820	0.757	0.728	
	[0.443,0.903]	[0.647,1.299]	[0.545,1.235]	[0.523,1.095]	[0.517,1.024]	
HH Income (Ref: Below $50K)						
$50-$75K	0.861	0.943	1.124	1.073	0.756	
	[0.610,1.216]	[0.679,1.309]	[0.774,1.633]	[0.756,1.523]	[0.540,1.057]	Ref.
$75K and above	0.859	0.849	0.813	0.969	0.867	
	[0.620,1.191]	[0.617,1.166]	[0.552,1.196]	[0.685,1.373]	[0.638,1.180]	
Employment Status (Ref: Currently working)						
Currently not working	0.879	0.929	1.038	0.827	0.699	
	[0.658,1.174]	[0.704,1.226]	[0.738,1.460]	[0.609,1.123]	[0.532,0.917]	Ref.
Household Size (Ref: 1)						
2	0.881	0.780	1.137	0.983	1.030	
	[0.643,1.207]	[0.572,1.065]	[0.786,1.644]	[0.700,1.381]	[0.759,1.398]	Ref.
3 and above	1.053	1.265	1.056	1.286	1.114	
	[0.733,1.513]	[0.903,1.770]	[0.679,1.643]	[0.879,1.881]	[0.789,1.573]	
Health Status						
	0.912	1.104	1.250	1.287**	1.089	
	[0.791,1.053]	[0.965,1.264]	[1.067,1.464]	[1.113,1.488]	[0.952,1.247]	Ref.
Constant	0.045***	0.031***	0.020***	0.149**	0.098***	
	[0.008,0.251]	[0.006,0.165]	[0.003,0.145]	[0.025,0.898]	[0.019,0.509]	
n			3,446			
Pseudo R^2			0.023			

Exponentiated coefficients (i.e., adjusted relative-risk ratios); 95% confidence intervals in brackets; * p < 0.10, ** p < 0.05, *** p < 0.01. All p values were adjusted for multiple hypothesis tests using Holm's method. Note the 95% CIs were not adjusted for multiple hypothesis tests.

Appendix Table A-5: Multinomial logistic regressions predicting class membership in the *COVID-19 panel study* for the latent class model that is *one class smaller* than the chosen number of latent classes. The reference group is "Stayers". Reported coefficients are adjusted relative-risk ratios (ARRR), which measure the proportionate change in the relative risk of belonging to a certain class, compared to the "stayers" class, when the predictor variable changes by one unit, while controlling for other predictor variables.

	Non-responders	Early attritors	Mid-wave attritors	Good responders	Stayers
Big-5 Personality Traits					
Conscientiousness Score	0.981	0.954***	0.974	0.952***	
	[0.957,1.005]	[0.936,0.971]	[0.955,0.993]	[0.935,0.969]	
Openness Score	1.031*	1.036***	1.023	1.005	
	[1.010,1.053]	[1.020,1.053]	[1.005,1.040]	[0.989,1.020]	
Extroversion Score	1.002	1.029***	1.016	1.014	Ref.
	[0.982,1.022]	[1.013,1.044]	[1.000,1.033]	[0.999,1.030]	
Neuroticism Score	1.019	1.020	1.020	1.016	
	[0.997,1.042]	[1.002,1.037]	[1.002,1.039]	[0.999,1.033]	
Agreeableness Score	1.003	1.022	1.006	1.022	
	[0.979,1.027]	[1.003,1.041]	[0.986,1.026]	[1.004,1.041]	
Panel Tenure (Ref: Less than 1 year)					
1 year and above	2.229***	0.968	1.504**	0.750*	Ref.
	[1.617,3.073]	[0.786,1.193]	[1.166,1.939]	[0.613,0.916]	
Hispanic (Ref: No)					
Yes	3.442***	1.300	1.152	1.596***	Ref.
	[2.576,4.599]	[1.019,1.657]	[0.869,1.527]	[1.262,2.017]	
Race & Ethnicity (Ref: White only)					
Black only	1.487	1.465	0.892	1.271	
	[0.968,2.284]	[1.077,1.992]	[0.606,1.312]	[0.922,1.753]	Ref.
Others	1.185	1.031	0.858	1.042	
	[0.848,1.655]	[0.794,1.339]	[0.632,1.166]	[0.807,1.345]	
Gender (Ref: Female)					
Male	0.856	1.105	0.960	1.034	Ref.
	[0.662,1.106]	[0.911,1.341]	[0.779,1.184]	[0.855,1.252]	
Age Group (Ref: 18-44)					
45-64	0.565***	0.341***	0.484***	0.508***	
	[0.424,0.753]	[0.274,0.424]	[0.384,0.609]	[0.411,0.629]	Ref.
65+	0.638	0.234***	0.327***	0.449***	
	[0.430,0.948]	[0.166,0.329]	[0.231,0.462]	[0.330,0.613]	
Education (Ref: GED or high school)					
Some College	0.790	0.904	0.842	0.931	
	[0.578,1.079]	[0.710,1.151]	[0.652,1.086]	[0.736,1.179]	Ref.
College and above	0.734	0.678*	0.580***	0.761	
	[0.520,1.037]	[0.518,0.887]	[0.435,0.773]	[0.586,0.989]	
HH Income (Ref: Below $50K)					
$50-$75K	0.802	0.858	1.021	1.010	Ref.

	[0.567,1.133]	[0.663,1.110]	[0.772,1.349]	[0.791,1.291]	
$75K and above	0.889	0.799	1.058	0.814	
	[0.648,1.219]	[0.627,1.017]	[0.814,1.376]	[0.640,1.034]	
Employment Status (Ref: Currently working)					
Currently not working	0.987	0.762	0.868	0.795	Ref.
	[0.748,1.304]	[0.614,0.946]	[0.688,1.095]	[0.643,0.983]	
Household Size (Ref: 1)					
2	0.897	1.028	1.163	1.213	
	[0.632,1.273]	[0.777,1.361]	[0.859,1.574]	[0.920,1.599]	Ref.
3 and above	1.160	1.344	1.251	1.462	
	[0.815,1.651]	[1.019,1.771]	[0.918,1.703]	[1.106,1.933]	
Health Status					
Self-report of health	0.902	0.940	1.064	0.965	Ref.
	[0.787,1.034]	[0.846,1.044]	[0.950,1.192]	[0.871,1.069]	
Constant	0.032***	0.096***	0.064***	0.352*	
	[0.006,0.163]	[0.028,0.329]	[0.017,0.243]	[0.106,1.172]	
n			5,743		
Pseudo R^2			0.060		

Exponentiated coefficients (i.e., adjusted relative-risk ratios); 95% confidence intervals in brackets; * p < 0.10, ** p < 0.05, *** p < 0.01. All p values were adjusted for multiple hypothesis tests using Holm's method. Note the 95% CIs were not adjusted for multiple hypothesis tests.

Appendix Table A-6: Multinomial logistic regressions predicting class membership in the *COVID-19 panel study* for the latent class model that is *one class larger* than the chosen number of latent classes. The reference group is "Stayers". Reported coefficients are adjusted relative-risk ratios (ARRR), which measure the proportionate change in the relative risk of belonging to a certain class, compared to the "stayers" class, when the predictor variable changes by one unit, while controlling for other predictor variables.

