

AN EXPERIMENTAL EVALUATION OF AN ONLINE INTERVIEW SCHEDULER: EFFECTS ON FIELDWORK OUTCOMES¹

KATHERINE MCGONAGLE*

NARAYAN SASTRY

In recent years, household surveys have expended significant effort to counter well-documented increases in direct refusals and greater difficulty contacting survey respondents. A substantial amount of fieldwork effort in panel surveys using telephone interviewing is devoted to the task of contacting the respondent to schedule the day and time of the interview. Higher fieldwork effort leads to greater costs and is associated with lower response rates. A new approach was experimentally evaluated in the 2017 wave of the Panel Study of Income Dynamics (PSID) Transition into Adulthood Supplement (TAS) that allowed a randomly selected subset of respondents to choose their own day and time of their telephone interview through the use of an online appointment scheduler. TAS is a nationally representative study of US young adults aged 18–28 years embedded within the world's longest running panel study, the PSID. This paper experimentally evaluates the effect of offering the online appointment scheduler on fieldwork outcomes, including number of interviewer contact attempts and interview sessions, number of days to complete the interview, and response rates. We describe panel study

KATHERINE MCGONAGLE is Research Professor at the Institute for Social Research, Survey Research Center, University of Michigan, Ann Arbor, MI, USA. NARAYAN SASTRY is Research Professor at the Institute for Social Research, Survey Research Center and Population Studies Center, at the University of Michigan, Ann Arbor, MI, USA

This work was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development [P01-HD087155]; and the National Science Foundation [SES 1623864].

*Address correspondence to Katherine McGonagle, 426 Thompson St., Box 1248, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA 48106-1248. E-mail: kmcgon@umich.edu.

1. The authors gratefully acknowledge Rachel LeClere and Shonda Kruger-Ndiaye for leading the field operations of the TAS study and for research assistance.

members' characteristics associated with uptake of the online scheduler and examine differences in the effectiveness of the treatment across sub-groups. Finally, potential cost-savings of fieldwork effort due to the online appointment scheduler are evaluated.

KEYWORDS: Contact strategies; Data collection; Fieldwork effort; Nonresponse; Panel study; Response rate; Young adults.

1. INTRODUCTION

An experiment to evaluate the impact of an online appointment scheduler on fieldwork effort was conducted during the 2017 wave of the biennial Panel Study of Income Dynamics (PSID) Transition into Adulthood Supplement (TAS), a nationally representative supplement of US young adults who are members of families that participate in the PSID. In line with the experience of other household surveys, the amount of fieldwork effort to achieve TAS response rate goals has risen in recent waves of the study. To counteract this trend, TAS has explored various approaches including a new strategy to encourage respondents to make their own interview appointment using an online scheduler. With its nationally representative sample embedded in an ongoing panel study, TAS offers a valuable platform to evaluate the effects of a new contact strategy on fieldwork outcomes.

Major technological developments over the past quarter century—such as the emergence of cell phones and text messaging, caller identification and call blocking technologies, and households ending landline telephone service—have made contacting and interviewing respondents more difficult (Williams and Brick 2018). Behavioral, social, and cultural changes have accompanied these developments, altering communication norms. For instance, fewer people feel obligated to answer all telephone calls and especially those from a caller whose number they do not recognize. There are also changing norms and preferences for talking by telephone, a rise in telemarketing calls, alternative ways to keep in touch with friends and family (such as social media and text messaging), and smaller average household sizes, making successful contacts by telephone less likely.

These technological and social changes have directly affected the ability of survey organizations to efficiently contact study participants and complete an interview. Recent studies have shown that rates of noncontact in ongoing surveys are growing (Williams and Brick 2018; Beullens, Loosveldt, Vandenplas, and Stoop 2018; de Leeuw, Hox, and Luiten 2018). For instance, de Leeuw et al. (2018) found that noncontact rates in major US and European labor force studies are accelerating even faster than refusal rates. Other studies have found that growth in noncontact rates is lower than in refusal rates (Williams and Brick 2018; Beullens et al. 2018), but suggest this may be due to increased fieldwork effort expended by survey organizations to make contact with

respondents (Beullens et al. 2018). Consistent with this hypothesis, Williams and Brick (2018) report evidence of increased effort over subsequent waves for a small set of face-to-face household surveys, noting that published data on level of effort is scant and generally unavailable. Nevertheless, there has been a well-documented decline in household survey response rates over the past quarter century (National Research Council 2013; Williams and Brick 2018; Beullens et al. 2018; de Leeuw et al. 2018), attributable to direct refusals and difficulty in making contact with respondents (Groves and Couper 2012).

While fieldwork effort is increasing steadily for household surveys, actual interviews are typically conducted in one or two sessions. Thus, the majority of attempts to contact respondents made by telephone interviewers in household surveys are to set appointments, although, once set, appointments are often broken by respondents, necessitating additional calls to make new appointments. With fixed resources, increases in the number of contact attempts leads to lower and potentially more select response rates. The number of contact attempts made by interviewers in telephone-administered surveys is thus a key driver of data collection costs.

