Survey Research Methods (2023)
Vol. 17, No. 4, pp. 411-427
https://doi.org/10.18148/srm/2023.v17i4.8172
European Survey Research Association

Transitioning to a Mixed-Mode Study Design in a National Household Panel Study: Effects on Fieldwork Outcomes, Sample Composition and Costs

Katherine A. McGonagle[1] · Narayan Sastry[1]
[1]University of Michigan, Institute for Social Research, Survey Research Center

The U.S. Panel Study of Income Dynamics (PSID) made a planned transition to a web-first mixed-mode data collection design in 2021 (web and computer-assisted telephone interviewing [CATI]), following nearly five decades of collecting data primarily using CATI with professional interviewers. To evaluate potential effects of mode on fieldwork outcomes, two sequential mixed-mode protocols were introduced using an experimental design. One protocol randomized sample families to a "web-first" treatment, which encouraged response through an online interview, followed by an offer of telephone to complete the interview; a second protocol randomized sample families to a "CATI-first" treatment until the last phase of fieldwork when the option to complete a web interview was offered. This paper examines the comparative effects of the two protocols on fieldwork outcomes, including response rates, interviewer contact attempts, fieldwork duration, and cost. Comparisons are also made with fieldwork outcomes and characteristics of non-responding sample members from the prior-wave when a traditional telephone design was used. We found that the web-first design compared to the CATI-first design led to comparably high response rates, and faster interview completion with lower effort and cost. With some notable exceptions, compared to the prior wave, the mixed-mode design reduced effort and had generally similar patterns of non-response among key respondent subgroups. The results provide new empirical evidence on the effects of mixing modes on fieldwork outcomes and costs and contribute to the small body of experimental evidence on the use of mixed-mode designs in household panel studies.

Keywords: sequential mixed-mode survey; household panel study; contact protocols; nonresponse error; response rate; survey costs

1 Introduction

In recent years, there has been a marked transformation in the collection of survey data from a reliance on interviewer-administered modes (telephone, in-person) to the use of mixed-mode designs that include web-based interviewing. This shift has been in response to technological, secular, and cultural changes making the collection of data by telephone and in-person modes increasingly difficult and expensive (Dillman, 2017). Recent experimental evidence drawn from cross-sectional surveys has generally found positive effects of mixed-mode designs on various data collection outcomes, including response rates, nonresponse bias, and survey costs (Hox, De Leeuw, and Klausch, 2017; Olson et al., 2021a; Tourangeau, Brick, Lohr, and Li, 2017). A handful of longitudinal studies have adopted mixed-mode designs, but very few have done so experimentally and the research evidence remains limited regarding the effects on survey outcomes and costs (Jäckle, Lynn, and Burton, 2015; Jäckle, Gaia, and Benzeval, 2017). Key gaps in the literature on mixed-mode designs in longitudinal studies include identifying best protocols for the assignment and sequencing of respondents to various modes (Jäckle et al., 2017) and the comparative costs of different data collection protocols (Couper and McGonagle, 2019; Olson et al., 2021b).

This paper reports on the experimental introduction of a sequential mixed-mode design during the 2021 wave of the Panel Study of Income Dynamics (PSID), a long-standing, nationally representative household panel study of U.S. families. PSID has used interviewer-administered interviewing (primarily telephone) since 1973. A key design feature of our analysis is the random assignment of families to one of two sequential mixed-mode protocols (telephone-

Corresponding author: Katherine A. McGonagle, University of Michigan, Institute for Social Research, Survey Research Center, Michigan, USA (Email: kmcgon@umich.edu)

first and web-first) in order to examine differences in data collection outcomes. Each protocol encouraged the completion of the interview in the mode of assignment using mode-specific contact protocols and eventually offered the other mode as an option for those who preferred it or were unable to complete the interview in the original mode. The main strengths of our analysis are the experimental assignment of initial mode and embedding the experiment within a nationally representative study that oversamples socio-demographic groups of interest thereby allowing us to examine outcomes by factors such as age and socioeconomic status. Our findings add critical information to the small but growing body of evidence from longitudinal studies on the use of mixed-mode designs.

We examine two research questions. First, what are the comparative effects of each of the two sequential mixed-mode protocols on fieldwork outcomes, including response rates, interviewer contact attempts, fieldwork duration, and cost? As part of this question, we further examine the characteristics of completed interviews in each randomized group, including mode changes, device of completion, and the number of sessions in each mode.

Second, we examine how fieldwork outcomes and the sample characteristics of non-responding families in the current wave (2021) compare to those of the prior wave (2019), which relied upon the use of a primary mode (telephone) and an established interviewer-administered design used in many panel studies over many decades. We use a difference-in-differences analysis that compares results for a sequential mixed-mode approach with those from the prior wave when only telephone was administered. To understand how the shift in study design may affect sample composition, we use multinomial logistic regression to compare basic demographic and socioeconomic characteristics of respondents who participated in both the 2019 and 2021 waves with nonrespondents from each wave.

The analysis draws on an experimental design embedded in a nationally representative household panel study of U.S. families to evaluate two different mixed-mode data collection protocols against the traditional telephone design and provides new empirical evidence on the effects of mixing modes on fieldwork outcomes and costs.

2 Background and Literature Review

Cross-sectional surveys have used mixed-mode designs for more than two decades, providing the basis for a large body of empirical findings on the use of multiple different modes (mail, telephone, web) and the effectiveness of different protocols (e.g., Hox et al., 2017; Tourangeau et al., 2017). Longitudinal studies are beginning to incorporate mixed-mode designs as an adaptive response to the increasing challenges of the traditional survey environment (see Dillman, 2017; 2009; Olson et al., 2021a). In particular, making contact with respondents in telephone and face-to-face modes is increasingly difficult and response rates have steadily declined (Beullens, Looseveldt, Vanderplas, and Stoop, 2018; De Leeuw, Hox, and Luiten, 2018; Groves and Couper, 2012). Yet, due to the inherent complexity of longitudinal studies and their need to maintain consistency in the collection of data over time, the adoption of mixed-mode designs has been slow. It has typically followed a long period of careful planning and gradual implementation, and research evidence from these transitions remains limited (De Leeuw, 2005; Voorpostel et al., 2021a).

Only a handful of household panel studies based on nationally representative samples of the full age range have experimentally evaluated their transition to the use of mixed modes. The findings from these studies point toward generally positive outcomes on response rates, survey costs, and sample composition. For example, Bianchi, Biffignandi, and Lynn (2017) found no negative effects on attrition or nonresponse bias over three waves during a gradual transition of the U.K. Understanding Society study from a primarily in-person design to a mixed-mode design (web, in-person). Relative costs for the mixed-mode group fell each wave and were significantly lower by the third wave compared to in-person interviewing. The Swiss Household Panel (SHP) added a web mode during its recruitment of a refreshment sample of families and found that while first wave response rates for web were slightly lower, there was no difference in rates of attrition or across key demographic characteristics of respondents by the second wave compared to telephone interviewing (Voorpostel et al., 2021b). The German Socioeconomic Panel Study found no evidence of nonresponse bias in a quasi-experiment that evaluated the impact of switching panel members from personal interviewing to self-administered web data collection (Voorpostel et al., 2021b).