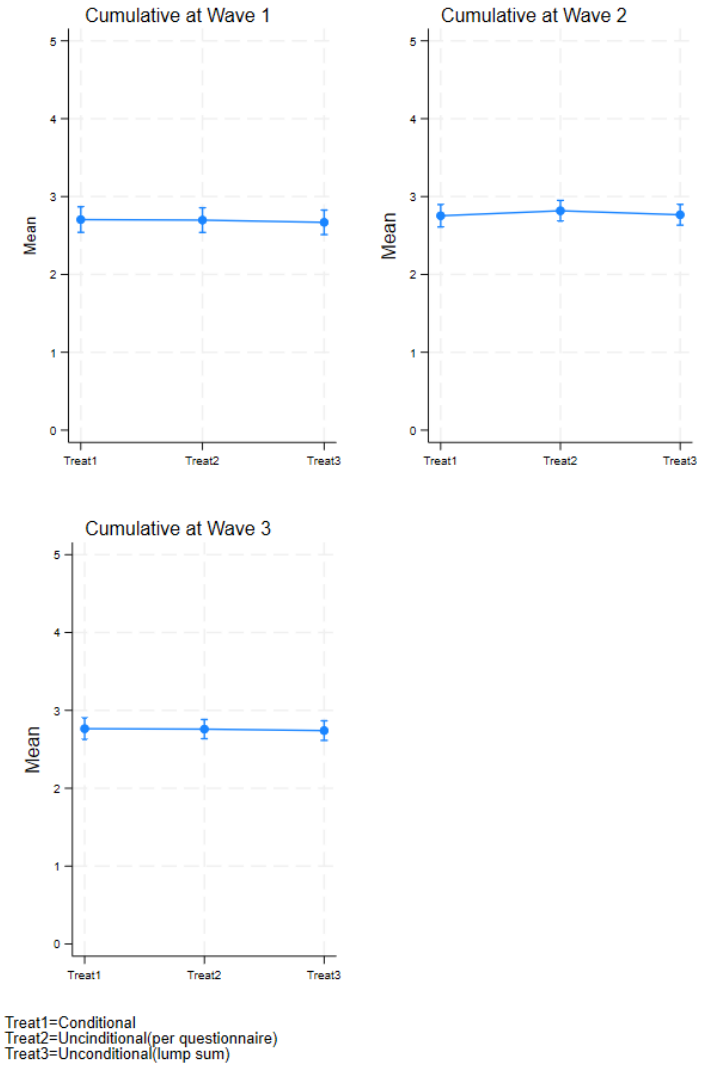
	Non-responders	Early attritors	Low Responders	Gradual attritors	Mid-wave attritors	Good responders	Stayers
Big-5 Personality Traits							
Conscientiousness Score	0.977	0.968	0.937***	0.953***	0.979	0.964***	
	[0.951,1.004]	[0.943,0.993]	[0.914,0.961]	[0.933,0.974]	[0.957,1.002]	[0.947,0.981]	
Openness Score	1.030	1.033*	1.041***	1.025	1.027	1.012	
	[1.006,1.054]	[1.010,1.057]	[1.018,1.064]	[1.005,1.044]	[1.007,1.047]	[0.997,1.027]	
Extroversion Score	0.998	1.038**	1.028	1.015	1.007	1.009	Ref.
	[0.975,1.021]	[1.016,1.060]	[1.006,1.050]	[0.996,1.033]	[0.988,1.026]	[0.995,1.024]	
Neuroticism Score	1.013	1.029	1.031	1.015	1.010	1.021	
	[0.988,1.039]	[1.005,1.054]	[1.007,1.055]	[0.995,1.036]	[0.989,1.031]	[1.005,1.038]	
Agreeableness Score	1.003	1.020	1.040**	1.004	0.994	1.024	
	[0.976,1.031]	[0.994,1.047]	[1.013,1.068]	[0.982,1.026]	[0.972,1.017]	[1.006,1.042]	
Panel Tenure (Ref: Less than 1 year)							
1 year and above	4.473***	0.946	0.768	0.960	1.463	0.831	Ref.
	[2.877,6.954]	[0.706,1.268]	[0.581,1.017]	[0.748,1.234]	[1.085,1.972]	[0.681,1.014]	
Hispanic (Ref: No)							
Yes	4.777***	1.298	1.484	1.226	0.936	1.753***	Ref.
	[3.464,6.590]	[0.921,1.830]	[1.067,2.064]	[0.918,1.638]	[0.657,1.332]	[1.393,2.206]	
Race & Ethnicity (Ref: White only)							
Black only	1.634	1.491	2.015***	0.767	0.852	1.377	
	[0.999,2.671]	[0.971,2.290]	[1.369,2.965]	[0.490,1.198]	[0.547,1.327]	[1.008,1.881]	Ref.
Others	1.430	0.898	0.954	0.863	0.764	1.141	
	[0.995,2.056]	[0.613,1.315]	[0.655,1.389]	[0.627,1.187]	[0.529,1.105]	[0.893,1.459]	
Gender (Ref: Female)							
Male	0.710	1.401	0.986	0.883	1.020	1.036	Ref.
	[0.529,0.954]	[1.070,1.835]	[0.752,1.293]	[0.699,1.116]	[0.804,1.295]	[0.861,1.248]	
Age Group (Ref: 18-44)							
45-64	0.580**	0.326***	0.422***	0.344***	0.571***	0.532***	Ref.

	[0.418,0.805]	[0.239,0.446]	[0.312,0.570]	[0.266,0.445]	[0.437,0.745]	[0.432,0.654]	
65+	0.733	0.291***	0.295***	0.195***	0.409***	0.468***	
	[0.473,1.138]	[0.182,0.464]	[0.185,0.470]	[0.127,0.299]	[0.280,0.599]	[0.347,0.632]	
Education (Ref: GED or high school)							
Some College	0.749	1.112	0.848	0.655**	1.071	1.002	
	[0.526,1.066]	[0.783,1.578]	[0.617,1.166]	[0.495,0.867]	[0.794,1.443]	[0.793,1.267]	Ref.
College and above	0.712	0.862	0.478***	0.534***	0.697	0.838	
	[0.482,1.052]	[0.586,1.268]	[0.329,0.695]	[0.391,0.728]	[0.497,0.978]	[0.648,1.084]	
HH Income (Ref: Below $50K)							
$50-$75K	0.860	0.872	1.082	0.905	0.849	1.150	
	[0.577,1.282]	[0.610,1.246]	[0.766,1.529]	[0.662,1.236]	[0.612,1.177]	[0.908,1.457]	Ref.
$75K and above	1.093	0.697	0.882	0.934	0.924	0.825	
	[0.764,1.562]	[0.496,0.980]	[0.628,1.239]	[0.700,1.247]	[0.687,1.244]	[0.653,1.043]	
Employment Status (Ref: Currently working)							
Currently not working	1.147	0.668	0.742	0.792	0.959	0.793	
	[0.838,1.571]	[0.489,0.914]	[0.552,0.997]	[0.613,1.024]	[0.736,1.249]	[0.645,0.976]	Ref.
Household Size (Ref: 1)							
2	0.822	1.148	0.983	1.104	1.177	1.136	
	[0.555,1.218]	[0.768,1.717]	[0.675,1.430]	[0.782,1.559]	[0.841,1.648]	[0.871,1.483]	Ref.
3 and above	1.074	1.511	1.089	1.340	1.280	1.443	
	[0.721,1.599]	[1.015,2.248]	[0.748,1.586]	[0.953,1.884]	[0.903,1.815]	[1.103,1.888]	
Health Status							
Self-report of health	0.899	0.962	0.890	0.997	1.085	0.982	
	[0.770,1.050]	[0.829,1.115]	[0.770,1.029]	[0.880,1.130]	[0.954,1.235]	[0.888,1.085]	Ref.
Constant	0.019***	0.015***	0.048***	0.338	0.066***	0.152***	
	[0.003,0.119]	[0.003,0.088]	[0.009,0.260]	[0.079,1.455]	[0.014,0.306]	[0.047,0.494]	
n			5,743				
Pseudo R^2			0.061				

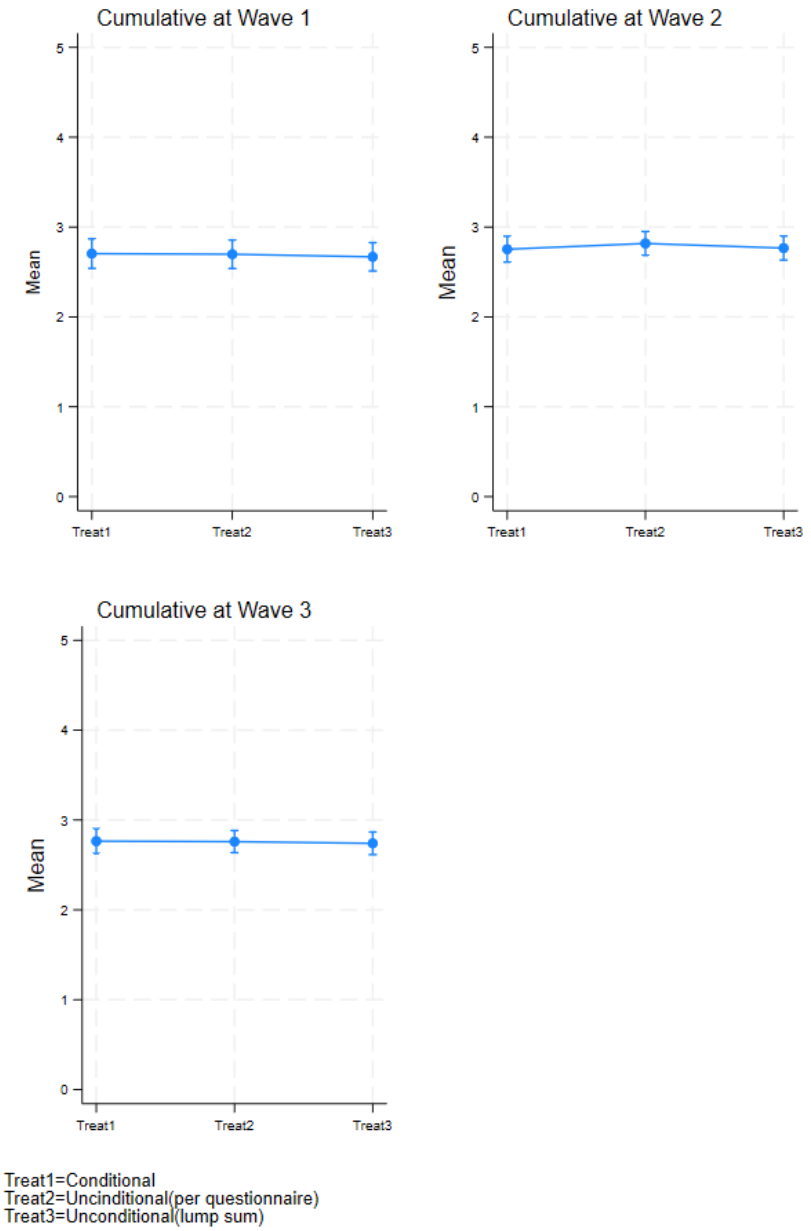
Exponentiated coefficients (i.e., adjusted relative-risk ratios); 95% confidence intervals in brackets; * p < 0.10, ** p < 0.05, *** p < 0.01. All p values were adjusted for multiple hypothesis tests using Holm's method. Note the 95% CIs were not adjusted for multiple hypothesis tests.