A variety of strategies to increase the likelihood of successful respondent contact and reduce the number of attempts have been undertaken by household panel studies (Burton, Laurie, and Lynn 2006; Watson and Wooden 2009; Schoeni, Stafford, McGonagle, and Andreski 2013). For instance, study materials are sent to respondents (McGonagle, Schoeni, and Couper 2013) informing them that interviewers will be calling, the numbers from which they will be calling, and the goals and scientific value of the study—all of which make it more likely that an interested respondent will answer the call. Survey organizations have trained interviewers to address respondent concerns (Groves and McGonagle 2001), used customized caller ID settings to display the study name or fieldwork organization (Callegaro, McCutcheon, and Ludwig 2010), and have sent respondents text messages in conjunction with interviewer calls (De Bruijne and Wijnant 2014; Dal Grande, Chittleborough, Campostrini, Dollard, and Taylor 2016). A variety of incentive strategies have been implemented, including time-delimited monetary incentives (Fomby, Sastry, and McGonagle 2017; Freedman, McGonagle, and Couper 2018) offering respondents additional payments for completing their interview within specific time frames during fieldwork. Moreover, optimal call windows for individual respondents have been identified using prior wave paradata (Lipps 2012; Kreuter, Mercer, and Hicks 2014). These strategies are designed to increase the efficiency of fieldwork by making it more likely that respondent contact is made and the interview is completed. Nonetheless, survey organizations are limited in the availability of levers to address the growing effort needed to make contact with respondents.

This paper describes an experiment designed to evaluate a new approach for facilitating respondent contact using an online interview appointment scheduler. Online schedulers have become ubiquitous in daily life for scheduling

various activities and events, such as setting appointments for a medical visit or auto repair and making restaurant or hotel reservations. To the best of our knowledge, there has been no prior use of online appointment schedulers in interviewer-administered panel surveys.

We examine four research questions.

- (I) Drawing on experimental data, we first examine the effect of offering the online appointment scheduler on four fieldwork outcomes: the number of interviewer contact attempts and interview sessions, and the number of days to complete an interview and response rates. As part of this question, we examine whether there are differences in the effects of the treatment on these fieldwork outcomes by key respondent characteristics, including gender, age, and prior wave study eligibility. Based on prior findings of differential responsiveness to financial incentives by economic characteristics of study participants (Laurie and Lynn 2009; McGonagle et al. 2013), we also specifically examine whether lower family income is related to the effectiveness of the online scheduler, potentially by accelerating receipt of the study incentive payment.
- (II) Second, we use nonexperimental data to assess fieldwork outcomes for the subset of respondents in the treatment group who took up the offer of the scheduler.
- (III) Third, we examine the respondent characteristics that predict the use of the online scheduler to set an appointment.
- (IV) Finally, we examine the cost effectiveness of offering the online scheduler, using both the experimental data to compare costs in the treatment group with the control group, and descriptive data to assess the costs among respondents in the treatment group who used the scheduler.

2. METHODS

2.1 Study Design of the PSID TAS

The experiment to evaluate an online scheduler was included in the 2017 wave of the PSID TAS. The TAS is a survey of young adults 18–28 years of age who belong to families that participate in the US PSID. The PSID is a longitudinal household panel study that has collected data on economic, social, and health behavior from a nationally representative sample of US families since 1968 (see McGonagle, Schoeni, Sastry, and Freedman 2012). The study follows the original 1968 panel members and their adult children as they grow up and form their own economically independent families. Interview data have been collected annually from 1968 to 1997, and biennially from 1999 onwards. The primary mode of data collection is via computer-assisted telephone interview by interviewers employed by the Survey Research Center (SRC) at the University of Michigan. TAS was launched in 2005 to study the causes and

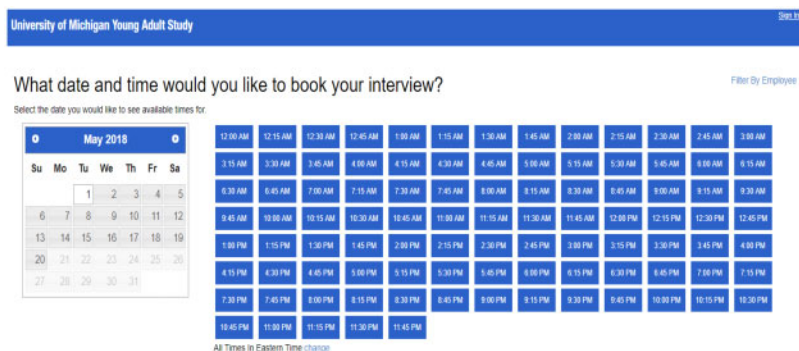


Figure 1. Online Appointment Scheduler Display.

consequences of life course transitions in young adulthood and in recognition of the extended process of the transition from adolescence to adulthood. The TAS interview captures multiple elements of young adulthood development, including employment and schooling transitions, relationships, and family formation. Eight biennial waves of TAS have been collected through 2019, with response rates ranging from 92 percent to 87 percent (McGonagle and Sastry 2015; Institute for Social Research 2019).