Several panel studies that collect data from specific age groups or cohorts have reported positive fieldwork outcomes following an experimental assessment of mixed modes. For example, Monitoring the Future (MTF), a nationally representative study of U.S. high school graduates, implemented an experimental design to assess a transition to web data collection compared to mailed paper questionnaires over two waves (2014 and 2016) with a sample of young adults aged 19–20 years. Response rates were higher, costs were lower, and nonresponse bias for a set of social and demographic characteristics was reduced for those randomized to the web mode compared to the mail condition at baseline and again two years later (Patrick et al., 2017; Patrick et al., 2019), which led the study to institutionalize a sequential web-first protocol beginning in 2018 (Patrick et al., 2020). The PSID-Transition into Adult-

Table 1

Random Assignment of Reinterview Cases to Treatment Conditions

	N	%
Web-First (WF)	3179	40
CATI-First (TF)	4759	60
Total	7938	100

hood Supplement (PSID-TAS), a nationally representative panel study of young adults aged 18–28 years in PSID families, found that respondents offered a sequential mixed-mode design (web-first, followed by telephone) compared to a traditional telephone-only design were more likely to complete the interview and with lower fieldwork effort and cost (Sastry and McGonagle, 2022). The U.S. National Longitudinal Study of Adolescent to Adult Health (Add Health; Harris et al., 2019) implemented a series of experiments during its transition from in-person interviewing to web self-administration with a follow-up component (mail, in-person), at Wave V (2016–2018) when the original respondents were aged 34–42 years. The weighted response rate from the mixed-mode design was nearly as high as the traditional in-person design and achieved significant (although unspecified) cost savings (Biemer et al., 2022).

Positive effects on fieldwork outcomes have also been found in at least two panel studies of older adults that have implemented mixed-mode designs experimentally. For example, the Age 55 Survey of the U.K. National Child Development Study (NCDS) was an early adopter of mixed-mode interviewing by introducing a sequential mixed-mode design in 2013. NCDS respondents at age 55 assigned to a web-first protocol with telephone follow-up had higher overall participation rates and no meaningfully different sample characteristics compared to a randomized telephone-only group (Goodman et al., 2022). An experiment conducted by the U.S. Health and Retirement Study (HRS), a nationally representative study of older adults found that respondents who reported using the internet in the prior wave and were randomized to a web-first condition were slightly more likely to participate and had significantly fewer interviewer contact attempts compared to internet users who were assigned to a telephone-only group (Ofstedal et al., 2019). In sum, the handful of existing longitudinal studies that have conducted experimental evaluations of the transition to mixed-mode generally point to positive effects on response rates, nonresponse bias, and survey costs, but the number of such studies is small, the sample composition of the studies is variable, and the evidence for cost savings is particularly limited. Our paper responds to the call for longitudinal studies to evaluate mixed-mode designs experimentally and to share the results, with the goal to enhance our understanding of optimizing the collection of survey data in a panel context (De Leeuw, 2018; Jäckle et al., 2017; Voorpostel et al., 2021a).

3 Methods

3.1 *PSID Study Design*

The Panel Study of Income Dynamics (PSID) is a longitudinal study of a nationally representative sample of U.S. families that started in 1968. Families in the PSID were interviewed annually until 1997 and biennially since 1997 (see McGonagle et al., 2012 for more information). Data collection occurs in odd-numbered years between about March 1 and December 31 over the course of approximately 44 weeks. The study interviews one adult respondent in each family, typically the individual who is most knowledgeable about the family finances. The interview takes about 80 minutes on average and collects a variety of data on economic, health, and social behavior. Families are initially offered a post-paid monetary incentive to participate, equal to approximately $1 per minute for the mean interview duration (i.e., $80 in 2021). The study also offers additional monetary incentives to some respondents to encourage response, including higher incentives at the end of fieldwork. Telephone interviews are conducted by professional interviewers employed by the Survey Research Center (SRC) at the University of Michigan. In recent years, the wave-to-wave reinterview response rate (i.e., the current wave response rate among families who participated in the prior wave) has been at least 90 percent (PSID User Guide 2019). In 2021, the study made a planned transition to a mixed-mode design, offering a random set of families the opportunity to complete their interview on web or by telephone.

3.2 *Mixed-Mode Design*

A mixed-mode experimental study design was implemented for 7938 families that had completed interviews in the prior wave in 2019 (i.e., reinterview families). Reinterview families were randomly assigned to either a sequential "Web-First (WF)" or [computer-assisted] telephone interview "CATI-First (TF)" treatment condition (Table 1). With a goal of achieving approximately 30 percent of all completed reinterviews on web in the initial transition wave, 40 percent of reinterview cases were randomly assigned to WF (N = 3179) and the remaining 60 percent to TF (N = 4759). There were no differences in socio-demographic characteristics of families by randomized treatment con-

dition (see Appendix). Families that were nonresponse in 2019, new to the study in 2021, comprised of a single institutionalized individual, or interviewed by a proxy in 2019 required a telephone interviewer to administer additional, complex questions to update or create their household roster and residential location information; these families were not included in the experiment. Data collection occurred over nine months between March 19, 2021 and December 31, 2021.

3.3 Fieldwork Protocols

As in recent prior waves, PSID in 2021 implemented an intensive fieldwork strategy using postal mailings, telephone calls, emails, and text messages to encourage study participation. An advance letter was sent through postal mail, email, and text message to all families providing general information about the upcoming study. The information sent to families in the WF treatment group described the new opportunity to complete the interview online and included unique login credentials and instructions on how to access the web portal. To encourage web interviews among families in the WF treatment group, no mention was made of a telephone interview option. TF treatment group families were told that an interviewer would be calling them soon to complete their interview. No information was provided about the availability of a web instrument. Families in both experimental groups were offered the post-paid incentive of $80.

Once released to the field, the treatment groups followed slightly different protocols to encourage their participation. The TF contact protocol followed standard SRC contact protocols for telephone fieldwork, well established over multiple prior waves of PSID, comprising at least one weekly email and/or text message reminder, with the number gradually rising to approximately 2–3 weekly by the end of the fieldwork period. After about 24 weeks of fieldwork, approximately 16 percent of TF cases that had not yet completed their telephone interview (n = 753) were invited to complete their interview online.

In order to ensure that new online technical systems were working correctly, the WF cases were released in random replicates and slightly later in the field period than the TF cases. A main goal of the WF protocol was to encourage families to promptly complete the interview online and avert the need for more expensive follow-up by telephone interviewers. After one week of fieldwork, and based on a protocol that demonstrated the effectiveness of resending hard copies of web login credentials (Benzeval et al., 2017), the login instructions, credentials, and authenticated links were re-sent. Over the following seven weeks, WF families were sent at least one weekly reminder via email and/or text message. At approximately Week 5—with the goal of reducing the number of cases requiring more expensive telephone follow-up—a time-delimited offer of an additional incentive of $20 was made to all remaining WF families in exchange for completing an online interview within two weeks. Finally, after seven weeks of nonresponse, the protocol switched to weekly telephone calls by interviewers who attempted to contact respondents (directly or through voice-mail) to offer the opportunity to complete the interview either by telephone or online.

3.4 Variables

Outcome measures. We compare the following fieldwork outcomes for families randomized to each of the sequential mixed-protocols: *response rates*, calculated as the percentage of all eligible families completing an interview (based on definition RR6, AAPOR, 2016); *the number of interviewer contact attempts* that were required to complete an interview or reach the end of fieldwork for nonresponse cases, with separate measures for *telephone calls, emails, and text messages*, and *total contact attempts*, constructed as the sum of telephone calls, emails, and text messages; *fieldwork duration*, defined as the number of weeks from the date of release to the field to the date when a final fieldwork disposition status (either response, nonresponse, or nonsample) was assigned. These outcome variables are created for both the 2019 and 2021 waves in order to compare the outcomes for each of the mixed-mode protocols in 2021 with those obtained during 2019 when a telephone-only design was used.

Fieldwork costs are calculated across all cases, and separately for completed cases and nonresponse cases, as the sum of costs for interview contact attempts, telephone administration of the interview, and respondent incentives. We exclude one-time outlays due to the transition to mixed-mode data collection, such as developing and testing mixed-mode questionnaires, and modifications to software and technical systems for web interviewing.

