



INSTITUTE FOR SOCIAL RESEARCH • SURVEY RESEARCH CENTER
PANEL STUDY OF INCOME DYNAMICS
UNIVERSITY OF MICHIGAN

Technical Series Paper #19-03

Recent Developments in Web-Based Data Collection for Longitudinal Studies

October 23, 2019

Mick P. Couper
Katherine A. McGonagle
Institute for Social Research, University of Michigan

Acknowledgements: Funding for this research was provided by the Michigan Center on the Demography of Aging (P30-AG012846), the National Institute on Aging (R01-AG040213), the National Science Foundation (SES 1623864), and the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (R01-HD069609 and P01-HD087155).

Introduction

This report summarizes recent experiences with the introduction of web-based surveys in mixed-mode data collection for longitudinal surveys. The report primarily focuses on – but is not restricted to – studies funded by the National Institute on Aging (NIA). The report is based on updates presented at the Joint Meeting of the Network on Longitudinal Studies of Aging in the U.S. and Household Panel Surveys Network workshop held in November 2018, organized by the Michigan Center on the Demography of Aging (MiCDA; see <https://micda.psc.isr.umich.edu/>). This is supplemented with more recent information gleaned from presentations at subsequent conferences and meetings. It is thus a selective review of the current state of knowledge at a particular point in time and with a relatively narrow focus.

Background

Largely because of cost pressures, but also driven by the rising penetration of the Internet, large-scale surveys around the world are evaluating the feasibility of moving partly or fully from interviewer-administered modes to web-based data collection. This trend is particularly true of surveys traditionally conducted by mail (such as censuses). The precipitous decline in random digit dial (RDD) telephone survey response rates has similarly led to increased interest in Internet-based alternatives. The last decade has also seen the rise of probability-based online panels (such as the LISS panel in the Netherlands, the Understanding America Study and AmeriSpeak panels in the U.S., the GESIS panel in Germany, and others). However, our focus here is primarily on longitudinal surveys of the general population that have primarily used interviewer-administered methods (whether face-to-face or telephone), especially those that focus on aging and life course. Against this backdrop, this report briefly reviews recent developments in selected panel surveys and offers some summary assessment of progress made, lessons learned, and gaps in knowledge.

While web surveys have been widely used in many surveys of college-age populations, there is growing interest in using web-based data collection for older adults. In part this is driven by the rapid rise in Internet access among older adults. Pew Research recently reported a dramatic rise in Internet penetration among those aged 65 and older, from 12% in 2000 to 67% in 2016 (Pew Research Center, 2017). Nonetheless, a sizable proportion of this population remains offline (23%, compared with only 10% of all adults). Issues about cognitive and or physical limitations (sensorimotor skills, eyesight, etc.), along with concerns about familiarity and comfort with technology have slowed the growth of web-based data collection for older cohorts. Against this backdrop, the rapidly-rising costs of interviewer-administered surveys along with the growth in Internet and smartphone use has necessitated a more aggressive exploration of mixed-mode alternatives to traditional modes of data collection for aging and life course studies.

Panel Study of Income Dynamics (PSID)

Since 2014, PSID has been engaged in developmental activities for integrating web interviewing into the Core interview, conducted primarily by telephone since 1972. With support from the

National Institute on Aging, PSID successfully conducted off-year mixed-mode (web, paper) supplements with PSID respondents, achieving response rates of 67% in 2014 and 81% in 2016, respectively (weighted). These supplements provided key information on web take-up rates by PSID families, and experience with self-administered questionnaire design, development of mode assignment models, and nonresponse follow-up protocols (e.g., Freedman et al., 2017; McGonagle and Freedman, 2017).

During 2015-2016, with supplemental funding from the National Science Foundation, PSID drew on the existing research evidence to translate the 2015 Core CATI instrument to web using the Blaise 5 programming software. An initial phase tested two versions of the web instrument using a sample of approximately 1,000 non-PSID families drawn from an external internet panel comprising the same socio-demographic characteristics as PSID Core families. The goals were to confirm that the instrument was functioning as intended and to obtain insights to preferred questionnaire strategies by experimentally evaluating competing question design features that varied across each web version (McGonagle et al., 2017).

Following this initial phase, one version of the web instrument using design features determined to be optimal was developed. A second phase tested this version with a supplemental sample of 200 PSID respondents representing those likely to complete their interview online without excessive follow-up, although they were moderately difficult CATI cases, requiring an average of 14 contact attempts in the prior wave to complete their interview. The goals of this test were to evaluate 1) operational issues, including instrument length, the number of sessions needed to complete the interview, and whether respondents had technical problems and 2) data quality, including comparisons for key estimates between data obtained by web and the prior wave of CATI (for further details, see McGonagle et al., 2017).

Some key findings of the PSID 2016 test include:

- Overall 80% of those invited to the test attempted to login to the interview, and all who met eligibility criteria (families with family composition change since the last Core interview were excluded from the test) completed the interview;
- About one-third attempted the survey before the first reminder was mailed on day 10, and another 26% attempted before the second reminder was mailed on day 20. Thus, more than half (57%) attempted the web instrument with either no reminder or only 1 reminder;
- Respondents took an average of 1.7 sessions to complete the web instrument with more than half completing in one session (59%) and nearly all (92%) finishing in 3 or fewer sessions. Nearly 80% of respondents completed the instrument in one day, and 90% completed the instrument across 2 days;
- The vast majority of respondents (>95%) reported having a positive experience, finding it easy to login and complete the survey, agreeing the look of the survey was pleasant and the experience was enjoyable.

Constructing a respondent-friendly web version of the highly complex PSID household roster was the focus of a third phase of development during late 2017- spring of 2019. A key goal was to test programming rules for the selection of a new Reference Person, an occurrence in 24% of the cases, and for which CATI interviewers receive 1.5 days of intensive training due to the complexity of selection rules.