By design, the field periods of TAS and PSID overlap, with interviewers contacting TAS study participants once their family has completed the PSID interview. The vast majority of all eligible TAS cases are released to interviewers at the start of the TAS field period followed by smaller batches made available over a period of several months as PSID fieldwork wraps up.

The eligible sample for the 2017 wave of TAS (TAS-17) comprised 2,894 young adults and was fielded over an eight-month period between October 2017 and June 2018. Interviewers used telephone, email, and text messaging to make contact with respondents. A random sample of respondents was offered the use of an online scheduler to select their own interview appointment day and time without requiring a successful contact attempt from an interviewer. Numerous commercial online appointment schedulers are available. “Flexbooker” (flexbooker.com) was used in this study. The online scheduler website displayed a monthly calendar with available appointment days and times (figure 1). By the end of the field period, TAS-17 completed interviews with 2,530 young adults at an overall response rate of 87 percent. The average interview length was about 60 minutes. Cases released at the start of fieldwork had an overall response rate of 91 percent ($n = 1,863$ interviews), reflecting their longer opportunity to participate in the study (i.e., the full eight months) compared to those in subsequent releases, where the overall response rate was 79 percent ($n = 667$ interviews).

Save Time - Schedule Your Interview Online!

Appointments available 24 hours a day, 7 days a week

Choose to receive appointment reminders by email or text

[www.flexbooker.com/calendar/UofM]**Figure 2. Online Appointment Scheduler Information Provided on Post-It Note to Treatment Group.**

2.2 Experimental Design

Respondents in the first sample release ($n = 2,054$) were randomly assigned to either a treatment group ($n = 1,538$) and offered the use of the online scheduler or a control group ($n = 516$) and not offered the use of the online scheduler. The greater proportion of cases assigned to the treatment group reflected our expectation that the scheduler would effectively reduce fieldwork effort. However, as TAS was the first major SRC data collection project to use the scheduler and its impact on caseload management was uncertain, the treatment group was further divided through random draws into smaller batches and released for fieldwork on a staggered basis. All cases in the control group were made available to interviewers on the first day of data collection along with an equal number of cases randomly drawn from the larger treatment group. Our analyses compare fieldwork outcomes between the control group and this first batch of treatment group cases.

Prior to the start of data collection, a letter was mailed to all respondents describing the study goals, the incentive for completing the interview, and notifying them that an interviewer would making contact soon. A 3" × 3" yellow Post-It Note was prominently affixed to the front of the letter sent to the treatment group describing the availability and web address of the online scheduler with a message encouraging its use (figure 2; "Save Time—Schedule Your Interview Online! Appointments available 24 hours a day, 7 days a week. Choose to receive appointment reminders by email or text"). The letter sent to the control group did not include the Post-It Note or provide any reference to or information about the online scheduler.

Interviewers were instructed during the main study training and in subsequent team meetings to treat cases in each of the groups identically, particularly with regards to making the same number of contact attempts to each group.

2.3 Measures

2.3.1 Outcome measures. We examined four measures of fieldwork effort: (1) *The number of interviewer contact attempts* required to finalize the case, including separate measures for (i) telephone calls, (ii) emails and text messages

(“email/text”), and (iii) total attempts, constructed as the sum of telephone calls and emails and text messages; (2) *multiple sessions to complete an interview*, an indicator variable for whether or not more than one session is required to complete the interview (“yes” = 1, “no” = 0), (3) *fieldwork duration*, defined as the number of days from the release date of the case for an individual respondent (i.e., “case”) to receive a final fieldwork disposition; and (4) *response rates*, calculated as the percentage of eligible respondents completing an interview (based on definition RR6, AAPOR 2016). We also examine whether an appointment was made using the scheduler (“yes” = 1, “no” = 0) among respondents assigned to the treatment group.

2.3.2 Respondent demographic and socioeconomic characteristics. We examined differences in fieldwork outcomes and uptake rates using four key respondent characteristics from public use data files available through the PSID online Data Center (<http://psidonline.org>, last accessed 28 August 2020). The variables include: (i) gender (“female” = 1, “male” = 0), (ii) whether the respondent was eligible in a prior wave or was newly eligible in the current wave (“eligible in prior wave” = 1, “newly eligible” = 0), (iii) whether the age of the respondent was in the top third of the distribution or younger (“age 26–28” = 1, “age 18–25” = 0), and (iv) whether respondents reported family income in 2017 as below the median as (“yes” = 1) or at or above the median (“no” = 0).

