

APPENDIX G

Survey of Doctorate Recipients Dependent Interviewing Study (DIS) Pilot Test

Survey of Doctorate Recipients

Dependent Interviewing Study (DIS) Pilot Test

Final Findings Report

February 2021

Authors

Darby Steiger
Westat SDR Team

Submitted to:

National Science Foundation
2415 Eisenhower Ave
Alexandria, Virginia 22314
703.292.4486

Submitted by:

Westat
An Employee-Owned Research Corporation®
1600 Research Boulevard
Rockville, Maryland 20850-3129
(301) 251-1500

Table of Contents

1. Background	3
2. Data Collection Methodology	6
3. Results	8
3a. Methodological findings	8
3b. Response Distributions to Key Variables	15
3c. Key findings from analytic tables	18
3d. Response Analysis Survey Findings	26
3e. Meaningful Change Analysis	30
4. DI Decision Criteria.....	32
4a. Should 2021 SDR Proceed with Dependent Interviewing? YES.....	32
4b. If we do proceed with dependent interviewing, which approach should be used? DI-2 is recommended	39
4c. If we do proceed with DI, will it apply just to the longitudinal sample, or to the full continuing cohort?.....	40
4d. If we do proceed with DI, would we use the approach with all items tested in the Pilot Study, or a subset?	41

1. Background

The Dependent Interviewing Study (DIS) Pilot test is a quantitative test of dependent interviewing that involved assigning sample members to one of three experimental treatments to determine the best questionnaire design approach for the 2021 SDR. The DIS Pilot was designed to identify an appropriate follow-up instrument for the longitudinal component of SDR.

This experiment served three purposes:

- inform the 2021 SDR data collection instrument design,
- identify which of the three treatments yields high data quality and is also perceived positively by the SDR population, and
- fill a gap in the literature regarding the use of dependent interviewing for self-administered questionnaires.

The experiment included a sample of respondents to the 2019 SDR, as well as a small sample of respondents who participated in the 2015 or 2017 SDR cycles, but who did not participate in 2019. Including cases who have not responded in the most recent SDR cycle allowed us to gather information from respondents who may have a different perception of seeing their response data from 3 or more years ago displayed in the instrument they are asked to complete.

Cases were randomly assigned to one of three instrument design conditions:

- an abbreviated version of the usual SDR questionnaire (IND, or independent measure condition);
- a dependent interview questionnaire that uses a single stage task to collect updated information (DI-1 condition), covering the same content as the independent measures; or
- a dependent interview questionnaire that uses a two-part question to collect updated information (DI-2 condition), also covering the same content as the independent measures.

All treatments used a web survey instrument for data collection. The DIS Pilot did not include CATI or paper data collection.

Dependent interviewing is a technique increasingly used in panel surveys, in which substantive answers from prior cycles are “fed forward and used to tailor the wording and routing of questions or to include in interview edit checks” (Jackle, 2006). Dependent interviewing can reduce measurement error by reducing repetitiveness and burden, aiding respondent recall, reducing spurious change, and generally providing respondents with a sense of continuity over the course of their participation in the panel (Pascale and Mayer, 2004). Several major household surveys in the U.S. use a dependent interviewing approach,

including for the household rostering process (Panel Survey of Income Dynamics, National Longitudinal Survey of Youth), for tracking health conditions and health care providers (Medical Expenditure Panel Survey, Medicare Current Beneficiaries Survey), and to measure changes in labor force participation (Survey of Income and Program Participation, Panel Survey of Income Dynamics).

Much of the research on dependent interviewing has focused on testing “proactive” dependent interviewing (PDI) versus “reactive” dependent interviewing (RDI) (Sala and Lynn, 2004; Lynn et al, 2005; Hoogendoorn, 2004; Lugtig and Lensvelt-Mulders, 2014; Al Baghal, 2017). In RDI, respondents are asked the question, and data from previous surveys are used as edit checks. In PDI, respondents are provided with the answers to the prior wave within the survey question, and are asked whether this information is still correct. Several studies have used PDI for recall items, with results suggesting that PDI increases data quality by reducing the spurious change frequently found in panel surveys (Hoogendoorn 2004; Jackle 2009; Lynn and Sala 2006).

Nearly all of these studies are based on computer administered interviews with an interviewer asking the questions. The PDI approach typically uses two steps to ask the question, with the interviewer presenting the prior wave’s data and asking if this data is “still correct” (or some variation of this language). If the information is not still correct, they are asked to provide updated information. Only a small handful of studies, all conducted in Europe, have explored the use of DI in a web-based self-administered survey (Hoogendoorn, 2004; Lugtig and Lensvelt-Mulders, 2014; al Baghal, 2017). Hoogendoorn (2004) seems to be the only study that tested a one-stage approach to DI, in which web respondents are shown their prior wave responses and can edit the data on that screen, without having to answer yes/no if the information is still correct. NCSES agreed that this approach merits additional research, especially as household respondents have become accustomed to filling out electronic forms in which they can simply update prior information, rather than having to be asked if each data element is still correct. This was the impetus for the experiment embedded in the DIS Pilot test, namely, to test a one-stage DI approach versus a two-stage approach vs. no DI, or independent measures. The results of this experiment provide guidance about whether to implement a DI instrument in the 2021 SDR, how to structure DI in a self-administered survey, and add to the small body of literature on use of DI methods in self-administered forms.

Purpose of research

Because the literature available focuses primarily on interviewer-administered surveys, there is not an established best practice for applying dependent interviewing to a self-administered survey. In the 2017 SDR data collection, 84% of the completed surveys came from web self-administered questionnaires, and in the 2019 data collection, 93% of the surveys were completed via the web questionnaire. With an increasing proportion of the sample completing via a self-administered web instrument, this experiment was designed to test two possible dependent interviewing approaches for the web instrument that can be compared to the usual repeated measures version of the questionnaire, as well as to each other.

In addition to the IND, DI-1, or DI-2 survey, all respondents were asked to complete a response analysis survey (RAS). The RAS was designed to collect data on respondents' experience and reactions to the version of the Pilot survey they completed. Including the subjective measures regarding respondent perceptions is important to consider in deciding the approach for 2021 SDR given the need to maintain the cooperation of sampled members across many cycles of data collection.

The DIS Pilot Study addresses the following three research questions:

1. Does dependent interviewing reduce the time to administer the SDR as compared to the standard repeated measures approach currently used in the survey?
2. Does dependent interviewing affect response quality (e.g., item nonresponse) and the measurement of employment changes relative to independent measures?
3. What do respondents think of the experience of responding to a pre-filled web-based questionnaire?

The Pilot also evaluates whether the two dependent interviewing approaches differ in their results to these same research questions.

1. Is there a difference in administration time between DI-1 and DI-2 instrument designs? How does the administration time for each compare to the time to administer the independent measures approach currently used in the survey?
2. Is there a difference in response quality (e.g., item nonresponse) and measurement of employment changes between DI-1 and DI-2 instrument designs? How does the response quality (e.g., item nonresponse) and the measurement of employment changes for each DI approach compare to independent measures?
3. Are there differences in how DI-1 and DI-2 respondents think of their experience in responding to the pre-filled web-based questionnaire? How do the perceptions of each DI approach compare to independent measures?

Participants

3,900 cases were selected from the 2019 SDR production sample living within the US, and all were invited to participate in the Pilot test. Since the pilot content focuses on employment status and occupation, the Pilot only selected from sample members who reported working in their last cycle of participation. By definition, all cases eligible for the Pilot were included in the 2015 sample expansion. Of the total 3,900 sampled cases, 3,600 cases were selected from sample members who completed the 2019 survey and an additional 300 cases that did not participate in 2019 but last participated in 2015 or 2017.

Respondents who provided critical items only (CIO) were not be eligible for selection for the Pilot since the two dependent interview approaches required items beyond the CIO items to prefill the questionnaire. Including only non-CIO respondents helped to minimize differences between sample members assigned to the three treatment groups.

Table 1 below shows the selection from the past three SDR cycles.

Table 1. DIS Pilot sample allocation by cohort and year of last response

	Last responded: 2015	Last responded: 2017	Last responded: 2019
2015 cohort	100	100	3,600
2017 new cohort		100	
2019 new cohort & supplemental			

The 3,600 cases were selected among the 2019 eligible respondents after stratifying the cases into 8 cells defined by gender, race/ethnicity (minority, other) and cohort (new 2019 cohort, other). The sample allocation was finalized after considering the numbers available in each cell. A modified version of proportional allocation was used, so that cells in which the proportional allocation did not provide sufficient sample size were adjusted. Within each cell, cases were sorted by field of degree prior to drawing the sample by systematic sampling.

2. Data Collection Methodology

The field period for the DIS Pilot study began on October 13, 2020 corresponding with the survey invitation mailing, and closed on November 23, 2020. All data were collected via an online web survey. The web instrument was programmed to be a mobile aware survey that rendered in a user-friendly format on mobile devices

Of 3,900 sampled cases, 26 were dropped as ineligible prior to the start of data collection (discussed in section 3a.) The remaining 3,874 sample members received up to six communications requesting their participation in the DIS Pilot, sent via mail, email or both depending on the contact information available for each case. All sample cases followed the same contact strategy, regardless of the assigned instrument version. The content of these contacts can be found in Appendix A. Table 2 shows the DIS Pilot contact strategy.

Table 2: Sequence of DIS Pilot contacts by mode and timeline

Contact	Mode	Schedule	Who received?
Invitation letter	Mail	10/13/2020	All sample members with a valid mailing address
Invitation email	Email	10/16/2020	All sample members with a valid email address
Reminder postcard	Mail	10/16/2020	Sample members who have not yet responded, and for whom there is a valid mailing address (e.g., will exclude Postal Non-Deliverable, PND)
Reminder email	Email	10/21/2020	Sample members who have not yet responded, and for whom there is a valid email address
Non-response follow-up letter	Mail	11/4/2020	Sample members who did not respond to any prior contact, and for whom there is a valid mailing address; FedEx was used for 2019 nonrespondents and USPS for all others
Non-response follow-up email	Email	11/9/2020	Sample members who did not respond to any prior contact, and for whom there is a valid email address

All sampled cases received a preloaded \$30 debit card in the initial survey invitation mailing as a token of appreciation. Subsequent communications noted the prior mailing of the debit card. The language for all communications are included in Appendix A.

