Comparing and Combining List and Endorsement Experiments: Evidence from Afghanistan

Kosuke Imai

Princeton University
Joint work with G. Blair and J. Lyall

March 1, 2013
University of Michigan Political Science Department

Methodological Motivation

- Survey is used widely in social sciences
- Validity of survey depends on the accuracy of self-reports

- Sensitive questions $\Longrightarrow$ social desirability, privacy concerns
- Racial prejudice, corruption, support for political actors
- Lies and nonresponses $\Longrightarrow$ potential bias

- Survey "experiments" as a solution
  1. Randomization: Randomized response method
  2. Aggregation: List experiment (item count technique)
  3. Cueing: Endorsement experiment

- Two problems of indirect measures and proposed solutions:
  1. Measurement error $\Longrightarrow$ *comparing* two measures
  2. Statistical inefficiency $\Longrightarrow$ *combining* two measures

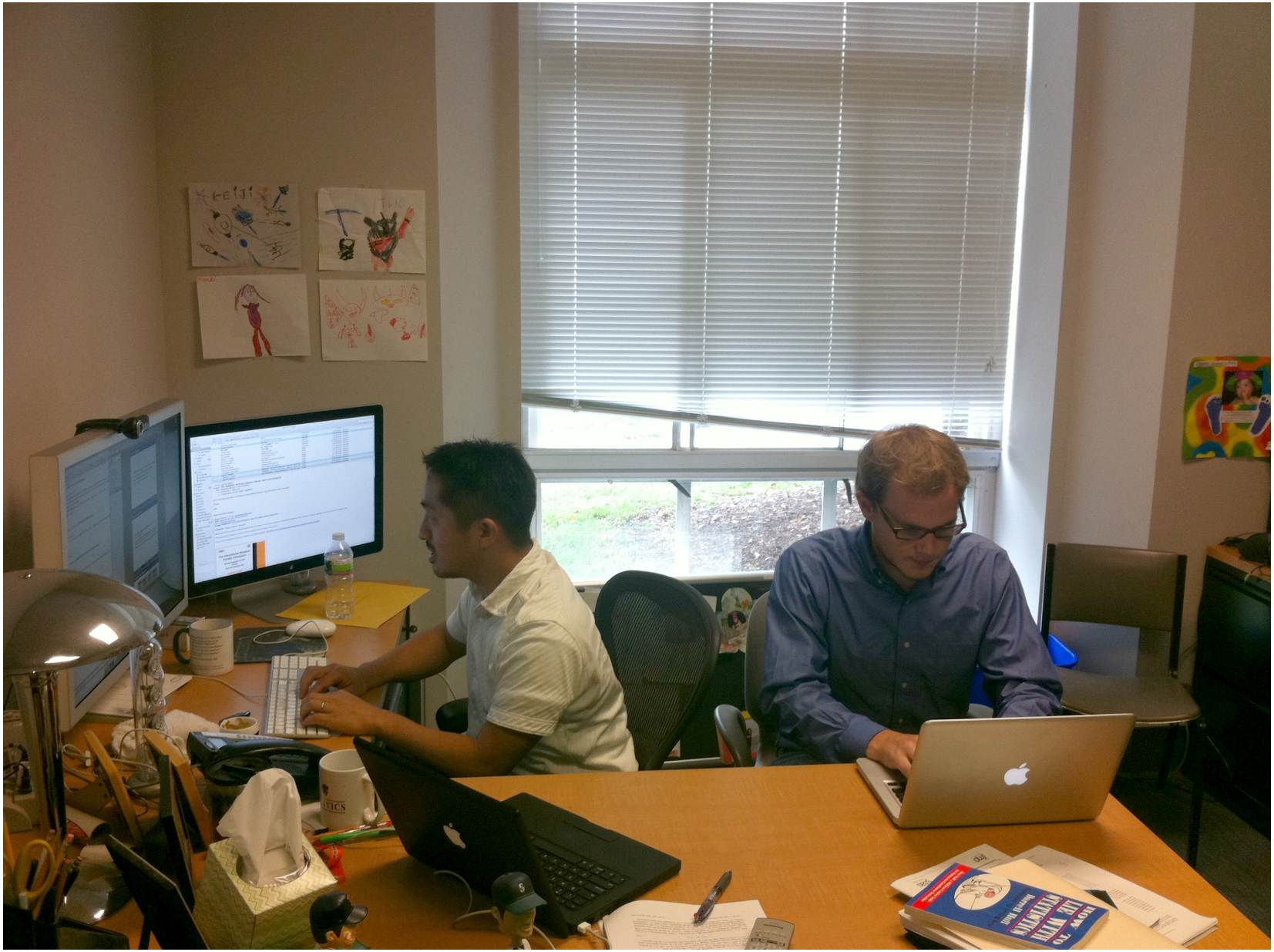
Theoretical and Substantive Motivation

- How do we measure "hearts and minds" 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