Appendix B: Chapter 3 Figures

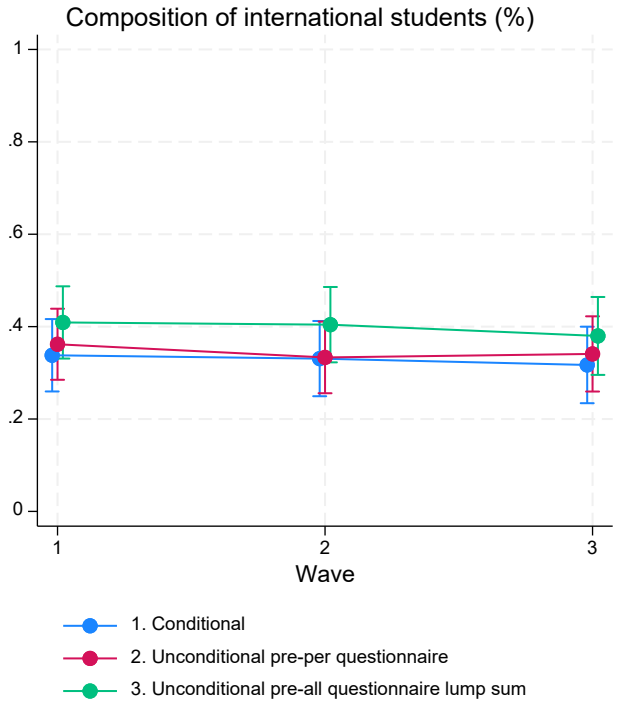
Appendix Figure B-1: Cumulative means of self-reported physical health (1 - Poor; 2 - Fair; 3 - Good; 4 - Very Good; 5 – Excellent). Standard errors are cluster robust.

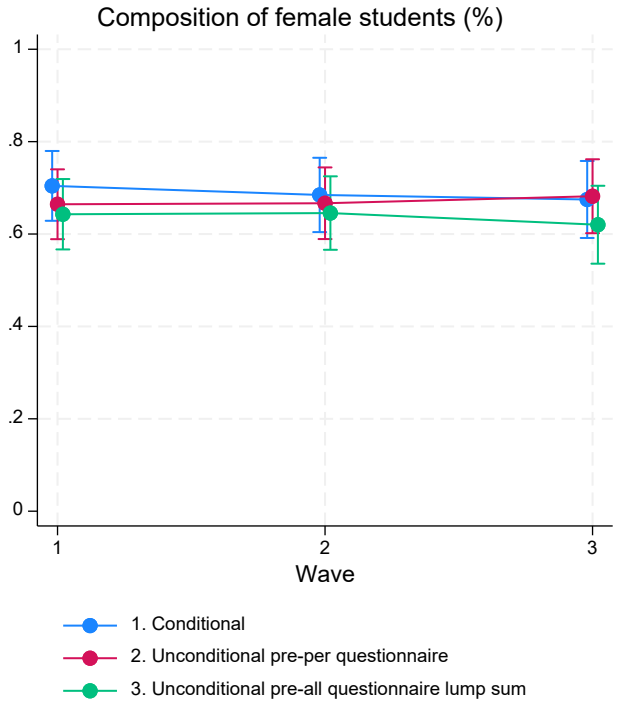


Appendix Figure B-2: Cumulative means of self-reported mental health (1 - Poor; 2 - Fair; 3 - Good; 4 - Very Good; 5 – Excellent). Standard errors are cluster robust.

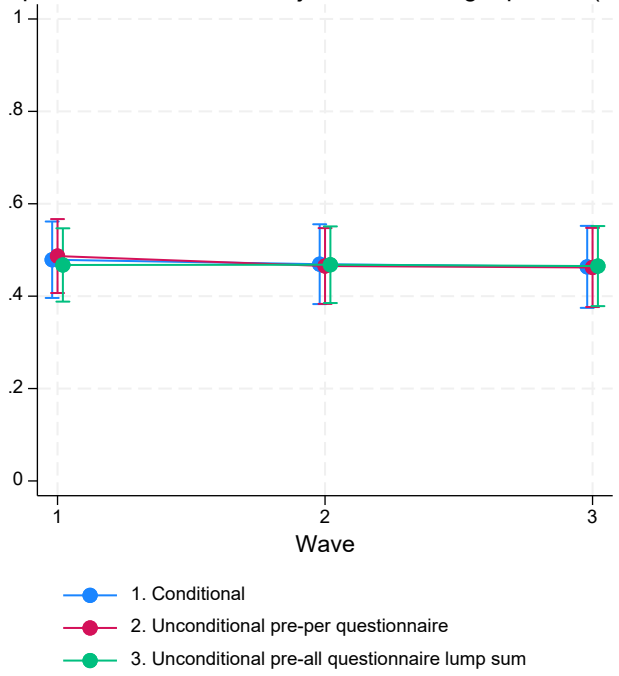


Appendix Figure B-3: Selective attrition analysis based on administrative data

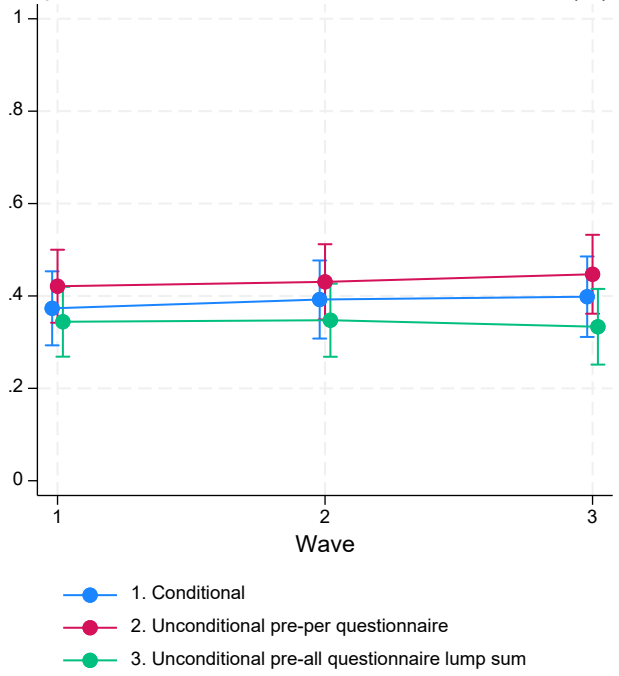




Composition of academic majors with low signup rates (%)



Composition of students admitted before Fall 2023 (%)



Appendix C: Chapter 4 Tables

Appendix Table C-1: Weighted sample demographic composition by survey frequency condition. Categorical variables are reported as percentages and Big-5 personality traits are reported as mean values.

	Overall (n=2,057)	High frequency: Control group (n=1,029)	Low frequency: Treated group (n=1,028)	p-value
Employment status				
Currently working	56.53	58.12	55.09	0.217
Gender				
Female	51.29	48.65	54.02	0.029
Education				
High school or GED	37.84	36.74	38.88	0.612
Some college	26.09	26.82	25.22	
College and above	36.08	36.44	35.90	
Age				
18-44	46.41	45.39	47.33	0.628
45-64	30.07	30.13	30.03	
65+	23.52	24.48	22.64	
HH income				
Less than 50K	47.16	47.52	46.83	0.780
50K-75K	17.55	17.88	17.03	
75K and above	35.29	34.60	36.14	
Race				
White only	73.42	74.83	71.96	0.229
Black only	13.77	12.49	15.09	
Others	12.81	12.68	12.94	
Hispanic				
Yes	17.94	18.09	17.79	0.880
HH Size				
1	30.14	30.49	29.95	0.744
2	36.98	37.65	36.38	
3 and above	32.88	31.86	33.66	
Household structure				
Households with children below 18	8.10	5.69	10.42	0.000
Personality traits				
Conscientiousness	34.32	34.25	34.42	0.658
Openness	34.87	34.71	35.03	0.424
Extraversion	24.75	24.78	24.73	0.882
Agreeableness	34.70	34.48	34.93	0.233

Neuroticism	21.98	22.04	21.88	0.675

Appendix Table C-2: Un-Weighted estimates of multivariable logit regression models predicting responses to UAS core surveys, non-core surveys, and combined core and non-core surveys. All standard errors (SEs) are cluster-robust treating individuals with repeated measures as clusters.

