The Treatment of Observation
How surveys affect survey-takers in over-researched populations

Jonathan Forney
Lead Methodologist, Forcier Consulting

Created for WAPOR, 2015
The Problem:
  Research fatigue
  General Intuitions and Specific Hypotheses

Methodology
  An Experimental Approach
  Efficacy of Design

Analysis and Findings
  Testing Key Hypotheses
  Regression Model Formulations
  Findings
The Problem:
Research fatigue

General Intuitions and Specific Hypotheses

Methodology
An Experimental Approach
Efficacy of Design

Analysis and Findings
Testing Key Hypotheses
Regression Model Formulations
Findings
Examples in the literature

- **Qualitative on fatigue and unkept promises:**
  - e.g. Mayssoun Sukarieh and Stuart Tannock, “On the Problem of Over-researched Communities: The Case of the Shatila Palestinian Refugee Camp in Lebanon,” *Sociology*.

- **Quantitative on fatigue and participation:**
  - e.g. Stephen R Porter, Michael E Whitcomb, and William H Weitzer, “Multiple surveys of students and survey fatigue,” *New Directions for Institutional Research*.

- **Quantitative on motivated underreporting:**
Research fatigue

Examples in the literature

▶ Qualitative on fatigue and unkept promises:
  ▶ e.g. Mayssoun Sukarieh and Stuart Tannock, “On the Problem of Over-researched Communities: The Case of the Shatila Palestinian Refugee Camp in Lebanon,” *Sociology*.

▶ Quantitative on fatigue and participation:
  ▶ e.g. Stephen R Porter, Michael E Whitcomb, and William H Weitzer, “Multiple surveys of students and survey fatigue,” *New Directions for Institutional Research*.

▶ Quantitative on motivated underreporting:
Examples in the literature

- **Qualitative on fatigue and unkept promises:**
  - e.g. Mayssoun Sukarieh and Stuart Tannock, “On the Problem of Over-researched Communities: The Case of the Shatila Palestinian Refugee Camp in Lebanon,” *Sociology*.

- **Quantitative on fatigue and participation:**
  - e.g. Stephen R Porter, Michael E Whitcomb, and William H Weitzer, “Multiple surveys of students and survey fatigue,” *New Directions for Institutional Research*.

- **Quantitative on motivated underreporting:**
The Problem:
  Research fatigue

General Intuitions and Specific Hypotheses

Methodology
  An Experimental Approach
  Efficacy of Design

Analysis and Findings
  Testing Key Hypotheses
  Regression Model Formulations
  Findings
General Intuitions and Specific Hypotheses

General Intuitions

- The condition of being repeatedly researched leads to apathy or even antipathy towards the research process.
- Repeated exposures to surveys may significantly affect survey-taking behaviors.
General Intuitions

▶ The condition of being repeatedly researched leads to apathy or even antipathy towards the research process.

▶ Repeated exposures to surveys may significantly affect survey-taking behaviors.
Specific Hypotheses

Research fatigue operationalized as very recent exposure to another survey.

H1 Research fatigue $\rightarrow$ lower willingness to participate/consent.
   Mechanism: Antipathy toward research process.

H2 Research fatigue $\rightarrow$ motivated underreporting.
   Mechanism: Antipathy toward research process.

H3 Research fatigue $\rightarrow$ less accurate responses to demanding questions.
   Mechanism: Less motivated than average.

H4 Research fatigue $\rightarrow$ more item non-response.
   Mechanism: Less motivated than average.
Specific Hypotheses

Research fatigue operationalized as very recent exposure to another survey.

H1  Research fatigue $\rightarrow$ lower willingness to participate/consent.

   Mechanism: Antipathy toward research process.

H2  Research fatigue $\rightarrow$ motivated underreporting.

   Mechanism: Antipathy toward research process.

H3  Research fatigue $\rightarrow$ less accurate responses to demanding questions.

   Mechanism: Less motivated than average.

H4  Research fatigue $\rightarrow$ more item non-response.

   Mechanism: Less motivated than average.
Specific Hypotheses

Research fatigue operationalized as very recent exposure to another survey.

**H1**  Research fatigue → lower willingness to participate/consent.

*Mechanism:* Antipathy toward research process.

