Improving Survey Methodologies for Sensitive Questions

Kosuke Imai

Princeton University

July 5, 2017
Seminar Talk at the University of Tokyo

Joint work with
Graeme Blair (UCLA), Jason Lyall (Yale), Bryn Rosenfeld (USC),
and Winston Chou (Princeton)
Introduction

▸ Challenge of measuring sensitive attitudes and behaviors
  ▸ social desirability bias
  ▸ non-response bias

▸ Racial prejudice, corruption, support for controversial political actors

▸ Indirect methods becoming increasingly popular
  ▸ list experiments: aggregation
  ▸ endorsement experiment: priming
  ▸ randomized response: randomization

▸ Development of statistical methods
  ▸ multivariate regression for each survey technique (Bullock, Imai and Shapiro 2010; Imai 2011; Blair & Imai 2012; Blair, Imai & Zhou 2015)
  ▸ using responses as predictors in outcome regression (Imai, Park & Greene 2015)

▸ Empirical validation studies
  ▸ validation against ground truth (Rosenfeld, Imai & Shapiro 2016)
  ▸ prediction of behavior (Hirose, Imai & Lyall 2017)
Two Ways to Improve Sensitive Question Survey Methods

1. Comparing and combining multiple measurements (Blair, Imai & Lyall 2014)
   - Agreement among multiple measurements $\rightsquigarrow$ more credible
   - Combining multiple measurements $\rightsquigarrow$ more powerful
   - Application: Hearts and minds in Afghanistan

2. Using auxiliary information (Chou, Imai, & Rosenfeld 2017)
   - Sometimes aggregate truths are available
     - Turnout rates and voting outcomes
     - Administrative records, e.g., crime and incarceration
   - Use auxiliary information to improve individual-level inference
   - Application: Mississippi anti-abortion referendum
How do we measure civilian attitudes in a conflict setting?

Current efforts in Afghanistan rely on direct questions:

1. USAID (TCAPF): “Who do you believe can solve your problems?”
2. ISAF (ANQAR): “Over the past 6 months, do you think the Taliban have grown stronger, grown weaker, or remained the same?”

Why are direct questions a bad idea?

1. Threats to enumerators and respondents
2. Nonresponse, social desirability bias
3. Interviews are public
4. Danger of selection bias in sampling locations (role of gatekeepers)

ANQAR (November-December 2011): 50% refusal rate
Surveying in the Heartland of Insurgency

Kosuke Imai (Princeton)

Sensitive Survey Questions

July 5, 2017 (U. Tokyo)
Script for the control group:

I’m going to read you a list with the names of different groups and individuals on it. After I read the entire list, I’d like you to tell me how many of these groups and individuals you broadly support, meaning that you generally agree with the goals and policies of the group or individual. Please don’t tell me which ones you generally agree with; only tell me how many groups or individuals you broadly support.

Karzai Government; National Solidarity Program; Local Farmers
Script for the treatment group:

I’m going to read you a list with the names of different groups and individuals on it. After I read the entire list, I’d like you to tell me how many of these groups and individuals you broadly support, meaning that you generally agree with the goals and policies of the group or individual. Please don’t tell me which ones you generally agree with; only tell me how many groups or individuals you broadly support.

Karzai Government; National Solidarity Program; Local Farmers; ISAF
<table>
<thead>
<tr>
<th>response value</th>
<th>Control Group</th>
<th>ISAF Treatment Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>frequency</td>
<td>proportion</td>
</tr>
<tr>
<td>0</td>
<td>188</td>
<td>20.5%</td>
</tr>
<tr>
<td>1</td>
<td>265</td>
<td>28.9</td>
</tr>
<tr>
<td>2</td>
<td>265</td>
<td>28.9</td>
</tr>
<tr>
<td>3</td>
<td>200</td>
<td>21.8</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>918</td>
<td></td>
</tr>
</tbody>
</table>
▶ Script for the control group:

A recent proposal calls for the sweeping reform of the Afghan prison system, including the construction of new prisons in every district to help alleviate overcrowding in existing facilities. Though expensive, new programs for inmates would also be offered, and new judges and prosecutors would be trained. How do you feel about this proposal?

