Testing Dependent Interviewing on a Self-Administered Survey

Jennifer Sinibaldi
Director, Research Program

National Center for Science and Engineering Statistics
Social, Behavioral and Economic Sciences
National Science Foundation
Overview

• Dependent Interviewing (DI) Experiment
  • Control group + 2 DI treatments
  • Respondent Experience questions
• Results: Did it seem to work?
  • Time/burden, measurement error, enjoyment
Background
Our Goal

• Explicitly treat survey as longitudinal
• Make better use of prior responses
  • More efficient
  • More consistent
• Better experience for respondents
The Proposed Solution

- “Proactive” Dependent Interviewing (DI) for some questions
  - Shorter questionnaire
  - More consistency across information that hasn’t changed
  - More pleasant respondent experience

- Try out 2 different formats of DI
  - 1-step (DI1) and 2-step (DI2)
  - Total of 3 versions of the questionnaire
Analysis: Research Questions

- Does DI reduce time to administer?
- Does DI appear to affect measurement error?
- Are respondents receptive to DI?
  - We need to ask follow-up questions → Respondent Experience Questions
- Overall, is one DI version superior?
What’s Novel about this?

- DI on self-administered questionnaire
- Motivation for doing DI is unique
- 2 styles of DI, 1 step approach (DI1) not typically used

- Literature:
  - Seam effect (e.g. Moore et al. 2006)
  - Web-based, Self-administered study
    - Lugtig and Lensvelt-Mulders 2014; Al Baghal 2017
  - Comparisons of Proactive vs. Reactive DI (e.g. Lynn & Jackle 2007)
  - 1 step approach (DI1) (Hoogendoorn 2004)
Examples of the Questions
The Two Forms of DI

- **DI 1: the 1-step approach**
  - Shows the question with the answer selected that R reported last time. Ask R to change it or indicate no change on the same screen.

- **DI 2: the 2-step approach**
  - Shows the question with the answer selected that R reported last time. Ask “Is this information correct?”.
  - If No, then 2nd screen asking standard question.
In your responses to the 2017 SDR, you reported the number of people who worked for your employer, as selected below.

Please update the information to indicate how many people worked for your principal employer during the week of September 1, 2020, counting all locations where this employer operates.

If the answer selected below was still correct during the week of September 1, 2020, please select the box “Information has not changed since 2017.”

- 10 or fewer employees
- 11 – 24 employees
- 25 – 99 employees
- 100 – 499 employees
- 500 – 999 employees
- 1,000 – 4,999 employees
- 5,000 – 24,999 employees
- 25,000 or more employees

☐ Information has not changed since 2017
Example Dependent Interviewing Question- DI2

In your responses to the 2019 SDR, you reported the number of people who worked for your employer, counting all locations where this employer operated, as selected below.

- 10 or fewer employees
- 11-24 employees
- 25-99 employees
- 100-499 employees
- 500-999 employees
- 1,000-4,999 employees
- 5,000-24,999 employees
- 25,000 or more employees

Was this information still correct as of the week of September 1, 2020?

- Yes
- No
Example Question- Control

Counting all locations where this employer operates, how many people work for your principal employer? Your best estimate is fine.

- 10 or fewer employees
- 11 – 24 employees
- 25 – 99 employees
- 100 – 499 employees
- 500 – 999 employees
- 1,000 – 4,999 employees
- 5,000 – 24,999 employees
- 25,000 or more employees
Respondent Experience Questions

- Topics: Perceived speed, enjoyment, question sensitivity
  - Ex: How confident are you that NCSES will protect your answers?
  - Ex: To what extent did you enjoy completing today’s survey?
  - Ex: How sensitive did you think the questions on this survey were?

- Some questions worded the same for all 3 groups

  **To what extent did you enjoy completing today’s survey?**
  - Did not enjoy at all
  - Enjoyed a little
  - Enjoyed somewhat
  - Enjoyed a great deal
Respondent Experience Questions

• Different question wording for DI vs. Control
  • **DI:** Do you think pre-filling some of your answers from [prior cycle] and asking you to confirm or update them made this survey...
  
  • **Control:** If this survey had pre-filled your most recent answers and asked you to confirm or update them, would that have made this survey...
    • Much more burdensome
    • A little more burdensome
    • Neither more or less burdensome
    • A little less burdensome
    • Much less burdensome
Respondent Experience Questions

• Questions for DI only

Were there any questions in today’s survey where you felt that the answer displayed from 2019 was “accurate enough” and you decided to leave it as-is, rather than updating it with potentially more accurate information?

