

The General Social Survey



Exploring Web and Face-to-face Sequential Data Collection Designs in the General Social Survey

GSS Methodological Report #140
August 2024

Authors:
Brian M. Wells
Leah Christian
René Bautista
Sara Lafia
Michael Davern



Exploring Web and Face-to-face Sequential Data Collection Designs in the General Social Survey

Prepared by Brian M. Wells, Leah Christian, René Bautista, Sara Lafia, and Michael Davern

Multimode data collection methodologies can help address survey challenges such as declining rates of participation, rising costs, and increasing needs for more timely data collection. However, the order of survey modes as part of a data collection protocol can have an impact on how effective a design can be. This brief explores how the sequential ordering of web and face-to-face (FTF) in a nationally representative survey may impact response rates, key trends, and overall costs.

In 2022, the General Social Survey (GSS) was fielded as a multimode study where respondents were randomly assigned to one of two data collection sequences in an experimental design. The first sequence used FTF as the primary mode and then invited all nonrespondents to complete the survey on the web (FTF-first). The second sequence started with a push-to-web (PTW) methodology and then invited a subsample of nonrespondents to complete the survey in a FTF interview (Web-first).

Our analyses found that both sequences produced comparable results and neither sequence achieved a better response rate. For costs, the Web-first sequencing was more cost effective per completed interview, but the PTW follow-up in the FTF-first sequence increased response rates at a lower cost and did not require subsampling of nonrespondents.

Background

Survey designs that rely on multiple modes of data collection have become increasingly utilized in recent decades to address declining response rates and rising data collection costs (de Leeuw, 2018; Olson et al., 2021). Using multiple survey modes has also been shown to be more representative of the population than using a single mode (Cornesse & Bosjnak, 2018). While multimode studies are common, there are many unanswered questions about which designs are best for a given set of conditions (e.g., population of interest, survey topic, timeline). This includes which data collection modes

to offer and whether to provide data collection modes concurrently or sequentially; and if sequentially, what order of mode administration would make the most sense.

Offering data collection modes in sequence can help improve response over a single mode alone by reducing coverage and nonresponse-related errors (Suzer-Gurtekin et al., 2018). Sequential designs also help to control survey costs. Less expensive survey modes (e.g., web) are typically used first in sequential designs to guarantee more cost savings (Wagner et al., 2014). However, long established surveys that have historically utilized more costly interviewer-administered modes (e.g., face-to-face)

may be hesitant to switch away from their traditional modes of collection because additional modes of administration can introduce new sources of survey error, such as sample imbalances or potentially mode effects (de Leeuw, 2018). The introduction of web as the primary data collection mode could also impact measurement properties (i.e., changes in response distributions) in historically interviewer-administered surveys (de Leeuw, 2018). This brief will examine how the sequential ordering of web and face-to-face (FTF) modalities impacts response rates, demographic and attitudinal outcomes, and overall costs in a nationally representative survey: the General Social Survey (Davern et al., 2024).

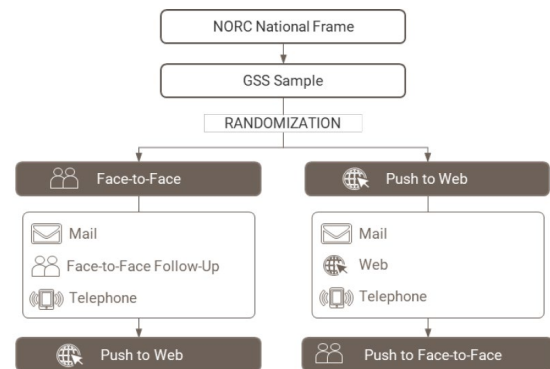
The 2022 GSS Design

The General Social Survey (GSS) is a series of nationally representative, cross-sectional surveys that has been conducted since 1972. The GSS monitors public opinion trends by providing a standard core of demographic, behavioral, attitudinal, and topical questions. The GSS aims to provide high-quality research data with minimal costs and latency.