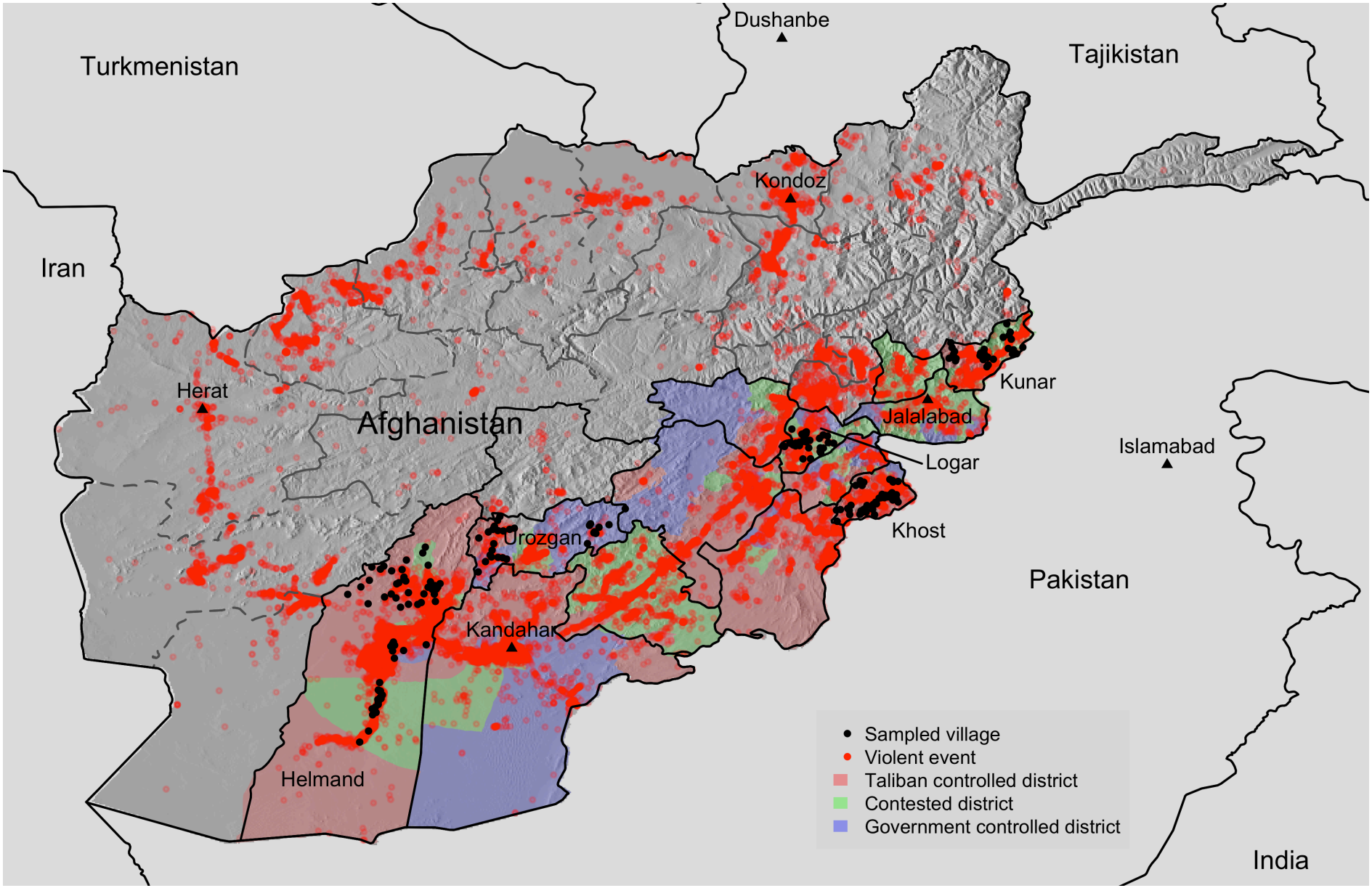
Public Nature of Interviews

Negotiated Access

A Battlefield in Princeton, New Jersey



Sampling in the Heartland of Insurgency



List Experiments

- 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

List Experiments

- 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

The Data from the List Experiment

response value	Control Group frequency	Control Group proportion	ISAF Treatment Group frequency	ISAF Treatment Group proportion
0	188	20.5%	174	19.0%
1	265	28.9	278	30.3
2	265	28.9	260	28.3
3	200	21.8	182	19.8
4			24	2.6
Total	918		918	