For completed interviews, we compare the following outcomes between the protocols: *mode of completion* (i.e., whether by web or telephone), *the number of sessions required to complete the interview*, with separate measures for *web sessions, telephone sessions,* and *total sessions* (the latter defined as the sum of web sessions and telephone sessions). For cases completing on web, we report on *completion device* (whether the interview was completed on a smartphone or on a personal computer or tablet).

Demographic and socioeconomic measures. Predictors of nonresponse in 2019 or 2021 included the following dummy variables (1 = yes, 0 = no): sex of respondent; age of respondent, coded as between 18 and 39 years of age,

40–59 years of age, and 60 years of age or greater; married or cohabiting versus single; total family income coded as less than or equal to 25th percentile, between 26th–89th percentile, and greater than 90th percentile; educational attainment, coded as less than 16 years or greater than 16 years; and PSID sample type, coded as cases who were selected from a national probability sample, a low-income oversample, or an immigrant refresher sample.

3.5 Analysis Methods

For the first research question examining the comparative effects of each of the two sequential mixed-mode protocols on fieldwork outcomes, we describe fieldwork outcomes separately by experimental condition and use t-tests to test mean differences and quantile regression to test median differences. We use the Kaplan-Meier estimator (Kaplan and Meier, 1958), a nonparametric statistic that calculates differences in time-to-completion, to provide estimates of time-to-completion of the interview across the entire field period.

The second research question—how fieldwork outcomes and the sample characteristics of non-responding families in the current wave using a sequential mixed-mode design compare to those of the prior wave using a fully interviewer-administered design—is assessed using a difference-in-differences analysis to compare *changes* in fieldwork outcomes between 2019 and 2021 for two groups: (1) those that received the web-first treatment in 2021 and (2) those that received the CATI-first treatment in 2021. Because all respondents in 2019 were interviewed by telephone with no option to complete the interview on web, this analysis affords a within-subject control for fieldwork outcomes in the previous wave, providing an unbiased estimate of the effect on fieldwork outcomes of switching to a web-first mixed- mode treatment compared to switching to a CATI-first mixed-mode treatment. The analysis creates "difference variables" by deriving the difference between each of the 2019 and 2021 outcome variables (i.e., the 2021 average is subtracted from the 2019 average for number of total attempts, telephone attempts, email/text attempts, and fieldwork days for each mode group). The statistical significance of the difference between the difference variables is evaluated (i.e., the difference-in-differences) using a t-test.

Our final analysis provides information on how respondent characteristics may be influenced by the shift from a telephone-only design in 2019 to a sequential mixed-mode design in 2021. We use multinomial logistic regression to examine how the set of socioeconomic and demographic respondent characteristics, identified above, predict membership in three mutually exclusive groups: respondents who complete an interview in both 2019 and 2021 (*the reference category*); non-respondents during the telephone-only wave in 2019; and non-respondents during the 2021 wave (all of whom completed the interview in 2019) when a mixed-mode design was used. The model provides separate intercept parameters and regression parameters for each logit. We test the statistical significance of the differences in the parameters predicting nonresponse during the telephone-only wave in 2019 (compared to the baseline category of response in both waves) against those predicting nonresponse in the mixed-mode wave in 2021 (compared to the baseline category of response in both waves) to provide information about characteristics that differentiate nonresponse in each wave.

4 Results

The first research question is evaluated by examining differences between the treatment conditions in response rates, completion mode and device, and the count of mode changes from the assigned mode (3.1). We then compare fieldwork effort and duration by treatment condition, and for completed interviews we describe the completion mode, completion device, and number of sessions by treatment condition (3.2). We finally compare fieldwork costs by treatment condition (3.3). For the second research question, we provide a comparison in fieldwork effort and duration between 2019 and 2021 across the two treatment groups (3.4) and describe socio-demographic characteristics that differentiate nonresponse during 2019 and 2021 (3.5).

4.1 Response Rates, Completion Mode and Device, and Mode Changing

The overall response rate for the cases in the experiment was 91% (Table 2). There was no difference in response rate by assigned treatment (91% for WF and 92% for TF, p = NS). The majority of respondents in both conditions completed the interview using the mode to which they were assigned: among those assigned to WF (N = 3179), 72% completed the interview on web and 28% completed on telephone. Among the WF cases completing on web (N = 2092), about half completed on a smartphone (51%) and half completed on a PC or tablet (48%). Among the cases randomized to TF (N = 4759), 86% completed their interview by telephone (N = 3728). Of the 15% completing on web, more completed on a smartphone (nearly 60%) than on a PC or tablet (nearly 40%).

There were very few sessions using the alternative mode by those completing interviews in their assigned mode. Among the cases assigned to TF who completed on CATI

Table 2

Response Rates and Completion Mode/Device/Use of Alternative Mode by Treatment Condition

	Total (N = 7938)		Web-First (N = 3179)		CATI-First (N = 4759)	
	N	%	N	%	N	%
Response Rate	7250	91	2891	91	4359	92
Completion Mode, among completed interviews						
1. Web	2723	38	2092	72	631	15
Completion Device						
Smartphone	–	–	1071	51	370	59
PC/Tablet	–	–	997	48	245	39
Missing	–	–	24	1	16	3
% with CATI session	–	–	–	4	–	58
2. CATI	4527	62	799	28	3728	86
% with web session	–	–	–	26	–	1

Table 3

Fieldwork Outcomes by Treatment Condition

	Web-First (N = 3179)	CATI-First (N = 4759)	Difference	Standard error	*p*
Number of total interviewer attempts					
Mean	21.9	29.5	–7.6	0.76	<0.01
Median	7.0	10.0	–3.0	0.52	<0.01
Number of telephone attempts					
Mean	9.4	16.3	–6.9	0.38	<0.01
Median	2.0	7.0	–5.0	0.13	<0.01
Number of email/text attempts					
Mean	12.2	12.9	–0.7	0.40	0.09
Median	5.0	3.0	2.0	0.24	<0.01
Number of fieldwork days					
Mean	74.5	108.7	–34.2	1.86	<0.01
Median	42.0	63.0	–21.0	1.40	<0.01

(N = 3,728), less than 1% (1%) ever opened the web instrument. The fact that most of the TF cases completed their interviews before being offered the option to complete the interview on web (after 6 months of fieldwork) is the main reason for the low use of web by those randomized to TF. Only 4% of WF cases completing the interview on web started the interview on CATI, despite the fact that more than half of these respondents received at least one call from interviewers offering them the opportunity to complete the interview on the telephone.

A greater number of those choosing a different mode from their initial assignment had sessions in both modes compared to those not changing modes. More than half of respondents assigned to TF who eventually completed the interview on web (58% of N = 631) started the interview with a telephone interviewer on CATI (half of these cases had exactly one session, with an average of 1.6 sessions). Only about one-quarter of WF cases who completed the interview on CATI (26% of N = 799) ever tried the web interview (the mean number of web sessions for those with at least one web session was 2.4 and the median was 2.0).