Model covariates included variables for (i) respondent self-reported racial identity (“white, non-Hispanic” = 1, “other” = 0), (ii) sample type for families who were part of the original SRC national probability sample (“src” = 1) or the original low-income oversample and immigrant refresher samples (“original/immigrant” = 0), and (iii) whether the respondent residence was in a metropolitan location as defined by the Beale-Ross Rural Urban Continuum Code (“yes” = 1, “no” = 0).

2.4. Analysis Strategy

We first confirm the randomization of respondents to the experimental conditions by estimating a propensity score for assignment to each group using logistic regression analysis with the following model covariates: gender, age, family income, race, urbanicity, whether part of the PSID original nationally representative “SRC” sample, part of the original PSID low-income sample, or part of an immigrant refresher sample, and whether the respondent was eligible in a prior wave or was newly eligible in the current wave.

We describe fieldwork outcomes separately by experimental condition and test mean differences using t-tests and univariate differences at quantiles of the distribution using quantile regression. We evaluate differences in outcomes by respondent characteristics using model-based estimates from interactions between the treatment group and respondent characteristics obtained from

Table 1. Respondent Characteristics by Experimental Condition

Respondent characteristic	Treatment group (<i>n</i> = 514)	Control group (<i>n</i> = 516)
% Female	48.8	51.5
Years of age		
Mean	23.1	22.9
Median	23.0	23.0
Family income (\$)		
Mean	70,931	71,727
Median	50,907	53,498
Race		
White, non-Hispanic	47.7	45.4
African-American, non-Hispanic	42.4	44.7
Other	9.9	9.9
% Urban residence	83.1	82.2
% Participated in prior wave	53.3	56.2
PSID family sample type		
% Original SRC sample	52.9	52.5
% Low-income sample	39.7	40.9
% Immigrant refresher sample	7.4	6.6

survival models. The survival models are parametric, with a generalized gamma specification chosen based on a comparison with other models (e.g., loglogistic, lognormal, and Weibull), for which the generalized gamma provides the best fit as assessed by the information criteria (Akaike 1981; Burnham and Anderson 2004).

We provide estimates of time-to-completion of the interview across the entire field period and by percentiles of case completion by experimental condition, with survival models using the Kaplan–Meier estimator (Kaplan and Meier 1958), a nonparametric statistic that calculates differences in time-to-completion.

Multivariate logistic regression is used to identify respondent characteristics predicting scheduler use among those assigned to the treatment group.

3. RESULTS

3.1 Sample Characteristics

Sociodemographic characteristics of the sample included in the experiment, by experimental condition, are shown in table 1. As would be expected due to random assignment, there are no statistically significant differences between the groups with respect to any of the characteristics shown in table 1. The test of

Table 2. Comparisons between Treatment Group and Control Group on Fieldwork Outcomes (N = 1,030)

	A		B	
	Treatment group	Control group	Treatment group by scheduler use	
	(N=514)	(N=516)	Yes (N=168)	No (N=346)
Fieldwork outcomes				
Number of interviewer attempts (mean)				
Telephone	11.6*	13.5	5.7***	14.5
Email/text message	6.2	6.3	3.9**	7.3
Multiple sessions to complete (%)	11.1	11.2	6.5*	13.3
Number of fieldwork days				
Mean	76.5*	87.6	38.3***	95.1
By the 25th percentile of case completion	19.0***	31.0	12.0***	33.0
By the 50th percentile of case completion	44.5*	56.0	17.0***	65.0
By the 75th percentile of case completion	104.0	121.5	33.5***	137.0
By the 90th percentile of case completion	233.0	237.0	97.0***	244.0
Response rate (%)	91.6	89.9	98.8***	88.2

NOTE.—Significance tests are for comparisons with the control group. All comparisons within treatment group by scheduler use (column B) are significant at $p \leq .001$.

* $p \leq .05$, ** $p \leq .001$, *** $p \leq .0001$.

the hypothesis that all coefficients in the propensity model are 0 was confirmed ($\chi^2(7)=7.4$, $p=0.40$), indicating that random assignment to the treatment group and control group was successful and the groups are balanced across model covariates.

3.2 Fieldwork Outcomes

The first research question compares the effect of offering the online scheduler with the control group across four fieldwork outcomes: the number of interviewer contact attempts, whether the interview was completed over multiple sessions, the number of days to complete fieldwork, and response rates (table 2 column A). These results provide unbiased estimates by considering all cases as assigned to their respective experimental condition regardless of whether or not the online scheduler was used.

3.2.1 Interviewer attempts and sessions. Those assigned to the treatment group required 2.0 fewer telephone calls on average to finalize the interview compared to the control group (11.6 v. 13.5 calls, respectively $p \leq .05$; [table 2](#), column A). The test of differences between the two groups at quantiles of the distributions of telephone calls and email/texts (results not shown) found that the treatment group received significantly fewer telephone calls at the 10th percentile (1.0 fewer calls, $p \leq .001$) and the 25th percentile of calls (2.0 fewer calls, $p \leq .0001$). There were no significant differences between the two groups after the 25th percentile of calls, indicating that the benefit of the online scheduler is greatest in reducing interviewer attempts early in the field period. There were no differences between the groups in email/text attempts at the mean (6.2 v. 6.3, respectively, $p = \text{NS}$) or across different percentiles of the distribution. There was also no difference between the two groups in needing multiple sessions to complete the interview (11.1 v. 11.2, respectively, $p = \text{NS}$).