	Coefficient (95% CI)		
	Core surveys	Non-core surveys	Core + Non-core surveys
	(1)	(2)	(3)
Frequency condition (Ref: Low frequency)			
High frequency	-0.058	-0.234*	-0.240*
	(-0.367,0.252)	(-0.486,0.018)	(-0.496,0.016)
Demographic Variables			
Employment status (Ref: Not working)			
Currently working	-0.189	-0.258**	-0.237*
	(-0.497,0.118)	(-0.508,-0.009)	(-0.493,0.019)
High frequency # Currently working	0.545***	0.432***	0.462***
	(0.140,0.951)	(0.108,0.757)	(0.130,0.795)
Gender (Ref: female)			
Male	-0.174	-0.055	-0.099
	(-0.389,0.040)	(-0.227,0.118)	(-0.275,0.078)
Education (Ref: High School or GED)			
Some College	0.411***	0.233**	0.283**
	(0.140,0.681)	(0.012,0.454)	(0.059,0.507)
College and above	0.518***	0.333***	0.386***
	(0.217,0.818)	(0.091,0.575)	(0.139,0.633)
Age (Ref: 18-44)			
45-64	-0.047	0.100	0.054
	(-0.297,0.202)	(-0.099,0.298)	(-0.149,0.257)
65+	0.038	0.084	0.069
	(-0.274,0.351)	(-0.165,0.333)	(-0.187,0.325)
Income (Ref: Less than 50K)			
50K-75K	-0.255	-0.034	-0.100
	(-0.561,0.051)	(-0.274,0.206)	(-0.349,0.149)
75K and above	-0.274*	-0.205*	-0.221*
	(-0.567,0.019)	(-0.435,0.024)	(-0.457,0.016)
Race (Ref: White only)			
Black only	-0.098	-0.240**	-0.192*
	(-0.380,0.184)	(-0.459,-0.021)	(-0.415,0.031)
Others	-0.297**	-0.221**	-0.243**
	(-0.566,-0.028)	(-0.441,-0.001)	(-0.468,-0.018)
Hispanic (Ref: Non-Hispanic)			
Yes	-0.260*	-0.370***	-0.330***
	(-0.552,0.031)	(-0.592,-0.149)	(-0.558,-0.101)
HH Size (Ref: One)			
Two	0.345***	0.234**	0.261**
	(0.090,0.599)	(0.030,0.437)	(0.051,0.470)
Three and above	0.374***	0.279**	0.301***

	(0.102,0.647)	(0.058,0.500)	(0.076,0.526)
Have children below 18 (Ref: No children at home)			
Have children at home	0.319	-0.085	0.026
	(-0.103,0.742)	(-0.380,0.211)	(-0.278,0.329)
Personality Traits			
Conscientiousness	0.040***	0.044***	0.041***
	(0.020,0.061)	(0.029,0.060)	(0.025,0.057)
Openness	0.022**	0.012*	0.015**
	(0.005,0.039)	(-0.001,0.026)	(0.001,0.029)
Extroversion	-0.009	-0.013*	-0.012*
	(-0.027,0.009)	(-0.027,0.001)	(-0.026,0.002)
Neuroticism	0.033***	0.019***	0.021***
	(0.016,0.049)	(0.008,0.031)	(0.009,0.033)
Agreeableness	0.014	0.006	0.007
	(-0.007,0.036)	(-0.010,0.022)	(-0.010,0.024)
Core survey wave (Ref: Wave=1)			
2	-2.494***		
	(-2.786,-2.202)		
3	-2.622***		
	(-2.911,-2.334)		
4	-2.730***		
	(-3.019,-2.441)		
5	-2.821***		
	(-3.109,-2.533)		
6	-2.922***		
	(-3.209,-2.634)		
7	-3.005***		
	(-3.293,-2.718)		
8	-2.988***		
	(-3.276,-2.700)		
9	-3.028***		
	(-3.317,-2.739)		
Core vs. Noncore (Ref: Non-core)			
Core surveys			0.523***
			(0.465,0.580)
High frequency # Core surveys			0.173***
			(0.072,0.274)
Constant	0.307	-1.812***	-1.914***
	(-0.595,1.210)	(-2.378,-1.246)	(-2.522,-1.306)
Observations	15,538	26,392	41,930
Pseudo R^2	0.089	0.038	0.050

* p < 0.10-** p < 0.05-*** p < 0.01. 95% confidence intervals in brackets.

Bibliography

Adachi, T., Tsunekawa, Y., & Tanimura, D. (2022). Association between the Big Five personality traits and medication adherence in patients with cardiovascular disease: A cross-sectional study. *PloS one*, *17*(12), e0278534.

Ajzen, I., & Fishbein, M. (1980). *Understanding Attitudes and Predictiing Social Behavior*.

Alattar, L., Messel, M., & Rogofsky, D. (2018). An Introduction to the Understanding America Study Internet Panel. *Soc. Sec. Bull.*, *78*, 13.

Alattar, L., Messel, M., Rogofsky, D., & Sarney, M. A. (2019). The use of longitudinal data on social security program knowledge. *Soc. Sec. Bull.*, *79*, 1.

Alcendor, D. J. (2020). Racial disparities-associated COVID-19 mortality among minority populations in the US. *Journal of clinical medicine*, *9*(8), 2442.

Almlund, M., Duckworth, A. L., Heckman, J., & Kautz, T. (2011). Personality psychology and economics. In *Handbook of the Economics of Education* (Vol. 4, pp. 1-181). Elsevier.

Angelucci, M., Angrisani, M., Bennett, D. M., Kapteyn, A., & Schaner, S. G. (2020). *Remote work and the heterogeneous impact of COVID-19 on employment and health*.

Angrisani, M., Finley, B., & Kapteyn, A. (2019a). Can Internet match high-quality traditional surveys? Comparing the Health and Retirement Study and its online version. In *The Econometrics of Complex Survey Data*. Emerald Publishing Limited.

Angrisani, M., Finley, B., & Kapteyn, A. (2019b). Can Internet match high-quality traditional surveys? Comparing the Health and Retirement Study and its online version. In *The Econometrics of Complex Survey Data* (Vol. 39, pp. 3-33). Emerald Publishing Limited.

Axelsson, M. (2013). Report on personality and adherence to antibiotic therapy: a population-based study. *BMC psychology*, *1*, 1-6.

Axelsson, M., Brink, E., Lundgren, J., & Lötvall, J. (2011). The influence of personality traits on reported adherence to medication in individuals with chronic disease: an epidemiological study in West Sweden. *PloS one*, *6*(3), e18241.

Ayala, L., Navarro, C., & Sastre, M. (2011). Cross-country income mobility comparisons under panel attrition: the relevance of weighting schemes. *Applied Economics*, *43*(25), 3495-3521.

Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., . . . Tourangeau, R. (2013). Summary report of the AAPOR task force on non-probability sampling. *Journal of Survey Statistics and Methodology*, *1*(2), 90-143.

Basic, E., & Rendtel, U. (2007). Assessing the bias due to non-coverage of residential movers in the German Microcensus Panel: an evaluation using data from the Socio-Economic Panel. *AStA Advances in Statistical Analysis*, *91*, 311-334.

Becker, R. (2023). The researcher, the incentive, the panelists and their response: The role of strong reciprocity for the panelists' survey participation. Survey Research Methods,

Bianchi, A., Biffignandi, S., & Lynn, P. (2016). *Web-CAPI sequential mixed-mode design in a longitudinal survey: effects on participation rates, sample composition and costs*.

Biddle, N., & Sollis, K. (2023). Determinants of participation in a longitudinal survey during the COVID-19 pandemic: the case of a low-infection country. *Longitudinal and Life Course Studies*, *14*(2), 275-293.

Biemer, P. P. (2011). *Latent class analysis of survey error*. John Wiley & Sons.

Biner, P. M., & Kidd, H. J. (1994). The interactive effects of monetary incentive justification and questionnaire length on mail survey response rates. *Psychology & Marketing*, *11*(5), 483-492.

Birkmeyer, J. D., Barnato, A., Birkmeyer, N., Bessler, R., & Skinner, J. (2020). The impact of the COVID-19 pandemic on hospital admissions in the United States: study examines trends in US hospital admissions during the COVID-19 pandemic. *Health Affairs*, *39*(11), 2010-2017.

Blumberg, S. J., Parker, J. D., & Moyer, B. C. (2021). National Health Interview Survey, COVID-19, and online data collection platforms: adaptations, tradeoffs, and new directions. *American Journal of Public Health*, *111*(12), 2167-2175.

Booker, C. L., Harding, S., & Benzeval, M. (2011). A systematic review of the effect of retention methods in population-based cohort studies. *BMC public health*, *11*, 1-12.

Booth, C., Wong, E., Brown, M., & Fitzsimons, E. (2024). Evaluating the Effect of Monetary Incentives on Web Survey Response Rates in the UK Millennium Cohort Study. *Survey Research Methods*, *18*(1), 47-58.