The household roster test occurred in the spring of 2019 using a small convenience sample of 103 respondents. The final response rate was 84%, and the average length of the interview was 7.6 minutes, approximately the same length as the CATI household roster section at 7.2 minutes. Examination of the data revealed that the roster programming rules worked correctly.

In parallel, during 2017-2019, a web version of the 2019 Core CATI instrument was programmed and tested, in anticipation of offering web as an option to approximately 30% of the Core PSID sample during main production in 2021. A test of the instrument including the household roster using a randomly selected sample of 250 PSID families is scheduled for October-December 2019.

The PSID Transition into Adulthood Supplement (TAS) is an ongoing study funded by *the Eunice Kennedy Shriver* National Institute of Child Health and Human Development of all young adults in PSID families ages 18-28. The study will implement a web-first design followed by nonresponse telephone calling in its 2019 wave. During 2018, a mixed-mode dress rehearsal was conducted using parallel web and CATI versions of the TAS 2017 instrument with 203 TAS respondents randomly drawn from a stratum of difficulty based on number of attempts to complete their prior wave CATI interview. All respondents were offered web-only for the first three weeks of data collection. The nonresponse follow-up period included telephone calls and approximately 30 contacts by mail, email, and SMS text message over the course of 13 weeks. A key goal of this test was to examine response rates by different subgroups of TAS respondent difficulty and the effectiveness of a pre-specified, intensive nonresponse protocol.

Key results from the TAS 2018 Mixed-mode test include:

- The study achieved an 88% response rate by the end of the 13-week field period, with response rates of 80%, 89%, and 90%, respectively, by level of prior wave difficulty from most to least;
- Nearly 60% of respondents completed the web interview within 3 weeks, and prior to the start of nonresponse telephone calling;
- Approximately 68% of the remaining cases who were called during the nonresponse telephone phase subsequently completed the interview, and 89% of them did so on the web;
- The average item missing data rate – calculated as the number of visited fields not answered – was low and comparable to CATI at approximately 3%.

Health and Retirement Study (HRS)

The Health and Retirement Study (HRS; see <http://hrsonline.isr.umich.edu/>) has been testing web-based data collection since 2007, when supplemental funds were obtained from the National Institute on Aging to administer between-wave or off-year surveys to panel members with Internet access. These off-year surveys continued through 2015. Response rates were similar to those achieved for off-year mail surveys of panel members, around 80% of eligible persons. This series of surveys enabled HRS to gain experience developing and implementing web surveys among an older population, explore issues of noncoverage, test new content and evaluate instrument design.

In the 2018 wave of data collection, HRS embedded a sequential mixed-mode experiment in the regular data collection. This followed a long development process, which included the conversion of the instrument from Blaise 4 to Blaise 5 (the mixed-mode version of the survey software used for HRS), and the introduction and evaluation of a new mixed-mode sample management system. Initial design and programming work began in 2013. Half of the HRS core content was then tested as an off-year web survey in 2015. Additional pilot and usability tests were conducted in 2017 to prepare for the launch of the mixed-mode design of the full HRS core instrument alongside the regular wave in 2018.

Eligibility criteria for the embedded experiment were as follows:

- Not in the half-sample assigned to the enhanced face-to-face interview in 2018
- Reported Internet access in the most recent core interview
- Completed at least one prior interview (i.e., not fresh cohorts)
- Not a proxy interview or in nursing home in most recent completed wave
- Most recent interview completed in English

This resulted in a total of 3,700 panelists eligible for the experiment. Of these, 2,250 were randomly selected for the web-first sample, with the balance of 1,450 being retained as a control sample.

Panel members in the web-first group were sent invitation letters with an \$80 prepaid incentive and survey URL, followed by an email invitation with link to survey (for those who provided an email address). A mix of mail and email reminders was sent over a period of several weeks before cases were switched to be called for a telephone interview. The control group received an advance letter with an \$80 prepaid incentive, but telephone interviews began shortly thereafter (as is the standard HRS protocol).

Preliminary results show an 80% overall response rate for the web-first sample, with 58% of sample persons completing the survey online with no phone follow-up, 4% completing online after one or more phone call, and 18% completing by phone or face-to-face (the standard protocol in the case of no phone number or noncontact). The control group yielded a response rate of 79.7%. Thus the response rate for the sequential mixed-mode group was marginally higher than in the interviewer-administered protocol, with almost three-quarters respondents

(over half of all sample persons) completing the survey online. Initial analyses of data quality indicators suggest slightly higher missing data rates for selected items (especially towards the end of the survey) in the Web-first group, but low rates of missing data overall for both experimental and group groups. Median interview length was slight longer for the Web-first protocol (109 minutes) than the control group (103 minutes). Further details can be found in Ofstedal et al. (2019).

Other U.S. longitudinal surveys discussed at MiCDA workshop

Several other surveys were discussed at the workshop, including the Survey of Income and Program Participation (SIPP; see <https://www.census.gov/sipp/>), the National Longitudinal Surveys of Youth (NLSY; see <https://www.bls.gov/nls/home.htm>), the National Health and Aging Trends Study (NHATS; see <https://www.nhats.org/>), the National Study of Caregiving (NSOC; see <https://www.nhats.org/scripts/QuickLinkNSOC.htm>), the National Social Life, Health, and Aging Project (NSHAP; see <http://www.norc.org/Research/Projects/Pages/national-social-life-health-and-aging-project.aspx>), Project Talent (see <http://www.projecttalent.org/>) and the Wisconsin Longitudinal Survey (WLS; see <https://www.ssc.wisc.edu/wlsresearch/>). None of these reported using the web as a primary mode of data collection at this time.