Identification Assumptions

1. No Design Effect: The inclusion of the sensitive item does not affect answers to control items
2. No Liars: Answers about the sensitive item are truthful

Endorsement Experiments

- 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
```

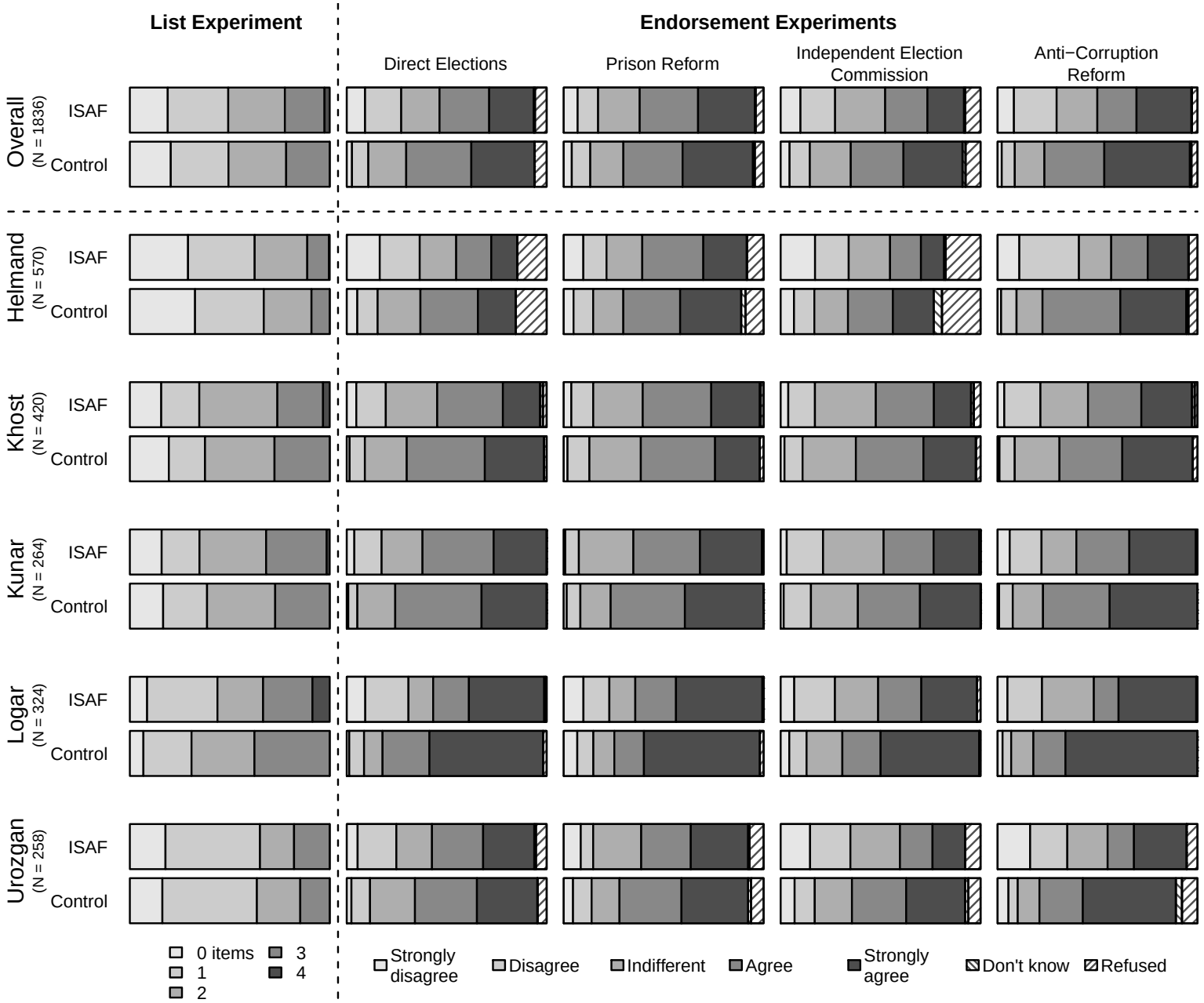
Endorsement Experiments

- 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