- Yes for one question
- Yes for more than one question
- No
Data Collection
Data Collection

- Survey of Doctorate Recipients (SDR)
- Sample 3,900 cases, mostly 2019 Rs
  - Stratified, random assignment to DI1, DI2, Control
- 38 SDR questions, not including R verification questions
  - 13 DI questions
    - 9 closed-ended (7 single response, 2 forced choice)
    - 4 open-ended
- 20 Respondent Experience questions
Data Collection

- Reference date: 1 Sept 2020
- $30 prepaid incentive
- AAPOR (unwtd) RR1: 65.6%
  - Across 3 groups: 64.2% - 66.7%
  - DI1 n = 832 | DI2 n = 863 | Control n = 879
Results
Research Question: Time to Administer

- Actual survey length (unweighted, trimmed)
  - DI1 15.0 min | DI2 14.0 min* | Control 15.4 min

- Could be influenced by format and number of Rs without changes

* 95% confidence level
**Research Question: Time to Administer**

- Actual survey length: **DI2** significantly faster
- Perceived survey length (wtd): **DI1 & DI2** significantly “faster”

<table>
<thead>
<tr>
<th>Response</th>
<th>DI1</th>
<th>DI2</th>
<th>Control**</th>
</tr>
</thead>
<tbody>
<tr>
<td>How would you rate the speed at which you completed today’s survey?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very slow</td>
<td>0.4%</td>
<td>1.4%</td>
<td>0.4%</td>
</tr>
<tr>
<td>Somewhat slow</td>
<td>7.6%</td>
<td>7.6%</td>
<td>16.4%</td>
</tr>
<tr>
<td>Somewhat fast</td>
<td>62.7%</td>
<td>59.3%</td>
<td>56.6%</td>
</tr>
<tr>
<td>Very fast</td>
<td>29.3%</td>
<td>31.7%</td>
<td>26.6%</td>
</tr>
</tbody>
</table>

*P<0.05, **P<0.01, ***P<0.001
Research Question: Measurement Error (1 of 6)

- Closed-ended questions: Overall Distributional Differences

<table>
<thead>
<tr>
<th>Question</th>
<th>DI1 vs. DI2</th>
<th>DI1 vs. Control</th>
<th>DI2 vs. Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employer type</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Employer size</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Educational institution</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Type of educational institution</td>
<td>n.s.</td>
<td>**</td>
<td>*</td>
</tr>
<tr>
<td>Faculty rank</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Work activity (forced check all)</td>
<td>‡‡ (1/14)</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Supervised others</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

*P<0.05, **P<0.01, ***P<0.001  Based on Pearson Design-Based Chi2
‡P<0.05, ‡‡P<0.01, ‡‡‡P<0.001  Based on Adjusted Wald F-Test (Bivariate variables only)
Research Question: Measurement Error (2 of 6)

- Closed-ended questions: Differences by Response Option

<table>
<thead>
<tr>
<th>Question</th>
<th>DI1 vs. DI2</th>
<th>DI1 vs. Control</th>
<th>DI2 vs. Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employer type</td>
<td>* (2/10)</td>
<td>* (3/10)</td>
<td>* (1/10)</td>
</tr>
<tr>
<td>Employer size</td>
<td>* (1/8)</td>
<td>* (1/8)</td>
<td>* (1/8)</td>
</tr>
<tr>
<td>Educational institution</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Type of educational institution</td>
<td>n.s.</td>
<td>* (2/6)</td>
<td>* (3/6)</td>
</tr>
<tr>
<td>Faculty rank</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Work activity (forced check all)</td>
<td>* (1/14)</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Supervised others</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

* 95% confidence level    Based on the Adjusted Wald F-Test
Research Question: Measurement Error (3 of 6)

- Open-ended questions

<table>
<thead>
<tr>
<th>Question</th>
<th>DI1 vs. DI2</th>
<th>DI1 vs. Control</th>
<th>DI2 vs. Control</th>
<th>DI1 vs. DI2</th>
<th>DI1 vs. Control</th>
<th>DI2 vs. Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employer Name</td>
<td>N.S</td>
<td>***</td>
<td>***</td>
<td>N.S</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Job Title</td>
<td>N.S</td>
<td>***</td>
<td>***</td>
<td>N.S</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Job Duties</td>
<td>N.S</td>
<td>***</td>
<td>***</td>
<td>N.S</td>
<td>***</td>
<td>**</td>
</tr>
</tbody>
</table>

*P<0.05, **P<0.01, ***P<0.001, N.S = Not Significant
Research Question: Measurement Error (4 of 6)

- Open-ended questions: Percent meaningful changes, of those who made changes to open-ended questions

<table>
<thead>
<tr>
<th>Question</th>
<th>DI1</th>
<th>DI2</th>
<th>Control***</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employer Name</td>
<td>90.4%</td>
<td>88.5%</td>
<td>47.1%</td>
</tr>
<tr>
<td>Job Title</td>
<td>89.5%</td>
<td>88.3%</td>
<td>53.6%</td>
</tr>
<tr>
<td>Job Duties</td>
<td>31.3%</td>
<td>51.1%</td>
<td>18.5%</td>
</tr>
</tbody>
</table>