While not explicitly discussed at the November network meeting, the National Longitudinal Study of Adolescent to Adult Health (AddHealth; see <https://www.cpc.unc.edu/projects/addhealth>) is also exploring a mixed-mode approach in Wave V (2016-2018), with participants in the 32-42 age range. While the first four waves involved a face-to-face interview of around 90 minutes, because of budget cuts, Wave V planned for a 50-minute web-mail survey, with a subsample of nonrespondents followed up face-to-face or by telephone. The sample was divided into four subsamples, with various experiments and quasi-experiments (including modular design, incentives, and propensity modeling) conducted over the data collection period. Initial results are summarized in Liao et al. (2019).

International surveys

Developments with respect to mixed-mode surveys in other countries were discussed at the workshop by Couper (2018). The discussion below updates this summary presentation with more recent developments.

Understanding Society, also known as the UK Household Longitudinal Survey (UKHLS; see <https://www.understandingsociety.ac.uk/>) has a longstanding program of research on mixed-mode data collection dating back to 2012 (see Jäckle, Lynn, and Burton, 2015). In part this is attributable to the presence of an Innovation Panel that permits testing of new methods before implementation the full sample (see <https://www.understandingsociety.ac.uk/documentation/innovation-panel>). Starting with wave 5 of the Innovation Panel (IP5), an increasing proportion of cases have been allocated to a web-first sequential mixed-mode design, while a random subset of cases has been allocated as

a control to face-to-face data collection. In that first round of testing, the web-first group yielded a household-level response rate of 76.7% (with 49.7% of respondents completing the survey online), slightly lower than the 80.3% for the FTF group. In IP 10, the household-level response rate for the mixed-mode group (80.8%) was higher than that for the FTF control (76.9%). The proportion of surveys completed online responses has increased to 73.3%.

Based on these successes, a web-first mixed-mode approach was introduced on Wave 8 of the *Understanding Society* main survey in 2018. A random 20% of the sample was “ring-fenced” for FTF data collection to serve as a control group. A series of adaptive tests were introduced over the course of Wave 8 (which is released monthly in replicate samples) to maximize uptake of the web option (see Carpenter and Burton, 2018). Overall, 37% of households allocated to the mixed-mode group fully completed all surveys (i.e., the household survey and all individual surveys) online, with 47% of households interviewed in the previous wave doing so. A combination of design features (incentives, reminders, etc.) tested over the first year of W8 increased the proportion of households where all adults completed their interview online from 19% to 42%.

The steady rollout of the mixed-mode protocol on the Innovation Panel has allowed ISER to evaluate many aspects of data quality between modes. These are documented in an extensive series of working papers available at <https://www.understandingsociety.ac.uk/research/working-papers>.

Also largely because of cost pressures, several of the birth cohorts in the UK have been exploring mixed-mode options. For example, the National Child Development Study (NCDS), focusing on the 1958 birth cohort, employed a sequential web-telephone sequential mixed-mode approach in the age 55 survey in 2013-4. All previous sweeps were conducted face-to-face (except for the age 46 survey which was by telephone). In the age 55 survey, one-seventh of the sample was allocated to a telephone-only protocol, while the balance was allocated to the mixed-mode protocol. Response rates were similar in the two groups (78%), with about 65% of respondents in the mixed-mode group completing the survey online. See <https://cls.ucl.ac.uk/wp-content/uploads/2017/07/NCDS-Age-55-Sweep-Technical-Report-FINAL.pdf> for further details. The age 62 sweep (planned for 2020) will revert to face-to-face data collection to include cognitive assessments and a separate nurse visit.

In contrast, the 1970 British Cohort Study (BCS70; see <https://cls.ucl.ac.uk/cls-studies/1970-british-cohort-study/>), has employed face-to-face data collection for all sweeps, including the age 50 sweep planned for 2020.

Next Steps (the Longitudinal Study of Young People in England, or LSYPE), which focuses on the cohort born in 1989-90, used face-to-face data collection in the first four sweeps but switched to sequential mixed-mode (web, telephone and face-to-face) in subsequent sweeps (ages 18-25) (see <https://cls.ucl.ac.uk/cls-studies/next-steps/>). In the age 25 sweep, conducted in 2015-6, all sample members who has ever participated in the study were included. An overall

response rate of 61% was obtained, with 61% of responses via the web, 9% by phone, and 30% face-to-face.

Finally, the Millennium Cohort Study (MCS), designed to study the lives of children born at the turn of the century, has used face-to-face data collection for the first six sweeps (or waves) of data collection (see <https://cls.ucl.ac.uk/cls-studies/millennium-cohort-study/>). In the seventh sweep, the study employed a sequential mixed-mode design, starting with web then switching to face-to-face. As of the time of writing, results are not yet available.

The Swiss Household Panel (SHP; see <https://forscenter.ch/projects/swiss-household-panel/>) tested the use of mixed-mode data collection in the pilot for the third refreshment sample in 2018. Sample cases were randomly assigned to one of three experimental groups:

- Group 1 (control): phone (if landline # available) and face-to-face (current SHP design); face-to-face cases will be pushed to phone in the second wave
- Group 2 (hybrid): sequential mixed-mode using phone and face-to-face at the household level (as in Group 1), and web at individual level; in the second wave, part will keep this design and part will switch to Group 3
- Group 3: web-only data collection.

Access to the Swiss population register makes this design feasible. Preliminary results are available from the first wave of data collection (see Voorpostel et al., 2019). Household-level responses rates (completed the household questionnaire) were 53% for group 1, 52% for group 2, and 47% for group 3, while individual-level response rates were 69%, 67% and 62% respectively. FORS is testing mixed-mode designs in several other longitudinal surveys.

