Using responsive and adaptive survey design to control data quality and costs

James Wagner
Overview

• Changing context of surveys
• Heterogeneity
• Responsive and adaptive survey design
  – Definition
  – Examples
• Logistics
• Way forward
**Challenge**: Decreasing Response

- US trends
- Similar experience in Europe – *de Leeuw, et al.*, 2018

*Williams and Brick, 2017*
Challenge:
Costs increasing for face-to-face surveys

• Over time, more effort required to achieve the same or worse results
• Costs go up

Contact Attempts per Complete

Williams and Brick, 2017
Opportunities

• Computerization
  – Allows us to **monitor field progress** in almost real-time
  – **Interventions** also possible
  – **More complex designs** possible

• **Nonresponse bias** vs nonresponse rates
  – What is the impact of design on estimates, not just response rates?
Recognition of Heterogeneity

• Survey design used to be “one-size-fits-all”
• Recent research looks at variation within samples
  – Tailoring the introduction
    • Groves and Couper, 1996
  – Leverage-Saliency theory
    • Groves, Singer, and Corning, 2000
    • Each sampled person has specific leverages
    • Survey makes these salient
  – Nonresponse bias analysis
    • Groves, 2006
    • Focus on impact on estimates
    • Naturally leads to examination of subgroups who respond under different designs
Recognition of Heterogeneity

• Language barriers to self-administered modes
  – Ahlmark, et al., 2015

• Incentives have differential impact
  – Groves, et al., 2004; Singer and Ye, 2013

• Differences in response to web surveys by age
  – Calinescu, et al., 2013; Börkan, 2010
Can We Utilize this Heterogeneity?

• Define important subgroups
• Vary the strategies across subgroups
• Optimize for cost and quality
• Example:
  – Web survey for those highly likely to respond
  – Face-to-face survey for those unlikely to respond with important differences
Context Matters

• What do we know about the sample before we begin?
• More observed characteristics means more information for forming subgroups
• Fewer observed characteristics... may need to learn about subgroups over time
• Two different approaches based on this distinction:
  – Adaptive Survey Design (*Schouten, Peytchev, and Wagner, 2017*)
  – Responsive Survey Design (*Groves and Heeringa, 2006*)
Responsive Survey Design

• *Groves and Heeringa (2006)*

• Arises from **uncertainty**
  – We do not know much about the sample ahead of time

• Differences within the sample are revealed across **phases**
  – Each phase constitutes a set of unique design features

• The goal is to design **complementary** phases
  – Biases of each phase “cancel” each other out
Example: Responsive Survey Design

- National Survey of Family Growth (NSFG)
- Phase 1: US Mail prenotification, $40 post-paid token of appreciation for main interview, interviewers have large workloads
- Phase 2: Priority Mail sent, $40 pre-paid and $40 post-paid token of appreciation, small workloads

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Phase 1</th>
<th>Phase 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$40</td>
<td>$80</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td>n=1,896</td>
<td>n=68</td>
</tr>
<tr>
<td>College degree or more</td>
<td>34</td>
<td>51**</td>
</tr>
<tr>
<td>Ever had an abortion</td>
<td>6</td>
<td>1**</td>
</tr>
<tr>
<td>Never had a live birth</td>
<td>41</td>
<td>60**</td>
</tr>
<tr>
<td>Ever had sex with a female</td>
<td>13</td>
<td>4**</td>
</tr>
<tr>
<td>Income $75,000+</td>
<td>17</td>
<td>25</td>
</tr>
<tr>
<td>Living in a multi-unit structure</td>
<td>38</td>
<td>24**</td>
</tr>
<tr>
<td><strong>Male</strong></td>
<td>n=1,432</td>
<td>n=70</td>
</tr>
<tr>
<td>Hispanic</td>
<td>20</td>
<td>37**</td>
</tr>
<tr>
<td>College degree or more</td>
<td>28</td>
<td>36</td>
</tr>
<tr>
<td>Never fathered a birth</td>
<td>57</td>
<td>64</td>
</tr>
<tr>
<td>Ever had sex with a male</td>
<td>7</td>
<td>1**</td>
</tr>
<tr>
<td>Income $75,000+</td>
<td>25</td>
<td>42**</td>
</tr>
<tr>
<td>Living in a multi-unit structure</td>
<td>37</td>
<td>26*</td>
</tr>
</tbody>
</table>

NSFG 2006-2010
Lepkowski, et al., 2013
* p<=0.10  
** p<=0.05
Adaptive Survey Design