Strongly agree; Agree; Indifferent; Disagree; Strongly disagree; Don’t Know; Refuse to answer
Script for the treatment group:

A recent proposal by ISAF calls for the sweeping reform of the Afghan prison system, including the construction of new prisons in every district to help alleviate overcrowding in existing facilities. Though expensive, new programs for inmates would also be offered, and new judges and prosecutors would be trained. How do you feel about this proposal?

Strongly agree; Agree; Indifferent; Disagree; Strongly disagree; Don’t Know; Refuse to answer
### Data from the Endorsement Experiments

#### Table: Endorsement Experiments Summary

<table>
<thead>
<tr>
<th>Region</th>
<th>Direct Elections</th>
<th>Prison Reform</th>
<th>Independent Election Commission</th>
<th>Anti-Corruption Reform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(N = 2754)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taliban</td>
<td>Strongly agree</td>
<td>Agree</td>
<td>Strongly disagree</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>ISAF</td>
<td>Strongly agree</td>
<td>Agree</td>
<td>Strongly disagree</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>Control</td>
<td>Strongly agree</td>
<td>Agree</td>
<td>Strongly disagree</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>Helmand</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(N = 855)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taliban</td>
<td>Strongly agree</td>
<td>Agree</td>
<td>Strongly disagree</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>ISAF</td>
<td>Strongly agree</td>
<td>Agree</td>
<td>Strongly disagree</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>Control</td>
<td>Strongly agree</td>
<td>Agree</td>
<td>Strongly disagree</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>Khost</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(N = 630)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taliban</td>
<td>Strongly agree</td>
<td>Agree</td>
<td>Strongly disagree</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>ISAF</td>
<td>Strongly agree</td>
<td>Agree</td>
<td>Strongly disagree</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>Control</td>
<td>Strongly agree</td>
<td>Agree</td>
<td>Strongly disagree</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>Kunar</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(N = 396)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taliban</td>
<td>Strongly agree</td>
<td>Agree</td>
<td>Strongly disagree</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>ISAF</td>
<td>Strongly agree</td>
<td>Agree</td>
<td>Strongly disagree</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>Control</td>
<td>Strongly agree</td>
<td>Agree</td>
<td>Strongly disagree</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>Logar</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(N = 486)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taliban</td>
<td>Strongly agree</td>
<td>Agree</td>
<td>Strongly disagree</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>ISAF</td>
<td>Strongly agree</td>
<td>Agree</td>
<td>Strongly disagree</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>Control</td>
<td>Strongly agree</td>
<td>Agree</td>
<td>Strongly disagree</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>Urozgan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(N = 387)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taliban</td>
<td>Strongly agree</td>
<td>Agree</td>
<td>Strongly disagree</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>ISAF</td>
<td>Strongly agree</td>
<td>Agree</td>
<td>Strongly disagree</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>Control</td>
<td>Strongly agree</td>
<td>Agree</td>
<td>Strongly disagree</td>
<td>Strongly disagree</td>
</tr>
</tbody>
</table>

#### Graph: Endorsement Experiments

- **Strongly agree**: Filled black bar
- **Agree**: Light gray bar
- **Indifferent**: Dark gray bar
- **Disagree**: White bar
- **Strongly disagree**: White bar with diagonal lines
- **Don't Know**: Black bar with diagonal lines
- **Refused**: White bar with horizontal lines
Descriptive Comparison: Overall

- Need for validation $\implies$ Multiple measurement strategy
- Two measures should give similar results

**Control Group**

- $\rho = 0.16$
- $\tau = 0.10$

**ISAF Treatment Group**

- $\rho = 0.52$
- $\tau = 0.43$
Descriptive Comparison: Question by Question

ISAF Treatment Group

List Experiment

Direct Elections (p < .01)

\[ \rho = 0.44, \tau = 0.37 \]

Prison Reform (p = 0.26)

\[ \rho = 0.12, \tau = 0.10 \]

Election Commission (p < .01)

\[ \rho = 0.44, \tau = 0.38 \]

Corruption Reform (p < .01)

\[ \rho = 0.50, \tau = 0.42 \]

Control Group

List Experiment

Endorsement Experiment

\[ \rho = 0.18, \tau = 0.14 \]

Endorsement Experiment

\[ \rho = 0.09, \tau = 0.08 \]

Endorsement Experiment

\[ \rho = 0.10, \tau = 0.07 \]

Endorsement Experiment

\[ \rho = 0.04, \tau = 0.03 \]
Comparing and Combining List and Endorse Experiments