While the German Socio-Economic Panel (SOEP; see <https://www.diw.de/en/soep>) has not used mixed-mode data collection involving the web in the core data collection. However, an add-on web survey was offered to participants in the 2017 Innovation Panel survey. Respondents with Internet access who expressed interest in completing online surveys and provided an email address were invited to a follow-up survey. Based on preliminary results, 25% of the gross sample, and 36% of those with Internet access completed the online survey. A similar approach was used in the 2018 Innovation Panel, with similar results (response rates of 24% and 36% respectively). In additional sequential web-telephone data collection has been tested on one sub-study of the SOEP, the Families in Germany (FiD) study. In the 2014 wave, panelists were screened first by telephone to determine their willingness to do the survey online. While 82% reported being willing to do the survey online, the web response rate was 47%. In the 2017 wave, phone was again used to complete the household roster and encourage online completion, while those who had previously responded online were invited to the web first. With an overall response rate of 78%, almost a third (30%) of all interviews were completed online.

Several other large-scale panel studies have no current plans to adopt mixed-mode data collection. These include the English Longitudinal Survey of Ageing (ELSA; see <https://www.elsa->

project.ac.uk/), the Household, Income and Labour Dynamics in Australia (HILDA) Survey (see <https://melbourneinstitute.unimelb.edu.au/hilda>), and the Survey of Health, Ageing and Retirement in Europe (SHARE; see <http://www.share-project.org/>). However, in the latter case, wave 6 (2015) was conducted using a sequential web-phone approach in the Netherlands because of a lack of funding. A 43% response rate was obtained from panel members, with 94% of responses via the web.

This brief and selective overview of international studies shows the variation in progress made in exploring sequential mixed-mode designs. This uneven progress is due to a number of factors, of which available funding and sponsor support for innovation are likely to be key.

Progress Made and Lessons Learned

The brief review of research summarized above suggests that we have learned a lot about what works and what does not work in implementing sequential mixed-mode design in longitudinal household studies. Here we summarize some of the key lessons learned from this research.

The first lesson is that switching from an interviewer-administered survey to a mixed-mode design involving web data collection is not a simple turnkey solution. It involves careful development and testing, especially if the goal is to minimize impacts to longstanding time series. We have not yet seen evidence of large-scale cost-savings from the switch (see next section), but the conversion to mixed-mode takes considerable upfront time and effort if it is to minimize negative effects on longitudinal data collections.

Second, after initial mixed results, we are now seeing successes in the conversion to mixed-mode data collection. Most notable among these is *Understanding Society*, which again benefited from a multi-year long process of research and development on the Innovation Panel before implementation in the mainstage survey. The 2018 Health and Retirement Study test also shows a lot of promise, again building on many years of development and testing. Similar positive results after a lengthy process of development and testing are reported by PSID.

However, it is clear from these examples that switching to a fully web-based data collection is not likely in the foreseeable future, especially for surveys of aging populations. The necessity of including interviewer administration for those unable or unwilling to complete the survey online remains clear. The key challenge for maximizing cost savings is how to maximize the proportion of the sample completing the survey online, thereby minimizing the need for more expensive follow-up efforts.

Third, the evidence on data quality differences is largely positive, with some notable exceptions (see later). Item missing data in HRS appears to be slightly higher among web respondents than telephone, but the rates are general low in both modes (Ofstedal et al., 2019). Similar results are found for *Understanding Society* (see Jäckle, Lynn, and Burton, 2015) and in PSID pilot tests to date (see McGonagle et al., 2017)

Response distributions are generally comparable across modes, but there are some exceptions (see, e.g., Cernat, Couper, and Ofstedal, 2016; Klausch, Schouten, and Hox, 2017; McGonagle et al. 2017; Schouten et al. 2013), some of which can't easily be predicted or explained by existing models of the response process (see d'Ardenne et al., 2017; Vannieuwenhuyze and Lynn, 2014). There is a lot of recent work on estimating mode effects (e.g., Cernat, Couper, and Ofstedal, 2016; Klausch, Schouten, and Hox, 2017; Schouten et al., 2013), and on calibrating or adjusting for mode effects (see Buelens and van der Brakel, 2015; Kolenikov and Kennedy, 2014; Vandenplas, Loosveldt, and Vannieuwenhuyze, 2016; Vannieuwenhuyze and Loosveldt, 2013). Developments in the area of identifying and dealing with mode effects in mixed-mode surveys continue, but the challenge of doing so in a longitudinal setting remains. Many analysts seem to be simply ignoring any measurement differences in practice.

Survey Length and Modular Design

Turning to more specific lessons learned, one of the initial concerns was that long surveys (such as are typically administered in large-scale longitudinal surveys) are not suitable for the web. Indeed, there are many prescriptions in the literature about exceeding 20 minutes (or even 12 or 10) for an online survey. In the longitudinal survey context, we have learned that motivated and cooperative respondents are willing to complete long surveys online. The median length of the *Understanding Society Wave 8* household questionnaire completed online is 17.1 minutes, with an additional median of 35.5 minutes for each adult in the household. The median length of the 2018 HRS survey completed online is about 109 minutes, while that of the PSID 2016 test instrument was approximately 74 minutes. Contrary to expectation, however, it appears that web surveys do not take less time than interviewer-administered surveys. The median response times for HRS 2018 were 109.1 minutes for web and 106.6 minutes for CATI. Similarly, median *Understanding Society W8* response times for the household questionnaire were 17.1 minutes for web and 14.2 minutes for CAPI, while those for the individual adult questionnaires were 35.5 minutes for web and 33.4 minutes for CAPI. Mean NCDS times from the age 55 sweep were 35.5 for web and 27.1 for phone.