As previously noted, participants were randomly assigned to one of the following experimental conditions:

- Independent measures questionnaire (IND)
- One-stage dependent interviewing questionnaire (DI-1)
- Two-stage dependent interviewing questionnaire (DI-2)

Each of the three instruments followed the introduction screens used in prior SDR cycles, starting with informed consent information and general navigation instructions. Continuing with the flow from prior SDR cycles, each instrument then asked the Sample Person Verification (SPV) items that ask about field and year of degree, as well as the doctoral awarding institution. Responses to these items allowed the SDR to verify that the respondent is the intended sampled person. The SPV portion of the instrument has historically used a two-stage dependent interview approach, prefilling data from the Doctorate Record File (DRF). The Pilot used the exact same SPV questions as asked in the production SDR survey for the IND and DI-2 version of the instrument. The DI-1 version of the instrument maintained the same content, but presented the questions in the one-stage format to keep internal consistency with the presentation of pre-filled information within the DI-1 instrument.

The look and feel of the instrument and methods of navigation matched that used in the 2019 data collection and was the same across the instruments. All three instruments were designed to render appropriately for mobile data collection.

Upon completion of the IND, DI-1, or DI-2 portion of the survey, the application automatically routed respondents to the Response Analysis Survey (RAS). Respondents

saw a transition statement at the start of the RAS so it was clear that this portion of the instrument was collecting their feedback on the test version of the SDR they just completed.

The instrument specifications for the IND, DI-1 and DI-2 instruments, including the introductory screens and Sample Person Verification (SPV) series, are included in Appendices B, C, and D respectively. The specifications for the RAS that followed each survey are shown in Appendix E.

Westat operated a toll-free help line and an email box for respondents to contact with questions or concerns throughout the data collection cycle.

3. Results

3a. Methodological findings

Final dispositions by condition

The SDR Pilot Study used a subset of 2019 SDR survey cycle final case outcome dispositions and added 3 additional codes meaningful for pilot analysis (10a, 10b, 10c). Table 1 presents response rate categories mapped to final case outcome dispositions.

Table 3a-1. Response Rate Category Mapped to Final Case Outcome Disposition

RR Calculation Category and Final Case Outcome	Total	DI1	DI2	IND
Eligible Respondents (R)	2,574	832	863	879
Partial - Main survey complete, RAS incomplete (10b)	13	2	5	6
Partial - Main survey complete, RAS complete, CI incomplete (10c)	16	4	6	6
Complete - Web (52)	2,545	826	852	867
Eligible Nonrespondents (NR)	1,282	446	424	412
Partial - Main survey incomplete (10a)	55	22	16	17
Hard refusal (92)	5	2	2	1
Non-response other (99)	1,222	422	406	394
Ineligible Respondents (IE)	27	9	10	8
Deceased (66)	1	0	1	0
Out of Scope, other (73)	26	9	9	8
Unknown Eligibility Nonrespondents (UE)	17	13	3	1
SPV failure (87)	17	13	3	1
Overall	3,900	1,300	1,300	1,300

SPV failures/ineligible (disposition = 87)

Similar to prior SDR cycles, the SDR Pilot Study included the Sample Person Verification (SPV) module that uses preloaded doctorate institution, field of study, and year earned information to verify that the person is the intended sampled member. Respondents who confirmed at least 2 of the 3 doctorate information were considered successfully located sample members and were directed to complete the rest of the survey. Otherwise, respondents who did not confirm at least 2 of the 3 doctorate information (item nonresponse or failed SPV) were considered ineligible and exited the survey. Unlike in the main SDR cycles, there was no SPV manual review protocol conducted to determine whether the intended sample member was reached using other survey data. This abbreviated SPV assessment protocol was implemented due to the shortened data collection period and due to the contents of the Pilot survey which did not collect all necessary information critical to the SPV manual review process as was done in the main survey.

Below is a summary of the 17 SPV Ineligible cases:

- 12 cases (mixed conditions) did not confirm the preloaded doctorate information but made only minor corrections to their doctorate information (institution, field of study, and year) that we are confident the intended respondent was reached.
- 4 cases (all DI1 condition) did not provide responses under any of the 3 doctorate confirmation questions and also did not provide corrections when prompted.
- The remaining case (DI1 condition) did not provide responses under 2 of the 3 confirmation questions and did not provide corrections when prompted. This case did not confirm the preloaded doctorate year and made a correction when prompted.

Completes (disposition = 52) vs. partials (disposition = 10a, 10b, 10c)

For the Pilot Study, Completes are defined as any case where the respondent passed the SPV criteria as described above, completed the entire survey from start to finish and submitted the survey submit button at the end of the survey. Partial completes are any case where the respondent broke-off at any point during the survey and did not return to complete the survey. Partial completes have the following subcategories:

- 10a – cases where the respondent broke off at any point in the main portion of the survey. These include both break-offs early in the SPV module where the SPV had not yet be assessed and cases where the respondent passed the SPV criteria but broke off at some point during the main survey
- 10b – cases where the respondent passed the SPV criteria, completed the main survey, but did not start or broke off at some point in the RAS portion of the survey.

- 10c – cases where the respondent passed the SPV criteria, completed the main survey and the RAS, but broke off at some point in the Contact Information section.

Other dispositions (refusals, unavailable during survey period, etc.)

Summaries of the remaining disposition statuses are below.

- 66 – Deceased
- 73 – Out of Scope. These records were flagged as ineligible for the following reasons prior to the start of Pilot Study data collection and were excluded from the mailing protocol:
 - Incorrectly sampled cases
 - NSF employee
 - International mailing address
- 92 – Hard refusal
- 99 – Non -response other. These records are cases where a mailing was sent but we did not receive a response.

Response rates by condition

Sample members were randomly assigned to one of three conditions as part of the sample selection process. Condition 1 cases were assigned to the DI1 interview, condition 2 cases were assigned to the DI2 interview, and condition 3 cases were assigned to the IND interview. While it was possible for a sample member to be shown a different version of an individual question from the one to which they were assigned, the dispositions presented in table 2 were tracked based on the initially assigned conditions.

Three sets of response rates are shown: for full completes, for completed surveys with no contact information, and for substantive completes. All three sets of response rates range from 64 to 68 percent overall and across conditions. Even though there is a nominal increase in response rates between DI-1 and IND for all three sets, p-values from pairwise comparisons only show statistical significance for a few of them. At an alpha= 0.1 level of significance, there are significant differences for all response rates between DI-1 and IND. At an alpha= 0.05 level of significance, the only difference that approaches significance is between DI-1 and IND, for substantive completes.

Table 3a-2. Response rates overall and by condition: SDR Pilot Study

Response rates overall and by condition	Response rate (%)	DI1 vs DI2 p-value	DI1 vs IND p-value	DI2 vs IND p-value
Full completes		0.27	0.09	0.57
Overall	65.7			
DI-1	64.0			
DI-2	66.0			
IND	67.1			
Completes, no contact info		0.24	0.08	0.57
Overall	66.1			
DI-1	64.3			
DI-2	66.5			
IND	67.6			
Substantive completes		0.19	0.05	0.54
Overall	66.5			
DI-1	64.4			
DI-2	66.9			
IND	68.0			

Demographics of respondents by condition

Table A1 in Appendix F shows that respondent characteristics including year of graduation, sex, race/ethnicity, age, citizenship, and SDR response year, are comparable across conditions. The majority of respondents obtained their doctorate degree after the year 2000 and are less than 50 years old. The gender distribution is balanced. As expected, most of the respondents in all conditions participated in the 2019 SDR.

Editing and coding by condition

For the 2020 SDR Pilot, the edits implemented to prepare the data for analysis and coding required some modifications. Editing conducted in the Pilot was focused on preparing the data for occupation coding. Because of this, edits which were unrelated to variables used in the occupation coding process were omitted. Additionally, imputation was not conducted in the pilot, eliminating the need for edits that set variables for imputation, and because the web was the only mode of data collection, edits were modified to only include web allowed values. Finally, the Pilot survey contained several new coronavirus specific variables for which data cleaning edits were modified or created.

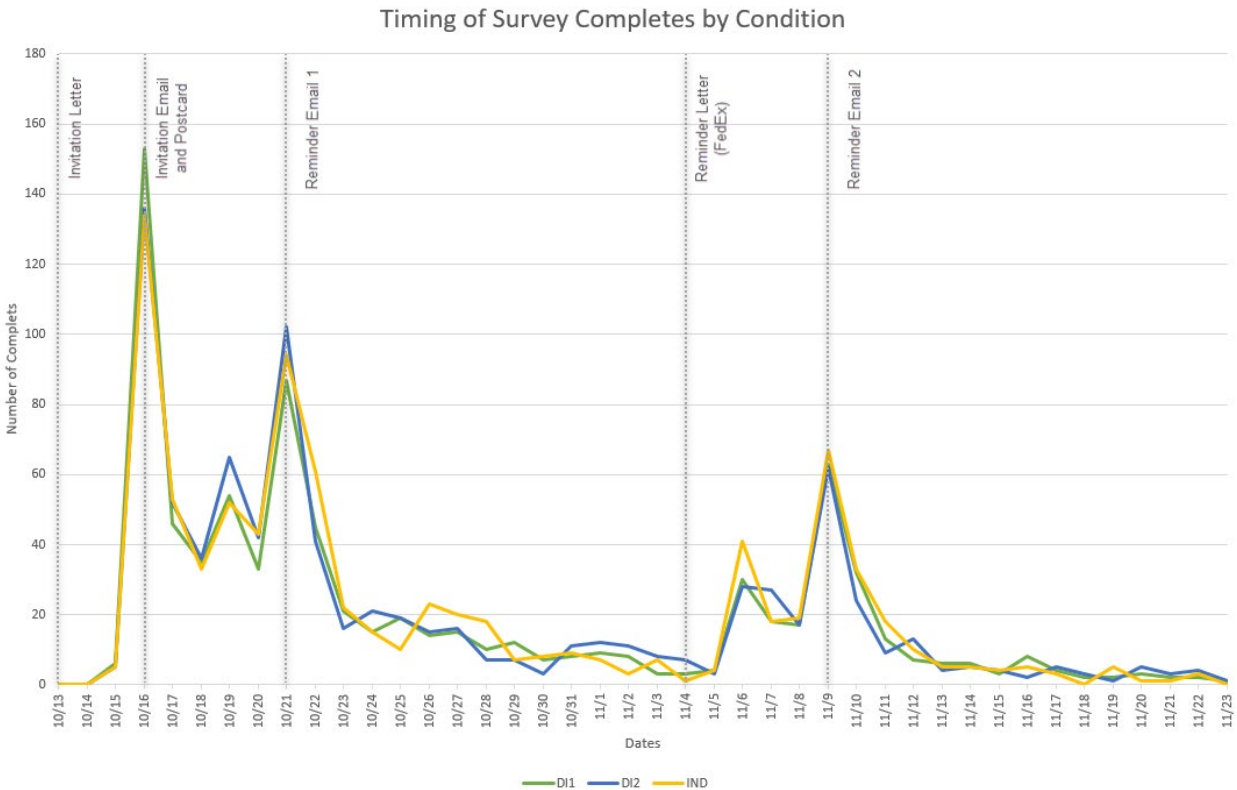
Pilot 2020 coding was limited to only include occupation coding. Institution, other specify, and field of study coding were not performed. For occupation coding, the same coding rules and guidelines were followed.