4.2 Fieldwork Outcomes by Treatment Condition

Interviewer contact attempts. Fieldwork effort during 2021 in the form of contact attempts by interviewers was substantially lower for cases randomized to WF than for TF cases (Table 3). The mean number of total interviewer attempts was 7.6 fewer for WF cases (21.9 attempts) compared to TF cases (29.5 attempts); the difference between the conditions at the median was smaller but still significant at 3.0 fewer

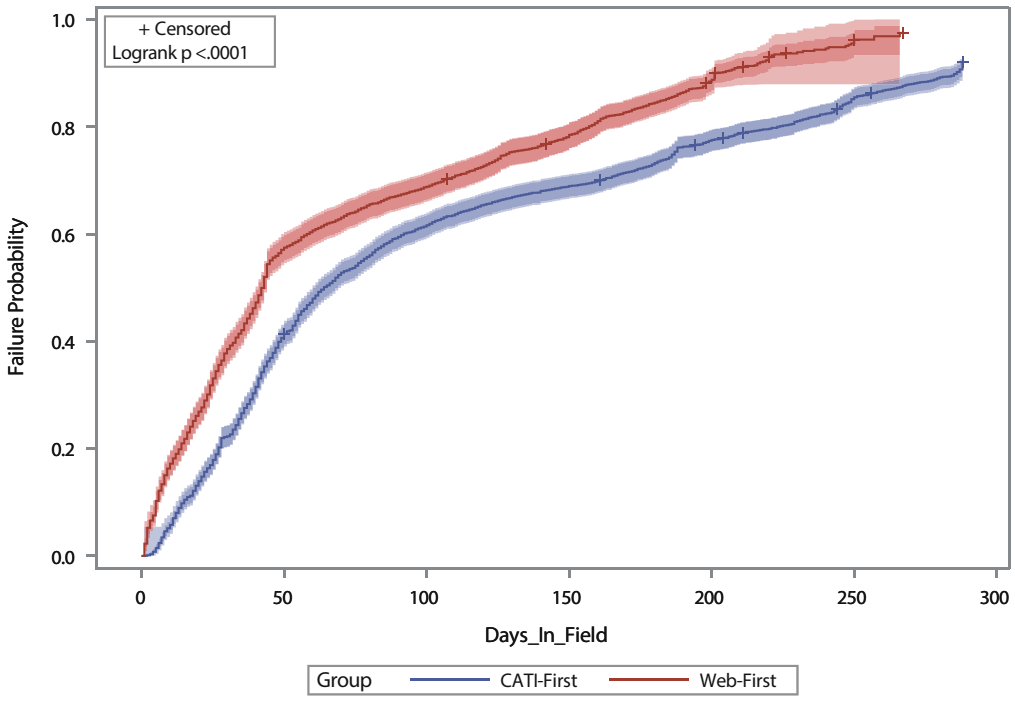

Fig. 1

Kaplan-Meier Estimate of the Cumulative Distribution Function of Duration to PSID-2021 Completion by CATI-First and Web-First

calls for WF cases (7.0 attempts) than TF cases (10.0 attempts). The lower number of total interviewer attempts for the WF cases was almost entirely due to fewer telephone calls compared to TF cases (nearly 7 fewer telephone calls on average, and 5 fewer calls at the median). The median number of email and text messages was significantly higher for WF cases (5.0 attempts) than TF cases (3.0 attempts). Notably, 45 percent of WF interviews were completed during the first seven weeks of fieldwork before interviewer reminder calling began and hence had zero telephone attempts (not shown in table).

Fieldwork duration. Time to fieldwork completion is illustrated by the Kaplan-Meier estimate of the cumulative distribution function for the duration to a completed case by treatment condition (Fig. 1). Fieldwork duration is adjusted for the exact date on which each case was released to the field and incomplete cases are censored on the final day of fieldwork. The figure highlights the considerably faster completion time for WF cases (solid line) compared to than those assigned to the TF treatment condition, dashed line; Wilcoxon test statistic for the equality of the cumulative distribution functions is $\chi^2(1) = 292.7$, $p < 0.01$). Cases randomized to WF had significantly fewer average and median days in the field than TF cases (Table 3). Average time to interview completion was about 74 days for WF cases and about 108 days for TF cases, a difference of more than a month. The median time to completion for WF cases was about 3 weeks earlier than TF cases (42 days vs 63 days, respectively).

Number of sessions by completion mode and device. As shown in Table 4, respondents who completed the interview in the mode of assignment had slightly fewer average sessions than those who changed mode. The average number of total sessions for WF cases was 2.7 for the majority who completed on web and 2.2 for those switching to telephone, a small but significant difference of 0.5 fewer sessions for telephone; there was no difference at the median. Web sessions timed out after 10 minutes of inactivity, which likely contributed to the slightly higher number of sessions. The fewest average number of sessions was for the majority of TF cases that completed on CATI (1.8) and the highest average number of sessions was for TF cases that changed to web (3.4), which had sessions in both modes. As would be expected, interviews completed on web had significantly more sessions using web than did interviews completed on telephone, and telephone interviews had significantly more sessions using telephone than did interviews completing on web. Among cases completing on web, those using a smartphone had slightly more average total sessions and web sessions than those using a PC/tablet; this difference was small and not statistically significant at the median.

Table 4

Fieldwork Outcomes and Completion Mode/Device/Sessions by Treament Condition for Completed Interviews

Completion Mode	Web-First (N = 2891)									
	Web (N = 2092)						CATI (N = 798)			
	Smartphone (n = 1071)	PC/Tablet (n = 997)	Difference	Standard error	P[a]	All Web Cases	All CATI Cases	Difference	Standard error	P[b]
Number of total sessions										
Mean	2.9	2.5	0.4	0.10	<0.01	2.7	2.2	0.5	0.09	<0.01
Median	2.0	2.0	0.0	–	1.0	2.0	2.0	0.0	–	1.0
Number of web sessions										
Mean	2.8	2.5	0.4	0.10	<0.01	2.7	0.6	2.1	0.07	<0.01
Median	2.0	2.0	0.0	–	1.0	2.0	0.0	2.0	0.11[c]	<0.01
Number of CATI sessions										
Mean	0.1	0.0	0.1	0.01	0.54	0.1	1.6	–1.6	0.04	<0.01
Median	0.0	0.0	0.0	–	1.0	0.0	1.0	–1.0	0.04[c]	<0.01
Completion Mode	CATI-First (N = 4355)									
	Web (N = 631)						CATI (N = 3724)			
	Smartphone (n = 370)	PC/Tablet (n = 245)	Difference	Standard error	p[a]	All Web Cases	All CATI Cases	Difference	Standard error	p[b]
Number of total sessions										
Mean	3.4	3.3	0.1	0.21	0.55	3.4	1.8	1.6	0.10	<0.01
Median	3.0	3.0	0.0	–	1.0	3.0	1.0	2.0	0.18	<0.01
Number of web sessions										
Mean	2.5	2.2	0.3	0.17	0.08	2.4	0.0	2.4	0.08	<0.01
Median	2.0	2.0	0.0	–	1.0	2.0	0.0	2.0	0.03[c]	<0.01
Number of CATI sessions										
Mean	0.9	1.0	0.2	0.12	0.17	1.0	1.7	0.8	0.05	<0.01
Median	1.0	1.0	0.0	–	1.0	1.0	1.0	0.0	–	1.0

N = 1 Web-First and n = 4 CATI-First cases are missing data on sessions

Device type is missing for n = 24 Web-First cases and n = 16 CATI-First cases

[a] Test of difference between Smartphone cases and PC/Tablet cases for cases completing on Web

[b] Test of difference between Overall Web cases and Overall CATI cases

[c] Standard error is estimated from nearest estimable quantile which in all cases have values that match the observed medians

Table 5

Comparison of variable costs of Web-First and CATI-First fieldwork

	Web-First	CATI-First
Cases		
Total	3179	4759
Completed	2,891	4359
Non-response	288	400
Response rate	91%	92%
Interviewer contact attempts		
Mean contact attempts (completed cases)	17.7	24.8
Mean contact attempts (non-response cases)	64.4	79.7
Mean total cost (all cases) at $7.20/attempt[a]	$158	$212
Interviewer survey administration by telephone		
Completed interviews (N) administered by telephone	28%	86%
Mean telephone interview length (minutes)	92.5	90.0
Mean total cost at $26.24/hour[b]	$10	$31
Respondent incentives		
Mean cost (across all cases)	$104	$101
Grand total mean variable cost-per-case		
All cases[c]	$272	$344
Completed cases[d]	$253	$322
Non-response cases[e]	$464	$574
Total variable survey costs[f]	$2,159,136	$2,722,734
Web-First vs. CATI-First difference ($)	$563,598	
Web-First vs. CATI-First difference (%)	26%	

[a] Mean total cost for contact attempts is equal to mean contact attempts by response status (non-response vs. completed) weighted by the fraction of cases in each group multiplied by the cost per contact. For WF, mean total cost is ((17.7 · 91%) + (64.4 · (100% – 91%))) × $7.20 = $158; for CF, mean total cost is ((24.8 · 92%) + (79.7 · (100% – 92%))) × $7.20 = $212.