Survey data collection methods have evolved over the course of the GSS. For most of its 50-year history, GSS data collection has been based on area probability sampling methods using primarily FTF data collection methodologies, using a small number of phone interviews to aid with nonresponse. In response to the COVID-19 pandemic in 2021, the GSS shifted to a remote-only survey using mail push-to-web (PTW) and supplementing with telephone interviews where necessary. In mail-PTW (also known as web-push), a survey invitation is mailed to an address inviting them to complete the survey online (Dillman, 2017). Other large-scale studies such as the American National Election Studies (ANES) have explored a similar approach with encouraging results (DeBell et al., 2018).

Learning from the 2021 GSS experience, NORC fielded the 2022 GSS as a multimode experimental study with both FTF interviews and web questionnaires as the primary modes of data collection, with a small number of supplementary phone interviews. More specifically, the 2022 GSS included a data collection sequence experiment

Figure 1. GSS 2022 Sequential Design Overview



transposing the order of FTF and PTW as the first mode of survey administration. The experiment sought to evaluate how best to combine FTF and PTW methods to minimize coverage, measurement and nonresponse error, and survey costs.

The mode assignment and nonresponse follow-up (NRFU) contact protocol was as follows (Figure 1). The GSS sample was randomly assigned to one of two experimental conditions, FTF-first or Web-first, which determined if participants would be offered an in-person interview or web questionnaire first. In the FTF-first sequence, nonrespondents were offered the option to complete the survey online via the web. For the Web-first sequence, only a subsample of nonrespondents (~20 percent) were offered the option to complete the survey in person. Telephone was used in both sequences at the interviewer's discretion, when it was deemed important to offer a telephone option.

The design allowed the GSS to compare both sequences because they were administered at the same time. The experiment compared the two data collection sequences by examining the impacts of the multimode design on costs and data quality. The experiment also allowed for an examination of whether the order of survey data collection modes affected response rates, demographic representativeness, and all attitudinal differences.

Our Research

Our primary analysis examines the differences between the FTF-first and Web-first completes. We compared both the full data collection sequence

samples and the sample achieved by the first primary mode offered. For demographics, we used a two-sample test of proportions accounting for complex sample design and used pre-raking design weights to prevent such adjustments from influencing our conclusions. For substantive responses, we used a two-sample test of proportions accounting for the complex sample design and used weights that accounted for nonresponse and adjusted to U.S. population estimates (i.e., raked weights). In total, we examined 177 attitudinal variables across a range of subject domains, selecting a subset of the most common and popular GSS Data Explorer variables.

Findings

RESPONSE RATES AND RESPONSE MODES BY EXPERIMENTAL CONDITION

Overall, both full data collection sequences achieved around a 50 percent weighted response rate (AAPOR RR3; AAPOR, 2016; Davern et al., 2024). The Web-first condition (whose NRFU mode is F2F) obtained 2,004 completes compared to 1,540 for the FTF-first condition (which has a NRFU component based on a

Table 1. Responses rates and response modes by experimental condition

Response Mode	Web-first Sequence		FTF-first Sequence		Total Completes	
	n	%	n	%	n	%
Web	1,208	60%	428	28%	1,636	46%
Face-to-Face	619	31%	954	62%	1,573	44%
Phone	100	5%	106	7%	206	6%
Mixed Mode	77	4%	52	3%	129	4%
Total Completes	2,004		1,540		3,544	
Sampled Households	11,168		3,844		15,012	
Response Rate	50.8%		49.9%		50.5%	

Note: Column percentages by data collection sequence are unweighted. Weighted response rates are calculated using WTSSNRPS excluding AmeriSpeak oversample cases and are AAPOR RR3.