**H2**  Research fatigue → motivated underreporting.

*Mechanism:* Antipathy toward research process.

**H3**  Research fatigue → less accurate responses to demanding questions.

*Mechanism:* Less motivated than average.

**H4**  Research fatigue → more item non-response.

*Mechanism:* Less motivated than average.
## Specific Hypotheses

Research fatigue operationalized as very recent exposure to another survey.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Statement</th>
<th>Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>Research fatigue $\rightarrow$ lower willingness to participate/consent.</td>
<td>Antipathy toward research process.</td>
</tr>
<tr>
<td>H2</td>
<td>Research fatigue $\rightarrow$ motivated underreporting.</td>
<td>Antipathy toward research process.</td>
</tr>
<tr>
<td>H3</td>
<td>Research fatigue $\rightarrow$ less accurate responses to demanding questions.</td>
<td>Less motivated than average.</td>
</tr>
<tr>
<td>H4</td>
<td>Research fatigue $\rightarrow$ more item non-response.</td>
<td>Less motivated than average.</td>
</tr>
</tbody>
</table>
Specific Hypotheses

Research fatigue operationalized as very recent exposure to another survey.

H1 Research fatigue $\rightarrow$ lower willingness to participate/consent.
   Mechanism: Antipathy toward research process.

H2 Research fatigue $\rightarrow$ motivated underreporting.
   Mechanism: Antipathy toward research process.

H3 Research fatigue $\rightarrow$ less accurate responses to demanding questions.
   Mechanism: Less motivated than average.

H4 Research fatigue $\rightarrow$ more item non-response.
   Mechanism: Less motivated than average.
An Experimental Approach

The Problem:
Research fatigue
General Intuitions and Specific Hypotheses

Methodology
An Experimental Approach
Efficacy of Design

Analysis and Findings
Testing Key Hypotheses
Regression Model Formulations
Findings

Jonathan Forney
Lead Methodologist, Forcier Consulting
Treatment of Observation
An Experimental Approach

Quasi-Experimental Design

- **Treatment** = Survey averaging 47 minutes on various topics, based on Gallup World Poll survey
- **Observation** = Survey averaging 28 minutes

Treatment Group: T – O

Control Group: _ – O
An Experimental Approach

Quasi-Experimental Design

- Treatment = Survey averaging 47 minutes on various topics, based on Gallup World Poll survey
- Observation = Survey averaging 28 minutes

Treatment Group: T – O

Control Group: _ – O
Considerations

▶ Study should be double-blind:
  ▶ Neither enumerators nor participants can know they are participating in experiment.

▶ Both “treatment” and “observation” surveys should be believable.

▶ Context:
  ▶ Urban neighborhoods in Juba, South Sudan.
An Experimental Approach

Logistics of Treatment

Simple Randomization of Treatment:

- Map enumeration areas.
- Number households.
- Attempt to treat all even-numbered households within EAs.

How to make sure treated individuals within households can be found again?

Limit target population to female head of household.

Jonathan Forney
Lead Methodologist, Forcier Consulting
An Experimental Approach

Logistics of Treatment

Simple Randomization of Treatment:

▶ Map enumeration areas.

▶ Number households.

▶ Attempt to treat all even-numbered households within EAs.

How to make sure treated individuals within households can be found again?

Limit target population to female head of household.
An Experimental Approach

Logistics of Treatment

Simple Randomization of Treatment:

- Map enumeration areas.
- Number households.
- Attempt to treat all even-numbered households within EAs.

How to make sure treated individuals within households can be found again?

Limit target population to female head of household.
An Experimental Approach

Logistics of Treatment

Simple Randomization of Treatment:

- Map enumeration areas.
- Number households.
- Attempt to treat all even-numbered households within EAs.

How to make sure treated individuals within households can be found again?

Limit target population to *female head of household*.

Jonathan Forney

Lead Methodologist, Forcier Consulting
An Experimental Approach

Logistics of Treatment

Simple Randomization of Treatment:

- Map enumeration areas.
- Number households.
- Attempt to treat all even-numbered households within EAs.

How to make sure treated individuals within households can be found again?