[b] Mean total cost for interview administration is equal to the average telephone interview length multiplied by the fraction of completed cases interviewed by telephone multiplied by the hourly interview rate. For WF, mean total cost is (92.5/60 · 28% · 91% · $26.24) = $10; for CF, mean total cost is (90/60 · 85% · 92% × $26.24) = $31.

[c] Grand total cost is the sum of costs for contact attempts, interview administration, and respondent incentives. For WF, total cost is $158 + $10 + $104 = $272; for CF, total cost is $212 + $31 + $101 = $343.

[d] Grand total cost for completed cases is the sum of costs for contact attempts for completed cases plus the sum of interview administration and respondent incentives divided by the response rate. For WF, cost per case is 17.7 · $7.20 + ($10 + $104) / 91% = $253; for CF, cost per case is 24.8 · $7.20 + ($31 + $101) / 92% = $322.

[e] Grand total cost for non-completed cases is based on the cost of interview contact attempts. For WF, cost per case is 64.4 · $7.20 = $464; for CF, cost per case is 79.7 · $7.20 = $574.

[f] Total survey costs are equal to the total number of cases (3179 + 4759 = 7938) multiplied by the grand total mean cost-per-case for each group.

4.3 Fieldwork Cost Comparison between Treatment Conditions

Table 5 presents results of a comparison between the WF and TF groups on fieldwork costs in three categories: interviewer contact attempts, interview administration, and respondent incentive payments.

The mean total cost for interviewer contact attempts is calculated as the weighted average of contact attempts, across completed and nonresponse cases, multiplied by the mean cost per contact attempt of US$7.20. The cost per contact attempt is based on actual interviewer hourly costs (wages plus fringe benefits) and the estimated productivity of four contact attempts per hour, which is independent of the contact type (telephone, email, or text message) because each requires approximately the same amount of interviewer time. Interviewers spend time to dial and leave a voice-mail message or compose, review, and send a text message or email message, and time to review summaries of the case history (e.g., descriptive respondent information and prior contact information, including the number, dates, and outcomes of prior contact attempts), and to update case history notes for each new attempt.

The second cost component is interviewer time spent conducting the interview, which for each group is based on the subset of cases completing a telephone interview, the average interview length, and hourly interviewer costs of $26.24. Mean total cost per case for interviewer telephone

Table 6

Differences in Fieldwork Outcomes by Wave and by 2021 Mode Assignment

	PSID-2021 Mode Asignment								
	Web-First (N = 3018)			CATI-First (N = 4509)					
Fieldwork outcome	2019	2021	Difference	2019	2021	Difference	Difference-in-differences	Standard error	*p*
Number of total interviewer attempts									
Mean	17.1	21.3	4.2	16.5	28.4	11.9	7.7	0.78	<0.01
Median	9.0	7.0	–2.0	9.0	9.0	1.0	–3.0	0.30	<0.01
Number of telephone attempts									
Mean	11.4	9.0	–2.3	11.1	15.8	4.7	7.0	0.41	<0.01
Median	7.0	2.0	–3.0	6.0	6.0	1.0	–4.0	0.12	<0.01
Number of email/text attempts									
Mean	5.7	11.8	6.1	5.4	12.5	7.1	1.0	0.39	0.02
Median	2.0	5.0	1.0	2.0	3.0	1.0	0.0	–	1.0
Number of fieldwork days									
Mean	91.8	71.4	–20.4	90.7	105.0	14.3	34.7	2.21	<0.01
Median	65.0	41.0	–19.0	65.0	61.0	9.0	–28.0	1.75	<0.01

Sample consists of 2021 treatment cases that completed an interview in 2019. All cases in 2019 were interviewed on CATI. The median difference as presented in the columns "Difference" and "Difference-in-differences" is not always equivalent to the mathematical difference of the two medians

administration is much lower for WF cases at $10 compared to $31 due to the greater number of telephone interviews for the TF group. The final cost component is respondent incentives for completed interviews for which the mean total cost per case is similar across the groups.

The grand total mean variable cost per case is $272 for WF and $344 for TF. The $72 higher mean cost for the TF cases is due to differences in mean total costs for interview contact attempts (+$54), interviewer survey administration by telephone (+$21), and respondent incentives (–$3). The mean total costs per case are shown separately for completed and nonresponse cases by group. The costs for each type of case are substantially higher for TF cases than for WF cases for both nonresponse cases (+24%) and completed cases (+27%). Across both groups, the average total cost per case is nearly twice as high for nonresponse cases as for completed cases.

The final panel shows the hypothetical total variable fieldwork costs of conducting the survey exclusively using the designs of WF ($2.16 million) or TF ($2.72 million). The hypothetical savings from using WF compared to TF are $0.56 million or about 26 percent.

4.4 Comparison in Fieldwork Effort and Duration between 2021 and 2019

Comparing interviewer attempts during the mixed-mode wave of 2021 with the prior wave of 2019 when all cases were offered the single mode of telephone (with no option for self-administered web) to complete their interview provides a control from which to evaluate the overall impact of changing to a mixed-mode design. In 2019, the number of total telephone attempts was substantially lower than in 2021 at about 11.0 on average (results not presented in table). As shown in Table 6, the cases randomized to WF and TF in 2021 had a similar number of 2019 overall contact attempts (17.1 and 16.5, respectively) and telephone attempts (11.4 and 11.1, respectively). Of particular note is the difference between the 2021 treatment conditions in the average number of telephone attempts between the waves. From 2019 **to** 2021, WF cases had 2.3 fewer calls on average (3.0 fewer at the median) while TF cases had nearly 5.0 more calls from 2019 to 2021 (1.0 more at the median), a difference-in-difference of 7.0 calls between the treatment conditions ($p < 0.01$). The higher number of telephone attempts in 2021 was entirely due to effort devoted to cases in the TF treatment condition, with the added option to complete the interview on web mitigating this increase in fieldwork effort.

The cases randomized to WF and TF also had similar average fieldwork durations during 2019 (91.8 days and 90.7 days, respectively). Between 2019 and 2021, WF cases had about 20 fewer field days on average (19 fewer at the median) while TF cases had about 14 more days on average (9 days fewer at the median), a statistically significant difference-in-difference of nearly 35 days.