Timing of response by condition

The chart below shows the count of returns by day, per condition, with the date of each contact identified with a vertical line. The timing of completed surveys relative to the SDR contact is virtually the same across the three conditions. For each condition, the largest counts of returns occurred on the days emails with the link to the survey went out to sample members. The invitation email went out four days after the invitation mailing, and Westat received the largest number of returns that day, for each condition (153 DI-1, 136 DI-2, and 134 IND). The next highest count of returns on a single day corresponded with the reminder email sent on day nine of the contact protocol, for each condition (87 DI-1, 102 DI-2, and 94 IND).

Within three days of both of the first class mailings, for each condition, we observed an increase in the number of completed surveys, but not quite as large as the jump in completes observed on the days when emails were sent. For all conditions, just over half of the final count of completed surveys were received within a day of the reminder email, or nine days from the invitation letter mailing and just about a week from the invitation email. The remaining half of completed surveys came in over the next month.

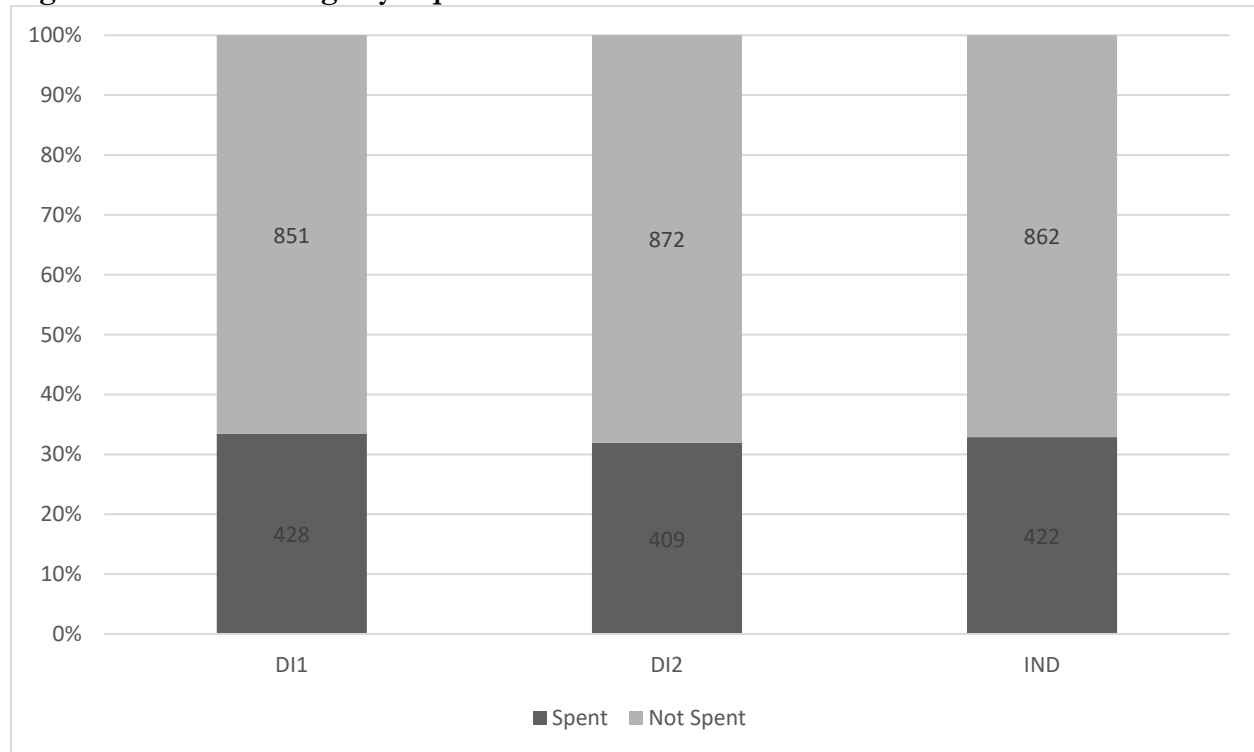
Figure 1. Timing of Survey Completes by Condition



Incentive use by condition

A total of 3,844 survey members with a valid mailing address received a \$30 Visa prepaid card in their invitation letter. Based on the data as of December 21, of the 3,844 cards sent with the invitation letter, 1,259 (33%) people used the card and 2,585 (67%) had not used the card. The roughly 30/70 split was shown across all three conditions, as shown in Figure 1 below.

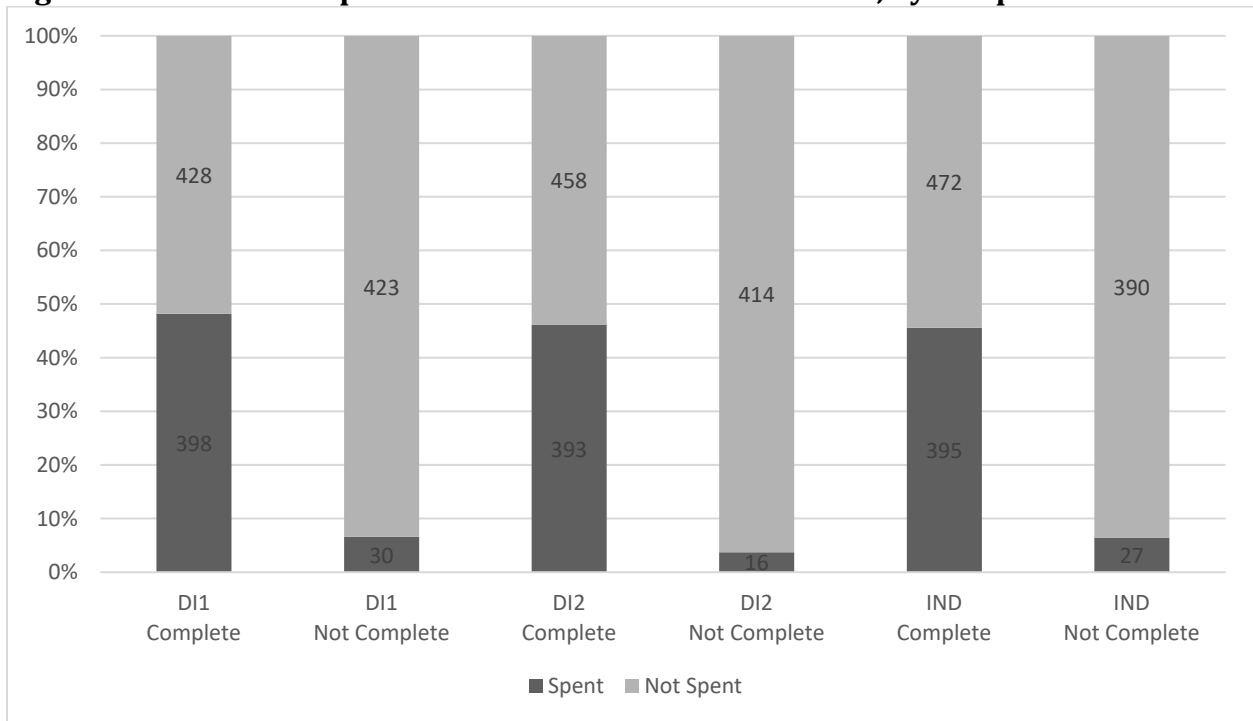
Figure 2. Incentive usage by experimental condition



Of the 3,844 sample members who were mailed an incentive card, 2,544 ¹(67.4%) completed the survey, (826 DI-1, 851 DI-2, and 867 IND). Of the 2,544 who completed the survey and received a prepaid card, only 1,186 (46.6%) used the card of which 398 DI1, 393 DI2, and 395 IND (Of the 3,844 selected survey members who were mailed an incentive card, 1,300 (33.8%) did not complete the survey, of which 453 DI-1, 430 DI-2, and 417 IND. Of the 1,300 who did not complete the survey, only 73 (5.6%) used the card, of which 30 DI-1, 16 DI-2, and 27 IND .

¹ One of the 2,545 completes did not have a valid address and was not mailed an incentive. The sample member completed the survey via the link sent with an email survey invitation.

Figure 3. Count of sample members who used the incentive, by completion status



Note that 54 respondents contacted the help desk and requested a re-mail of the card. The case management system counts the 54 re-mailed cards as part of the 3,844 cards initially mailed. Westat transferred the funds from the initial card to the re-mailed card so the funds could not be spent twice. However, those 54 cards show in the system as used because of the transfer of funds to the replacement card.

3b. Response Distributions to Key Variables

This section presents the distribution of responses on key variables in the DIS Pilot Test instrument, showing very few differences by condition. Detailed tables with statistical testing are presented in Appendix F, and a summary of the results is presented in Table 3b.