Limit target population to female head of household.
An Experimental Approach

EA Map

Jonathan Forney
Lead Methodologist, Forcier Consulting

Treatment of Observation
The Problem:
- Research fatigue
- General Intuitions and Specific Hypotheses

Methodology
- An Experimental Approach

Efficacy of Design

Analysis and Findings
- Testing Key Hypotheses
- Regression Model Formulations
- Findings
Total number of respondents in Treatment Survey: 167
Total number of respondents in Final Observation Survey: 574
Total number of treated respondents who were re-captured in the Final Observation Survey: 30
Efficacy: Overview

Total number of respondents in Treatment Survey: 167
Total number of respondents in Final Observation Survey: 574

Total number of treated respondents who were re-captured in the Final Observation Survey: 30
Efficacy: Overview

Total number of respondents in Treatment Survey: 167
Total number of respondents in Final Observation Survey: 574
Total number of treated respondents who were re-captured in the Final Observation Survey: 30
**Table: Reported Timeframe of Most Recent Survey Experience**

<table>
<thead>
<tr>
<th>Timeframe</th>
<th>Not Treated</th>
<th>Potentially Treated</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Surveyed</td>
<td>214 74.56%</td>
<td>204 76.12%</td>
<td>418</td>
</tr>
<tr>
<td>&gt; 6 months</td>
<td>3 1.05%</td>
<td>1 0.37%</td>
<td>4</td>
</tr>
<tr>
<td>5-6 months</td>
<td>1 0.35%</td>
<td>2 0.75%</td>
<td>3</td>
</tr>
<tr>
<td>4-5 months</td>
<td>3 1.05%</td>
<td>3 1.12%</td>
<td>6</td>
</tr>
<tr>
<td>3-4 months</td>
<td>11 3.83%</td>
<td>6 2.24%</td>
<td>17</td>
</tr>
<tr>
<td>2-3 months</td>
<td>18 6.27%</td>
<td>22 8.21%</td>
<td>40</td>
</tr>
<tr>
<td>1-2 months</td>
<td>31 10.80%</td>
<td>22 8.21%</td>
<td>53</td>
</tr>
<tr>
<td>&lt; 1 month</td>
<td>6 2.09%</td>
<td>8 2.99%</td>
<td>14</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>287 100.00%</td>
<td>268 100.00%</td>
<td>555</td>
</tr>
</tbody>
</table>
**Table: Reported Frequency of Survey Participation**

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Not Treated</th>
<th>Potentially Treated</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Surveyed</td>
<td>214 (74.56%)</td>
<td>204 (76.12%)</td>
<td>418</td>
</tr>
<tr>
<td>1</td>
<td>45 (15.68%)</td>
<td>36 (13.43%)</td>
<td>81</td>
</tr>
<tr>
<td>2</td>
<td>19 (6.62%)</td>
<td>22 (8.21%)</td>
<td>41</td>
</tr>
<tr>
<td>3</td>
<td>7 (2.44%)</td>
<td>5 (1.87%)</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>2 (0.70%)</td>
<td>0 (0.00%)</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>0 (0.00%)</td>
<td>1 (0.37%)</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>287 (100.00%)</td>
<td>268 (100.00%)</td>
<td>555</td>
</tr>
</tbody>
</table>
The Problem:

Research fatigue
General Intuitions and Specific Hypotheses

Methodology

An Experimental Approach
Efficacy of Design

Analysis and Findings

Testing Key Hypotheses
Regression Model Formulations
Findings
Hypothesis 1: Higher Refusal Rates

Table: Refusals - Row-wise Percentages

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Not Treated</th>
<th>Potentially Treated</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consent</td>
<td>144 51.25%</td>
<td>137 48.75%</td>
<td>418</td>
</tr>
<tr>
<td>Refuse</td>
<td>22 46.81%</td>
<td>25 53.19%</td>
<td>81</td>
</tr>
<tr>
<td>Total</td>
<td>166 50.61%</td>
<td>162 49.39%</td>
<td>328*</td>
</tr>
</tbody>
</table>

*Smaller sample size due to enumerator error.