Börsch-Supan, A., Krieger, U., & Schröder, M. (2013). Respondent incentives, interviewer training and survey participation. *München: Munich Center for the Economics of Ageing (MEA)*.

Bosnjak, M., Dannwolf, T., Enderle, T., Schaurer, I., Struminskaya, B., Tanner, A., & Weyandt, K. W. (2018). Establishing an open probability-based mixed-mode panel of the general population in Germany: The GESIS panel. *Social Science Computer Review*, *36*(1), 103-115.

Bradburn, N. (1978). Respondent burden. Proceedings of the Survey Research Methods Section of the American Statistical Association,

Bradley, V. C., Kuriwaki, S., Isakov, M., Sejdinovic, D., Meng, X.-L., & Flaxman, S. (2021a). Unrepresentative big surveys significantly overestimated US vaccine uptake. *Nature*, *600*(7890), 695-700.

Bradley, V. C., Kuriwaki, S., Isakov, M., Sejdinovic, D., Meng, X., & Flaxman, S. (2021b). Unrepresentative Big Surveys Significantly Overestimate US Vaccine Uptake.

Bruine de Bruin, W. (2021). Age differences in COVID-19 risk perceptions and mental health: Evidence from a national US survey conducted in March 2020. *The Journals of Gerontology: Series B*, *76*(2), e24-e29.

Cabrera-Álvarez, P., & Lynn, P. (2023). *Short-term impact of increasing the value of unconditional and conditional incentives in Understanding Society*.

Cabrera-Álvarez, P., & Lynn, P. (2024). Text Messages to Facilitate the Transition to Web-First Sequential Mixed-Mode Designs in Longitudinal Surveys. *Journal of Survey Statistics and Methodology*, *12*(3), 651-673.

Callegaro, M., Baker, R. P., Bethlehem, J., Göritz, A. S., Krosnick, J. A., & Lavrakas, P. J. (2014). *Online panel research: A data quality perspective*. John Wiley & Sons.

Cantor, D., O'Hare, B. C., & O'Connor, K. S. (2008). The use of monetary incentives to reduce nonresponse in random digit dial telephone surveys. *Advances in telephone survey methodology*, 471-498.

Castiglioni, L., Pforr, K., & Krieger, U. (2008). The effect of incentives on response rates and panel attrition: Results of a controlled experiment. *Survey Research Methods*, *2*(3), 151-158.

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of classification*, *13*, 195-212.

Chang, L., & Krosnick, J. A. (2009). National surveys via RDD telephone interviewing versus the internet comparing sample representativeness and response quality. *Public Opinion Quarterly*, *73*(4), 641-678.

Chappell, A., Smith, K., & Strough, J. (2022). OLDER AGE IS (STILL) ASSOCIATED WITH MENTAL HEALTH BENEFITS IN JUNE–JULY 2021 OF THE COVID-19 PANDEMIC. *Innovation in Aging*, *6*(Suppl 1), 511.

Cheng, A., Zamarro, G., & Orriens, B. (2020). Personality as a predictor of unit nonresponse in an internet panel. *Sociological Methods & Research*, *49*(3), 672-698.

Chin, A., Couper, M. P., & Beckett, D. (2021). Attrition in a longitudinal online randomized controlled trial with prospective homeowners. *Field Methods*, *33*(1), 3-24.

Chun, Y., Roll, S., Miller, S., Lee, H., Larimore, S., & Grinstein-Weiss, M. (2023). Racial and ethnic disparities in housing instability during the COVID-19 pandemic: the role of assets and income shocks. *Journal of Economics, Race, and Policy*, *6*(2), 63-81.

Church, A. H. (1993). Estimating the effect of incentives on mail survey response rates: A meta-analysis. *Public Opinion Quarterly*, *57*(1), 62-79.

Cornesse, C., Felderer, B., Fikel, M., Krieger, U., & Blom, A. G. (2022). Recruiting a probability-based online panel via postal mail: experimental evidence. *Social Science Computer Review*, *40*(5), 1259-1284.

Costa, P., & McCrae, R. (1999). A five-factor theory of personality. *Handbook of personality: Theory and research*, *2*(01), 1999.

Costa, P. T., & McCrae, R. R. (1988). Personality in adulthood: a six-year longitudinal study of self-reports and spouse ratings on the NEO Personality Inventory. *Journal of Personality and Social Psychology*, *54*(5), 853.

Couper, M. P., Kapteyn, A., Schonlau, M., & Winter, J. (2007). Noncoverage and nonresponse in an Internet survey. *Social Science Research*, *36*(1), 131-148.

Couper, M. P., & McGonagle, K. (2019). Recent developments in web-based data collection for longitudinal studies. *Panel Study of Income Dynamics Technical Series Paper #19-03. Ann Arbor: Institute of Social Research, University of Michigan.*

Couper, M. P., & Ofstedal, M. B. (2009). Keeping in contact with mobile sample members. *Methodology of longitudinal surveys*, *183*, 203.

Crawford, S. D., Couper, M. P., & Lamias, M. J. (2001). Web surveys: Perceptions of burden. *Social Science Computer Review*, *19*(2), 146-162.

Cubel, M., Nuevo-Chiquero, A., Sanchez-Pages, S., & Vidal-Fernandez, M. (2016). Do personality traits affect productivity? Evidence from the laboratory. *The Economic Journal*, *126*(592), 654-681.

Das, M., Kapteyn, A., & Bosnjak, M. (2018). Open probability-based panel infrastructures. *The Palgrave handbook of survey research*, 199-209.

de la Haye, K., Saw, H.-W., Miller, S., de Bruin, W. B., Wilson, J. P., Weber, K., . . . Kapteyn, A. (2023). Ecological risk and protective factors for food insufficiency in Los Angeles County during the COVID-19 pandemic. *Public health nutrition*, *26*(10), 1944-1955.

de Leeuw, E., Hox, J., & Luiten, A. (2018). International nonresponse trends across countries and years: An analysis of 36 years of labour force survey data. *Survey Methods: Insights From the Field*, 1-11.

Dillman, D. A. (1978). *Mail and telephone surveys: The total design method* (Vol. 19). Wiley New York.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: the tailored design method*. John Wiley & Sons.

Dutwin, D., & Buskirk, T. D. (2021). Telephone sample surveys: dearly beloved or nearly departed? Trends in survey errors in the era of declining response rates. *Journal of Survey Statistics and Methodology*, *9*(3), 353-380.

Earp, M., Kaplan, R., & Toth, D. (2022). Modeling the relationship between proxy measures of respondent burden and survey response rates in a household panel survey. *Journal of Official Statistics*, *38*(4), 1145-1175.

Edwards, P., Roberts, I., Clarke, M., DiGuiseppi, C., Pratap, S., Wentz, R., & Kwan, I. (2002). Increasing response rates to postal questionnaires: systematic review. *Bmj*, *324*(7347), 1183.

Edwards, P. J., Roberts, I., Clarke, M. J., DiGuiseppi, C., Woolf, B., & Perkins, C. (2023). Methods to increase response to postal and electronic questionnaires. *Cochrane database of systematic reviews*(11).

Eggleston, J. (2024). Frequent Survey Requests and Declining Response Rates: Evidence from the 2020 Census and Household Surveys. *Journal of Survey Statistics and Methodology*, smae022.

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2022). The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *Assessment*, *29*(2), 136-151.

Engstrom, C., & Sinibaldi, J. (2024). Reducing Burden in a Web Survey through Dependent Interviewing. *Journal of Survey Statistics and Methodology*, *12*(1), 60-79.

Felderer, B., Müller, G., Kreuter, F., & Winter, J. (2018). The effect of differential incentives on attrition bias: evidence from the PASS wave 3 incentive experiment. *Field Methods*, *30*(1), 56-69.

Fitzgerald, J., Gottschalk, P., & Moffitt, R. A. (1998). An analysis of sample attrition in panel data: The Michigan Panel Study of Income Dynamics. In: National Bureau of Economic Research Cambridge, Mass., USA.

Fomby, P., Sastry, N., & McGonagle, K. A. (2017). Effectiveness of a time-limited incentive on participation by hard-to-reach respondents in a panel study. *Field Methods*, *29*(3), 238-251.

Frankel, L. L., & Hillygus, D. S. (2014). Looking beyond demographics: Panel attrition in the ANES and GSS. *Political Analysis*, *22*(3), 336-353.

Friedman, H. S., & Kern, M. L. (2014). Personality, well-being, and health. *Annual review of psychology*, *65*(1), 719-742.

Gaia, A. (2017). The effect of respondent incentives on panel attrition in a sequential mixed-mode design. *Understanding Society Working Paper 2017-03, Colchester: University of Essex*.

Galesic, M. (2006). Dropouts on the web: Effects of interest and burden experienced during an online survey. *Journal of Official Statistics*, *22*(2), 313.

Gelman, A., Stevens, M., & Chan, V. (2003). Regression modeling and meta-analysis for decision making: A cost-benefit analysis of incentives in telephone surveys. *Journal of Business & Economic Statistics*, *21*(2), 213-225.