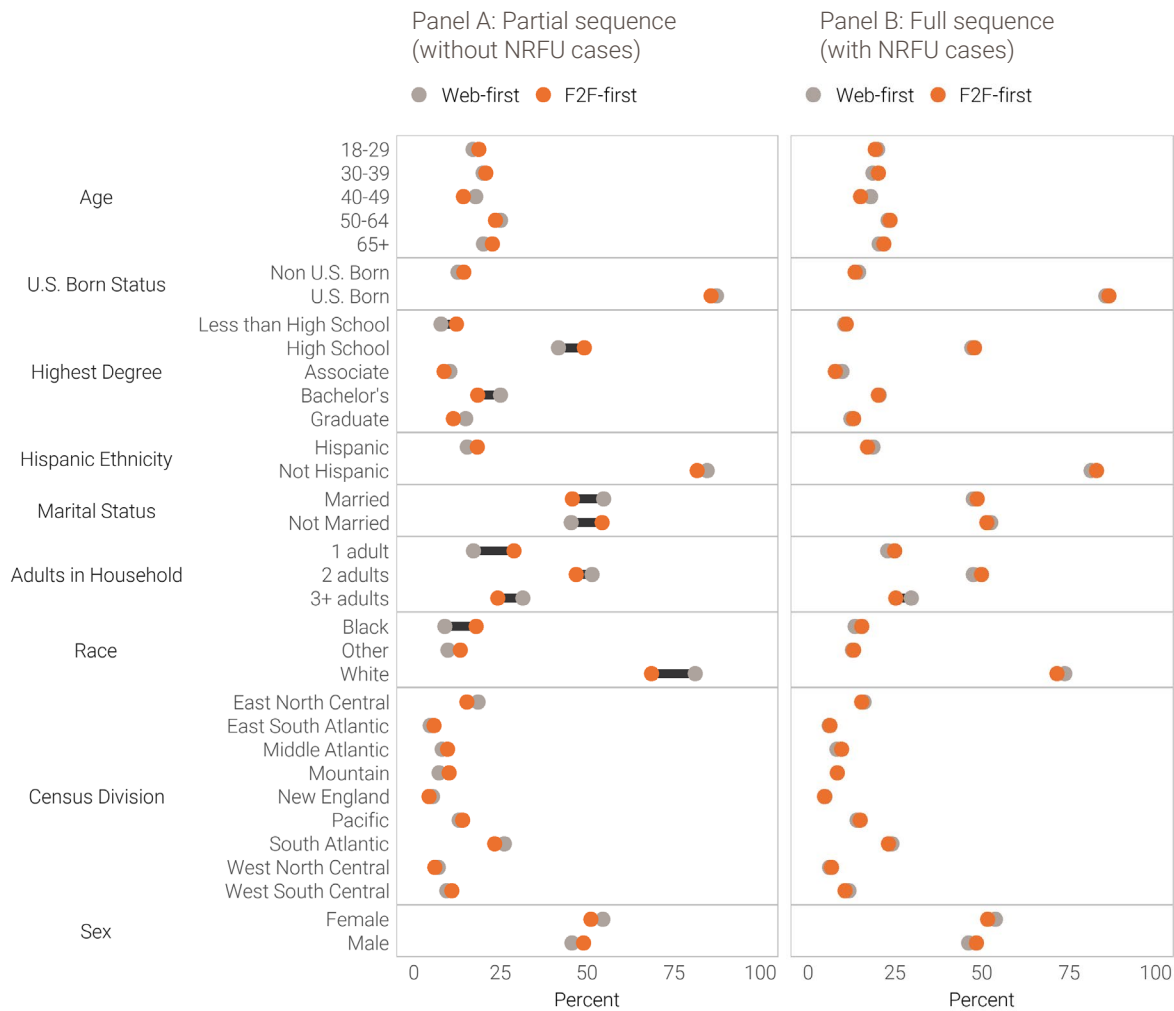
web survey) (**Table 1**). The Web-first condition achieved this higher number of completes while achieving an equivalent response rate, but it did require far more sampled households to do so. Between the two sequences, we saw that most respondents completed the survey via the first mode offered. Sixty percent of the Web-first sample completed via web whereas 62 percent in the FTF-first sequence completed a FTF interview. Approximately 30 percent of respondents completed in the NRFU mode offered (i.e., web survey for FTF-first, or FTF interviewing for Web-first). About five to seven percent completed on the phone. Less than five percent used a combination of survey modalities (e.g., people starting the survey F2F and finishing on the phone).

DEMOGRAPHIC CHARACTERISTICS BY EXPERIMENTAL CONDITION

For demographic analysis, we compare responses obtained with the first mode offered in each sequence (i.e., either web or FTF) prior to the NRFU. In that comparison, we see nine significant differences between the two experimental conditions in relation to the respondent's race, education, marital status, and in the number of adults in the household (**Figure 2, Panel A**). Succinctly, the web completes from the Web-first experimental condition include more White persons, college educated, and married respondents, and three or more adult households, relative to the F2F-first condition.