There were no statistically significant differences in the effects of the treatment on interviewer attempts or sessions by respondent characteristics. However, the treatment effect on interviewer attempts was larger for respondents with lower family income (compared to higher family income; results not shown). In particular, while the mean interviewer attempts needed to finalize an interview does not typically vary by family income in the control group (both lower and higher income respondents in the control group had a mean of nearly twenty attempts), the treatment led to 3.1 fewer attempts for lower income respondents compared to 0.8 fewer attempts for higher income respondents. These effects may have achieved statistical significance with more statistical power.

3.2.2 Number of fieldwork days and response rates. Respondents offered the use of the online scheduler required significantly fewer mean days of overall fieldwork than the control group ([table 2](#), column A; 76.5 v. 87.6 days, respectively $p \leq .05$). The difference in fieldwork days was especially large early in the field period, with only 19.0 days needed for the treatment group to complete 25 percent of all its cases compared to 31.0 days needed by the control group ($p \leq .0001$), and 44.0 days for the treatment group to complete half of all its cases compared to 56.0 days for the control group ($p \leq .05$). There was no difference between the groups in the number of days needed to reach the 75th percentile of fieldwork completion or later.

Kaplan–Meier completion rates by experimental condition underscore the beneficial impact of the online appointment scheduler early in the field period ([figure 3](#)). [Figure 3](#) displays the interview completion probability by experimental condition. As indicated by the statistically significant difference of the Kaplan–Meier estimator ($p = .016$), the treatment group (dark gray) completed the interview at a significantly faster rate overall than the control group (light gray). The figure depicts completion rates at different percentiles of the

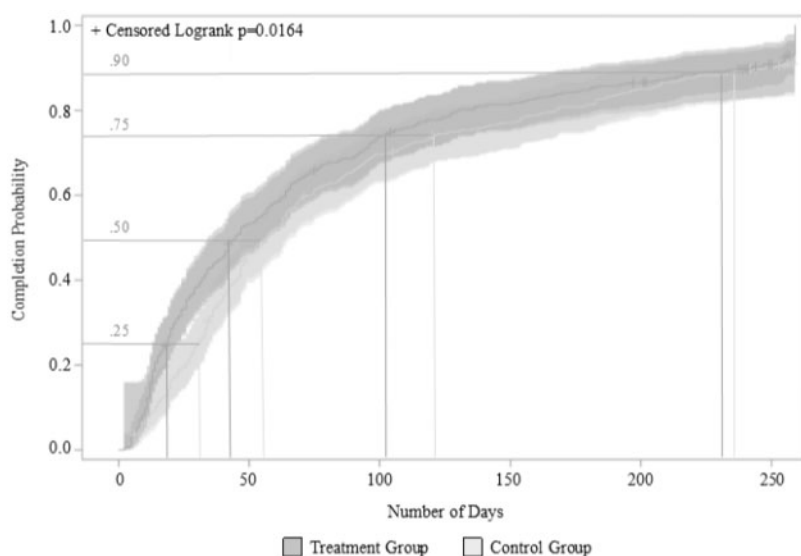


Figure 3. Timing of Completing the Interview by Experimental Condition.

distribution, showing that the gap in completion rates between the groups is especially large up to the 50th percentile, after which it starts to narrow. The treatment group achieves completion of 75 percent of all its cases sooner than the control group, but this difference is not statistically significant. By the end of the eight-month field period, there is no difference between the groups in response rate (91.6 percent *v.* 89.9 percent, $p=NS$).

There were no statistically significant differences in the effects of the treatment on median duration to interview completion by respondent characteristics. However, the treatment effect was larger when comparing subgroups based on prior-wave eligibility and family income, and statistical significance might have been achieved with a larger sample size. Returning respondents typically complete their interview faster than newly eligible respondents, and the treatment reduced the median number of days of fieldwork more for these returning respondents (i.e., 36.9 days in the treatment group compared to 52.4 days in the control group, a reduction of 15.5 days) compared to newly eligible respondents (i.e., 55.9 days in the treatment group compared to 66.8 days in the control group, a reduction of 10.9 days). Similarly, respondents from lower income families typically have shorter fieldwork duration. These families were more responsive to the treatment, which reduced median fieldwork duration by 16.1 days (i.e., 39.5 days in the treatment group compared to 55.6 days in the control group) compared to a reduction of 10.0 days among higher income families (i.e., 50.9 days in the treatment group compared to 60.9 days in the control group; results not shown).