• More information available about the sample
• Possible to identify subgroups in the sample before data collection
• Prior experimentation with design alternatives
• Use targeted designs for each subgroup
• Optimize for cost and quality
Example: Adaptive Survey Design

• Dutch Labor Force Survey
  – *Schouten, et al., 2017*

• Create 9 strata using 5 most relevant auxiliary variables:

<table>
<thead>
<tr>
<th>Registered Unemployed</th>
<th>Young Household Member and Employed</th>
</tr>
</thead>
<tbody>
<tr>
<td>65+ Households without employment</td>
<td>Non-Western and Employed</td>
</tr>
<tr>
<td>Young Household Members without Employment</td>
<td>Western and Employed</td>
</tr>
<tr>
<td>Non-Western without Employment</td>
<td>Large Households</td>
</tr>
<tr>
<td>Western without Employment</td>
<td></td>
</tr>
</tbody>
</table>
### Example: Adaptive Survey Design

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>W</strong></td>
<td>23.2</td>
<td>23.6</td>
<td>15.5</td>
<td>10.8</td>
<td>27.9</td>
<td>27.7</td>
<td>17.5</td>
<td>36.7</td>
<td>22.4</td>
</tr>
<tr>
<td><strong>TS</strong></td>
<td>12.2</td>
<td>31.4</td>
<td>8.5</td>
<td>4.7</td>
<td>19.7</td>
<td>13.3</td>
<td>7.2</td>
<td>18.1</td>
<td>21.2</td>
</tr>
<tr>
<td><strong>TE</strong></td>
<td>20.8</td>
<td>41.3</td>
<td>15.2</td>
<td>8.6</td>
<td>31.1</td>
<td>23.8</td>
<td>14.3</td>
<td>33.3</td>
<td>37.5</td>
</tr>
<tr>
<td><strong>F</strong></td>
<td>43.5</td>
<td>53.5</td>
<td>42.2</td>
<td>34.1</td>
<td>45.1</td>
<td>45.3</td>
<td>35.9</td>
<td>46.7</td>
<td>54.6</td>
</tr>
<tr>
<td><strong>FE</strong></td>
<td>52.4</td>
<td>58.3</td>
<td>51.0</td>
<td>41.2</td>
<td>51.2</td>
<td>54.9</td>
<td>46.0</td>
<td>56.8</td>
<td>61.4</td>
</tr>
<tr>
<td><strong>W→TS</strong></td>
<td>28.3</td>
<td>41.0</td>
<td>20.2</td>
<td>13.9</td>
<td>36.3</td>
<td>34.0</td>
<td>20.8</td>
<td>44.5</td>
<td>23.1</td>
</tr>
<tr>
<td><strong>W→TE</strong></td>
<td>32.8</td>
<td>48.4</td>
<td>23.8</td>
<td>17.5</td>
<td>42.1</td>
<td>41.1</td>
<td>25.8</td>
<td>52.1</td>
<td>24.4</td>
</tr>
<tr>
<td><strong>W→FS</strong></td>
<td>46.3</td>
<td>57.7</td>
<td>38.6</td>
<td>32.7</td>
<td>50.0</td>
<td>51.0</td>
<td>39.3</td>
<td>58.9</td>
<td>50.0</td>
</tr>
<tr>
<td><strong>W→FE</strong></td>
<td>49.8</td>
<td>58.3</td>
<td>43.4</td>
<td>36.6</td>
<td>52.6</td>
<td>54.7</td>
<td>44.3</td>
<td>62.0</td>
<td>54.2</td>
</tr>
</tbody>
</table>
Example: Adaptive Survey Design

• From these estimates, it is possible to use optimization techniques to assign strategies to the strata
  – Maximize some quality measure
  – Subject to other quality constraints (response rate, balance indicator, or other – more on this in next section)
  – Subject to cost constraint
    • Need cost estimates for each strategy, ideally for each strategy/stratum combination
Logistics

• ASD and RSD presuppose technical and administrative structure to implement

• Current systems not built for ASD/RSD

• May need to start with existing systems, build designs that can be accommodated
  – Then add features to survey design and improve systems

• Management: Start with training
  – Start slow and grow
Way Forward

• Identify risks
• Identify available resources
• Prepare a plan
  – ASD: Subgroups, matched to designs
  – RSD: Complementary design phases
• Implement
• Document, learn, extend...
Thank you!

jameswag@umich.edu
References (1)


References (2)