When we compare demographics for the two experimental conditions including NRFU cases (that is, when we compare the full data collection sequences), there was only one significant difference in demographics (**Figure 2, Panel B**). The full Web-first sequence saw more households with three or more adults than the full FTF-first sequence. The differences previously observed for race, education, and marital status all come within about two percentage points between the two sequences, which can be attributed to sampling error. Overall, the addition of the nonresponse follow-up to each experimental condition (either F2F or web), significantly helped to balance each sample so that they were more similar to each other.

Figure 2. Demographic differences by experimental condition with and without nonresponse follow-up (NRFU)



Note: Proportions estimated using design weights (i.e., pre-raking). Statistically significant response categories at $p < 0.05$ for the partial sequence: less than high school, high school, bachelor’s, married, not married, 1 adult, 3+ adults, Black, and White. Statistically significantly response categories at $p < 0.05$ for the full sequence: 3+ adults.

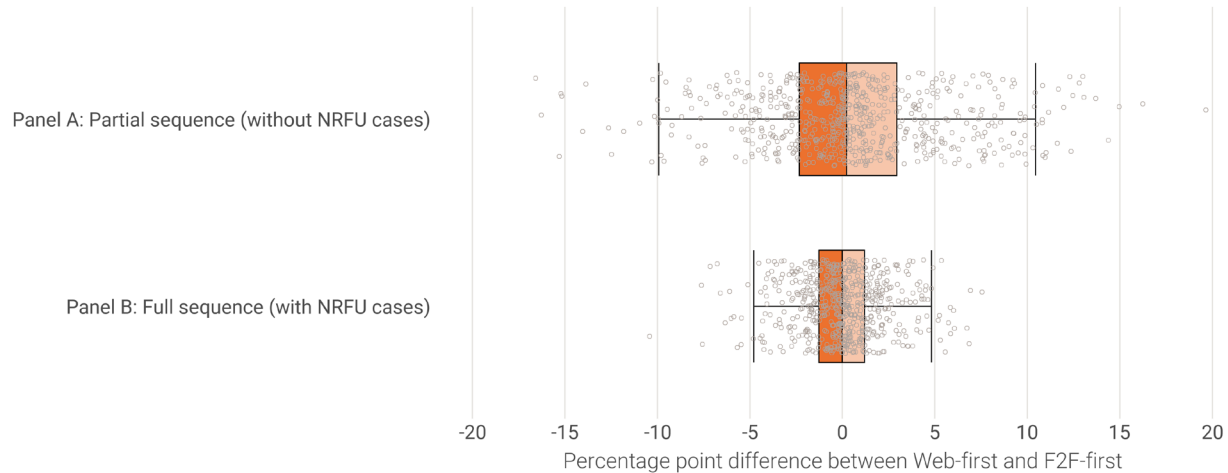
SURVEY RESPONSES BY EXPERIMENTAL CONDITION

When we analyze attitudinal data as if it were a single-mode study (i.e., excluding NRFU cases collected in the secondary survey mode), we find that response distributions are not equivalent. Looking at 614 response categories (of 177 variables) after only the first mode of each sequence (**Figure 3, Panel A**), there are large differences between Web-first and FTF-first. While the middle half of the differences presented in the analysis without NRFU cases are within three percentage points, there are many other survey questions with differences above ten percentage points. However, when cases collected as part of NRFU are added to

the analysis (**Figure 3, Panel B**), the estimates between the two sequences suggest far more comparable results with 83 percent of differences within three percentage points. The reduction in the range of differences shows how introducing the NRFU modes helps to minimize differences between the two sequences.