3.3 Fieldwork Effort among Respondents Using the Scheduler

The second research question describes fieldwork effort among the subset of respondents in the treatment group who used the scheduler. Nearly 1/3 (32.7 percent) of respondents in the treatment group who were offered the scheduler used it to set an interview appointment. Fieldwork effort for those using the scheduler was dramatically lower than the control group, with 7.8 fewer telephone calls on average (5.7 v. 13.5 calls, respectively, $p \leq .0001$), 5.0 fewer at the median (2.0 v. 7.0 calls, respectively, $p \leq .0001$), and 2.4 fewer emails/text messages on average (3.9 v. 6.3 emails/texts, respectively, $p \leq .0001$; with no difference at the median) to complete the interview (table 2, column B). Those using the scheduler were less likely to require multiple interview sessions compared to the control group (6.5 percent v. 11.2 percent, respectively, $p \leq .05$). Fieldwork duration was substantially shorter among those using the scheduler at the mean (38 days compared to 87 days, respectively, $p \leq .0001$) and median (17 days compared to 56 days, respectively, $p \leq .0001$). Finally, interviews were completed by the vast majority of those who made an appointment with the online scheduler, with a response rate nearly 10.0 percentage points higher than the control group (98.8 percent v. 89.9 percent, $p \leq .0001$).

3.4 Respondent Characteristics Predicting Online Scheduler Use

The third research question examines the respondent characteristics that predict the use of the online scheduler to set an appointment. Results of multivariate logistic regression models show that scheduler use is 2.7 times higher among females than males ($p \leq .0001$), 2.6 times higher among respondents who were eligible in the prior wave compared to those who were new to the study ($p \leq .0001$), and 1.5 times higher among respondents in the top third of the age distribution ($p \leq .05$; table 3). Various nonlinear forms of total family income and cut-points for low income were explored and had consistently non-significant effects on scheduler use. There were no other significant main effects or higher-order interaction effects of respondent characteristics on scheduler use. Additional covariates for urban residence, race (white v. other), and indicators for membership in the original PSID sample frames were included in the model and had nonsignificant effects on scheduler use.

3.5 Cost Estimates

The final research question addresses the cost-implications of the online scheduler. A basic estimate of the costs of fieldwork effort associated with contact attempts by experimental conditional was generated (table 4). We have estimated that interviewers are able to make four contact attempts per hour, and that each attempt type (telephone, email, text message) requires approximately

Table 3. Predictors of Scheduler Use within the Treatment Group (*n* = 514)

Respondent characteristic	Made appointment using scheduler Odds ratio
Female	2.7**
Prior wave eligibility	2.6**
Oldest age (26–28 years)	1.5*
Income below 50th percentile	1.2 ^{ns}

* $p \leq .05$, ** $p \leq .0001$.

Table 4. Cost Estimates of Field Effort by Experimental Condition

Cost parameters	A		B	
	Treatment group	Control group	Treatment group by scheduler use	
			Yes	No
Number of cases	514	516	168	346
Average cost per interviewer attempt		\$5.76		
Total interviewer attempts (mean)	17.8	19.8	9.6	21.8
Total cost	\$52,699	\$58,849	\$9,290	\$43,447
Cost per case	\$103	\$114	\$55	\$126

the same amount of time, including time spent reviewing interviewer notes about the sample person (e.g., best times to make contact) and results of prior contact attempts (e.g., number and timing of prior contact attempts), and time spent to dial and leave a voice mail message, or type, review, and send a text message or email message. Using the average hourly wage of an interviewer (\$23), a per-attempt cost of \$5.76 was derived. Total costs were derived by multiplying the cost per attempt and number of cases by the average number of total attempts (i.e., the sum of attempts by telephone, email, and text message) in each group. A limitation of this analysis is that costs that are difficult to estimate such as temporary increases in interviewer and supervisory hours for learning about and managing the new system are not included. On average, the treatment group realized modest savings of approximately 10 percent compared to the control group (\$52,699 v. \$58,849, a difference of \$6,150). Among those using the scheduler, the cost savings were much more substantial, with the cost per case less than half the cost of the control group (\$55 v. \$114) and those who did not use the scheduler (\$126).

4. DISCUSSION

This paper describes an experimental evaluation of a new strategy designed to reduce fieldwork effort in a nationally representative panel study of young adults. Several key findings emerged. First, the group of respondents offered the online scheduler required fewer telephone calls and days of fieldwork to complete the interview compared to a control group. These respondents completed their interviews significantly faster than the control group and did so especially early in the field period. While the majority of all interviews were finalized within three months, it took another five months of fieldwork to reach study response rate goals, by which point both groups achieved high overall response rates. The faster rate of interview completion by respondents who scheduled their own appointments freed up interviewer resources to work on more difficult cases, facilitating overall production efficiency. These findings suggest that making an online scheduler available to all respondents at the start of the field period may accelerate data collection and reduce the overall number of weeks needed to collect data.