Table 3b-1. Summary of statistical testing of differences in key variables by condition

Table #	Description	DI-1 vs. DI-2	DI-1 vs. IND	DI-2 vs. IND
23	Employment status	NS	NS	NS
24	Employer type			
	In a non-incorporated business, professional practice, or farm	NS	NS	NS
	In an incorporated business, professional practice, or farm	NS	NS	NS
	In a for-profit company or organization	0.0242	0.0278	NS
	In a non-profit organization (including tax-exempt and charitable organizations)	NS	NS	NS
	In a local government in the U.S. (e.g., city, county, school district)	NS	NS	NS
	In a U.S. state government (including U.S. state colleges/universities)	NS	NS	NS
	In the U.S. military service, active duty or Commissioned Corps (e.g., USPHS, NOAA)	NS	NS	0.0079
	In the U.S. federal government (e.g., civilian employee)	NS	NS	NS
	In a non-U.S. government (at any level)	0.0033	0.0017	NS
	Other	NS	NS	NS
25	Employer size	NS	NS	NS
26	Educational institution	NS	NS	NS
27	Type of educational institution			
	Preschool, elementary, middle, or secondary school or system	NS	NS	NS
	Two-year college, community college, or technical institute	NS	0.0476	NS
	Four-year college or university, other than a medical school	NS	0.0186	NS
	Medical school (including university-affiliated hospital or medical center)	NS	NS	NS
	University-affiliated research institute	NS	NS	NS
	Other	NS	NS	NS
28	Faculty rank			
	Not applicable; no ranks designated at this institution	NS	NS	NS
	Not applicable; no ranks designated for my position	NS	NS	NS
	Professor	NS	NS	NS
	Associate Professor	NS	NS	NS
	Assistant Professor	0.0204	NS	NS
	Instructor	NS	NS	NS
	Lecturer	NS	NS	NS
	Other	NS	NS	NS
29	Supervision of others	NS	NS	NS
30	Work activities:	NS	NS	NS
	Accounting, finance, contracts	NS	NS	NS

Table #	Description	DI-1 vs. DI-2	DI-1 vs. IND	DI-2 vs. IND
	Basic research - study directed toward gaining scientific knowledge primarily for its own sake	0.0038	NS	NS
	Applied research - study directed toward gaining scientific knowledge to meet a recognized need	0.0369	NS	NS
	Development - using knowledge gained from research for the production of materials, devices	NS	NS	NS
	Design of equipment, processes, structures, models	NS	NS	NS
	Computer programming, systems or applications development	NS	NS	NS
	Human resources - including recruiting, personnel development, training	NS	NS	NS
	Managing or supervising people or projects	NS	NS	NS
	Production, operations, maintenance (e.g., chip production, operating lab equipment)	NS	NS	NS
	Professional services (e.g., health care, counseling, financial services, legal services)	NS	NS	NS
	Sales, purchasing, marketing, customer service, public relations	NS	0.0279	NS
	Quality or productivity management	NS	NS	NS
	Teaching	NS	NS	NS
	Other activity	NS	NS	0.0220

3b1. Current employment status (Table 23)

All respondents sampled for the DIS Pilot Test were working at the time of their prior cycle response. We did not expect to see any differences in employment status by condition, since condition was randomly assigned. Indeed, there are no statistically significant differences.

3b2. Employer type (Table 24)

Respondents in DI-1 were significantly less likely than DI-2 or IND respondents to say they worked at a for-profit company or organization, and were more likely to say they worked for a non-US government entity ($p < 0.05$). IND respondents were significantly more likely to say they worked in the U.S. military ($p < 0.05$).

3b3. Size of employer (Table 25)

When asked to report the number of people who work for their primary employer, no significant differences were detected by interviewing condition, with more than six in ten working for employers with more than 1,000 employees (Table 25).

3b4. Educational institutions (Table 26)

When asked whether or not their primary employer is an educational institution, no significant differences were detected by interviewing condition, with approximately 45 percent in each condition working for an educational institution (Table 26).

3b5. Type of educational institution (Table 27)

Among those who work for an educational institution, there were no differences in the type of educational institution they work for across the two dependent interviewing conditions, however, DI-1 respondents were significantly more likely than IND respondents to be working at two-year institutions or four-year colleges and universities (Table 27).

3b6. Faculty rank (Table 28)

The distribution of faculty ranks was almost entirely equivalent across interviewing condition, with roughly two-thirds having a rank of professor, associate professor, or assistant professor. DI-1 respondents were significantly less likely to report a rank of Assistant Professor than DI-2 respondents (Table 28).

3b7. Supervision of others (Table 29)

Just under half of all respondents across the interviewing conditions have supervisory responsibilities, with no significant differences detected by condition (Table 29).

3b8. Job activities (Table 30)

Respondents were presented with a list of 14 job activities and were asked to report which ones they spent at least 10 percent of their time performing during a typical week at their job. A few significant differences emerged, with those in the DI-1 condition significantly less likely than those in DI-2 to report doing basic research or applied research (Table 29). Those assigned to DI-1 were significantly more likely to report being a sales, purchasing, marketing, customer service, or public relations role than those in the IND condition. Those in the IND condition were significantly more likely to report doing some “other” activity than those assigned to the DI-2 condition.

3b9. Comparison of employer and job to prior cycle reference date, EMSMI (Table 31)

Respondents were asked if they were working for the same employer and in the same type of job during the pilot test reference period as they were during the prior cycle reference period. The response distribution was similar across the three conditions, with roughly

eight in ten saying they had the same employer and same type of job as previously reported.

3c. Key findings from analytic tables

The DIS Pilot Study aimed to answer a number of research questions concerning the impact of implementing a dependent interviewing approach on response burden, response quality, and respondent reactions to being shown their pre-filled answers. This section presents the findings on response burden and response quality. All detailed tables appear in Appendix F, and a summary of tables with significance testing is presented in Table 3c.

Table 3c-1. Summary of statistical testing from analytic tables by condition

Table #	Description	DI-1 vs. DI-2	DI-1 vs. IND	DI-2 vs. IND
1	Average survey length	0.0414	NS	0.0002
1a	Average survey length among employed respondents making a change to data	NS		
2	Any change in employer	NS		
	Any change in job title	NS		
	Any change in job duties	<.0001		
2a	Distribution of having same employer/same job when occupation code changed	NS	0.0098	<.0001
	Distribution of having same employer/same job when occupation code did not change	0.0175	NS	0.0233
2b	Average number of characters in verbatim responses for job title and job duties among those typing in a response	NS		
2c	Count of change flags			
	0	<.0001	<.0001	<.0001
	1-3	<.0001	<.0001	<.0001
	4-6	0.0393	NS	0.0204
	7-9	NS	<.0001	<.0001
	10-14	0.0003	<.0001	<.0001
	15-19	0.0236	<.0001	<.0001
	20 or more	NS	<.0001	<.0001
2d	Frequency making a change, by type of item			
	SPV items	<.0001	<.0001	NS
	Text response items	<.0001	<.0001	<.0001
	Single response items	0.0186	<.0001	<.0001
	Grid item	<.0001	<.0001	<.0001
4c	Frequency of moving backwards in the instrument			
	0	NS	NS	NS
	1-3	NS	NS	NS
	4-6	NS	NS	NS
	7-9	NS	NS	NS
	10-14	NS	NS	NS
	15-19	0.0499	NS	NS
	20 or more	NS	NS	NS
6a	Rates of change in employment status from previous SDR response	NS	NS	NS

6b	Rates of change in employer name from previous SDR response	NS	<.0001	<.0001
6c	Rates of change in principle job title from previous SDR response	--	--	--
6d	Rates of occupation change from previous SDR response	NS	<.0001	<.0001

3c1. Response burden (Tables 1 and 1a)

As shown in Table 1, the average length of the DIS Pilot Study was 14.70 minutes for DI-1, 14.05 minutes for DI-2 and 15.25 minutes for IND. These are all statistically significantly different from each other. The bulk of this burden was the time to complete the survey, with roughly the same amount of time spent for DI-1 (8.08 minutes) and DI-2 (7.76 minutes), and slightly more time spent on IND (9.77 minutes). As anticipated, the RAS took less than 5 minutes to complete, averaging 4.35 minutes for DI-1, 4.12 minutes for DI-2 and 3.35 minutes for IND. As expected, average timings were longer for respondents who are currently employed than for those not currently working.

Table 1a breaks out burden levels by whether or not any changes were made to prior cycle responses. One would expect burden to increase for respondents who are making updates to their prior cycle responses, and one might also expect that it would take slightly longer to make changes in a two-stage format than in a one-stage format. Indeed, the average timing was 2.55 minutes longer for DI-1 for those who made a change compared to those who did not; the average timing was 3.07 minutes longer for DI-2 respondents who made changes. Among those making changes, there were no significant burden differences between DI-1 and DI-2.

Perceived burden results are presented in Table 19 and discussed in section 3d13.

3c2. Rate of making changes to key open-ended employment items (Table 2)

Table 2 shows rates of entering edits to text-field employment variables by the version of the instrument respondents were exposed to for that measure. Note that even if respondents were assigned to a DI condition, if there was no prior cycle response for that measure, or if they indicated they had changed jobs (after reporting employer information), they were directed to the IND version of the question.

Editing rates are statistically equivalent across the two DI conditions. However, DI-1 is generating significantly more changes in job duties. This may be due to the ease of making edits on the DI-1 screen, whereas in DI-2, the sample member must first indicate that the information is no longer correct, and then enter the response on a blank screen, which may be perceived as more burdensome.

All edits are counted as changes in this table, even if respondents changed a character or added a word to their response. (See section 3e for a discussion of meaningful change.)

3c3. Consistency between reported job change and occupation coding change (Table 2a)

Table 2a examines the level of consistency between the EMSMI variable (asking if the respondent has the same employer and type of job as in the prior cycle) and whether a change in OCC (occupation) code was measured. One would expect that respondents with a change in occupation code would have been more likely to report in EMSMI that their job had changed.

The table suggests a relatively low level of consistency between the two variables. For DI-1, 32 percent of those who had a measured OCC code change reported that they had a different job in EMSMI, compared to 38 percent in DI-2 and 20 percent in IND. DI-1 and DI-2 results are statistically comparable to each other, but both are significantly different from IND.

Conversely, when OCC code remained consistent, respondents were much more likely to say that their job had not changed (91% DI-1, 95% DI-2, 93% IND). DI-1 results are significantly different from DI-2, and DI-2 are significantly different from IND.

These results suggest that there may be significant measurement error in the way respondents are interpreting the concept of a “different job” in EMSMI.

3c4. Level of Effort in Entering Verbatim Text Responses (Table 2b)

Table 2b compares the number of characters provided in verbatim text responses for job title and job duties. While there were no differences between DI-1, DI-2 and IND, those in the DI-1 condition provided marginally significantly more information for their job duties than DI-2 (p-value=0.0885). As evidenced in Table 35 (section 3e4), those in the DI-1 condition tended to provide more meaningful edits to their job duties than those in the IND condition.

3c5. Overall Changes Made Throughout the Instrument (Table 2c)

Table 2c summarizes the rate at which respondents changed any of their data between their prior cycle response and the Pilot Test on the items that used dependent interviewing. “Changes” include modifying a response to a closed-ended question, or changing even a single character in an open-text field. Notably, those in the IND condition were not shown their prior cycle responses, so one would expect more changes to be made to their data in the absence of this information. Indeed, 97 percent of IND respondents changed their answer to at least one variable in the study, compared to roughly 68 percent of DI-1 respondents and 52 percent of DI-2 respondents. While DI-2 respondents were the least likely to make at least one change, the average number of changes made was not meaningfully different between DI-1 (2.00) and DI-2 (1.91), but both were significantly lower than IND (10.60).

3c6. Changes Made to Types of Survey Items (Table 2d)

Table 2d presents the average number of change flags that were triggered for different types of survey items throughout the instrument, including the SPV section, verbatim text response items, single response items, and grid questions. In assessing the performance of dependent interviewing, we wanted to be sure to test the approach with different types of questions to be able to assess how DI might affect survey estimates.