*P<0.05, **P<0.01, ***P<0.001
Research Question: Measurement Error (5 of 6)

- Any questions where answer was accurate enough although could have updated it? (among DI conditions only)

<table>
<thead>
<tr>
<th>Response</th>
<th>DI1</th>
<th>DI2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes, For 1 Question</td>
<td>11.4%</td>
<td>13.0%</td>
</tr>
<tr>
<td>Yes, for Multiple Questions</td>
<td>21.8%</td>
<td>21.9%</td>
</tr>
<tr>
<td>No</td>
<td>66.8%</td>
<td>65.1%</td>
</tr>
</tbody>
</table>

*P<0.05, **P<0.01, ***P<0.001 Based on Adjusted Wald F-Test
Research Question: Measurement Error (6 of 6)

- Any questions where the answer displayed was no longer true, and left as-is? (among DI conditions only)

<table>
<thead>
<tr>
<th>Response</th>
<th>DI1</th>
<th>DI2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes, For 1 Question</td>
<td>3.8%</td>
<td>5.2%</td>
</tr>
<tr>
<td>Yes, for Multiple Questions</td>
<td>3.3%</td>
<td>1.0%</td>
</tr>
<tr>
<td>No</td>
<td>92.9%</td>
<td>93.8%</td>
</tr>
</tbody>
</table>

*P<0.05, **P<0.01, ***P<0.001 Based on Adjusted Wald F-Test
Research Question: What do Rs think of DI? (1 of 3)

- How sensitive do you think the questions in this survey were?

<table>
<thead>
<tr>
<th>Response</th>
<th>DI1</th>
<th>DI2</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>How Sensitive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not at all</td>
<td>34.2%</td>
<td>39.0%</td>
<td>35.7%</td>
</tr>
<tr>
<td>A Little</td>
<td>36.3%</td>
<td>37.2%</td>
<td>35.6%</td>
</tr>
<tr>
<td>Somewhat</td>
<td>24.1%</td>
<td>16.7%</td>
<td>24.1%</td>
</tr>
<tr>
<td>Very</td>
<td>5.3%</td>
<td>7.1%</td>
<td>4.7%</td>
</tr>
</tbody>
</table>

*P<0.05, **P<0.01, ***P<0.001
Research Question: What do Rs think of DI? (2 of 3)

- Do you think pre-filling some of your answers from 201X and asking you to confirm or update them (made/would make) this survey...?

<table>
<thead>
<tr>
<th>Response</th>
<th>DI1</th>
<th>DI2</th>
<th>Control***</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Filling Burdensome?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Much More</td>
<td>0.1%</td>
<td>0.2%</td>
<td>0.9%</td>
</tr>
<tr>
<td>A Little More</td>
<td>1.6%</td>
<td>1.0%</td>
<td>6.0%</td>
</tr>
<tr>
<td>Neither</td>
<td>10.9%</td>
<td>11.2%</td>
<td>21.7%</td>
</tr>
<tr>
<td>A Little Less</td>
<td>24.3%</td>
<td>26.2%</td>
<td>42.2%</td>
</tr>
<tr>
<td>Much Less</td>
<td>63.1%</td>
<td>61.5%</td>
<td>29.2%</td>
</tr>
</tbody>
</table>

*P<0.05, **P<0.01, ***P<0.001
To what extent did you enjoy completing today’s survey?

<table>
<thead>
<tr>
<th>Response</th>
<th>DI1</th>
<th>DI2</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td>12.4%</td>
<td>13.7%</td>
<td>16.7%</td>
</tr>
<tr>
<td>A Little</td>
<td>33.7%</td>
<td>35.5%</td>
<td>35.0%</td>
</tr>
<tr>
<td>Somewhat</td>
<td>43.7%</td>
<td>40.6%</td>
<td>41.7%</td>
</tr>
<tr>
<td>A Great Deal</td>
<td>10.3%</td>
<td>10.2%</td>
<td>6.6%</td>
</tr>
</tbody>
</table>

*P<0.05, **P<0.01, ***P<0.001
Conclusion
Research Questions

• Does DI reduce time to administer? Yes
• Does DI appear to affect measurement error? Small indications of both improvement and increase
• Are respondents receptive to DI? Not a strong positive or negative sentiment
• Overall, is one DI version superior? DI2 slightly preferred

Conclusion: Not seeing dramatic results (+/-) which seems to indicate a smooth transition to DI.
Thank You!

Jennifer Sinibaldi
Director, Research Program
National Center for Science and Engineering Statistics
jsinibal@nsf.gov

https://ncses.nsf.gov/about/research
https://ncses.nsf.gov/about/funding-opportunities