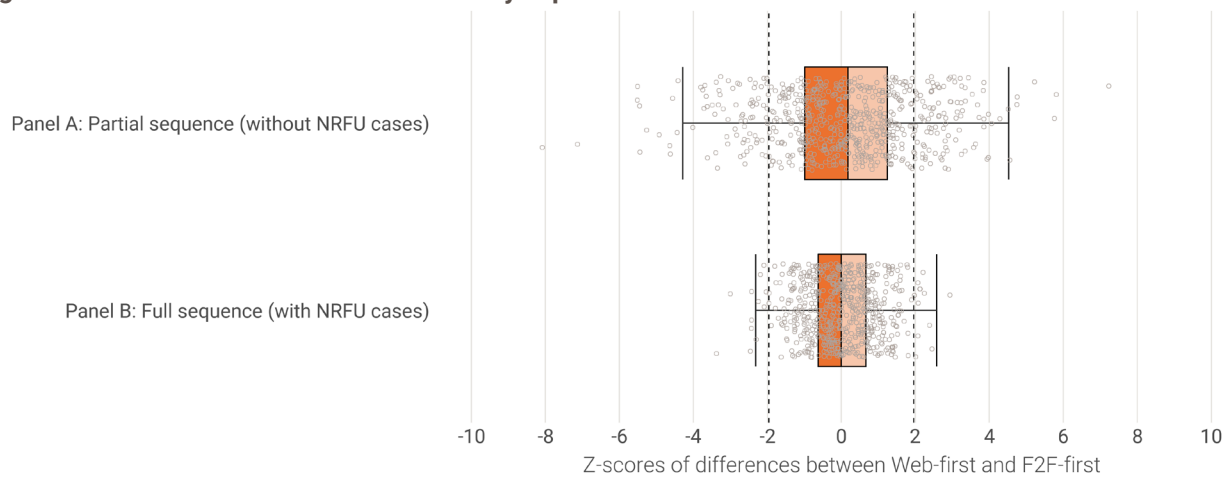
To further illustrate this last point, and to make estimate differences more comparable, we standardized survey response differences (**Figure 4**). The addition of the second data collection mode at the NRFU stage reduced the number of significant differences. When examining the estimates with only the initially offered mode, thirty percent of the individual-level responses were statistically different

Figure 3. Percentage point differences of estimates by experimental condition



Note: Weighted using WTSSNRPS.

Figure 4. Z-scores of estimate differences by experimental condition



Note: Weighted using WTSSNRPS. Dashed line represents a standard 95% confidence level ($z = 1.96$).

at the 0.05 alpha level. In contrast, less than four percent of response differences were significantly different when we considered the full data collection sequence, which is just as likely by random chance at an alpha level of 0.05%.

RELATIVE COSTS

While the sequence order of web and FTF minimally impacts the difference on attitudinal estimates or demographic composition, the order of survey mode does impact costs. Our cost modeling found that the cost per completed interview for a FTF-first complete is about 50 percent larger than the cost per completed interview for a Web-first complete.

Accordingly, the Web-first sequence proved to be more cost effective than FTF-first obtaining 60 percent of completes by the most affordable survey mode. Adjusting for the greater number of completes in the Web-first condition, this results in an over 30 percent relative cost savings compared to the FTF-first condition. The PTW follow-up for nonrespondents in the FTF-first sequence significantly increased the response rate at a lower cost than the historical in-person follow-up. These results illustrate that optimal data collection sequence is a function of the relative costs of PTW and FTF based on fixed set-up costs and variable costs per case.

Discussion

The findings of this experiment showcase the promise of push-to-web (PTW) or face-to-face (FTF) data collection sequences for large scale studies, such as the General Social Survey. The comparability in weighted response rates, demographic composition and survey estimates suggest that the data collection sequence of PTW and FTF have minimal impact on survey participation or attitudinal estimation when a full multimode sequence is used.

Despite the similarities in both sequences, we did observe some differential benefits for each sequence. The FTF-first sequence was able to make effective use of a more affordable data collection mode (i.e., web surveys) for all nonrespondents which did not require subsampling for nonresponse follow-up, simplifying statistical weighting. The Web-first sequence, however, minimized early data collection costs by obtaining a majority of final responses quickly and reducing the need for in-person follow-up. However, the Web-first design required subsampling of cases to reduce data collection costs for the nonresponse FTF interviewing. In the case of the 2022 GSS, about 20 percent of nonrespondents were subsampled for follow up. Adjusting the subsampling rate of a nonresponse follow-up could variably change the cost of in-person data collection.

We note that due to some difficulty in recruiting field staff in some locations, the FTF-first sequence needed to expedite the PTW nonresponse follow-up for some cases. This limited the extent to which the FTF completes could be obtained in this sequence but did facilitate swifter completion in understaffed areas. The final analytic weights do adjust for area nonresponse correcting for these differences.