- Refusal rate for not treated = 13.3%
- Refusal rate for potentially treated = 15.4%
Hypothesis 1: Higher Refusal Rates

Table: Refusals - Row-wise Percentages

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Not Treated</th>
<th>Treated</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consent</td>
<td>144 51.25%</td>
<td>137 48.75%</td>
<td>418</td>
</tr>
<tr>
<td>Refuse</td>
<td>22 46.81%</td>
<td>25 53.19%</td>
<td>81</td>
</tr>
<tr>
<td>Total</td>
<td>166 50.61%</td>
<td>162 49.39%</td>
<td>328*</td>
</tr>
</tbody>
</table>

*Smaller sample size due to enumerator error.

- Refusal rate for not treated = 13.3%
- Refusal rate for potentially treated = 15.4%
Testing Key Hypotheses

Hypothesis 1: Higher Refusal Rates

Table: Refusals - Row-wise Percentages

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Not Treated</th>
<th>Potentially Treated</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consent</td>
<td>144 51.25%</td>
<td>137 48.75%</td>
<td>418 100%</td>
</tr>
<tr>
<td>Refuse</td>
<td>22 46.81%</td>
<td>25 53.19%</td>
<td>81 100%</td>
</tr>
<tr>
<td>Total</td>
<td>166 50.61%</td>
<td>162 49.39%</td>
<td>328* 100%</td>
</tr>
</tbody>
</table>

*Smaller sample size due to enumerator error.

- Refusal rate for not treated = 13.3%
- Refusal rate for potentially treated = 15.4%
Testing Key Hypotheses

Hypothesis 2: Motivated Underreporting

- Test: Two modules with filter questions that invite motivated underreporting.
  - One module placed in middle of survey and one module at end of survey.

- Regression Models:
  - Logistical regression for binary responses to individual filter questions.
  - OLS regression and ordered logit regressions for summed scores across a given module.
    - Summing binary responses: \( Q_1 + Q_2 + Q_3 \ldots \) etc.
    - Produces ordinal variable.
Hypothesis 2: Motivated Underreporting

- **Test:** Two modules with filter questions that invite motivated underreporting.
  - One module placed in middle of survey and one module at end of survey.

- **Regression Models:**
  - Logistical regression for binary responses to individual filter questions.
  - OLS regression and ordered logit regressions for summed scores across a given module.
    - Summing binary responses: Q1 + Q2 + Q3...etc.
    - Produces ordinal variable.
Testing Key Hypotheses

Hypothesis 3: Inaccurate Responses

▶ Test: Two modules with high recall demand questions:
  ▶ Food consumption.
  ▶ Household consumption.

▶ Regression Models:
  ▶ OLS regression and ordered logit
Testing Key Hypotheses

Hypothesis 3: Inaccurate Responses

- Test: Two modules with high recall demand questions:
  - Food consumption.
  - Household consumption.

- Regression Models:
  - OLS regression and ordered logit
Testing Key Hypotheses

Hypothesis 4: Higher Item Non-response

- **Test:** All modules include opportunity for respondent to express “Don’t know” or “Refuse to respond.”

- **Regression Models:**
  - Ordered logit and negative binomial regressions of non-response:
    - Summed item non-response counts per module (ordinal variable).
    - Total item non-response count in survey (resembles over-dispersed count).
Hypothesis 4: Higher Item Non-response

- Test: All modules include opportunity for respondent to express “Don’t know” or “Refuse to respond.”

- Regression Models:
  - Ordered logit and negative binomial regressions of non-response:
    - Summed item non-response counts per module (ordinal variable).
    - Total item non-response count in survey (resembles over-dispersed count).
The Problem:
Research fatigue
General Intuitions and Specific Hypotheses

Methodology
An Experimental Approach
Efficacy of Design

Analysis and Findings
Testing Key Hypotheses
Regression Model Formulations
Findings
Variables Included

Independent variables:

- Exposure to treatment survey.
- Exposure to other surveys.
  - Frequency.
  - Timeframe (how recent).

Control variables:

- Demographic controls.
- Enumerator fixed effects.

Formulations:

- Individual models.
- Combined models.
- Models with and without adjustments for clustering.
Variables Included

Independent variables:
- Exposure to treatment survey.
- Exposure to other surveys.
  - Frequency.
  - Timeframe (how recent).

Control variables:
- Demographic controls.
- Enumerator fixed effects.