Table 7

Results of Multinomial Logistic Regression: Characteristics Predicting Non-Response in 2019 (N = 649, telephone-only) vs. 2021 (N = 688, mixed-mode) Compared to Response in Both Waves (N = 6920)

	Non-response in 2019 (vs. Response in 2019 and 2021)			Non-response in 2021 (vs. Response in 2019 and 2021)			Test of parameter difference Non-Response in 2019 vs. NonResponse in 2021		
Variables	β	Standard Error	*p*	β	Standard Error	*p*	β of Difference	Standard Error of Difference	*P*[a]
Male reference person (vs. female reference person)	0.71	0.12	<0.01	0.08	0.12	0.50	0.63	0.16	<0.01
Age 18–39 years (vs. 40–59 years)	0.57	0.09	<0.01	0.31	0.09	<0.01	0.26	0.12	0.03
Age 60 years and older (vs. 40–59 years)	–0.28	0.12	0.02	–0.56	0.12	<0.01	0.28	0.16	0.09
< 16 years of education (vs 16 years or more)	0.20	0.10	0.04	0.60	0.11	<0.01	–0.40	0.14	<0.01
Total family income ≤ 25th percentile (vs. 25th–50th percentile)	–0.07	0.11	0.50	0.39	0.10	<0.01	–0.46	0.14	<0.01
Total family income ≥ 90th percentile (vs. 25th–50th percentile)	0.40	0.14	<0.01	0.39	0.14	<0.01	0.01	0.19	0.97
Single mantal status (vs mamed)	0.69	0.11	<0.01	0.01	0.12	0.96	0.68	0.15	<0.01
Immigrant refresher sample member (vs. original oversample or national sample member)	0.71	0.11	<0.01	0.50	0.11	<0.01	0.21	0.14	0.14
Intercept	–3.66	0.16	<0.01	–3.03	0.17	<0.01	–0.63	0.22	<0.01
Likelihood Ratio = 298.8 (16). p ≤ 0.01									
N		649			688				

Total N = 8257. Reference category for the equation is Response in both 2019 and 2021. Parameters are maxium likelihood estimates of generalized logits

[a]p-value for Wald's X2-square test (df = 1) of difference between parameter for Non-response in 2019 vs. Non-response in 2021

4.5 Predictors of Response Status 2019 and 2021

The final research question examines the issue of whether sample characteristics of non-respondents changed with the introduction of a sequential mixed-mode design in 2021. The analysis uses multinomial logistic regression to identify socioeconomic and demographic characteristics associated with nonresponse in 2019, when all interviews were administered by telephone, and in 2021, during the shift to mixed-mode design, by comparing each of these two groups with a third group—sample members who completed interviews in both 2019 and 2021.

Each column of Table 7 presents individual parameter estimates from the model that predicts nonresponse in 2019 and nonresponse in 2021 compared to the base outcome of responding in both waves. All of the variables described in "Analysis Methods" above were retained in the regression models. We find that the overall fit of the model based on the likelihood ratio test is significant (Wald Chi-square = 298.8 (16), $p < 0.01$); that is, there is a statistically significant difference between the model with the explanatory characteristics included and a baseline model based only on the intercept ("constant"). We find that with some exceptions, the patterns of predictors of nonresponse in each wave are generally similar, although the magnitude and level of significance varies. Significant predictors of higher nonresponse in 2019 (compared to responding in both waves) include being male, younger than age 40, a recent immigrant, single marital status, less education, and high income, while being older than age 60 relative to being middle-aged is associated with a lower likelihood of nonresponse. As in 2019, higher nonresponse in 2021 is significantly associated with being young, low education, high income (relative to income between the 25th and 90th percentiles) and immigrant status, and being older reduces the likelihood of nonresponse relative to being middle-aged; however, there

is no significant effect of higher nonresponse in 2021 of being male or single marital status. We also see that, unlike in 2019, higher nonresponse in 2021 is associated with less income. The estimated parameters for being male, young, single and having less education and family income differ significantly between the waves. We find that compared to the telephone-only wave of 2019, the effects on nonresponse are mixed. Four of the six characteristics are related to lower nonresponse in 2021. The parameters for being male, young, older, and single on nonresponse are significantly lower in 2021, while those for lower education and family income are significantly higher.

A variety of alternative specifications for age and income were evaluated, including percentile cut-points for income (50^{th} and 75^{th}) and cut-points by decades for age as well as continuous and non-linear terms to capture variation in nonresponse by wave. We also assessed the presence of second-order interactions between the covariates (although our statistical power was limited to do so). We found no evidence of variation in the effects of these specifications or interactions between the covariates on nonresponse by wave.

5 Discussion

This paper reports on fieldwork outcomes in a long-running, nationally representative panel study of U.S. families during a transition from a long-standing study design based on traditional interviewer-administered telephone data collection to the use of sequential mixed-mode protocols with web and CATI. Our analysis was based on the random assignment of families to either a web-first or CATI-first treatment. The experimental design supported a rigorous evaluation in the transition wave of two sequential mixed-mode protocols and an unbiased comparison to fieldwork outcomes obtained during the prior wave that used a telephone-only design. We review several key findings and limitations, discuss their implications, and identify next steps for future research.

Our first research question examined the comparative effects of the web-first sequential protocol against the CATI-first design on fieldwork outcomes and costs. We found that the wave-to-wave reinterview response rates obtained from the two sequential designs were similar, and compared favorably to the historically high prior wave response rates based on the interviewer-administered telephone design. The majority of respondents completed the interview using the mode to which they were randomized, and those who did so had fewer sessions to complete the interview than those who changed mode. For the WF group, the high uptake of web in the inaugural wave indicated a strong preference for completing the interview online. The relatively small share of respondents who were invited to web and completed the interview on telephone also appeared to have a strong preference for telephone, as very few of them ever tried to access the web portal. The high overall response rate by the TF group is consistent with the strong response rates historically obtained in the PSID using a telephone-based design. Nonetheless, a substantial number of the TF families received months of fieldwork by interviewers and only completed the interview at the end of fieldwork when offered web. These results suggest that successfully identifying respondents with strong preferences for CATI would allow us to eliminate the web-first phase entirely, and increase efficiency. However, assigning respondents to telephone mode based on prior wave mode completion would ignore possible shifts over time towards preference for web and may have substantial error. In the panel context, tracking and analyzing prior wave telephone completion among web-first cases in subsequent waves can help us understand these types of shifts.

Our results also point to the use of multiple modes to accommodate respondents' various preferences for completing the interview as a key tool for maximizing response rates. In a mixed-mode design, a protocol that pushes the majority to the less expensive mode (i.e., web), followed by a fairly quick subsequent offer of the alternative mode (telephone) for those still active yielded high response rates and cost savings. These findings are consistent with those of other panel studies that have found comparably high response rates for sample members randomized to a mixed-mode design with those receiving standard single-mode protocols (Bianchi et al., 2017; Biemer et al., 2022).

Consistent with recent research showing the high and increasing use of mobile devices to respond to web surveys (Antoun et al., 2017), we found that more than half of respondents who completed the interview on web did so on a smartphone, despite the substantial interview length. The high use of smartphones to complete the interview combined with the high (and growing) ownership rate across all age groups offers a powerful tool for engaging respondents in a mixed-mode environment, including presenting opportunities for the design of new contact protocols using reminders sent via text messages with embedded links. Optimizing the use of smartphones for online completion fits within a general framework of gaining participation by enhancing respondent autonomy and making survey completion convenient (Groves and Couper, 2012). While several studies have suggested that measurement error is not a primary concern for web surveys completed on mobile devices (see Antoun et al., 2017; Couper et al., 2017), web design for smartphone use remains a high priority given the potential of this device for data collection (De Leeuw, 2018).

We also found that the WF sequential protocol saved fieldwork resources compared to the TF design. Over the course of the fieldwork period, respondents assigned to WF had significantly fewer interviewer contact attempts than those assigned to TF. A primary reason for this difference was the rapid completion of WF cases, 45 percent of whom completed the interview in the first 7 weeks of fieldwork—precluding the need for field interviewers to make any calls to these cases. A comparison with the fieldwork effort of the same cases during the 2019 CATI wave showed that while overall contact attempts were higher in 2021, web cases actually received fewer calls on average, while TF cases received more. One reason for the higher overall effort needed in 2021 may be the continuing upward secular trend in fieldwork effort over time due to difficulties contacting respondents using interviewer-administered protocols (e.g., Beullens et al., 2018; De Leeuw et al., 2018; National Research Council, 2013; Williams and Brick, 2017), a situation exacerbated by challenges conducting fieldwork during the global pandemic (e.g., Burton, Lynn, and Benzeval, 2020; Sastry, McGonagle, and Fomby, 2020). The introduction of a web option as part of a mixed-mode survey design helped stem the increased fieldwork effort and maintain high response rates (at lower cost).