The SPV items reflect the respondent's first exposure to the dependent interviewing style. Note that the respondents in the IND condition and the DI-2 condition were presented with the traditional SDR presentation of the SPV items, which uses a DI-2 approach, displaying their prior cycle response and asks if the information was still correct. As expected, for the SPV items, we find essentially the same rate of change and average number of changes for DI-2 and IND. DI-1 produces a significantly higher rate of change. As discussed in section 3e, most of these changes were non-meaningful, with the bulk of the edits being respondents adding or removing details about their PhD field, rather than substantively changing the information.

For the text response items, all IND respondents (99.88%) triggered change flags for at least one item, editing an average of 4.27 items. Again, this is not unexpected, since respondents could not see their prior responses. And also, as one might expect, we see a significantly higher rate of change for DI-1 items than for DI-2 items. With an editable text response in DI-1, it is easier for respondents to make changes than in DI-2, where they first need to indicate that the information is no longer correct, and then need to type in the entire response on the subsequent screen. Since edits are so much easier to make in DI-1, we would expect to be picking up more non-meaningful change in DI-1. In fact, this is what we see, for example, with the job duties variable, with seven in ten respondents in DI-1 providing a non-meaningful change, compared to 55 percent in DI-2 (see Table 35).

For the single-response items, which include employer type, employer size, academic employer, and supervising others, we should not expect to see differences among the three conditions. However, IND respondents were about three times more likely to make changes to at least one single-response item than either DI-1 or DI-2. This could be a potential sign of confirmation bias with the DI approach, or conversely, an improvement in the quality of information collected. DI-1 respondents were also significantly more likely to make an edit to at least one of these variables than DI-2.

Finally, for the grid style of questions, we analyzed the item about work activities that presents 14 activities in a yes/no grid format, and found that IND respondents were again at least 4 times more likely to make a change to at least one of the items in the grid than were DI-1 and DI-2 respondents. As a point of comparison, between the 2017 and 2019 production SDR collections, 81 percent of respondents made a change to at least one item in this grid. In the Pilot, 93 percent made at least one change in the IND condition, compared to 21 percent in DI-1 and 10 percent in DI-2. In spite of DI-1 respondents being twice as likely to change at least one item in the grid as DI-2, the average number of

changes made within the grid was very similar across the two conditions (0.35 items changed in DI-1, 0.62 items changed in DI-2).

3c6a. Frequency of making a meaningful change to text-response items (Table 2e)

Table 2e summarizes the rate at which respondents made “meaningful” changes to any of their data between their prior cycle response and the Pilot Test on the open-text items that used dependent interviewing. A detailed discussion of meaningful change analysis is presented in section 3e below, but it is important to note that the definition of “meaningful” is inherently subjective.

Notably, those in the IND condition were not shown their prior cycle responses, so one could expect more meaningful changes to be made to their data in the absence of this information. Indeed, for employer name, job title, and job duties, those in the IND condition were significantly more likely to make meaningful edits to their prior cycle responses. It is certainly reasonable that we are seeing more meaningful changes for the IND condition because of increased noise in the IND measurements, with the respondent more likely to refer to the same thing differently after a 1-year gap. So the direction of the difference is not unexpected.

3c7. Item Nonresponse (Table 3)

Table 3 displays item nonresponse rates by the version of the question to which respondents were exposed, rather than the condition to which they were assigned. Some DI-1 and DI-2 respondents were presented with the IND version of the question when they had missing data from the prior cycle, or if they had indicated that they had changed employers since the prior cycle. Thus significance testing is not performed on this table. Item nonresponse was extremely low across all three of the questionnaire conditions. There were a handful of items where item nonresponse exceeded 3 percent, most of which were text-response items. The condition(s) in which item nonresponse was above 3 percent are mentioned in parentheses.

- Other reason for leaving principal employer (DI-1)
- Month and year of leaving principal employer (DI-1, DI-2)
- Other type of employer (DI-1)
- Other type of academic position (DI-1, DI-2)
- Other type of educational institution (IND)
- Other type of work activity (DI-1, DI-2, IND)
- Salary (DI-1, IND)
- Salary range (DI-1, DI-2, IND)
- Other reason salary impacted by COVID-19 (DI-1, DI-2)

In addition, while none of the closed-ended RAS items garnered more than 3 percent nonresponse, each of the text-response RAS items generated considerable item nonresponse in all three conditions.

3c8. Data Quality: Backing up Behaviors in DI-2 (Table 4a)

One concern in the DI-2 design is confirmation bias, namely that respondents would indicate that something had changed in the item being measured, but when asked to input the corrected information, respondents would change their mind and back up to say that nothing had changed in order to avoid the added burden of having to re-answer the question. This proved to be an extremely rare phenomenon, with only 4 percent of respondents doing this at any point during the instrument. This backing up behavior never occurred for more than 2 percent of respondents in any of the items.

Another concern about confirmation bias is that respondents might initially indicate that there was no change to the measure when perhaps there actually was. To detect this, we looked for behaviors in which respondents initially said no change, and then later backed up and modified their answer to indicate that there had been a change. This occurred for fewer than 5 percent of respondents.

3c9. Data Quality: Making a Change but Indicating “No Change” in DI-1 (Table 4b)

One concern in the DI-1 design is that participants might make a change to the prior cycle response for a measure, but also mark the box at the bottom of the screen to indicate that the information has not changed since the prior cycle year. Each time this type of error occurred, respondents were shown an error message and were asked to re-enter their information. This could be an indication of user error, or an indication that they are updating the information but that it is only editorial, and not a substantive change. Fewer than 10 percent of respondents ever engaged in this behavior, and nearly all of them only made this error one time.

3c10. Frequency of Backing Up (Table 4c)

At any point during the survey, respondents could back up in the instrument to either review or change a prior response. The distribution of backing-up behavior was very similar across the three conditions. Respondents in each condition backed up between one and two times and roughly half of the respondents in each condition backed up at least one time during the instrument.

3c11. Average length of gap between last job and reference date (Table 5)

For respondents who reported a job change or are currently not employed, Table 5 presents the average number of days between the prior job and the Pilot Test reference date of September 1, 2020. With small numbers of cases to analyze on this measure, the results were generally comparable between conditions.

3c12. Rates of change in employment status (Table 6a)

Consistent with the findings that DI-1 respondents were significantly more likely to be unemployed as of the Pilot Test reference date (Table 23), we see marginally significantly higher rates of change in employment status for DI-1 than DI-2 or IND (Table 6).

3c13. Rates of change in employer name (Table 6b)

Among those who were employed as of the reference date, we did not see any differences in rates of change of employer name for DI-1 and DI-2, but we did see roughly three times more change for those in the IND condition. As noted in the meaningful change analysis, this was heavily due to respondents entering slightly different information than they did in the prior cycle that was not deemed to be meaningful.

3c14. Rates of change in principal job title (Table 6c)

Table 6c presents the rate of change in job title, overall and by key respondent characteristics. Because some DI-1 and DI-2 respondents were exposed to the IND version of the item (if they were missing prior cycle data or had changed employers), we do not have significance testing on this table. However, one can see there is no difference between DI-1 and DI-2 rates of change in job title, but more than four times more change in the IND condition. As noted in the meaningful change analysis, this was heavily due to IND respondents entering slightly different information that was not deemed to be meaningful than they did in the prior cycle.

3c15. Rates of Occupation Code Change from 2019 (Table 6d)

Table 6d presents the rates of occupation codes changing based on responses to a collection of SDR survey items and our coding process. We see that rates of occupation code changes are consistent between DI-1 (21.19%) and DI-2 (22.89%), and are both significantly lower than IND (36.10%). As a point of comparison, 40% of working respondents in 2019 had an occupation code change from 2017, and 38% had an occupation code change between 2015 and 2017. This could be an indication that dependent interviewing is leading to an underestimate of change in occupation, or it could be that it is actually improving estimates by reducing spurious change, especially given that the two DI approaches are so similar to each other.

Additionally, table 6e shows the distribution of occupation codes by condition, as well as from the full set of 2019 working respondents. The codes have been aggregated to the broader job categories shown to respondents and used by coders in selecting a “best” code. The distribution of occupation codes within these job categories for each of the pilot conditions does not meaningfully differ from the 2019 unweighted distribution. The similarity of the distributions suggests that the occupation code changes that occurred between 2019 and the pilot in the IND version and similarly between the 2017 and 2019 SDR production cycles may reflect only changes at the detailed code levels, and do not affect the higher level job categories. The lower rate of change in occupation code in the DI versions may just reflect a reduction in the changes that occur at the more detailed level. If the more detailed occupation codes are not the primary level for analysis, the statistical difference in occupation code changes between the DI methods and IND may not be meaningful.

3c16. Breakoff Rates (Table 21)

Table 21 presents the rates of breaking off the survey before completion. As shown in the table, there was only one breakoff, occurring in the IND condition at the job title question. There were a small handful of breakoffs in the RAS in each condition, and also a few during the contact information module.

3d. Response Analysis Survey Findings

In addition to the main questionnaire items, respondents were asked to complete a response analysis survey (RAS). The RAS collected data on respondents' experience and reactions to the questionnaire version they completed. Including the subjective measures regarding respondent perceptions is important to consider in deciding the approach for 2021 SDR given the need to maintain the cooperation of sampled members across many cycles of data collection. Table 3d below shows the high-level summary of findings from the RAS responses. In general, respondents in the DI-1 and DI-2 conditions reported similar responses to the RAS.

Table 3d-1. Summary of statistical testing from RAS by condition

Table #	Description	DI-1 vs. DI-2	DI-1 vs. IND	DI-2 vs. IND
7	Perceived length	NS	<.0001	<.0001
8	Perceived similarity to other SDR surveys	NS	<.0001	0.0019
9	Level of enjoyment	NS	NS	NS
10	Perceived sensitivity	0.0026	0.6097	0.0015
11	Level of confidence in protection of data	NS	NS	NS
12	Recall of completing prior cycle SDR	NS		
13	Recall of prior cycle SDR answers	NS		
14	Prior experience with dependent interviewing	NS		
15	Reactions to dependent interviewing:			
	Surprised	NS	<.0001	<.0001
	Confused	0.0041	<.0001	<.0001
	Appreciative	NS	<.0001	<.0001
	Comfortable	NS	<.0001	<.0001
	Annoyed	NS	<.0001	<.0001
	Concerned	NS	<.0001	<.0001
	Relieved	NS	<.0001	<.0001
16	Impact of dependent interviewing on accuracy	NS	<.0001	<.0001
17	Acknowledgement of confirmation bias			
	Yes for one question	0.0043		
	Yes for more than one question	NS		
	No	NS		
18	Additional reports of confirmation bias			
	Yes for one question	NS		
	Yes for more than one question	0.0419		
	No	NS		
19	Perceived burden with dependent interviewing	NS	<.0001	<.0001
20	Overall reaction to idea of dependent interview	NS	<.0001	<.0001

3d1. Perceived Length (Table 7)

Table 7 presents respondents' perceptions of how quickly they felt they were able to complete the survey. DI-1 and DI-2 garnered similar reactions, with more than 9 out of 10 indicating that they survey went "somewhat" or "very" fast. Those in the IND condition were significantly less likely to say the survey was somewhat or very fast (83%).