- Formal comparison and integration
- What is the probability of supporting ISAF?
  1. List: prob. of saying yes to the sensitive item
  2. Endorsement: prob. of endorsement having a positive effect on support for policy
- These probabilities should be similar!
- They can be estimated with a new multivariate regression method
- Endorsement and list experiments can even be combined for a joint analysis
List Experiments Framework

- \( N \) respondents
- \( J \) control items
- \( T_i \): binary treatment indicator (1 = treatment, 0 = control)
- \( V_i \): pre-treatment covariates
- \( Y_i \): outcome variable

Define a type of each respondent by
- total number of yes for \( J \) control items \( Y_i(0) \)
- truthful answer to the sensitive item \( Z_i^* \): \( Y_i(1) = Z_i^* + Y_i(0) \)
- A total of \((2 \times (J + 1))\) types

<table>
<thead>
<tr>
<th>( Y_i )</th>
<th>Treatment group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>(3,1)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>(2,1) (3,0)</td>
<td>(3,1) (3,0)</td>
</tr>
<tr>
<td>2</td>
<td>(1,1) (2,0)</td>
<td>(2,1) (2,0)</td>
</tr>
<tr>
<td>1</td>
<td>(0,1) (1,0)</td>
<td>(1,1) (1,0)</td>
</tr>
<tr>
<td>0</td>
<td>(0,0)</td>
<td>(0,1) (0,0)</td>
</tr>
</tbody>
</table>
List Experiments Framework

- $N$ respondents
- $J$ control items
- $T_i$: binary treatment indicator ($1 = \text{treatment}, 0 = \text{control}$)
- $X_i$: pre-treatment covariates
- $Y_i$: outcome variable

Define a type of each respondent by
- total number of yes for $J$ control items $Y_i(0)$
- truthful answer to the sensitive item $Z_i^*$: $Y_i(1) = Z_i^* + Y_i(0)$
- A total of $(2 \times (J + 1))$ types

<table>
<thead>
<tr>
<th>$Y_i$</th>
<th>Treatment group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>(3,1)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>(2,1)</td>
<td>(3,0)</td>
</tr>
<tr>
<td>2</td>
<td>(1,1)</td>
<td>(2,0)</td>
</tr>
<tr>
<td>1</td>
<td>(0,1)</td>
<td>(1,0)</td>
</tr>
<tr>
<td>0</td>
<td>(0,0)</td>
<td>(0,0)</td>
</tr>
</tbody>
</table>

- Joint distribution of $(Y_i(0), Z_i^*)$ is identified
Statistical Modeling for List Experiments

- Model for sensitive item: e.g., probit regression

\[ \Pr(Z^*_i = 1 \mid V_i) = \Phi(V_i^\top \delta) \]

- Model for control items given the response to sensitive item: e.g., binomial or beta-binomial probit regression

\[ \Pr(Y_i(0) = y \mid V_i, Z^*_i = z) = J \times \Phi(V_i^\top \psi_z) \]

- Maximum likelihood with the EM algorithm or Bayes with MCMC
Endorsement Experiments Framework

- $N$ respondents
- $J$ policy questions
- $Y_{ij} \in \{0, 1\}$: response of respondent $i$ to policy $j$ (can be ordinal)
- $T_{ij} \in \{0, 1\}$: random endorsement of policy $j$ for respondent $i$
- $V_i$: Covariates measured prior to the treatment
Multiple questions $\implies$ item response theory

$$\Pr(Y_{ij} = 1 \mid T_i = t) = \Phi(\alpha_j + \beta_j(x_i + ts_{ij}^*))$$

- $\alpha_j$: average popularity of policy $j$
- $\beta_j$: how much policy $j$ differentiates pro- and anti-reform respondents
- $x_i$: “ideal point” = how pro-reform respondent $i$ is
- $s_{ij}^*$: endorsement effect

Support level:

$$s_{ij} = \begin{cases} s_{ij}^* & \text{if } \beta_j \geq 0 \\ -s_{ij}^* & \text{otherwise} \end{cases}$$

such that $\frac{\partial}{\partial s_{ij}} \Pr(Y_{ij} = 1 \mid T_{ij} = 1) > 0$
Key quantity: Probability of being a supporter

List experiments:

\[ \Pr(Z_i^* = 1 \mid V_i) = \Phi(V_i^\top \gamma) \]