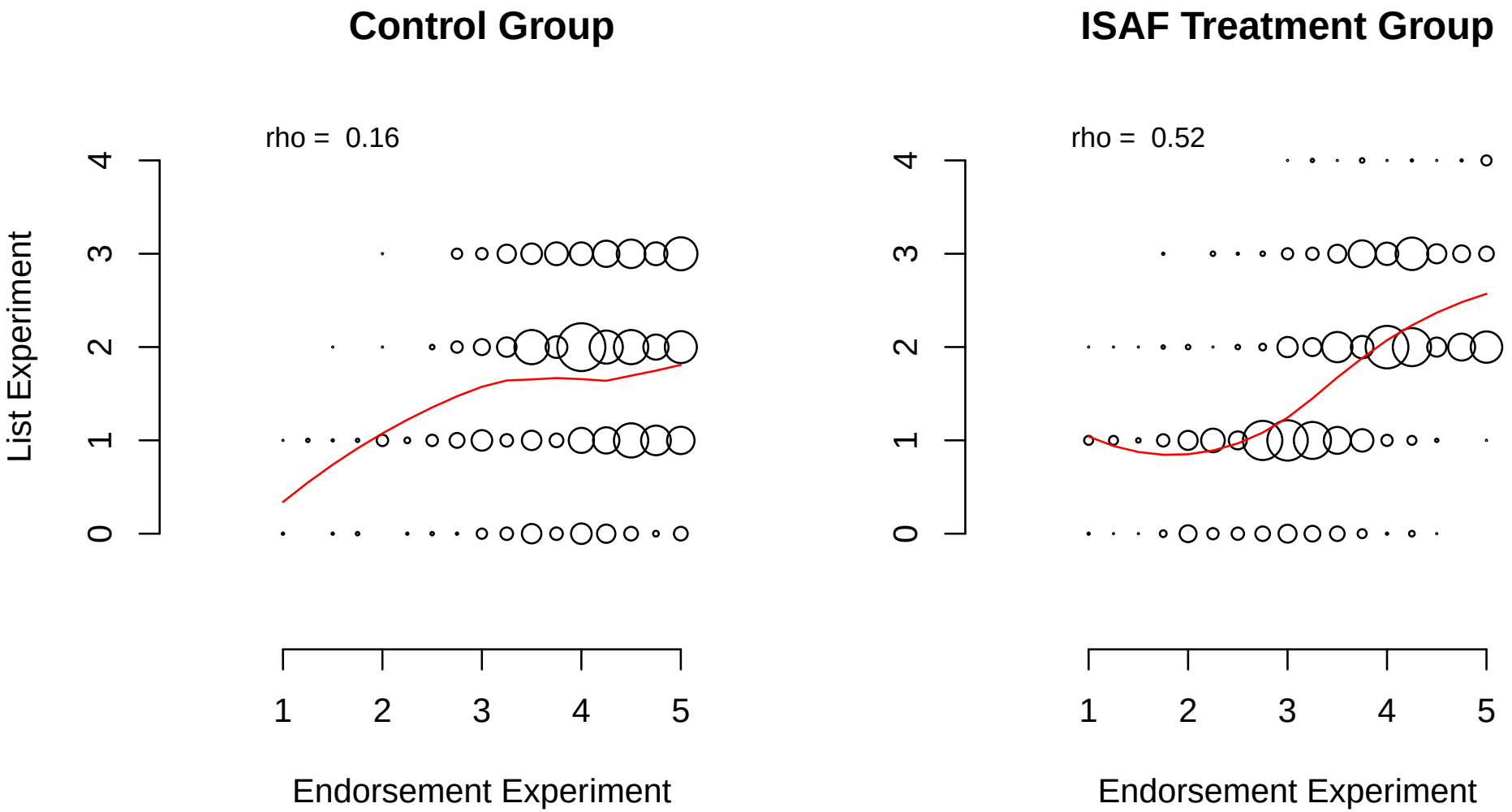
Assumption and Interpretation

- More indirect than list experiments
- Easier to implement but harder to interpret

- When can we interpret endorsement effects as support (or affinity) for endorser?
  - Endorsements have no influence on respondents' interpretation of policy questions. No learning