3d2. Perceived Similarity to other SDR Surveys (Table 8)

Table 8 presents respondents' perceptions of how similar or dissimilar the Pilot Study was to other SDR surveys. DI-1 and DI-2 garnered similar reactions, with more than 9 out of 10 indicating that they surveyed was "very" or "somewhat" similar. Those in the IND condition were significantly less likely to say the survey was very or somewhat similar, though the difference may not be seen as meaningful (91%).

3d3. Level of Enjoyment of Pilot Survey (Table 9)

Participants were asked to rate the extent to which they enjoyed completing the survey. DI-1 and DI-2 garnered similar reactions, with more than half indicating that they enjoyed the survey "a great deal" or "somewhat." Those in the IND condition provided significantly, though perhaps not meaningfully, lower levels of enjoyment (Table 9).

3d4. Perceived Sensitivity of Pilot Survey (Table 10)

Participants were asked to rate how sensitive the survey questions were. One might expect respondents in the dependent interviewing conditions to feel the questions were more sensitive, since their prior responses were being displayed on the screen. While the DI-1 and IND conditions garnered similar responses, DI-2 respondents were significantly less likely to feel the questions were sensitive than either DI-1 or IND (Table 10).

3d5. Level of Confidence in NCSSES Protecting Survey Responses (Table 11)

If the dependent interviewing approach is to be used, it is important that respondents feel confident that their data is being protected by NCSSES. Indeed, regardless of condition, roughly 89 percent of respondents across all conditions are very or somewhat confident that NCSSES will protect their answers, with more than half each condition saying they are very confident (Table 11).

3d6. Recall of Completing Prior Cycle SDR (Table 12)

In implementing a dependent interviewing approach with a survey that is only conducted biannually, it is possible that respondents could be surprised by seeing their previous responses on the screen if they do not recall completing the prior cycle's survey. In fact, more than 9 out of 10 respondents in the dependent interviewing conditions do recall completing the prior cycle of SDR (Table 12), with no significant difference between conditions.

3d7. Recall of Prior Cycle SDR Answers (Table 13)

While nearly all DI respondents recall their past participation in SDR, as shown in Table 13, they are less likely to recall their specific responses to questions. With no significant differences by DI condition, roughly 6 in 10 say they recall what their responses were.

3d8. Prior Experience with Dependent Interviewing (Table 14)

DI-1 and DI-2 respondents were asked if they had ever participated in a survey in which historical information had been pre-filled for them to confirm or update. Only 4 in ten respondents reported this experience, regardless of condition (Table 14).

3d9. Reactions to Dependent Interviewing (Table 15)

Respondents were shown a set of seven possible reactions that they may have experienced in seeing their pre-filled answers on the screen. Those in the IND condition were asked the extent to which they might have these reactions if the survey had pre-filled their answers, while DI-1 and DI-2 respondents were asked to assess the extent of their actual reactions.

As shown in Table 15, there was only one significant difference between DI-1 and DI-2 respondents in terms of their reactions to the survey, with DI-1 respondents slightly more likely to express confusion than DI-2 respondents. Among DI-1 and DI-2 respondents, roughly 9 out of ten were neither confused nor annoyed, more than 8 out of ten were not concerned, and more than half were not surprised. Roughly 7 in ten were comfortable with the approach, more than 6 in ten were appreciative, and roughly one-third were relieved.

IND respondents, however, had significantly different reactions to the idea of dependent interviewing on all measures, expressing more negativity about the approach on each measure.

3d10. Impact of Dependent Interviewing on Accuracy (Table 16)

Those in the DI-1 and DI-2 conditions were asked if pre-filling their answers helped their answers to be more or less accurate, and those in the IND condition were asked how they thought this approach might affect their accuracy (Table 16). As seen with the reactions to the survey, again there were no significant differences in the impact of DI on accuracy for DI-1 and DI-2 respondents, with roughly 6 in ten saying it made their answers more accurate, and most others saying it had no impact on their accuracy. IND respondents, however, were much more likely to say that dependent interviewing could make their answers less accurate (18%).

3d11. Acknowledgment of Confirmation Bias in Dependent Interviewing (Table 17)

Those in the DI-1 and DI-2 conditions were asked if there were any questions in the survey where they felt the prior cycle response was “accurate enough”, and rather than updating it to make it more accurate, they left the prior response as-is (Table 17). While there could be some social desirability bias in acknowledging that they did not provide fully accurate responses, more than 3 in ten respondents in each condition did admit to doing this at some point in the survey, with DI-2 respondents significantly more likely to acknowledge that this behavior occurred on one question.

Respondents were asked to explain why they did not update answers that were “close enough.” Some of those open-ended responses are provided for illustrative purposes, with key phrases provided in bold.

- my field was called Algebraic Topology, not Topology Foundations, but **I guess it's close enough.** (DI-2)
- The question regarding how many hours I work each week had the answer pre-populated. I glanced at it, felt it looked about right, and moved on. **If it had not been pre-populated, I might have given it a bit more thought.** (DI-2)
- **It was easier than making it more accurate.** (DI-2)
- The answers are accurate enough and I didn't feel a need to change. **I would only change it if the information was entirely inaccurate.** (DI-2)
- For the percentage of time that I spend on various activities, it can be hard to really calculate, as it varies from week to week and from season to season. **I didn't think that I had significantly better data, so I left the choices as they were.** But I might have done differently from scratch. (DI-1)
- I had written in 2019 in that my main job was research/program evaluation and supervising "10" staff. I only have 9 supervisees right now, but that number fluctuates over time, so I figured 10 was close enough and decided to leave the pre-populated response. **I didn't think the detail mattered.** (DI-1)
- As long as the previous info is roughly the same, I leave the answer as-is even though it may not be very accurate, since **it is more convenient.** (DI-1)
- My job description is pretty much all over the place. I wear a lot of hats. I felt like there were things I put in last year that I would have forgotten to put in this year and updating one item didn't seem necessary. **It was good enough for this survey in my mind.** (DI-1)

3d12. Additional Reports of Confirmation Bias in Dependent Interviewing (Table 18)

Those in the DI-1 and DI-2 conditions were also asked if there were any questions in the survey where they felt the prior cycle response was “wrong”, and rather than updating it to make it accurate, they left the prior response as-is (Table 18). Only about 5 percent of respondents in each condition acknowledged doing this at some point in the survey, with

DI-1 respondents significantly more likely to say it happened at more than one question (2.6% vs. 1.3%).

3d13. Perceived Burden of Dependent Interviewing (Table 19)

DI-1 and DI-2 respondents were asked if the dependent interviewing approach made the survey more or less burdensome; IND respondents were asked if they thought this approach would make the survey more or less burdensome. There were no significant differences between DI-1 and DI-2, with 87 percent of respondents in each condition saying that it made the survey much less or a little less burdensome (and roughly 63 percent saying much less burdensome). Those in the IND condition were significantly less positive about the idea, but still more than 7 in ten felt it would make the survey at least a little less burdensome.

3d14. Overall Reaction to the Idea of Dependent Interviewing (Table 20)

All respondents were asked if they thought pre-filling answers was a good or a bad idea. There were no significant differences between DI-1 and DI-2, with 90 percent of respondents in each condition saying that dependent interviewing was a very or somewhat good idea. Those in the IND condition were significantly less positive about the idea, but still two-thirds thought it was a very or somewhat good idea.

3e. Meaningful Change Analysis

For text response items in the DI-1 condition, respondents were asked to edit their prior cycle information on the screen, or to mark if the information from the prior cycle was still correct. Any edits that were made to the prior cycle information, even minor spelling corrections, were flagged as changes. Likewise for DI-2 text response items, respondents were shown their prior cycle response and were asked if the information was still correct as of the reference date. If not, they were taken to a new screen to type in the updated information. Regardless of the nature of the change, if it did not identically match the prior cycle response, it was flagged as a change. Finally, for IND, respondents were asked the survey measure with no presentation of the prior cycle response. Thus even if nothing had changed since the prior cycle, they may have entered the response slightly differently from the prior cycle, which would have triggered a change flag. To better understand the nature of changes that were made to prior cycle responses, this section explores whether the changes were determined to be meaningful or non-meaningful. Non-meaningful edits are defined as making spelling or grammatical changes, as well as edits that add or remove detail from the prior cycle response without substantially changing the answer. Examples of this are a prior cycle response for job title of “Anthropology Professor” being edited to “Anthropological Professor,” or a prior cycle response for PhD Institution of “CUNY Graduate School and University Center” being edited to “CUNY Graduate Center.”

3e1. PhD field of study (Table 32)

Very few changes were made to the PhD institution name in the SPV module. While the numbers are still relatively small, a higher number of DI-1 respondents made edits to their PhD field of study (n=85), compared to approximately 30 in each of the DI-2 and IND conditions. While there was no significant difference in rates of providing non-meaningful edits to field of study, nearly all of these edits were non-meaningful (Table 32).

3e2. Employer name (Table 33)

IND respondents were much more likely to make an edit to their employer name (N=383, compared to 110 for DI-1 and 132 for DI-2). While there were no differences in meaningful changes for DI-1 and DI-2, those in the IND condition were 4 to 5 times more likely than DI-1 and DI-2 to make a non-meaningful change to the information they had provided in the prior cycle.

3e3. Job title (Table 34)

Again, IND respondents were much more likely to make an edit to their employer name (N=582, compared to 208 for DI-1 and 221 for DI-2). While there were no differences in meaningful changes for DI-1 and DI-2, those in the IND condition were four times more likely than DI-1 and DI-2 to make a non-meaningful change to the information they had provided in the prior cycle.

3e4. Job duties (Table 35)

Nearly all IND respondents provided a different response to their job duties than they had in the prior cycle (N=835), and nearly all of those changes were determined to be non-meaningful edits (80%). While significantly less than IND, those in the DI-1 condition were significantly more likely to make non-meaningful changes to their prior cycle job duties (69%) than those in the DI-2 condition (55%).