Gerry, C. J., & Papadopoulos, G. (2015). Sample attrition in the RLMS, 2001–10: Lessons for longitudinal analysis and an application in health. *Economics of Transition*, *23*(2), 425-468.

Gidengil, C. A., Parker, A. M., & Zikmund-Fisher, B. J. (2012). Trends in risk perception and intention to be vaccinated for H1N1 influenza: a longitudinal study of the first year of the pandemic. *American Journal of Public Health*, *102*(4), 672.

Göritz, A. S. (2006). Incentives in web studies: Methodological issues and a review. *International Journal of Internet Science*, *1*(1), 58-70.

Göritz, A. S., & Neumann, B. P. (2016). The longitudinal effects of incentives on response quantity in online panels. *Translational Issues in Psychological Science*, *2*(2), 163.

Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American sociological review*, 161-178.

Greenberg, P., & Dillman, D. (2023). Mail communications and survey response: A test of social exchange versus pre-suasion theory for improving response rates and data quality. *Journal of Survey Statistics and Methodology*, *11*(1), 1-22.

Groves, R. M., & Couper, M. P. (2012). *Nonresponse in household interview surveys*. John Wiley & Sons.

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: a meta-analysis. *Public Opinion Quarterly*, *72*(2), 167-189.

Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly*, *68*(1), 2-31.

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-saliency theory of survey participation: description and an illustration. *Public opinion quarterly*, *64*(3), 299-308.

Hansson, I., Berg, A. I., & Thorvaldsson, V. (2018). Can personality predict longitudinal study attrition? Evidence from a population-based sample of older adults. *Journal of Research in Personality*, *77*, 133-136.

Hays, R. D., Liu, H., & Kapteyn, A. (2015). Use of Internet panels to conduct surveys. *Behavior research methods*, *47*(3), 685-690.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics*, *24*(3), 411-482.

Henderson, M., Wight, D., Nixon, C., & Hart, G. (2010). Retaining young people in a longitudinal sexual health survey: a trial of strategies to maintain participation. *BMC medical research methodology*, *10*, 1-11.

Hibbing, M. V., Cawvey, M., Deol, R., Bloeser, A. J., & Mondak, J. J. (2019). The relationship between personality and response patterns on public opinion surveys: The big five, extreme response style, and acquiescence response style. *International Journal of Public Opinion Research*, *31*(1), 161-177.

Holbrook, A. L., Krosnick, J. A., & Pfent, A. (2007). The causes and consequences of response rates in surveys by the news media and government contractor survey research firms. *Advances in telephone survey methodology*, 499-528.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65-70.

Hu, M., Kirlin, J. A., West, B. T., He, W., Ong, A. R., Zhang, S., & Zhang, X. (2020). Estimation of underreporting in diary surveys: an application using the National Household Food Acquisition and purchase survey. *Journal of Survey Statistics and Methodology*, *8*(5), 877-902.

Hu, M., Melipillán, E. R., West, B. T., Kirlin, J. A., & Paniagua, I. (2020). Response patterns in a multi-day diary survey: implications for adaptive survey design. Survey Research Methods,

Jäckle, A., & Lynn, P. (2008). Respondent incentives in a multi-mode panel survey: cumulative effects on nonresponse and bias. *Survey Methodology*, *34*(1), 105-117.

Jäckle, A., Lynn, P., & Burton, J. (2015). Going online with a face-to-face household panel: Effects of a mixed mode design on item and unit non-response. *Survey Research Methods*, *9*(1), 57-70.

James, N. (2023). Respondents for Nearly Three Decades: How do loyal sample members differ from others? *Survey Research Methods*, *17*(1), 15-36.

James, T. (1997). Results of the wave 1 incentive experiment in the 1996 survey of income and program participation. *Proceedings of the Survey Research Section of the American Statistical Association*, 834-839.

Jang, M., & Vorderstrasse, A. (2019). Socioeconomic status and racial or ethnic differences in participation: web-based survey. *JMIR research protocols*, *8*(4), e11865.

Jeffery, M. M., D'onofrio, G., Paek, H., Platts-Mills, T. F., Soares, W. E., Hoppe, J. A., . . . Melnick, E. R. (2020). Trends in emergency department visits and hospital admissions in health care systems in 5 states in the first months of the COVID-19 pandemic in the US. *JAMA internal medicine*, *180*(10), 1328-1333.

Jin, H., & Kapteyn, A. (2022). Relationship Between Past Survey Burden and Response Probability to a New Survey in a Probability-Based Online Panel. *Journal of Official Statistics*, *38*(4), 1051-1067.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). Big five inventory. *Journal of Personality and Social Psychology*.

John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives.

Kaminska, O., McCutcheon, A. L., & Billiet, J. (2010). Satisficing among reluctant respondents in a cross-national context. *Public Opinion Quarterly*, *74*(5), 956-984.

Kaplan, R. L., Holzberg, J., Eckman, S., & Giesen, D. (2022). Preface Overview of the Special Issue on Respondent Burden. *Journal of Official Statistics*, *38*(4), 929-938.

Kapteyn, A., Angrisani, M., Bennett, D., de Bruin, W. B., Darling, J., Gutsche, T., . . . Schaner, S. (2020). Tracking the effect of the COVID-19 pandemic on the lives of American households. *Survey Research Methods*, *14*(2), 179-186.

Kapteyn, A., Angrisani, M., Darling, J., & Gutsche, T. (2024). The Understanding America Study (UAS). *BMJ open*, *14*(10), e088183.

Kennedy, C., Hatley, N., Lau, A., Mercer, A., Keeter, S., Ferno, J., & Asare-Marfo, D. (2020). *Assessing the Risks to Online Polls From Bogus Respondents*.

Keusch, F., Struminskaya, B., Antoun, C., Couper, M. P., & Kreuter, F. (2019). Willingness to participate in passive mobile data collection. *Public Opinion Quarterly*, *83*(S1), 210-235.

Kim-Mozeleski, J., SN, P. M., Trapl, E., Perzynski, A., Tsoh, J., & Gunzler, D. (2023). Food Insecurity Trajectories in the US During the First Year of the COVID-19 Pandemic. *Preventing Chronic Disease*, *20*, E03-E03.

Ko, K. M., Moon, S.-J., Koh, J. H., Pae, C.-U., & Min, J.-K. (2020). Contribution of personality traits, psychological factors, and health-related quality of life to medication adherence in patients with rheumatoid arthritis. *Yonsei medical journal*, *61*(5), 406.

Kraemer, F., Silber, H., Struminskaya, B., Sand, M., Bosnjak, M., Koßmann, J., & Weiß, B. (2024). Panel Conditioning in A Probability-based Longitudinal study: A Comparison of Respondents with Different Levels of Survey Experience. *Journal of Survey Statistics and Methodology*, *12*(1), 36-59.

Kreuter, F., Yan, T., & Tourangeau, R. (2008). Good item or bad—can latent class analysis tell?: the utility of latent class analysis for the evaluation of survey questions. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *171*(3), 723-738.

Kristman, V. L., Manno, M., & Côté, P. (2005). Methods to account for attrition in longitudinal data: do they work? A simulation study. *European journal of epidemiology*, *20*, 657-662.

Krosnick, J., MacInnis, B., Suh, A., & Yeager, D. (2013). Assessment of survey accuracy through a multi-modes national field experiment, presented at the American Association for Public Opinion Research national conference", Boston, May 2013
In.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology*, *5*(3), 213-236.

Lauren, B. N., Silver, E. R., Faye, A. S., Rogers, A. M., Woo-Baidal, J. A., Ozanne, E. M., & Hur, C. (2021). Predictors of households at risk for food insecurity in the United States during the COVID-19 pandemic. *Public health nutrition*, *24*(12), 3929-3936.

Laurie, H. (2007). *The effect of increasing financial incentives in a panel survey: an experiment on the British Household Panel Survey, wave 14*.

Laurie, H., & Lynn, P. (2009). The use of respondent incentives on longitudinal surveys. *Methodology of longitudinal surveys*, 205-233.

Lepkowski, J. M., & Couper, M. P. (2002). Nonresponse in the second wave of longitudinal household surveys. *Survey nonresponse*, 259-272.

Lima, M. P., Machado, W. d. L., & Irigaray, T. Q. (2018). Predictive factors of treatment adherence in cancer outpatients. *Psycho-Oncology*, *27*(12), 2823-2828.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*(3), 767-778.

Lugtig, P. (2014). Panel attrition: Separating stayers, fast attriters, gradual attriters, and lurkers. *Sociological Methods & Research*, *43*(4), 699-723.

Lugtig, P., Das, M., & Scherpenzeel, A. (2014). Nonresponse and attrition in a probability-based online panel for the general population. *Online Panel Research: A Data Quality Perspective*, 135-153.

Luiten, A., Hox, J., & de Leeuw, E. (2020). Survey nonresponse trends and fieldwork effort in the 21st century: Results of an international study across countries and surveys. *Journal of Official Statistics*, *36*(3), 469-487.