In the PSID 2016 test, the average length of the web instrument was about 82 minutes (~74 median) compared to about 72 minutes in CATI (~73 median). Two reasons for the higher average duration in web were explored. First, examination of time stamp data revealed periods of time when respondents "paused" in the web instrument for between 2 and 15 minutes (the time-out limit). About two-thirds of respondents paused at least once, with a total average duration of all pauses of about 16 minutes. Second, about half of web respondents used at least one, and usually multiple, assistive documents (e.g., tax returns); those doing so had average web durations of 24 minutes longer than those not using documents. The influence of document use on interview length was greater in web than in CATI, with an increase on average interview length of about 27% in web compared to about 11% in CATI (for further details, see McGonagle et al., 2017).

Because of concerns about the length of survey instruments, a number of researchers have proposed modular designs, or chunking the survey into smaller pieces completed at different

times. However, emerging evidence suggests that this results in lower overall response rates. For example, Peytchev et al. (2019) tested a 30-minute module versus two 15-minute modules in National Postsecondary Student Aid Study. They obtained a response rate of 59.5% for the single module and 53.1% for the two-module version. Toepoel and Lugtig (2018) found higher start rates when the survey was split into three parts, but also higher breakoff rates, with the net result of a slightly lower completion rate. After experimenting with a module design in the 2017 European Values Survey in Switzerland, FORS similarly concluded that “it is better to field a whole 1 hour web-survey, rather than splitting it into two parts.” (FORS preliminary presentation to the Scientific Board). Liao et al. (2019) tested a modular approach (along with several other experiments) in a pilot study for Wave V of AddHealth. They found that the submission rate for the modular design was lower than for a single questionnaire, and that respondents tended to prefer one long submission rather than two smaller ones. The PSID 2016 test found that respondents randomly assigned to a condition informing them of the total interview length (60-80 minutes) took significantly more days to start the interview, but significantly fewer days to complete once started, compared to a group told that the interview consisted of three parts of 20-30 minutes (McGonagle et al., 2017). Finally, the use of a modular design was experimentally evaluated in the TAS-18 mixed-mode pilot study (TAS presentation to the PSID Board). Respondents informed at the start of the web survey that the total interview length was 60-80 minutes had significantly higher response rates than those told the interview had two parts, each 30-40 minutes. While tests of modular designs continue, these early results suggest that modular design may not be a better alternative, or that we haven’t yet found the best way to implement modular designs.

One consequence of administering lengthy instruments online is that the surveys are often completed in multiple sessions. For instance, among respondents who completed the HRS 2018 survey online, 41% completed the survey in a single session, while 19% took four or more sessions to complete the survey. In comparison, 51% of CATI respondents completed the survey in one session while 12% took four or more sessions. In the PSID 2016 test, 59% completed the web instrument in one session and 8% needed 3 or more sessions, compared to CATI where 74% completed the survey in one session and less than 2% took 3 or more sessions. Other surveys have reported similar observations.

This suggests that it is important to make it easy for respondents to suspend and resume the survey. Given the findings on modular designs, this approach may be especially important for long surveys. Developing reminder messages that encourage respondents to login to the web portal and start the interview, as well as messages that offer encouragement and assistance to those who started the interview but suspended it is an approach being tried in PSID and TAS.

The above discussion should not be taken to imply that we could increase the length of surveys without consequence. Surveys already impose a high burden on respondents, and everything should be done to reduce rather than increase already-lengthy surveys.

While web surveys may require more sessions to complete compared to CATI, a potential benefit may be the shorter time period needed to complete online interviews, leading to

shorter production periods overall compared to interviewer-administered modes. For example, the TAS 2018 mixed mode pilot completed 60% of its web interviews within 3 weeks, compared to 12 weeks to complete the same proportion of CATI interviews during the 2017 wave. McGonagle and Sastry (2019) report positive results of a test of providing an online scheduler for respondents to schedule CATI interviews, reducing production time and call attempts.

Mode of invitation

The evidence is clear that using email invitations to supplement mail invitations is effective at increasing the proportion of respondents who complete the survey online (see Cernat and Lynn, 2018; Patrick et al., 2018). These findings point to the importance of collecting valid contact information, including email addresses, from panel members in prior waves. However, those who provide a valid email address tend to be more cooperative than those who don't (see Carpenter and Burton, 2018).

Mode assignment

Several studies have examined the impact of providing individuals with their self-reported preferred survey mode on mixed-mode data collection outcomes (e.g., Olson et al. 2012). Propensity models have also been used in *Understanding Society* to identify panel members who are most likely to respond via the web and to target those panelists in sequential mixed-mode data collection (see Kaminska and Lynn, 2017). Freedman et al. (2017) found that a simple model using demographic variables and information on Internet use yielded highly accurate prediction estimates for the assignment of respondents to initial mode and targeted data collection protocols. This work also identified a group with no strong predicted mode preference who required substantially more field effort compared to those with high mode propensity. A majority of individuals in this group were young, single, female, with less than a college education, and used a smartphone to complete the web interview. With the availability of covariates to predict mode, panel studies may especially benefit from developing models of likely mode and implementing targeted protocols. Additional research is needed to develop protocols for individuals with no particular mode propensity, especially given the need to engage the participation of youth in ongoing household studies.

Dependent interviewing and updating household composition

Another concern that has been raised relates to the use of dependent interviewing in longitudinal survey. The concern is that because one has less control over who completes the survey (especially in household as opposed to individual surveys) there is a risk of revealing information reported by another household member in a subsequent wave. An experiment in the second mixed-mode wave of the *Understanding Society* Innovation Panel (see Burton, 2015) found that compared to face-to-face respondents, web respondents expressed significantly greater concerns about the confidentiality of their data (22.5% face-to-face, vs.

32.8% web) and were more likely to say that dependent interviewing made them less willing to participate in the future (2.1% face-to-face, vs. 6.2% web). We know of no recent experimental work on this issue, and web surveys of individuals (e.g., *Understanding Society*) are using dependent interviewing with no reported negative consequences.