3e5. Relationship between meaningful changes to job duties and OCC coding (Table 36)

One way of understanding the impact of the meaningful change analysis is to explore how respondents who had changes in their occupation (OCC) code handled the job duties question. We hypothesize that those with an occupation change should have been more likely to report a meaningful change in their job duties, though a number of variables go into the occupation coding scheme. This hypothesis did not play out in the data. Among those who had a change in their OCC code, most in the DI-1 (49%) and DI-2 (44%) conditions had made no changes at all to their job duties, and an additional 24 percent made non-meaningful changes. Only 26 percent of DI-1 and 30 percent of DI-2 respondents

who had an OCC code change had made a meaningful edit to their job duties, a similar proportion to those in the IND condition

Conversely, among those who did not have a change in their occupation code, we would expect few meaningful changes to their job duties. This did play out in the data, with fewer than five percent of DI-1 and DI-2 respondents making meaningful changes to job duties, and most in fact making no changes at all. Similarly, nearly all IND respondents who had no change in OCC code had made non-meaningful edits to their job duties.

4. DI Decision Criteria

4a. Should 2021 SDR Proceed with Dependent Interviewing? YES

Based on the key findings in this report, we recommend that 2021 proceed with dependent interviewing, specifically the DI-2 approach. A summary of those findings is presented below in order to support the recommendation.

4a1. Burden assessment: DI is less burdensome than IND

Overall, the DI-2 interview was significantly shorter than the IND version; the DI-1 interview was not significantly shorter than IND, and was significantly longer than DI-2 (Table 1). However, given that 97 percent of IND respondents made at least one change to their prior cycle response (Table 2c), it is also important to compare timings to those who made changes to the dependent interviewing version of the instrument. The average timing of 15.80 minutes for IND respondents is comparable to the timings of 15.81 minutes for DI-1 and 15.92 minutes for DI-2 respondents who made at least one change to their prior cycle responses (Table 1a).

With few differences in actual burden when changes are being made to prior cycle responses, it is also important to assess perceptions of burden. Regardless of a DI-1 or DI-2 approach, those assigned to dependent interviewing were significantly more likely to believe that they were able to complete the survey “very or somewhat fast” and that dependent interviewing made the survey “much less or somewhat less burdensome” than IND respondents (Table 19).

4a2. Response Rate: DI is comparable to IND

As discussed in section 3a, response rates were not significantly different between the dependent interviewing conditions and the IND condition, though DI-1 tended to have the lowest response rates.

4a3. Data Quality: DI may produce some non-meaningful confirmation bias

There are several measures of data quality that should be assessed to determine how the DI approach performed compared to IND. These include item nonresponse and breakoffs, backing up behavior, and degree to which respondents may be subject to confirmation bias.

a. Item nonresponse and breakoffs

Item nonresponse and breaking off behaviors were both extremely low across all conditions (Tables 3 and 22). These are both indicators that dependent interviewing does not differ from IND in terms of data quality.

b. Backing up behaviors

Overall, backing up behaviors were equivalent across the three conditions (Table 4c), with roughly half of respondents backing up at least once during the instrument. One concern with the DI-2 approach is that respondents might change their minds about editing their data once they realized they would have to re-enter the information. This problem did not play out, with fewer than 5 percent of respondents engaging in backing up behavior to change answers from “yes change” to “no change” at any point during the instrument.

c. Confirmation bias

One concern about dependent interviewing is that respondents might accept their prior cycle responses as “accurate enough” and, rather than taking the time to update the information to be more accurate, would accept the prior cycle response. This is a concern if the prior cycle response is, in fact, incorrect for the current cycle, which would lead to so-called confirmation bias and result in estimates of change that are too low. However, independent interviewing can lead to “anti-confirmation bias”: when the concept targeted by the survey question is difficult to define accurately (e.g. job duties), it is possible for the respondent to provide an answer that is different from that given in the previous cycle, even though the underlying situation of the respondent did not actually change. In such situations, independent interviewing might lead to estimates of change that are too high. By showing the response from the previous cycle, dependent interviewing can help reduce the likelihood of non-meaningful change occurring in the data. There are 4 ways in which we have explored the extent to which confirmation bias (or anti-confirmation bias) occurred:

- 1) Rate of non-meaningful change on open-text responses. If confirmation bias were an issue, we would expect respondents to only make meaningful changes to their employer and job information, and avoid taking the time to make minor updates. While minor updates do not substantively change the survey findings, they can help improve the overall quality of the SDR dataset. There may be some evidence of this; we find that, among those who made an edit to their employer name or job title,

fewer than 15 percent in either DI-1 or DI-2 conditions made non-meaningful updates (Tables 33 and 34). However, more than half of respondents who made edits to their job duties made non-meaningful updates (Table 35). This suggests that respondents are willing to take the time to make edits to their data, even if it is just to make minor updates.

- 2) Rate of backing up and changing answers from “yes, change” to “no change” after realizing they would have to enter information from scratch. In the DI-2 condition, it is possible that respondents could demonstrate confirmation bias by initially indicating that their prior cycle response was no longer correct, but when realizing that they would have to take the extra step of re-answering the question (on a second screen), changed their answer to say that actually, there was no change. We found very little evidence of this, as shown in Table 4a. Fewer than 5 percent of respondents engaged in this behavior at any point during the instrument.
- 3) Rate of acknowledging confirmation bias. In the RAS portion of the instrument, DI-1 and DI-2 respondents were asked to acknowledge if they ever left answers as-is, rather than changing them, either because the prior cycle response was “accurate enough.” One might expect that DI-2 respondents were more likely to have engaged in these behaviors, again, because of the added burden of clicking onto a 2nd screen. We do see a marginally, but not significantly, higher rate of engaging in this behavior among DI-2 respondents (35% acknowledged this vs. 32% for DI-1). We also asked if they ever left an inaccurate answer as-is, rather than correcting it. This behavior was far less common, with only 7% of DI-1 and 5% of DI-2 respondents acknowledging this.
- 4) Statistical modeling of likelihood to report change in key SDR variables. Unweighted logistic regression was used to assess whether question presentation (IND vs. DI-1 and DI-2) resulted in shifts in the likelihood to report a change in a set of key SDR variables in the pilot study. The five variables evaluated in this analysis are:
 - NEDTP: A10. Employer type
 - TENSTA: A18. Tenure status
 - EMED: A14. Indicator for educational institution employer
 - OCPRT: A19. Principal job title
 - EMMAIN: A11. Main business or industry

The following 18 variables were considered as possible confounders for the effect of DI-1 and DI-2:

- CONDITION: From 2019 pilot data, indicators for 3 categories (1: DI1, 2: DI2, 3: IND)
- GENDER: From 2019 pilot data, indicators for 3 categories (1: DI1, 2: DI2, 3: IND)
- NSDRMEMTOD: From 2019 RUF. Field of study for first US S&E or health PhD (major group,

- TOD), with 8 categories.
- EMSECSM: From 2019 RUF. 3-category employer sector (1: Educational Institution, 2: Government, 3: Business/Industry)
- HCAPIN: From 2019 RUF, physical disabilities indicator for 2 categories (N: No, Y: Yes)
- URM19: From 2019 paradata, Under-represented minority (URM) flag from sample frame, with 2 categories (1: URM, 2: non-URM)
- COHORT19: From 2019 paradata, 2019 SDR sample cohort indicators for 3 categories (NEW: 2016/2017 graduates from SED, SUPP: 2015 supplemental panel, CONT: 2015/2017 continuing panel)
- CURCIT19: From 2019 paradata, Sample member's current citizenship, with 2 categories (1: U.S., 2: non-U.S.)
- YEARS_SINCE_PHD: Derived from 2019 RUF, indicators for 4 categories (1: <10, 2: 10-20, 3: 20-30, 4: >30)
- AGEGROUP: Derived from 2019 RUF, indicators for 5 categories (1: <3, 2: 35-44, 3: 45-54, 4: 55-64, 5: >64)
- RACETHM_R: Derived from 2019 RUF, indicators for 3 categories (1: Asian, 2: White, 4: Other)
- EVER_LOC19: From 2019 paradata, indicators of whether sample member ever needed locating in 2019, with 2 categories (Y: Needed locating, N: Locating not needed)
- DR19_R: Derived from 2019 paradata, with 5 categories (1: 25 percentile, 2: 50 percentile, 3: 75 percentile, 4: 100 percentile, 5: 0). Quantiles of DifferentiationRatio (Mean of the ratio of the maximum number of consecutive items with the same response, of all items in the grid set across grid sets).
- SL19_R: Derived from 2019 paradata, with 5 categories (1: 25 percentile, 2: 50 percentile, 3: 75 percentile, 4: 100 percentile, 5: 0). Quantiles of Straightlining (Mean of indicators (1, if all items in the set having the same response) and (0, if at least one answer differed of all items in the grid set) across grid sets).
- EDITSORE_DATA_R: Derived from 2019 paradata, with 5 categories (1: 25 percentile, 2: 50 percentile, 3: 75 percentile, 4: 100 percentile, 5: 0). Quantiles of number of edits made at all variables
- IMPscore_data_R: Derived from 2019 paradata, with 5 categories (1: 25 percentile, 2: 50 percentile, 3: 75 percentile, 4: 100 percentile, 5: 0). Quantiles of number of imputations made at all variables
- FINAL_INCENTIVE19: From 2019 paradata, indicators of final incentive offer for 4 categories (NONE: No incentive offer, EARLY: Early offer in Starting Phase, BOTH: Both early and late offers, LATE: Late offer made in Late Stage)
- IN_RR_GROUP19: From 2019 paradata, response categories that correspond to sample member outcomes at the beginning of the Interim Phase, with 4 categories (R, IE, NR, UE)

The table below shows the estimated odds ratios for the use of DI-1 and DI-2 vs. IND, their confidence intervals and the associated p-values. As the results show, all odds ratios are less than 1 and are statistically significantly so for at least one of the two DI conditions. Hence, the evidence shows that the use of DI results in lower estimates of the occurrence of changes in these SDR variables.

