Adaptive Design Strategies
Who do you target?
And what do you do with them?

Jaki S. McCarthy
Kathleen Ott
Melissa Mitchell
USDA National Agricultural Statistics Service
Adaptive Design in Surveys

• Different data collection strategies have always been used in survey data collection

• However, these are often unplanned or unknown by the survey designers
  – For example prioritization of cases by interviewers
  – Scheduling of phone interviewers
  – Mode switching by field staff
The challenge:

- How do we strategically allocate resources to data collection?

- Decisions on allocating resources should be made to reduce TSE, not just serve local level objectives.
Two NASS examples

- Agricultural Resource Management Survey (ARMS)
- Crops/Stocks Quarterly Survey
Our examples...

**ARMS**
- Collects farm financial information and costs associated with producing agricultural commodities
- Estimates at US, regional, and state level (for 15 states)
- Lengthy annual survey with historically low response rates (in the 60% range)
- Sample sizes typically >30k
- Data collection primarily in person over approx. 3 months

**Crops/Stocks**
- Collects crop acreage, inventory and production and grain and oilseed stocks
- Conducted in June, September, December and March
- Estimates made for US and state (major commodities and specialty crops differ by state)
- Sample sizes large > 65k
- Data collection period is short (~2 weeks), RR near 80%
- Most data collection is central call center CATI with limited telephone, online, mail and in person interviews
How is nonresponse handled?

**ARMS**
- Estimates at US, regional, and state level
- Calibration weighting is used to compensate for nonresponse (including bias) and measurement error
- Multiple calibration targets used based on known external population totals
- ARMS records reweighted to meet targets

**Crops/Stocks**
- Estimates made at US and state level
- Nonresponse adjustments are made based on strata
- NR strata defined by size in acres, grain storage capacity, and some specialty crops
“Impact Operations”

• Like many establishment surveys, farms are highly skewed
• Often a small number of operations will dominate an estimate – “impact operation”
• Or a small number of operations may be critical to nonresponse weighting
Managing data collection in ARMS

• Calibration targets are known in advance
• Must have minimum amount of target reported in order to use for weighting
• Operations large relative to calibration targets are “impact operations”
• Nonresponse propensity models available and can pre-identify likely “impact” nonrespondents
ARMS data collection strategies

• Target additional resources ($$$) to these identified records ("impact" + likely NRs)
  – INITIAL in person contact by field office director or other senior level staff
  – Data collection by experienced or supervisory interviewers only
  – Interviewer incentives for hard cases

• Ideally, monitor during data collection
Crops/Stocks data collection strategies

- Also have nonresponse propensity models
- Some nonresponse strata can be used to identify impact operations
  - Strata with large operations (e.g. 5000+ acres)
  - Specialty crops (potatoes)
- Additional resources can be targeted at “impact” likely nonrespondents
Data Collection Strategies for Crops/Stocks “Impact Operations”

• For those impact operations likely to be nonrespondents:
  – Managed by local field office;
  – Field enumerator phoned
  – With in person follow up ($$$)

• For those impact operations likely to be respondents:
  – Phoned by central phone center (¢)
  – At end of data collection, any of these “easy” cases still pending sent to field office for last attempts ($)

• Compared offices asked to follow these procedures to all other offices
So how did it work?
### Overall results – Crops/Stocks

<table>
<thead>
<tr>
<th>Crops/Stocks</th>
<th>Likely</th>
<th>Other</th>
<th>Difference (Actual – Predicted) %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refusals</td>
<td>REFUSAL</td>
<td>Other</td>
<td>Treatment</td>
</tr>
<tr>
<td>Sept All</td>
<td>21%</td>
<td>52-78%</td>
<td></td>
</tr>
<tr>
<td>Dec All</td>
<td>15%</td>
<td>46-73%</td>
<td></td>
</tr>
<tr>
<td>Sept Impact Operations</td>
<td>18%</td>
<td>53-79%</td>
<td>2.8</td>
</tr>
<tr>
<td>Dec Impact Operations</td>
<td>12%</td>
<td>47-74%</td>
<td>0.2</td>
</tr>
<tr>
<td>Inaccessible</td>
<td>NONCONTACT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sept</td>
<td>39%</td>
<td>34-75%</td>
<td></td>
</tr>
<tr>
<td>Dec</td>
<td>30%</td>
<td>29-71%</td>
<td></td>
</tr>
<tr>
<td>Sept Impact Operations</td>
<td>38%</td>
<td>32-74%</td>
<td>9.4</td>
</tr>
<tr>
<td>Dec Impact Operations</td>
<td>29%</td>
<td>26-70%</td>
<td>-5.1</td>
</tr>
</tbody>
</table>
# Overall Results -- ARMS

<table>
<thead>
<tr>
<th></th>
<th>Likely Nonrespondents</th>
<th>Others</th>
<th>Targeted Records</th>
<th>Control Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete</td>
<td>55.6%</td>
<td>73.1%</td>
<td>55.3%</td>
<td>55.2%</td>
</tr>
<tr>
<td>Refusal</td>
<td>36.9%</td>
<td>21.8%</td>
<td>36.2%</td>
<td>37.5%</td>
</tr>
<tr>
<td>Noncontact</td>
<td>4.9%</td>
<td>4.2%</td>
<td>4.8%</td>
<td>5.2%</td>
</tr>
<tr>
<td>Office Hold</td>
<td>2.9%</td>
<td>0.9%</td>
<td>3.7%</td>
<td>2.1%</td>
</tr>
</tbody>
</table>
So what happened?

• ARMS:
  – Field staff only minimally adhered to HQ directions
  – “we don’t like rewarding only some interviewers”
  – “office staff were not available to contact in person”
And what happened here?

• Crops/Stocks:
  – Sending cases to field interviewers doesn’t produce much improvement
  – It's hard to embed experimental procedures into operational surveys
    • No one wants to stick to “control” procedures that seem worse than current practice
    • No way to know what our treatment was being compared to

• Hard cases ARE hard!
Big lesson learned: Hard to test in operational environments

• We need to do a better job:
  – Selling the tests,
  – Making procedures “doable”
  – Monitoring what actually happens in data collection
  – Deciding how we know if we’ve made improvements

• Future plans:
  – Consider whether to continue a focus on “hard” cases or move to those more likely to be converted
  – Bring on smaller set of test states with firmer commitment to testing
  – Work from the top and the bottom (convince field staff and HQ staff they report to)
Adaptive Design is Hard

We’re trying as hard as we can already!

What’s Left?