A related concern is about updating the household roster or household composition. This activity is a key part of many longitudinal surveys, and involves feeding forward information from prior waves for review and update. The concern is that respondents may not be as successful performing this task in a self-administered survey than interviewers may be. With the goal of making it easier for respondents to make updates to the preloaded household roster, PSID is currently testing the use of an external “look-up” database linked to the web survey that provides each respondent a list of their relatives who also participate in the study. Respondents who have relatives moving in to their households since the prior wave can select them from the list to have the family member’s information added directly to the roster. Information on ease of technical use and privacy concerns will be collected during the test.

Initial results from the 2018 HRS suggest that web respondents are making as many changes (additions, subtractions, and corrections) as are telephone respondents. While additional analyses are underway, this suggests concerns about updating rosters for most respondents may be unfounded, but additional work should be done to examine whether there are respondent characteristics, including cognitive ability, household size, particular family composition changes, privacy concerns, etc., that may lead to difficulty completing rosters.

Gaps in knowledge and remaining challenges

While a lot of progress has been made in advancing our understanding of how best to implement sequential mixed-mode designs in longitudinal surveys, a number of significant gaps in knowledge and challenges remain. Here we highlight selected challenges.

As noted earlier, differences in response distributions are generally not large for most survey questions. However, there are some significant exceptions. One of these relates to measures of cognition. In analyses across several waves of HRS data involving interviewer-administration and web data collection, Ofstedal, McClain, and Couper (2019) found notable differences in cognitive performance, with respondents performing better (i.e., higher levels of cognitive performance) on the web. Al Baghal (2019) reports finding similar effects in *Understanding Society*. The challenge of finding mode-resistant or mode-equivalent measures of cognition remains.

A number of other specific measurement or questionnaire issues are in need of greater study with regard to mixed-mode implementation. These include complex financial questions (both income and expenditures), life history calendars, the collection of job descriptions to facilitate industry and occupation coding and other complex lookup tasks (like prescription drugs). McGonagle et al. (2017) found that while PSID respondents provided significantly less detail

(about one-third as much) in open-ended questions about their occupational activities in web than in the prior CATI wave – during which interviewers probed multiple times for job descriptors – the detail was sufficient for successful coding to 3-digit Census occupation codes. Across the two modes, 88% of cases were assigned identical first digits, 80% identical first and second digits, and 74% were identical on all three digits. These results indicate that open-ended information can be successfully collected through self-administered modes (and probing may be unnecessarily high in telephone).

Another challenge for the adoption of mixed-mode data collection is that of collecting biospecimens or conducting physical tests. Many panel surveys include extensive collection of biospecimens (including saliva, blood spots, whole blood, etc.) and in most cases these are administered by interviewers or (in some cases) nurses. While there have been some small scale efforts at self-administration of saliva tests using mailed kits (see, e.g., Gatny, Couper, and Axinn, 2013), we know of no effort to test the collection of biospecimens in a self-administered survey.

Longitudinal (and other) surveys increasingly make use of administrative data to supplement survey responses. This generally implies the need to obtain informed consent from participants to such linkage. Several studies have found that the administrative data linkage consent rates are significantly lower in among web respondents than those interviewed in-person, despite the fact that web respondents are generally more cooperative. For example, Jäckle et al. (2019a) reported rates of consent to linkage to Financial Conduct Authority (FCA) data by mode of 67% for CAPI and 47% for web, a difference of 19 percentage points. This difference increases with adjustments for selection into mode. Similar results were found for the Next Steps cohort study (Thornby et al., 2017) and for social media linkage requests (Al Baghal et al., 2019). Research is underway to understand the reasons for lower data linkage consent rates among web respondents in order to address this problem. For example, Jäckle et al. (2019b) found that linkage consent rates for web improve when the linkage request is brief and simple. Additional research is needed to address this serious drawback for mixed-mode surveys requiring consent for data linkage.

As we have noted earlier, initial concerns about verifying and updating household composition in panel studies seems to be unfounded. However, this is not the case for recruiting fresh cohorts using mixed-mode methods. Research on cross-sectional surveys points to considerable challenges of inviting households by mail using address-based sampling (ABS) to complete online screening surveys to identify eligible household members and select respondents for follow-up surveys. More promising results are being found in countries where population registers can be used to identify eligible persons. Both the Swiss Household Panel refreshment sample (Voorpostel et al., 2019) and the Generations and Gender Survey (GGG; see Lugtig et al., 2019) have tested web-first approaches in recruiting fresh samples. In the U.S., where such frames are not available to researchers, significant challenges remain, and panel refreshment is likely to remain interviewer-administered for some time.

A final and notable gap in knowledge relates to the costs of mixed-mode data collection. While cost savings are one of the ostensible reasons for the move to mixed-mode designs, we know remarkably little about the relative costs of mixed-mode data collection. Bianchi, Biffignandi, and Lynn (2017) report cost ratios of 0.85 in wave 5 (where web made up 23.8% of responses) and 0.89 in wave 7 (with 41.9% of responses by web), suggesting that increasing the proportion of web has not increased the costs saving of mixed-mode relative to face-to-face. Ofstedal et al. (2019) report an estimate of 3.2 fewer calls per case for the web-first group compared to the CATI group, but have not yet estimated the cost savings. Given the general reluctance of data collection agencies to reveal detailed financial information, the costs savings are largely unknown. Further, the increase in web responses may serve to make face-to-face follow-up less efficient, as interviewers have fewer cases to work and these cases may be more reluctant (requiring more effort). In addition, the initial costs of transitioning a complex survey to the web are often not considered. Considerable effort is involved in converting instruments designed for interviewer-administration into versions suitable for web-based self-administration (see, e.g., Betts and Cubbon, 2015; McGonagle et al., 2017). Without much evidence, we can speculate that while mixed-mode approaches may save money in the long run, the savings may not be as dramatic as hoped for, and the costs of making the transition may initially be higher.