RESP_VAR	EFFECT	ODDSRATIO	LOWERCL	UPPERCL	P-value
NEDTP	CONDITION (DI-1 vs IND)	0.396	0.3	0.522	0.0433
NEDTP	CONDITION (DI-2 vs IND)	0.272	0.202	0.367	<.0001
EMED	CONDITION (DI-1 vs IND)	0.415	0.237	0.728	0.0358
EMED	CONDITION (DI-2 vs IND)	0.537	0.322	0.896	0.4683
TENSTA	CONDITION (DI-1 vs IND)	0.454	0.312	0.659	0.0503
TENSTA	CONDITION (DI-2 vs IND)	0.417	0.287	0.607	0.0081
OCPRT	CONDITION (DI-1 vs IND)	0.137	0.109	0.173	<.0001
OCPRT	CONDITION (DI-2 vs IND)	0.138	0.11	0.173	<.0001
EMMAIN	CONDITION (DI-1 vs IND)	0.018	0.013	0.025	<.0001
EMMAIN	CONDITION (DI-2 vs IND)	0.018	0.013	0.025	<.0001

d. Rates of change by device type

Dependent interviewing should produce comparable rates of change regardless of what type of device is used to complete the instrument. Table 37 shows the percentage of respondents making at least one change to different types of items by device type, as well as the average number of changes made for each type of question. There were very few responses using tablets, so results focus on comparing desktop/laptop behaviors to mobile phone behaviors. Within the DI-1 condition, we see very similar response patterns for SPV items (29% change on PC vs. 26% change on mobile) and text response items (37% vs. 38%). Mobile phone respondents were slightly more likely to make edits to single response items (32%) than PC respondents (24%), and PC respondents were slightly more likely to make edits to the grid item (22%) than mobile respondents (16%), but none of these differences were statistically significant.

Slightly different patterns were detected with DI-2 respondents. Minimal differences were found between PC and mobile users on SPV items and single response items. PC users were somewhat more likely to make edits to text response items (25%) than mobile users (17%), and were also slightly more likely to make an edit to the grid item (11%) than mobile users (4%), but none of these differences were statistically significant.

These findings suggest that dependent interviewing performs similarly, regardless of device type.

4a4. Respondent Reactions to Dependent Interviewing: DI generates very positive reactions; IND respondents express some skepticism

Across the board, respondents assigned to dependent interviewing report the same or significantly more positive reactions to dependent interviewing. Regardless of DI approach, nine out of ten thought it was a very good or somewhat good idea to pre-fill answers, and 87 percent say it made the survey a little less or much less burdensome. Roughly six in ten say that DI made their answers more accurate (with only 1 percent saying it made their responses less accurate). More than half in each condition say they were not surprised by seeing their answers prefilled, nine out of ten were not confused, eight out of ten were not concerned, more than six in ten were very appreciative, and roughly seven in ten were very comfortable.

IND respondents, however, expressed more skepticism about the idea of dependent interviewing, having not experienced it during the pilot test instrument. Fewer than seven in ten say it is a very good or somewhat good idea, and roughly an equal proportion anticipate it would make the survey less burdensome. Just over one-third suspect it would make their answers more accurate (whereas nearly half anticipate it would have no impact). They were significantly less likely to express positive reactions to the idea (such as appreciation or relief), and significantly more likely to express concerns about the idea (such as surprise, confusion, annoyance, or concern).

Respondents were invited to provide open-ended comments about their reactions to dependent interviewing. Roughly one-third of DI-1 and DI-2 participants provided comments, along with roughly one-quarter of IND respondents. Westat staff coded those responses into positive, neutral, and negative sentiments. DI-1 and DI-2 respondents shared similar sentiment about dependent interviewing, with both garnering about 70% positive reactions, 5% neutral, and roughly 25% negative reactions. IND reactions were less positive, with about half expressing positive reactions, one-quarter neutral, and about one in five negative.

Some comments provided by respondents are shared for illustrative purposes:

Positive Sentiment

- It is actually nice to be reminded of what I answered last time and just to be given the opportunity to confirm or change the response. It makes filling out the survey much easier. (DI-2)
- It was really nice to not have to re-enter the info since it had not changed. Good job! (DI-1)
- I like being able to see my previous answers and update them. I've been in my job awhile and not much has changed. Also, I was surprised to see some of the detail I put in my answers previously. I'm glad I didn't need to recreate that. (DI-1)
- This is by far the easiest survey experience I can remember...I never felt some of the redundancy I usually do. (DI-2)
- I don't like to have to repeat inputting information you already have. Glad to have you populate and confirm information rather than me trying to populate. (IND)

- Having that data preloaded would be another indicator that you were legit--that you already knew many of the things you were asking and requiring less decisions from me about what/what not to divulge. And it saves time. (IND)

Negative Sentiment

Data Protections

- Because I am likely to go through these forms quickly, I would be worried that pre-filled answers would actually make [me] go more slowly because I would worry that the pre-filled information is incorrect. (IND)
- I suppose the only thing I worry about with any survey info is hacking. This info could be used to help establish a false identity based on me. (DI-1)
- I am always nervous giving out information that could be used for phishing purposes. (DI-2)
- Having the personal data prefilled in a survey makes me concerned about the data security and privacy. (DI-1)

Concerns about Confirmation Bias

- Pre-fills would probably reduce my level of effort. I usually read the entire list of options and try to find the closest match -- especially since my job is unique and I don't remember the details of past surveys. With pre-filled answers, I would probably just click through, thinking, "Yeah, that option's probably good enough." (IND)
- My only concern is that it is easier to gloss over answers without critically deciding if the answer has changed. (DI-2)
- I think having my previous answers available might make me pay less attention to my responses. (IND)
- Seeing my previous responses would initially surprise me. My concern would be that those responses (especially to behavior-type questions) might guide my initial gut reaction to the question. I would be more likely to be persuaded by my previous answer. (IND)

4a5. Response Distributions to DI and IND are Comparable

In determining whether to proceed with dependent interviewing, it is critical to assess the potential impact on survey estimates. If respondents are less likely to make changes to prior cycle information because of confirmation bias, this could impact longitudinal trends. Conversely, if respondents report a much higher rate of change than seen in the IND, trends may also be affected. As we see in tables 23-31, there were very few significant differences in the response distributions of key survey items. More specifically, DI-2 responses did not differ from IND for any of the items examined, and DI-1 differed from IND for only one item, employer type.

However, the rate of change in occupation code was significantly smaller in both DI approaches relative to IND. As context, the rate of change in occupation code between 2015 and 2017, and between 2017 and 2019 SDR production cycles was 37.7% and 39.9%

respectively. The IND rate of change in occupation code was similar at 35.8% despite the shorter elapsed time period. The two DI methods resulted in a rate of occupation code change of 20% (DI-1) and 22% (DI-2). While there is not a measure of truth available with this study design, prior research suggests that proactive dependent interviewing approaches reduce measurement error in reporting change (Mathiowetz, 2000; Sala, 2004), particularly in reporting occupation. It is also interesting that the rate of change in the Pilot IND condition, with a much shorter elapsed period of time, was comparable to the rate of change in the usual two-year elapsed time. This may be indicative of measurement error with occupation coding in the IND approach.

4b. If we do proceed with dependent interviewing, which approach should be used? DI-2 is recommended

4b1. Burden Assessment: DI-1 slightly longer than DI-2

Overall, DI-1 timings are significantly longer (15.21 minutes) than DI-2 (14.37 minutes) (Table 1). When respondents made changes to their prior cycle responses, there were no significant differences between the DI-1 and DI-2 approaches. We might have expected that DI-2 changes would have been more burdensome, since respondents needed to visit two screens to make a change, whereas in DI-1, the edits could be made on a single screen. However, this did not play out for any of the DI items. Timing differences for those who made changes at any point during the instrument were insignificant (Table 1a).

A significant difference is seen for those who did not make any changes to their prior cycle data (13.58 minutes for DI-1 vs. 12.75 minutes for DI-2). This could be a result of learning the functionality of the DI-1 approach, in which respondents need to mark a box indicating “no change” if the information has not changed. In the DI-1 condition, respondents were asked to either update the information on the screen, or to mark a checkbox indicating that the information had not changed since the prior cycle. Some respondents did not realize they had to mark the checkbox if nothing had changed, and would click the next button without doing anything on the screen. Respondents could receive up to two soft error messages if this occurred (across both in the SPV section and the survey items). In the SPV section, 295 respondents in the DI-1 condition (fully 34% of respondents) received an error message saying that they had left a screen blank rather than marking that no change had occurred. Within the survey itself, 154 DI-1 respondents triggered this soft error (18%). This suggests that some DI-1 respondents did not fully understand how to navigate the DI-1 functionality, which could have led to the increased burden for those making no changes.

4b2. Rate of Problematic Response Behavior: DI-1 higher than DI-2

As just indicated, roughly one-third of DI-1 respondents showed some confusion with how to navigate initial SPV screens when no changes had occurred. Conversely, when making a

change to the prior cycle response for a measure, approximately 10 percent of DI-1 respondents also marked the box at the bottom of the screen to indicate that the information had not changed since the prior cycle year. If this type of error occurred, respondents were shown an error message and were asked to re-enter their information. This could be an indication of user error, or an indication that they are updating the information but that it is only editorial, and not a substantive change. Nearly all of these respondents only made this error one time, but again it is an indication that DI-1 is eliciting user errors.

For the DI-2 condition, two types of problematic behaviors were analyzed. First, we analyzed the rate at which respondents backed up to change their answer to the “gate” question of whether or not the prior cycle response was still correct (Table 4a-1). Fewer than 4 percent of respondents backed up to change their answer from a “yes” to a “no” at any point during the instrument.

Second, we analyzed whether respondents initially indicated that there was no change to the measure when perhaps there actually was (Table 4a-2). To detect this, we looked for behaviors in which respondents initially said no change, and then later backed up and modified their answer to indicate that there had been a change. This occurred for fewer than 5 percent of respondents.

4b3. Unit and Item Nonresponse – No difference

As shown in chapter 3 methodological results, there was no significant difference in response rates between DI-1 and DI-2, though DI-1 did result in a slightly lower nominal response rate. Item nonresponse rates were low across the board, and not significantly different by condition.

4b4. RAS Results – No difference

As shown in tables 7 through 20, there were no significant differences in reactions to the two dependent interviewing approaches.

4b5. Response Distributions – No difference

As shown in tables 23-31, there were very few significant differences between DI-1 and DI-2 for frequency distributions on key survey measures.

4c. If we do proceed with DI, will it apply just to the longitudinal sample, or to the full continuing cohort?

We recommend implementing dependent interviewing with the full continuing cohort, due to the reduction in burden and the positive reactions to the DI approach.

4d. If we do proceed with DI, would we use the approach with all items tested in the Pilot Study, or a subset?

We recommend implementing dependent interviewing with all of the items that were tested in the Pilot Study. However, we do believe that further testing is merited for the job duties item in the DI-2 approach. For respondents who indicate the information is no longer correct, they could be shown their prior cycle job duties in an editable field, rather than being provided with a blank screen to enter the information from scratch.