Endorsement experiments:

\[ \Pr(s_{ij} > 0 \mid V_i) = \Phi(V_i^\top \lambda/\omega) \]

Compare the coefficients: \( \gamma \) and \( \lambda/\omega \)

Combine the two models: \( \gamma = \lambda/\omega \)
Overall Proportion of ISAF Supporters

List  Endorsement  Difference  Combined
(List − Endorse)

Kosuke Imai (Princeton)  Sensitive Survey Questions  July 5, 2017 (U. Tokyo)
Effects of Taliban and ISAF Victimization

-0.25 0.00 0.25 0.50

Effects on Probability of Supporting ISAF

Victimization by Taliban

Victimization by ISAF

Endorse

Combined

List

0.00

-0.25

Kosuke Imai (Princeton)
Estimate voting on anti-abortion referendum using:
- direct question
- list experiment (item/unmatched count technique)
- endorsement experiment
- randomized response

Validate estimates against official election outcome:
- sample from voter history file
- county-level voting recap reports for validation

Case selection:
- a poll conducted 24 hours before the election predicts 44% no votes
- the amendment was defeated 58% to 42%

Findings:
- direct question ↝ significant under-estimation though efficient
- indirect methods ↝ much less biased though more variable
- endorsement and randomized response ↝ least bias
Did you vote YES or NO on the Personhood Initiative, which appeared on the November 2011 Mississippi General Election ballot?

Voted Yes
Voted No
Did not vote
Don’t know
Refused
Bias of the Direct Question

Proportion of 'no' votes on Personhood

Actual vote

Direct question

Kosuke Imai (Princeton)
Here is a list of four things that some people have done and some people have not. Please listen to them and then tell me HOW MANY of them you have done in the past two years. Do not tell me which you have and have not done. Just tell me how many:

Discussed politics with family or friends
Cast a ballot for Governor Phil Bryant
Paid dues to a union
Given money to a Tea Party candidate or organization

(treatment) Voted ‘YES’ on the ‘Personhood’ Initiative

How many of these things have you done in the past two years?
Endorsement Experiment

We’d like to get your overall opinion of some people in the news. As I read each name, please say if you have a very favorable, somewhat favorable, somewhat unfavorable, or very unfavorable opinion of each person.

(control) Phil Bryant, Governor of Mississippi?

(treatment) Phil Bryant, Governor of Mississippi, who campaigned in favor of the ‘Personhood’ Initiative on the 2011 Mississippi General Election ballot?
To answer this question, you will need a coin. Once you have found one, please toss the coin two times and note the results of those tosses (heads or tails) one after the other on a sheet of paper. Do not reveal to me whether your coin lands on heads or tails. After you have recorded the results of your two coin tosses, just tell me you are ready and we will begin.

Now, please answer ‘yes’ if either your second coin toss came up heads or you voted ‘YES’ on the Personhood Initiative, which appeared on the November 2011 Mississippi General Election ballot.

Yes

No

Don’t know

Refused
Pooled Analysis

Estimated proportion of 'no' votes on Personhood

- Direct Question (n = 2,655)
- List Experiment (n = 1,352)
- Endorsement Experiment (n = 1,841)
- Randomized Response (n = 943)

- Unweighted
- Weighted
- Regression Adjusted

actual vote share

Kosuke Imai (Princeton)  Sensitive Survey Questions  July 5, 2017 (U. Tokyo)
County-level Analysis

Unweighted

Weighted

Regression Adjusted

Direct Question

List Experiment

Endorsement Experiment

Randomized Response

bias = 0.249
RMSE = 0.254

bias = 0.170
RMSE = 0.183

bias = 0.016
RMSE = 0.102

bias = −0.022
RMSE = 0.074

bias = 0.215
RMSE = 0.224

bias = 0.144
RMSE = 0.161

bias = −0.016
RMSE = 0.139

bias = 0.030
RMSE = 0.141

bias = −0.042
RMSE = 0.119

bias = 0.198
RMSE = 0.204

bias = 0.124
RMSE = 0.142

bias = 0.118
RMSE = 0.141

bias = 0.010
RMSE = 0.067
The Proposed Methodology for List Experiment

- List experiment can be analyzed by method of moments:

\[
\mathbb{E}(Y_i \mid T_i, X_i) = \underbrace{f(X_i, \gamma)}_{\text{Control Items}} + T_i \underbrace{g(X_i, \delta)}_{\text{Sensitive Trait}}
\]