The combination of web and FTF seem to assist in improving cooperation by obtaining people with different characteristics regardless of ordering in mode sequence. This finding supports previous work on how multimode surveys result in improved demographic representation. While our work touches on this, more research should be done to understand the mechanisms of selection and more fully consider how multimode sequences might change sampling composition and key trends relative to single mode studies.

Conclusions

Multimode studies need to be carefully designed and coordinated. Our findings from the 2022 GSS mode sequence experiment show that costs can be minimized without significant impacts to data quality while utilizing different variations of a web and FTF sequential data collection methodologies. The 2022 GSS experiment and corresponding analysis further support the benefits of surveys that include nonresponse follow-up leveraging multimode survey designs. The 2024 GSS is designed to replicate the 2022 experimental design, enabling a future examination of whether the design produces consistent results at two time points.

References

1. The American Association for Public Opinion Research (AAPOR) (2016), *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (9th ed.). Oakbrook Terrace, IL, USA: AAPOR. <https://aapor.org/wp-content/uploads/2022/11/Standard-Definitions20169theditionfinal.pdf>
2. Cornesse, C., & Bosnjak, M. (2018). Is there an association between survey characteristics and representativeness? A meta-analysis. *Survey Research Methods*, 12(1), 1-13. doi:10.18148/srm/2018.v12i1.7205 <https://doi.org/10.18148/srm/2018.v12i1.7205>
3. Davern, M., Bautista, R., Freese, J., Herd, P., & Morgan, S. L. (2024). *General Social Survey, 1972-2022 (Release 3a)* [Data set and code book]. NORC. <https://gss.norc.org/Documents/codebook/GSS%202022%20Codebook.pdf>
4. DeBell, M., Amsbar, M., Meldener, V., Brock, S., & Maisel, N. (2018). *Methodology report for the ANES 2016 time series study*. Palo Alto, CA, and Ann Arbor, MI: Stanford University and the University of Michigan. https://electionstudies.org/wp-content/uploads/2016/02/anes_timeseries_2016_methodology_report.pdf
5. de Leeuw, E. D. (2018). Mixed-mode: Past, present, and future. *Survey Research Methods* 12(2), 75-89. <https://doi.org/10.18148/srm/2018.v12i2.7402>

6. Dillman, D. A. (2017). The promise and challenge of pushing respondents to the web in mixed-mode surveys. *Survey Methodology*, 43(1), 3-30. <http://www.statcan.gc.ca/pub/12-001-x/2017001/article/14836-eng.htm>
7. Suzer-Gurtekin, Z. T., Valliant, R., Heeringa, S. G., & de Leeuw, E. D. (2018). Mixed-mode surveys: Design, estimation and adjustment methods. In T. P. Johnson, B. Pennell, I. A. L. Stoop, & B. Dorer (Eds.), *Advances in Comparative Survey Methodology* (pp. 409-430). John Wiley & Sons, Inc.
8. Wagner, J., Arrieta, J., Guyer, H., & Ofstedal, M. B. (2014). Does sequence matter in multimode surveys: Results from an experiment. *Field Methods*, 26(2), 141-155. <https://doi.org/10.1177/1525822X13491863>

ACKNOWLEDGEMENTS

We would like to thank NORC's Methodology and Quantitative Social Sciences Department for providing support for this report. We would like to thank Pamela Herd and Stephen Morgan for their review of the brief, Zachary Seeskin for the information regarding costs, and Akari Oya for assisting with last minute analyses. Finally, we would like to thank Colm O'Muircheartaigh, Ned English, and Steven Pedlow who included elements of this research in a presentation for the 2023 European Survey Research Association (ESRA) Conference. This experiment was conducted with support from the National Science Foundation grant award 2049169.

ABOUT NORC

NORC at the University of Chicago conducts research and analysis that decision-makers trust. As a nonpartisan research organization and a pioneer in measuring and understanding the world, we have studied almost every aspect of the human experience and every major news event for more than eight decades. Today, we partner with government, corporate, and nonprofit clients around the world to provide the objectivity and expertise necessary to inform the critical decisions facing society.