Treatment effects were stronger for respondents who had participated in a prior wave (who generally complete the interview more quickly) and for those with family income below the median (who require slightly fewer days of fieldwork but similar levels of contact attempts), although these effects did not achieve statistical significance due to a lack of statistical power. Establishing trust and familiarity with study procedures may underlie the stronger treatment effect among prior wave respondents. The finding that the treatment led to fewer interviewer attempts and fieldwork days for lower income respondents (compared to higher income respondents), thus accelerating receipt of the post-paid study incentive, fits with prior research pointing to greater responsiveness to financial study incentives by lower income individuals (Laurie and Lynn 2009; McGonagle et al. 2013; see Singer and Ye 2013).

About 1/3 of respondents who were offered the scheduler actually used it to set an interview appointment. These respondents were more likely to be female, older, and have participated in a prior wave. Those setting their own appointments had very high response rates and comparatively low field effort. As a result, the average dollar cost per completed interview for respondents making their own appointment was less than half that of other respondents (i.e., both respondents who were not offered the scheduler, and those who chose not to use it).

Notably, respondents who made their own appointments were twice as likely to complete the interview without rescheduling as those who were offered the scheduler but refrained from using it. This may be due to several factors including: the use of an on-screen calendar (v. setting an appointment on the telephone without a visual aid) enhanced the odds that the appointment was made on a day/time that worked best for the respondent; the scheduler may allow connection with other online calendars during the selection of the

interview appointment; and the potential of selecting nontraditional appointment times (v. reluctance to request a nontraditional time directly from an interviewer). These findings are consistent with theoretical frameworks that have been applied to survey participation (see [Singer and Ye 2013](#)), including social exchange theory which predicts that individuals are more likely to agree to a request when the rewards are perceived to exceed the costs ([Dillman, Smyth, and Christian 2009](#)). By facilitating the convenience of setting an interview time that worked well, and by reducing numerous interviewer contact attempts by telephone, perceived costs to participation may have been reduced.

Given the substantially lower field effort and data collection cost for respondents setting their own interview appointments, exploring ways to increase respondents' use of the scheduler seems worthwhile. As with other forms of technological adoption such as web-based interviewing, usage rates may increase organically over subsequent waves as respondents grow accustomed to the idea of scheduling their own appointments. As suggested by the leverage-saliency theory of survey participation ([Groves, Singer, and Corning 2000](#)), highlighting benefits of the scheduler through additional respondent messaging, and by making it accessible through familiar websites such as the study respondent webpages, may increase its use.

A key strength of this study is the random assignment of respondents to experimental conditions and the implementation of the experiment prior to the start of data collection, which means that the results are not confounded by respondents' prior interactions with interviewers.

However, a significant limitation is since interviewers could not be blinded to the treatment, it is possible that cases received differential treatment. While interviewers were instructed to make the same number of contact attempts for all cases, they could have deviated from these instructions and prioritized cases in the experimental condition because they perceived them as more amenable. In fact, respondents in the treatment group who did not use the scheduler received slightly more contact attempts than the control group suggesting that despite their training instructions, interviewers may have devoted more effort to these cases, although this difference was small and nonsignificant. It is probably more likely that this difference in contact attempts is due to the case composition of the control group, which includes a mixture of cooperative respondents who required lower mean contacts (and who would have used the scheduler had it been offered) and more difficult respondents who required higher mean contacts to complete their interview. The changing case composition of each group over the course of the field period—that is, the treatment group shrunk faster and became more difficult as cooperative cases using the scheduler completed their interviews—makes it difficult to evaluate this factor.

One additional limitation concerns the generalizability of the findings based on young adults to older age groups (i.e., individuals above age 28). However, given the high internet use among all adults and the pervasiveness of

commercial uses of online schedulers, there is reason to believe that the online scheduler would have similarly positive outcomes across a wide age range.

Despite these limitations, the results of this evaluation of the use of an online scheduler for data collection in a telephone-administered study seem promising and worthy of future exploration. By using a variety of strategies to increase fieldwork efficiency, the response rates of large, ongoing panel studies have been more stable than other types of surveys. Yet the growing field effort expended by these studies to achieve response rate goals may not be sustainable. The availability of an online interview appointment scheduler is another approach that may modestly lower fieldwork effort and data collection costs in panel studies. The benefits of the scheduler should be applicable to any interviewer-administered field effort that includes advance information provided to respondents prior to interviewer calling. This could include mixed-mode studies that retain a subset of respondents on telephone or through face-to-face visits or use interviewer-administered modes for nonresponse follow-up.