Formulations:
- Individual models.
- Combined models.
- Models with and without adjustments for clustering.
Variables Included

Independent variables:
- Exposure to treatment survey.
- Exposure to other surveys.
  - Frequency.
  - Timeframe (how recent).

Control variables:
- Demographic controls.
- Enumerator fixed effects.

Formulations:
- Individual models.
- Combined models.
- Models with and without adjustments for clustering.
Variables Included

Independent variables:
- Exposure to treatment survey.
- Exposure to other surveys.
  - Frequency.
  - Timeframe (how recent).

Control variables:
- Demographic controls.
- Enumerator fixed effects.

Formulations:
- Individual models.
- Combined models.
- Models with and without adjustments for clustering.
Variables Included

Independent variables:
- Exposure to treatment survey.
- Exposure to other surveys.
  - Frequency.
  - Timeframe (how recent).

Control variables:
- Demographic controls.
- Enumerator fixed effects.

Formulations:
- Individual models.
- Combined models.
- Models with and without adjustments for clustering.
The Problem:
- Research fatigue
- General Intuitions and Specific Hypotheses

Methodology
- An Experimental Approach
- Efficacy of Design

Analysis and Findings
- Testing Key Hypotheses
- Regression Model Formulations

Jonathan Forney
Lead Methodologist, Forcier Consulting

Treatment of Observation
No hypotheses find consistent and significant support across multiple modules and model specifications.

- Cannot rule out insufficient power for experimental treatment as explanation for inability to reject null hypotheses.

- Still, (while not significant) the signs of some effects are consistently the opposite of what hypotheses suggest:
  
  - Hypothesis 2: On average, treated individuals actually show slightly higher endorsement rates for filter questions, contrary to hypothesis 2.
  
  - Hypothesis 4: On average, treated individuals show slightly lower item non-response rates, contrary to hypothesis 4.
Summary

No hypotheses find consistent and significant support across multiple modules and model specifications.

- Cannot rule out insufficient power for experimental treatment as explanation for inability to reject null hypotheses.

- Still, (while not significant) the signs of some effects are consistently the opposite of what hypotheses suggest:
  - Hypothesis 2: On average, treated individuals actually show slightly higher endorsement rates for filter questions, contrary to hypothesis 2.
  - Hypothesis 4: On average, treated individuals show slightly lower item non-response rates, contrary to hypothesis 4.
Findings

Summary

No hypotheses find consistent and significant support across multiple modules and model specifications.

- Cannot rule out insufficient power for experimental treatment as explanation for inability to reject null hypotheses.

- Still, (while not significant) the signs of some effects are consistently the opposite of what hypotheses suggest:
  - Hypothesis 2: On average, treated individuals actually show slightly higher endorsement rates for filter questions, contrary to hypothesis 2.
  - Hypothesis 4: On average, treated individuals show slightly lower item non-response rates, contrary to hypothesis 4.
Lessons Learned

- Need a better strategy for ensuring that treated individuals can be re-captured in the (post-test) observation survey.

- Need to test hypothesis that main effect of over-researching populations is creating class of habitual survey-takers.
  - On average, treated individuals actually had longer (not shorter) survey completion times.

- Maybe this experiment did not produce *enough* fatigue to have a measurable effect:
  - Would ideally have multiple levels/doses of treatment.
Lessons Learned

- Need a better strategy for ensuring that treated individuals can be re-captured in the (post-test) observation survey.

- Need to test hypothesis that main effect of over-researching populations is creating class of habitual survey-takers.
  - On average, treated individuals actually had longer (not shorter) survey completion times.

- Maybe this experiment did not produce enough fatigue to have a measurable effect:
  - Would ideally have multiple levels/doses of treatment.

Jonathan Forney
Lead Methodologist, Forcier Consulting
Lessons Learned

▶ Need a better strategy for ensuring that treated individuals can be re-captured in the (post-test) observation survey.

▶ Need to test hypothesis that main effect of over-researching populations is creating class of habitual survey-takers.
  ▶ On average, treated individuals actually had longer (not shorter) survey completion times.

▶ Maybe this experiment did not produce *enough* fatigue to have a measurable effect:
  ▶ Would ideally have multiple levels/doses of treatment.