Because of the rapid take-up, fewer telephone attempts, and less time to administer the interview, survey costs examined in the current study for the web-CATI sequential design compared to the CATI-web design were on average about 26 percent lower, consistent with the handful of other studies that have provided cost comparisons (Bianchi et al., 2017; Sastry and McGonagle, 2022). One caveat is that our cost assessment was based on variable costs, including contact attempts, interview administration time, and respondent incentives. While these costs comprise the majority of fieldwork costs, they do not include other costs of a mixed-mode survey that are challenging to estimate, including one-time investments of developing, programming and testing a web instrument, ongoing maintenance of two parallel instruments (web and CATI), and the development and management of mixed-mode technical systems and fieldwork protocols during data collection, Over subsequent waves, as others have noted (Goodman et al., 2022), it seems likely that such costs will stabilize and potentially decline as baseline instruments and field protocols are continued or modified for the new wave.

The rapid take-up by families using web to complete their interview raises the issue of identifying the optimal length of fieldwork in a mixed-mode environment in order to maximize cost savings. In large household studies, traditional interviewer-administered survey designs have generally required lengthy fieldwork periods due to constraints in the number calls or visits that can be made by a limited number of fieldwork interviewers. Recent industry-wide labor shortages of field interviewers have worsened this situation (University of Michigan, 2022). Moreover, especially in panel studies, some respondents may be accustomed to a lengthy fieldwork period and delay their participation until the end, after absorbing costly interviewer resources. A direction for future research in panel studies is to evaluate the use of different data collection durations on fieldwork outcomes. For example, a web-first protocol combined with respondent messaging that sets an expectation for a shortened field period may achieve fieldwork goals while reducing overall costs and allow more intensive and expensive interviewer-administered fieldwork to be concentrated on high priority cases.

Another key advantage of a mixed-mode framework is enhancing sample heterogeneity. Consistent with the experience of other panel studies that have made recent shifts to mixed-mode designs (e.g., Patrick et al., 2019; Sastry and McGonagle, 2022), we found that expanding the choice of modes in 2021 was associated with a reduction in nonresponse among males, younger respondents, and those of single marital status. We also found that respondents in the oldest age category (i.e., age 60+ years) continued to have lower rates of nonresponse after the shift to mixed-mode in 2021. These respondents had significantly lower nonresponse in 2019 compared to middle-aged respondents, and even lower nonresponse in 2021 compared to 2019. An online interview was completed by more than half of this age group who were offered web at the start of fieldwork. The high online participation among older individuals is consistent with U.S. national estimates of more than 75 percent reporting using the internet in 2021 among individuals aged 65+ years, an increase of about 10 percentage points in just five years, compared to nearly all adults 18 to 29 years of age (99%) and those 30 to 49 years of age (98%; Pew Research Center, 2022). The opportunity of self-administration of the interview online may be especially appealing for older adults with physical health limitations who would otherwise find lengthy telephone sessions with an interviewer an arduous experience.

One result that warrants monitoring over time is the significant increase in nonresponse in 2021 compared to 2019 by those with less education and lower family income. Among non-respondents in 2019, 70.7 had less than 16 years of education and 21.6 percent had family income below the 25th percentile (when looking at the entire sample distribution by family income). By comparison, 79.5 percent of those not responding in 2021 had less education and 29.2 percent had lower family income, a modest but significant difference. Among panel study respondents accustomed to completing the interview on telephone, it is possible that lower socioeconomic status is associated with a reluctance to complete the interview in a new mode for

reasons of access or because the online mode raises concerns about participation.

There are several limitations of the current study that suggest directions for future research. One limitation is that attributing sample composition changes to the introduction of mixed-mode ignores the effect of the COVID-19 global pandemic, a major societal shock in the period between 2019 and 2021 with known differential health and economic effects on various population groups including those with lower socioeconomic status (e.g., Chang et al., 2021; Fletcher et al., 2021). An important direction for future research on the consequences of transitioning to mixed-mode study designs is monitoring changes over time in panel sample composition as attrition accumulates (Voorpostel et al., 2021b) and working toward an understanding of the mechanisms underlying these changes. In addition, future research should examine how contemporaneous mode preference influences nonresponse bias. In a mixed-mode environment where respondents have a choice of mode completion, are certain sociodemographic characteristics related to a preference for self-administered versus interviewer-administered modes? Having this information available from prior waves can facilitate the design of efficient fieldwork protocols (although mode preference may also change over time).

An additional limitation of the current study is the exclusive focus on families that provided an interview in the prior wave. A goal for future research is the design and testing of mixed-mode protocols targeting families that were unable or preferred not to participate in the prior wave. The availability of the web mode may in fact enhance participation for some of these families. A challenge is designing user-friendly self-administered instruments that facilitate accurate completion of the extensive family composition and residential information updates that are needed for families who have gaps in their participation.

In summary, our findings on the implementation of a mixed-mode design in a long-running nationally representative household panel study of U.S. families add significantly to the limited experimental research on mixing modes in longitudinal studies. The positive impact of a mixed-mode design on fieldwork outcomes and costs in the current study should be broadly encouraging for other panel studies who are considering the adoption of mixed modes, although more research is needed. We identified a set of topics for future research, including optimizing fieldwork protocols for smartphone design, the possibility of reducing fieldwork duration to account for the rapid completion of interviews by web, and studying sample composition changes over time due to changes in nonresponse patterns. A critical next step in the evaluation of this widespread use of new modes is understanding potential mode differences in survey item responses and their impact on economic and sociological measurement.

Acknowledgements This work was supported by the National Science Foundation [SES 1623864], the Eunice Kennedy Shriver National Institute of Child Health and Human Development [R01 HD069609], and the National Institute on Aging [R01 AG040213]. The authors are grateful for the contributions of Rachel Orlowski and Shonda Kruger-Ndiaye for implementing the experimental intervention and leading fieldwork activities. The deidentified data and code used for this article are available through the ICPSR PSID Data Repository (McGonagle and Sastry, 2023).

References

Antoun, C., Couper, M.P., & Conrad, F.G. (2017). Effects of mobile versus pc web on survey response quality: A crossover experiment in a probability web panel. *Public Opinion Quarterly*, *81*(S1), 280–306. https://doi.org/10.1093/poq/nfw088.

Benzeval, M., Bianchi, A., Brewer, M., et al. (2017). *Understanding society innovation panel wave 9: results from methodological experiments*. Understanding society working paper 2017-07. Colchester: Institute for Social and Economic Research.

Beullens, K., Loosveldt, G., Vandenplas, C., & Stoop, I. (2018). Response rates in the European Social Survey: Increasing, decreasing, or a matter of fieldwork efforts? *Survey Methods: Insights from the Field*. https://surveyinsights.org/?p=9673.

Bianchi, A., Biffignandi, S., & Lynn, P. (2017). Web-face-to-face mixed-mode design in a longitudinal survey: Effects on participation rates, sample composition, and costs. *Journal of Official Statistics*, *33*(2), 385–408.

Biemer, P.P., Harris, K.M., Burke, B.J., Liao, D., & Halpern, C.T. (2022). Transitioning a panel survey from in-person to predominantly web data collection: results and lessons learned. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *185*(3), 798–821.

Burton, J., Lynn, P., & Benzeval, M. (2020). How understanding society: the UK household longitudinal study adapted to the COVID-19 pandemic. *Survey Research Methods*, *14*(2), 235–239. https://doi.org/10.18148/srm/2020.v14i2.7746.