Summary

The last several years have seen significant progress in moving toward mixed-mode data collection for longitudinal surveys. These achievements have come in part through substantial investments on the part of funding agencies. The value of these nationally-important data time-series requires careful evaluation before making major changes to data collection processes. The progress in our knowledge of these methods points to the value of careful development and testing. Progress is incremental, and does not occur overnight. It also points to the value of sharing information and lessons learned across longitudinal (and indeed cross-sectional) surveys, both nationally and internationally. Collectively we have advanced – and continue to advance – our knowledge about what works best and what may not work as we transition to mixed-mode data collection.

There are still a number of gaps in our knowledge, and remaining challenges to overcome. As reviewed above, these include:

- Maximizing the proportion of panelists who complete the survey fully online, to maximize the costs savings and minimize measurement error effects.
- Maximizing informed consent to administrative record linkage.
- Identifying sources of measurement differences between modes and finding ways to adjust for such differences in analysis.
- Finding ways to administer cognitive assessments in a mixed-mode environment.

Work is already underway on several of these fronts. We are confident that the remaining challenges will be addressed in time. But there is sufficient evidence from this review of recent work to substantiate Annette Jäckle's recent blog post: "...as a field we have moved away from questioning whether we should use mixed mode data collection – to asking how best to design mixed mode surveys. Going forward, the challenge for mixed modes surveys is how to combine modes to the best advantage and how to address non-response bias and selectivity across different modes." (<https://www.closer.ac.uk/news-opinion/blog/news-opinion2017mixed-modes-stay/>). We look forward to the expansion of knowledge that the next few years will bring, as we move toward a mixed-mode world while preserving the longitudinal comparability of our data.

References

Al Baghal, T. (2019), "The Effect of Online and Mixed-Mode Measurement of Cognitive Ability." *Social Science Computer Review*, 37 (1): 89-103.

Al Baghal, T., Sloan, L., Jessop, C., Williams, M.L., and Burnap, P. (2019), "Linking Twitter and Survey Data: The Impact of Survey Mode and Demographics on Consent Rates Across Three UK Studies." *Social Science Computer Review*, first published online February 26, 2019.

Betts, P., and Cubbon, B. (2015), "Adapting Labour Force Survey Questions from Interviewer-administered Modes for Web Self-completion in a Mixed-mode Design." *Statistical Journal of the IAOS*, 31 (4): 629-644.

Bianchi, A., Biffignandi, S., and 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.

Buelens, B., and van der Brakel, J.A. (2015), "Measurement Error Calibration in Mixed-Mode Sample Surveys." *Sociological Methods and Research*, 44 (3): 392-426.

Burton, J. (ed.) (2013), "Understanding Society Innovation Panel Wave 5: Results from Methodological Experiments." Colchester: University of Essex, ISER: Understanding Society Working Paper Series No. 2013-06.

Carpenter, H., and Burton, J. (2018), "Adaptive Push-to-web: Experiments in a Household Panel Study." Colchester: University of Essex, ISER: Understanding Society Working Paper Series No. 2018-05.

Cernat, A., Couper, M.P., and Ofstedal, M.B. (2016), "Estimation of Mode Effects in the Health and Retirement Study using Measurement Models." *Journal of Survey Statistics and Methodology*, 4 (4): 501-524.

Cernat, A., and Lynn, P. (2018), "The Role of E-mail Communications in Determining Response Rates and Mode of Participation in a Mixed-mode Design." *Field Methods*, 30 (1): 70-87.

Couper, M.P. (2018), "Mixed-Mode Data Collection: Selected International Experiences." Presentation at the Joint Meeting of the Network on Longitudinal Studies of Aging in the U.S. and Household Panel Surveys Network, Washington, DC, November.

Couper, M.P. (2019), "Online Surveys: Opportunities and Challenges." Keynote presentation at the NCRM/ONS conference on The Future of Online Data Collection in Social Surveys, Southampton, June; see <https://www.ncrm.ac.uk/research/datacollection/>

d'Ardenne, J., Collins, D., Gray, M., Jessop, C., and Pilley, S. (2017), "Assessing the Risk of Mode Effects. Review of Proposed Survey Questions for Waves 7-10 of Understanding Society." ISER, University of Essex: Understanding Society Working paper no. 2017-04.

Freedman, V.A., McGonagle, K., and Couper, M.P. (2018), "Use of a Targeted Sequential Mixed Mode Protocol in a Nationally Representative Panel Study." *Journal of Survey Statistics and Methodology*, 6 (1): 98-121.

Gatny, H.H., Couper, M.P., and Axinn, W.G. (2013), "New Strategies for Biosample Collection in Population-Based Social Research." *Social Science Research*, 42: 1402-1409.

Jäckle, A., Beninger, K., Burton, J., and Couper, M.P. (2019a), "Understanding Data Linkage Consent in Longitudinal Surveys." In P. Lynn (ed.), *Advances in Longitudinal Survey Methodology*. New York: Wiley, in press.

Jäckle, A., Burton, J., Couper, M.P., Crossley, T.F., and Walzenbach, S. (2019b), "Consent to HMRC Data Linkage Experiment." Colchester: University of Essex, ISER: Understanding Society Working Paper Series No. 2019-03.

Jäckle, A., Lynn, P., and 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): 55-70.

Kaminska, O., and Lynn, P. (2017), "The Implications of Alternative Allocation Criteria in Adaptive Design for Panel Surveys." *Journal of Official Statistics*, 33 (3): 781-799.