Lynn, P. (2018). Tackling panel attrition. *The Palgrave handbook of survey research*, 143-153.

MacInnis, B., Krosnick, J. A., Ho, A. S., & Cho, M.-J. (2018). The accuracy of measurements with probability and nonprobability survey samples: Replication and extension. *Public Opinion Quarterly*, *82*(4), 707-744.

Malhotra, N., & Krosnick, J. A. (2007). The effect of survey mode and sampling on inferences about political attitudes and behavior: Comparing the 2000 and 2004 ANES to Internet surveys with nonprobability samples. *Political Analysis*, *15*(3), 286-323.

Manfreda, K. L., Bosnjak, M., Berzelak, J., Haas, I., & Vehovar, V. (2008). Web surveys versus other survey modes: A meta-analysis comparing response rates. *International journal of market research*, *50*(1), 79-104.

Martin, E., & Winters, F. (2001). Money and motive: effects of incentives on panel attrition in the survey of income and program participation. *Journal of Official Statistics*, *17*(2), 267.

Maslovskaya, O., Struminskaya, B., & Durrant, G. (2022). The future of online data collection in social surveys: Challenges, developments and applications. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *185*(3), 768-772.

Mavletova, A., & Couper, M. P. (2016). Device use in web surveys: The effect of differential incentives. *International journal of market research*, *58*(4), 523-544.

McCarthy, J. S., Beckler, D. G., & Qualey, S. M. (2006). An Analysis of the Relationship Between Survey Burden And Nonresponse: If we bother them more, are they less cooperative? *Journal of Official Statistics*, *22*(1), 97.

McCrae, R. R., & Costa Jr, P. T. (1994). The stability of personality: Observations and evaluations. *Current directions in psychological science*, *3*(6), 173-175.

McCrae, R. R., Costa Jr, P. T., Martin, T. A., Oryol, V. E., Rukavishnikov, A. A., Senin, I. G., . . . Urbánek, T. (2004). Consensual validation of personality traits across cultures. *Journal of Research in Personality*, *38*(2), 179-201.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of personality*, *60*(2), 175-215.

McFall, S. (2013). Understanding Society: UK household longitudinal study: Cognitive ability measures. *Institute for Social and Economic Research, University of Essex*.

McGonagle, K., Couper, M., & Schoeni, R. F. (2011). Keeping track of panel members: an experimental test of a between-wave contact strategy. *Journal of Official Statistics*, *27*(2), 319.

McGonagle, K. A. (2020). The Effects of an Incentive Boost on Response Rates, Fieldwork Effort, and Costs across Two Waves of a Panel Study. *Methoden, daten, analysen*, *14*(2), 241.

McGonagle, K. A., Sastry, N., & Freedman, V. A. (2023). The effects of a targeted "early Bird" incentive strategy on response rates, Fieldwork effort, and costs in a National panel study. *Journal of Survey Statistics and Methodology*, *11*(5), 1032-1053.

McLauchlan, C., & Schonlau, M. (2016). Are final comments in web survey panels associated with next-wave attrition? Survey Research Methods,

McMaster, H. S., LeardMann, C. A., Speigle, S., & Dillman, D. A. (2017). An experimental comparison of web-push vs. paper-only survey procedures for conducting an in-depth health survey of military spouses. *BMC medical research methodology*, *17*, 1-9.

Mercer, A., Caporaso, A., Cantor, D., & Townsend, R. (2015). How much gets you how much? Monetary incentives and response rates in household surveys. *Public Opinion Quarterly*, *79*(1), 105-129.

Messer, B. L., & Dillman, D. A. (2011). Surveying the general public over the internet using address-based sampling and mail contact procedures. *Public Opinion Quarterly*, *75*(3), 429-457.

Molloy, G., O'carroll, R., & Ferguson, E. (2014). Conscientiousness and medication adherence: a meta-analysis. *Annals of Behavioral Medicine*, *47*(1), 92-101.

Montagni, I., Cariou, T., Tzourio, C., & González-Caballero, J.-L. (2019). "I don't know","I'm not sure","I don't want to answer": a latent class analysis explaining the informative value of nonresponse options in an online survey on youth health. *International Journal of Social Research Methodology*, *22*(6), 651-667.

Morales, D. X., Morales, S. A., & Beltran, T. F. (2021). Racial/ethnic disparities in household food insecurity during the COVID-19 pandemic: a nationally representative study. *Journal of racial and ethnic health disparities*, *8*, 1300-1314.

Muthén, L. K., & Muthén, B. O. (1998-2017). Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén.

Na, L., Yang, L., Mezo, P. G., & Liu, R. (2022). Age disparities in mental health during the COVID19 pandemic: The roles of resilience and coping. *Social science & medicine*, *305*, 115031.

Newman, J. C., Des Jarlais, D. C., Turner, C. F., Gribble, J., Cooley, P., & Paone, D. (2002). The differential effects of face-to-face and computer interview modes. *American Journal of Public Health*, *92*(2), 294-297.

Nicoletti, C., & Buck, N. (2003). Explaining Contact and Refusals in the British and German Household Panels. *Report on Attrition Analysis and Item Non-response, Chintex Working Paper*(16).

NORC at the University of Chicago. (2022). *TECHNICAL OVERVIEW OF THE AMERISPEAK® PANEL NORC'S PROBABILITY-BASED HOUSEHOLD PANEL*. https://amerispeak.norc.org/content/dam/amerispeak/research/pdf/AmeriSpeak%20Technical%20Overview%202019%2002%2018.pdf

Nourazari, S., Davis, S. R., Granovsky, R., Austin, R., Straff, D. J., Joseph, J. W., & Sanchez, L. D. (2021). Decreased hospital admissions through emergency departments during the COVID-19 pandemic. *The American journal of emergency medicine*, *42*, 203-210.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural equation modeling: A multidisciplinary Journal*, *14*(4), 535-569.

Olsen, R. (2018). Panel attrition. *The Palgrave handbook of survey research*, 509-517.

Olsen, R. J. (2018). Respondent attrition versus data attrition and their reduction. *The Palgrave handbook of survey research*, 155-158.

Olson, K., & Witt, L. (2011). Are we keeping the people who used to stay? Changes in correlates of panel survey attrition over time. *Social Science Research*, *40*(4), 1037-1050.

Ono, M., Schneider, S., Junghaenel, D. U., & Stone, A. A. (2019). What affects the completion of ecological momentary assessments in chronic pain research? An individual patient data meta-analysis. *Journal of medical Internet research*, *21*(2), e11398.

Paperwork Reduction Act Guide. (2011). https://www.opm.gov/about-us/open-government/digital-government-strategy/fitara/paperwork-reduction-act-guide.pdf

Pashazadeh, F., Cernat, A., & Sakshaug, J. W. (2021). The effects of biological data collection in longitudinal surveys on subsequent wave cooperation. *Advances in Longitudinal Survey Methodology*, 100-121.

Patrick, M. E., Couper, M. P., Laetz, V. B., Schulenberg, J. E., O'Malley, P. M., Johnston, L. D., & Miech, R. A. (2018). A sequential mixed-mode experiment in the US National Monitoring the Future Study. *Journal of Survey Statistics and Methodology*, *6*(1), 72-97.

Perez-Arce, F., Angrisani, M., Bennett, D., Darling, J., Kapteyn, A., & Thomas, K. (2021). COVID-19 vaccines and mental distress. *PloS one*, *16*(9), e0256406.

Perlis, T. E., Des Jarlais, D. C., Friedman, S. R., Arasteh, K., & Turner, C. F. (2004). Audio-computerized self-interviewing versus face-to-face interviewing for research data collection at drug abuse treatment programs. *Addiction*, *99*(7), 885-896.

Petrolia, D. R., & Bhattacharjee, S. (2009). Revisiting incentive effects: evidence from a random-sample mail survey on consumer preferences for fuel ethanol. *Public Opinion Quarterly*, *73*(3), 537-550.

Peycheva, D., Calderwood, L., Wong, E., & Silverwood, R. (2023). Effects of a time-limited push-to-web incentive in a mixed-mode longitudinal study of young adults. *Survey Research Methods*, *17*(2), 147-157.

Porter, S. R., Whitcomb, M. E., & Weitzer, W. H. (2004). Multiple surveys of students and survey fatigue. *New directions for institutional research*, *2004*(121), 63-73.

Radler, B. T., & Ryff, C. D. (2010). Who participates? Accounting for longitudinal retention in the MIDUS national study of health and well-being. *Journal of aging and health*, *22*(3), 307-331.

Raifman, M. A., & Raifman, J. R. (2020). Disparities in the population at risk of severe illness from COVID-19 by race/ethnicity and income. *American journal of preventive medicine*, *59*(1), 137-139.

Rennert-May, E., Leal, J., Thanh, N. X., Lang, E., Dowling, S., Manns, B., . . . Ronksley, P. E. (2021). The impact of COVID-19 on hospital admissions and emergency department visits: a population-based study. *PloS one*, *16*(6), e0252441.

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological science*, *2*(4), 313-345.