- We simply add moment conditions of the form

\[
\mathbb{E}[g(X_i, \delta)] = h \\
\mathbb{E}[g(X_i, \delta) \mid G_i = k] = h_k
\]

- A similar strategy works for randomized experiment

- (Testable) Assumption: Same parameters solve all moment conditions
  \[\leadsto\] Constant parameters across groups.
Proposed Methodology for Endorsement Experiment

1. Use precinct-level indicators as covariates

\[ s_{ij}^{\text{indep.}} \sim \mathcal{N}(\lambda^\top X_i, \omega^2), \]

2. Assume the prior for precinct random effects \( \lambda_r \):

\[ \lambda_r^{\text{indep.}} \sim \mathcal{N}(\mu_{\lambda_{\text{county}[r]}}, \sigma_{\text{county}[r]}^2), \]

where \( \text{county}[r] \) denotes the county containing precinct \( r \).

3. Choose \( \mu_{\lambda_{\text{county}[r]}} \) to match auxiliary information

- We assume the following conjugate prior for \( \omega^2 \):

\[ \omega^2 \sim \kappa/\chi^2_{\nu} \]

- The marginal prior for \( s_{ij}^* \) is a \( t \)-distribution:

\[ s_{ij}^* \mid G_i = k^{\text{indep.}} \sim t_{\nu}(\mu_{\lambda_k}, \sigma_k^2). \]
## Efficiency Comparison with Direct Questioning

<table>
<thead>
<tr>
<th></th>
<th>List Experiment</th>
<th>Randomized Response</th>
<th>Endorsement Experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=1,325</td>
<td>N=818</td>
<td>N=1,841</td>
</tr>
<tr>
<td>s.e.</td>
<td>ratio</td>
<td>s.e.</td>
<td>ratio</td>
</tr>
<tr>
<td>Direct questioning</td>
<td>0.017</td>
<td>0.021</td>
<td>0.289</td>
</tr>
<tr>
<td>No auxiliary info.</td>
<td>0.067</td>
<td>3.963</td>
<td>0.040</td>
</tr>
<tr>
<td>With auxiliary info.</td>
<td>0.019</td>
<td>1.150</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Auxiliary Information Improves List Experiment

Without Auxiliary Information

- Bias = -0.005
- RMSE = 0.102
- Cor = 0.653

With Auxiliary Information

- Bias = 0.042
- RMSE = 0.098
- Cor = 0.774
Auxiliary Information Improves Endorsement Experiment

\[ \text{bias} = 0.136 \]
\[ \text{RMSE} = 0.212 \]
\[ \text{cor} = 0.249 \]

\[ \text{bias} = -0.005 \]
\[ \text{RMSE} = 0.102 \]
\[ \text{cor} = 0.653 \]
Without Auxiliary Information

**Actual**

- Bias: 0.015
- RMSE: 0.11
- Correlation: 0.639

**Actual**

- Bias: 0.042
- RMSE: 0.098
- Correlation: 0.774
Auxiliary Information Improves Multivariate Inference

Estimated proportion of 'no' votes on Personhood

<table>
<thead>
<tr>
<th>Gender</th>
<th>Party</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Republican</td>
<td>No Higher</td>
</tr>
<tr>
<td>Female</td>
<td>Democrat</td>
<td>Higher</td>
</tr>
</tbody>
</table>

- Direct Questioning
- List Experiment
- List Experiment with Auxiliary Information
- Randomized Response
- Randomized Response with Auxiliary Information

Kosuke Imai (Princeton)  Sensitive Survey Questions  July 5, 2017 (U. Tokyo)  38 / 39
Concluding Remarks

- Direct question is severely biased
- All indirect methods reduce bias:
  - Endorsement and randomized response $\rightsquigarrow$ least bias
  - List experiment $\rightsquigarrow$ ceiling/floor effects, design effects
  - Ease of implementation: list $>$ endorse $>$ randomized response
- But, they are inefficient: bias-variance tradeoff
- Two ways to improve indirect question methods:
  1. Use of multiple-measurement strategies when truth is not available
  2. Use aggregate-level truth to improve individual-level estimates
- Open-source software:
  - list for list experiment (Blair, Imai & Park)
  - endorse for endorsement experiment (Shiraito & Imai)
  - rr for randomized response (Blair, Imai & Zhou)