REFERENCES

- Akaike, H. (1981), "Likelihood of a Model and Information Criteria," *Journal of Econometrics*, 16, 3–14.
- Beullens, K., G. Loosveldt, C. Vandenplas, and I. Stoop (2018), "Response Rates in the European Social Survey: Increasing, Decreasing, or a Matter of Fieldwork Efforts?," *Survey Methods: Insights from the Field*, Available at <https://surveyinsights.org/?p=9673>.
- Burnham, K. P., and D. R. Anderson (2004), "Multimodel Inference: Understanding AIC and BIC in Model Selection," *Sociological Methods & Research*, 33, 261–304.
- Burton, J., H. Laurie, and P. Lynn (2006), "The Long-Term Effectiveness of Refusal Conversion Procedures on Longitudinal Surveys," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169, 459–478.
- Callegaro, M., A. L. McCutcheon, and J. Ludwig (2010), "Who's Calling? The Impact of Caller ID on Telephone Survey Response," *Field Methods*, 22, 175–191.
- Dal Grande, E., C. R. Chittleborough, S. Campostrini, M. Dollard, and A. W. Taylor (2016), "Pre-Survey Text Messages (SMS) Improve Participation Rate in an Australian Mobile Telephone Survey: An Experimental Study," *PLOS One*, 11, e0150231.
- De Bruijne, M., and A. Wijnant (2014), "Improving Response Rates and Questionnaire Design for Mobile Web Surveys," *Public Opinion Quarterly*, 78, 951–962.
- de Leeuw, E., J. Hox, and A. Luiten (2018), "International Nonresponse Trends across Countries and Years: An Analysis of 36 Years of Labour Force Survey Data," *Survey Insights: Methods from the Field*, Available at <https://surveyinsights.org/?p=10452>.
- Dillman, D. A., J. D. Smyth, and L. M. Christian (2009), *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method* (3rd ed.), New York: John Wiley.
- Fomby, P., N. Sastry, and K. A. McGonagle (2017), "Effectiveness of a Time-Limited Incentive on Participation by Hard-to-Reach Respondents in a Panel Study," *Field Methods*, 29, 238–251.
- Freedman, V. A., K. A. McGonagle, and M. P. Couper (2018), "Use of a Targeted Sequential Mixed Mode Protocol in a Nationally Representative Panel Study," *Journal of Survey Statistics and Methodology*, 6, 98–121.
- Groves, R. M., E. Singer, and A. Corning (2000), "Leverage-Saliency Theory of Survey Participation: Description and an Illustration," *The Public Opinion Quarterly*, 64, 299–308.

- Groves, R. M., and K. A. McGonagle (2001), "A Theory-Guided Training Protocol Regarding Survey Participation," *Journal of Official Statistics*, 17, 249–265.
- Groves, R. M., and M. P. Couper (2012), *Nonresponse in Household Surveys*, New York: John Wiley.
- Institute for Social Research (2019), *Panel Study of Income Dynamics, Transition into Adulthood Supplement 2017: User Guide*, Ann Arbor: University of Michigan.
- Kaplan, E. L., and P. Meier (1958), "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, 53, 457–481.
- Kreuter, F., A. Mercer, and W. Hicks (2014), "Increasing Fieldwork Efficiency through Prespecified Appointments," *Journal of Survey Statistics and Methodology*, 2, 210–223.
- Laurie, H.L., and P. Lynn (2009), "The Use of Respondent Incentives on Longitudinal Surveys," in *Methodology of Longitudinal Surveys*, ed. P. Lynn, pp. 205–233, New York: Wiley.
- Lipps, O. (2012), "A Note on Improving Contact Times in Panel Surveys," *Field Methods*, 24, 95–111.
- McGonagle, K. A., and N. Sastry (2015), "Cohort Profile: The Panel Study of Income Dynamics' Child Development Supplement and Transition into Adulthood Study," *International Journal of Epidemiology*, 44, 415–422.
- McGonagle, K. A., R. F. Schoeni, and M. P. Couper (2013), "The Effects of a Between-Wave Incentive Experiment on Contact Update and Production Outcomes," *Journal of Official Statistics*, 29, 261–276.
- McGonagle, K. A., R. F. Schoeni, N. Sastry, and V. A. Freedman (2012), "The Panel Study of Income Dynamics: Overview, Recent Innovations, and Potential for Life Course Research," *Longitudinal and Life Course Studies*, 3, 268–284.
- National Research Council (2013), "Nonresponse in Social Science Surveys: A Research Agenda," in *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*, eds. Tourangeau R. and T. J. Plewes, Washington, DC: The National Academies Press.
- Schoeni, R. F., F. P. Stafford, K. A. McGonagle, and P. Andreski (2013), "Response Rates in National Panel Surveys," *The Annals of the Academy of Political and Social Science*, 645, 60–87.
- Singer, E., and C. Ye (2013), "The Use and Effects of Incentives in Surveys," *The Annals of the American Academy of Political and Social Science*, 645, 112–141.
- 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: AAPOR.
- Watson, N., and M., Wooden (2009), Identifying factors affecting longitudinal survey response. In *Methodology of longitudinal surveys*, ed. Lynn, Peter, 157–83. Chichester, UK: John Wiley & Sons.
- Williams, D., and J. M. Brick (2018), "Trends in U.S. Face-to-Face Household Survey Nonresponse and Level of Effort," *Journal of Survey Statistics and Methodology*, 6, 186–211.