Klausch, T., Schouten, B., and Hox, J.J. (2017), "Evaluating Bias of Sequential Mixed-mode Designs Against Benchmark Surveys." *Sociological Methods and Research*, 46 (3): 456-489.

Kolenikov, S., and Kennedy, C. (2014), "Evaluating Three Approaches to Statistically Adjust for Mode Effects." *Journal of Survey Statistics and Methodology*, 2 (2): 126-158.

Liao, D., Biemer, P.P., Harris, K.M., Burke, B.J., and Halpern, C.T. (2019), "Transitioning from In-Person Mode to Web-Mail Mixed Mode in a Panel Survey." Paper presented at the annual meeting of the American Association for Public Opinion Research, Toronto, May.

Lugtig, P., Toepoel, V., Emery, T., Cabaco, S., Lück, D., Schumann, A., and Naderi, R. (2019), "A Push-to-Web Experiment for Cross-National Surveys: Effects on Nonresponse Rates, Bias and Measurement." Paper presented at the European Survey Research Association conference, Zagreb, July.

McGonagle, K.A. and Sastry, N. (2019), "Schedule Your Own Telephone Interview Time Online? Experimental Evidence of the Effects of an Online Interview Scheduler on Fieldwork Outcomes." Paper presented at the European Survey Research Association, Zagreb, Croatia, July 15-19, 2019.

McGonagle, K.A. (2018), "Conversion to Web and Mixed mode in PSID." Paper presented at the Joint Meeting of the Network on Longitudinal Studies of Aging in the U.S. and Household Panel Surveys Network, Washington, DC, November.

McGonagle, K.A. and Freedman, V.A. (2017), "The Effects of a Delayed Incentive on Response Rates, Response Mode, Data Quality, and Sample Bias in a Nationally Representative Mixed Mode Study." *Field Methods*, 29 (3): 221-237.

McGonagle, K.A., Freedman, V.A., Griffin, J., and Dascola, M. (2017), "Web Development in the PSID: Translation & Testing of a Web Version of the 2015 PSID Telephone Instrument." Ann Arbor: Institute for Social Research, PSID Technical Series Paper #17-02, https://psidonline.isr.umich.edu/Publications/Papers/tsp/2017-02_Trans_Test_Web_Tele_Instr.pdf

Ofstedal, M.B. (2018), "Transitioning a Panel Survey to Web: Experiences from the Health and Retirement Study." Paper presented at the Joint Meeting of the Network on Longitudinal Studies of Aging in the U.S. and Household Panel Surveys Network, Washington, DC, November.

Ofstedal, M.B., Couper, M.P., Hupp, A.L., Gatward, R., and Weir, D. (2019), "Introducing Web to a Panel Survey: The Health and Retirement Study." Paper presented at the European Survey Research Association conference, Zagreb, July.

Ofstedal, M.B., McClain, C.A., and Couper, M.P. (2019), "Measuring Cognition in a Multi-mode Context." In P. Lynn (ed.), *Advances in Longitudinal Survey Methodology*. New York: Wiley, in press.

Patrick, M.E., Couper, M.P., Laetz, V.B., Schulenberg, J.E., O'Malley, P.M., Johnston, L., and Miech, R.A. (2018), "A Sequential Mixed Mode Experiment in the U.S. National Monitoring the Future Study." *Journal of Survey Statistics and Methodology*, 6 (1): 72-97.

Pew Research Center (2017), "Tech Adoption Climbs Among Older Adults." Washington, DC: Pew Research Center, <https://www.pewinternet.org/2017/05/17/tech-adoption-climbs-among-older-adults/>

Peytchev, A., Peytcheva, E., Conzelmann, J.G., Wilson, A., and Wine, J. (2019), "Modular Survey Design: Experimental Manipulation of Survey Length and Monetary Incentive Structure." *Journal of Survey Statistics and Methodology*, first published online April 12, 2019.

Schouten, B., van den Brakel, J., Buelens, B., van der Laan, J., and Klausch, T. (2013), "Disentangling Mode-Specific Selection and Measurement Bias in Social Surveys." *Social Science Research*, 42 (6): 1555-1570.

Thornby, M., Calderwood, L., Kotecha, M., Beninger, K., and Gaia, A. (2017), "Collecting Multiple Data Linkage Consents in a Mixed Mode Survey: Evidence and Lessons Learnt from Next Steps." London: Institute of Education, Centre for Longitudinal Studies Working paper 2017/13.

Toepoel, V., and Lugtig, P. (2018), "Modularization in an Era of Mobile Web: Investigating the Effects of Cutting a Survey Into Smaller Pieces on Data Quality." *Social Science Computer Review*, first published online, July 1, 2018.

Vandenplas, C., Loosveldt, G., and Vannieuwenhuyze, J.T.A. (2016), "Assessing the Use of Mode Preference as a Covariate for the Estimation of Measurement Effects between Modes. A Sequential Mixed Mode Experiment." *Methods, Data, Analyses*, 10 (2): 119-142.

Vannieuwenhuyze, J.T.A., and Loosveldt, G. (2013), "Evaluating Relative Mode Effects in Mixed-Mode Surveys: Three Methods to Disentangle Selection and Measurement Effects." *Sociological Methods and Research*, 42 (1): 82-104.

Vannieuwenhuyze, J.T.A., and Lynn, P. (2014), "Measurement Effects Between CAPI and Web Questionnaires in the UK Household Longitudinal Study." Colchester: University of Essex, ISER: Understanding Society Working Paper Series No. 2014-01.

Voorpostel, M., Tillmann, R., Monsch, G.-A., Antal, E., Kuhn, U., and Lipps, O. (2019), "Results of the First Wave of the Mixed Mode Pilot for the Swiss Household Panel." Paper presented at the FORS lunch seminar, 12th February.