Rodgers, W. L. (2011). Effects of increasing the incentive size in a longitudinal study. *Journal of Official Statistics*, *27*(2), 279-299.

Rolstad, S., Adler, J., & Rydén, A. (2011). Response burden and questionnaire length: is shorter better? A review and meta-analysis. *Value in Health*, *14*(8), 1101-1108.

Rossen, L. M., Branum, A. M., Ahmad, F. B., Sutton, P., & Anderson, R. N. (2020). Excess deaths associated with COVID-19, by age and race and ethnicity—United States, January 26–October 3, 2020. *Morbidity and Mortality Weekly Report*, *69*(42), 1522.

Ryu, E., Couper, M. P., & Marans, R. W. (2006). Survey incentives: Cash vs. in-kind; face-to-face vs. mail; response rate vs. nonresponse error. *International Journal of Public Opinion Research*, *18*(1), 89-106.

Saßenroth, D. (2013). *The impact of personality on participation decisions in surveys: A contribution to the discussion on unit nonresponse*. Springer Science & Business Media.

Sastry, N., Fomby, P., & McGonagle, K. A. (2021). Effects on Panel Attrition and Fieldwork Outcomes from Selection for a Supplemental Study: Evidence from the Panel Study of Income Dynamics. *Advances in Longitudinal Survey Methodology*, 74-99.

Satherley, N., Milojev, P., Greaves, L. M., Huang, Y., Osborne, D., Bulbulia, J., & Sibley, C. G. (2015). Demographic and psychological predictors of panel attrition: Evidence from the New Zealand Attitudes and Values Study. *PloS one*, *10*(3), e0121950.

Schneider, S., May, M., & Stone, A. A. (2018). Careless responding in internet-based quality of life assessments. *Quality of Life Research*, *27*(4), 1077-1088.

Schoeni, R. F., Stafford, F., McGonagle, K. A., & Andreski, P. (2013). Response rates in national panel surveys. *The Annals of the American academy of political and social science*, *645*(1), 60-87.

Schonlau, M., Van Soest, A., Kapteyn, A., & Couper, M. (2009). Selection bias in web surveys and the use of propensity scores. *Sociological Methods & Research*, *37*(3), 291-318.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461-464.

Singer, E. (2002). The use of incentives to reduce nonresponse in household surveys. *Survey nonresponse*, *51*(1), 163-177.

Singer, E. (2011). Toward a benefit-cost theory of survey participation: Evidence, further tests, and implications. *Journal of Official Statistics*, *27*(2), 379-392.

Singer, E., Van Hoewyk, J., Gebler, N., & McGonagle, K. (1999). The effect of incentives on response rates in interviewer-mediated surveys. *Journal of Official Statistics*, *15*(2), 217.

Singer, E., & Ye, C. (2013). The use and effects of incentives in surveys. *The Annals of the American Academy of Political and Social Science*, *645*(1), 112-141.

Sinibaldi, J., & Karlsson, A. Ö. (2017). The effect of rest period on response likelihood. *Journal of Survey Statistics and Methodology*, *5*(1), 70-83.

Soto, C. J. (2019). How replicable are links between personality traits and consequential life outcomes? The life outcomes of personality replication project. *Psychological Science*, *30*(5), 711-727.

StataCorp. (2023). Stata 18 User's Guide. College Station, TX: Stata Press.

Stone, A. A., Broderick, J. E., Schwartz, J. E., Shiffman, S., Litcher-Kelly, L., & Calvanese, P. (2003). Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction. *Pain*, *104*(1-2), 343-351.

Stone, A. A., Schneider, S., & Smyth, J. M. (2023). Evaluation of pressing issues in ecological momentary assessment. *Annual Review of Clinical Psychology*, *19*(1), 107-131.

Stone, A. A., Schneider, S., Smyth, J. M., Junghaenel, D. U., Couper, M. P., Wen, C., . . . Goldstein, S. (2024). A population-based investigation of participation rate and self-selection bias in momentary data capture and survey studies. *Current Psychology*, *43*(3), 2074-2090.

Stone, A. A., Schneider, S., Smyth, J. M., Junghaenel, D. U., Wen, C., Couper, M. P., & Goldstein, S. (2023). Shedding light on participant selection bias in Ecological Momentary Assessment (EMA) studies: Findings from an internet panel study. *PloS one*, *18*(3), e0282591.

Sturgis, P., & Brunton-Smith, I. (2023). Personality and survey satisficing. *Public Opinion Quarterly*, *87*(3), 689-718.

Toepoel, V. (2012). Effects of incentives in surveys. *Handbook of survey methodology for the social sciences*, 209-223.

Toepoel, V., & Funke, F. (2018). Sliders, visual analogue scales, or buttons: Influence of formats and scales in mobile and desktop surveys. *Mathematical Population Studies*, *25*(2), 112-122.

Tourangeau, R. (1999). Remembering what happened: Memory errors and survey reports. In *The science of self-report* (pp. 41-60). Psychology Press.

Tourangeau, R. (2000). *The psychology of survey response*. Cambridge University Press.

Tourangeau, R. (2020). How Errors Cumulate: Two Examples. *Journal of Survey Statistics and Methodology*, *8*(3), 413-432.

Trabilsy, M., Ortiz, K., & Camacho-Rivera, M. (2023). COVID-19-Associated Food Insecurity and Mental Health Symptoms Among Latinx Adults in the United States. *Hispanic Health Care International*, 15404153231208136.

Trappmann, M., Gramlich, T., & Mosthaf, A. (2015). The effect of events between waves on panel attrition. Survey Research Methods,

Tzamourani, P. (2000). An experiment with promised contribution to charity as a respondent incentive on a face-to-face survey. *Survey Methods Newsletter*, *20*(2), 13-15.

Uhrig, S. (2008). *The nature and causes of attrition in the British Household Panel Study*.

Understanding America Study. (2023). *WEIGHTING PROCEDURE*. https://uasdata.usc.edu/index.php?r=eNpLtDKyqi62MrFSKkhMT1WyLrYyslwwskuTcjKT9VLyk0tzU_NKEksy8_NS8svzcvITU0BqgMrzEnMhys2tlEIdgxXCUzPTM0oy89IVAoryk1NTSotS9QpS0pSsawFtoCFw

Valentino, N. A., Zhirkov, K., Hillygus, D. S., & Guay, B. (2020). The consequences of personality biases in online panels for measuring public opinion. *Public Opinion Quarterly*, *84*(2), 446-468.

Versey, H. S. (2021). The Impending Eviction Cliff: Housing Insecurity During COVID-19. In (Vol. 111, pp. 1423-1427): American Public Health Association.

Walsh, E., & Brinker, J. K. (2015). Temporal considerations for self-report research using short message service. *Journal of Media Psychology*.

Ward, M., Meade, A. W., Allred, C. M., Pappalardo, G., & Stoughton, J. W. (2017). Careless response and attrition as sources of bias in online survey assessments of personality traits and performance. *Computers in Human Behavior*, *76*, 417-430.

Watson, N., & Wooden, M. (2009). Identifying factors affecting longitudinal survey response. *Methodology of longitudinal surveys*, *1*, 157-182.

West, B. T., Zhang, S., Wagner, J., Gatward, R., Saw, H.-W., & Axinn, W. G. (2023). Methods for improving participation rates in national self-administered web/mail surveys: Evidence from the United States. *PloS one*, *18*(8), e0289695.

Westra, A., Sundukchi, M., & Mattingly, T. (2015). Designing a Multipurpose Longitudinal Incentives Experiment for the Survey of Income and Program Participation. Annual Conference of the Federal Committee on Statistical Methodology, Washington, DC,

Wheeler, K., Wagaman, A., & McCord, D. (2012). Personality traits as predictors of adherence in adolescents with type I diabetes. *Journal of child and adolescent psychiatric nursing*, *25*(2), 66-74.

Williams, D., & Brick, J. M. (2018). Trends in US face-to-face household survey nonresponse and level of effort. *Journal of Survey Statistics and Methodology*, *6*(2), 186-211.

Wong, E. (2020). Incentives in longitudinal studies. *London: Centre for Longitudinal Studies*.

Wood, M., & Kunz, S. (2014). *CAWI in a mixed mode longitudinal design*.

Wright, A. J., & Jackson, J. J. (2023). Do changes in personality predict life outcomes? *Journal of Personality and Social Psychology*.

Yan, T., Fricker, S., & Tsai, S. (2020). Response burden: What is it and what predicts it? *Advances in questionnaire design, development, evaluation and testing*, 193-212.

Yan, T., & Williams, D. (2022). Response burden–Review and conceptual framework. *Journal of Official Statistics*, *38*(4), 939-961.

Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A., & Wang, R. (2011). Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples. *Public Opinion Quarterly*, nfr020.

Zabel, J. E. (1998). An analysis of attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an application to a model of labor market behavior. *Journal of Human Resources*, 479-506.

Zagorsky, J. L., & Rhoton, P. (2008). The effects of promised monetary incentives on attrition in a long-term panel survey. *Public opinion quarterly*, *72*(3), 502-513.