Chang, H.Y., Tang, W., Hatef, E., et al. (2021). Differential impact of mitigation policies and socioeconomic status on COVID-19 prevalence and social distancing in the United States. *BMC Public Health*, *21*, 1140–1150. https://doi.org/10.1186/s12889-021-11149-1.

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

Couper, M. P., Antoun, C., & Mavletova, A. (2017). Mobile web surveys. In P. P. Biemer, E. Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, N. C. Tucker & B. T. West (Eds.), *Total survey error in practice*. https://doi.org/10.1002/9781119041702.ch7.

De Leeuw, E. (2018). Mixed-mode: past, present, and future. *Survey Research Methods*, *12*(2), 75–89.

De Leeuw, E. (2005). To Mix or Not to Mix Data Collection Modes in Surveys. *Journal of Official Statistics*, *21*(2), 233–255.

Dillman, D. A. (2017) The promise and challenge of pushing respondents to the web in mixed-mode surveys. *Survey Methodology*, 12–001-X(43), 1.

Dillman, D. A., Smyth, J. D. & Christian, L. M. (2009) *Internet, mail, and mixed-mode surveys: The tailored design method*, 3rd edition. Hoboken, NJ: Wiley.

Fletcher, M., Espey, J., Grossman, M. K., et al. (2021). Social vulnerability and county stay-at-home behavior during COVID-19 stay-at-home orders, United States, April 7 – April 20, 2020. *Annals of Epidemiology*, *64*, 76–82.

Goodman, A., Brown, M., Silverwood, R. J., Sakshaug, J. W., Calderwood, L., Williams, J., & Ploubidis, G. G. (2022). The impact of using the Web in a mixed-mode follow-up of a longitudinal birth cohort study: Evidence from the National Child Development Study. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 1–29.

Groves, R. M., & Couper, M. P. (2012). *Nonresponse in household surveys*. New York: John Wiley.

Harris, K. M., Halpern, C. T., Whitsel, E. A., et al. (2019). Cohort profile: the national longitudinal study of adolescent to adult health (add health). *International Journal of Epidemiology*, *48*, 1415–1415. https://doi.org/10.1093/ije/dyz115.

Hox, J., De Leeuw, E., & Klausch, T. (2017). Mixed mode research: issues in design and analysis. In P. Biemer, E. De Leeuw & S. Eckman, et al. (Eds.), *Total survey error in practice* (pp. 511–530). New York: Wiley.

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.

Jäckle, A., Gaia, A., & Benzeval, M. (2017). *CLOSER Resource Report: "Mixing modes and measurement methods in longitudinal studies."*. London: UCL Institute of Education.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*, 457–481.

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

McGonagle, K., & Sastry, N. *Fieldwork analysis data*. Ann Arbor: Inter-university Consortium for Political and Social Research. https://doi.org/10.3886/E184610V3.

McGonagle, K. A., Schoeni, R. F., Sastry, N., & Freedman, V. A. (2012). The panel study of income dynamics: overview, recent innovations, and potential for life course research. *Longitudinal and Life Course Studies*, *3*(2), 268–284.

National Research Council (2013). Nonresponse in social science surveys: a research agenda. In R. Tourangeau & T. J. Plewes (Eds.), *Panel on a research agenda for the future of social science data collection. Committee on national statistics, division of behavioral and social sciences and education, national research council*. Washington, DC: The National Academies Press.

Ofstedal, M. B., Couper, M. P., Hupp, A. L., Gatward, R., & Weir, D. (2019). *Introducing web to a panel survey: the health and retirement study*. Zagreb, July 15–19. Paper presented at the meeting of the European Survey Research Association.

Olson, K., Smyth, J. D., Horwitz, R., Keeter, S., Lesser, V., Marken, S., Mathiowetz, N. A., McCarthy, J. S., O'Brien, E., Opsomer, J. D., & Steiger, D. (2021a). Transitions from telephone surveys to self-administered and mixed-mode surveys: AAPOR task force report. *Journal of Survey Statistics and Methodology*, *9*(3), 381–411.

Olson, K., Wagner, J., & Anderson, R. (2021b). Survey costs: where are we and what is the way forward? *Journal of Survey Statistics and Methodology*, *9*(5), 921–942. https://doi.org/10.1093/jssam/smaa014.

Patrick, M. E., Couper, M. P., Laetz, V. B., Schulenberg, J. E., O'Malley, P. M., Johnston, L. D., & Miech, R. A. (2017). A sequential mixed mode experiment in the U.S. National Monitoring the Future Study. *Journal of Survey Statistics and Methodology*, *6*(1), 72–97.

Patrick, M. E., Couper, M. P., Bohyun, J., Laetz, V., Schulenberg, J. E., Johnston, L. D., Bachman, J., & O'Malley, P. M. (2019). Two-year follow-up of a sequential mixed-mode experiment in the U.S. national Monitoring the Future Study. *Survey Practice*, *12*(1).

Patrick, M.E., Couper, M.P., Parks, M.J., Laetz, V., & Schulenberg, J.E. (2020). Comparison of a web-push survey research protocol with a mailed paper and pencil protocol in the Monitoring the Future panel survey. *Addiction*, *116*(1), 191–199.

Pew Research Center (2022). Share of those 65 and older who are tech users has grown in the past decade." Washington, D.C. https://pewrsr.ch/3HZd2ao. Accessed 13 Jan 2022.

Sastry, N., & McGonagle, K.A. (2022). Switching from telephone to web-first data collection: results from the transition into adulthood supplement to the panel study of income dynamics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 1–22.

Sastry, N., McGonagle, K.A., & Fomby, P. (2020). Effects of the COVID-19 crisis on survey fieldwork: experience and lessons from two major supplements to the U.S. Panel Study of Income Dynamics. *Survey Research Methods*, *14*(2), 241–245.

The American Association for Public Opinion Research (2016). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*, 9th edition. AAPOR.

Tourangeau, R., Brick, M., Lohr, S., & Li, J. (2017). Adaptive and responsive survey designs: a review and assessment. *Journal of the Royal Statistical Society*, *180*, 203–223.

University of Michigan (2022). *Challenges in interviewer hiring and retention*. Meeting of the Michigan Center on the Demography of Aging, Network on Longitudinal Studies of Aging in the U.S. and Household Panel Surveys Network, Ann Arbor, 7 Nov 2022.

Voorpostel, M., Roberts, C., & Ghoorbin, M. (2021a). Integrating online data collection in a household panel study: effects on second-wave participation. *Survey Methods: Insights from the Field*. https://surveyinsights.org/?p=15709.

Voorpostel, M., Lipps, O., & Roberts, C. (2021b). Mixing modes in household panel surveys: Recent developments and new findings. In P. Lynn (Ed.), *Advances in longitudinal survey methodology*. https://doi.org/10.1002/9781119376965.ch9.

Williams, D., & Brick, J.M. (2017). Trends in U.S. face-to-face household survey nonresponse and level of effort. *Journal of Survey Statistics and Methodology*, *6*, 186–211.

Appendix

Table A1

Respondent characteristics by treatment group

Respondent characteristic	Web-First (N = 3179)	CATI-First (N = 4759)
% Male Reference Person	68.4	69.1
Age		
Mean	48.8	48.7
Median	46.0	46.0
% Married	54.0	54.5
% < 16 years of education	69.6	69.7
Total family income (USD$)		
Mean	$82,277	$84,549
Median	$60,100	$60,200
Race/Ethnicity		
% White, non-Hispanic	48.2	47.9
% African-American, non-Hispanic	37.7	37.2
% Other race	14.1	14.9
Total	100.0	100.0
Sample Type		
% National probability sample	54.7	54.8
% Low-income sample	32.8	32.6
% Immigrant refresher sample	12.5	12.6
Total	100.0	100.0

All comparisons between the groups based on paired t-tests are statistically non-signifcant