  - I'm a hardcore Democrat but don't know much about this traditionally democratic policy. You now tell me even a Republican supports it and so the policy must be really good

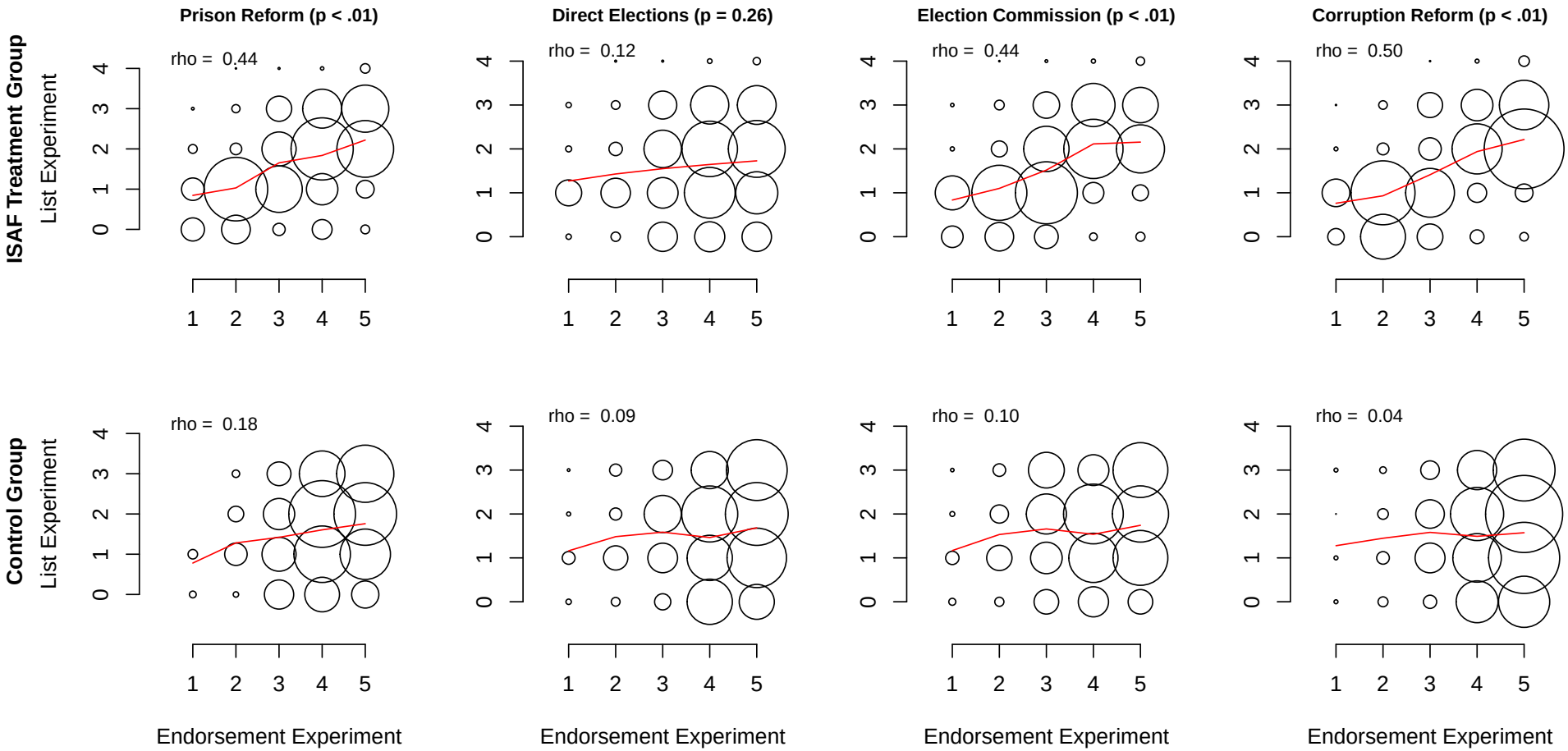
- Some considerations when designing endorsement experiments:
  1. Policies must belong to the same policy dimension
  2. Endorsements must be credible
  3. Few respondents with extreme views

Descriptive Comparison: Overall

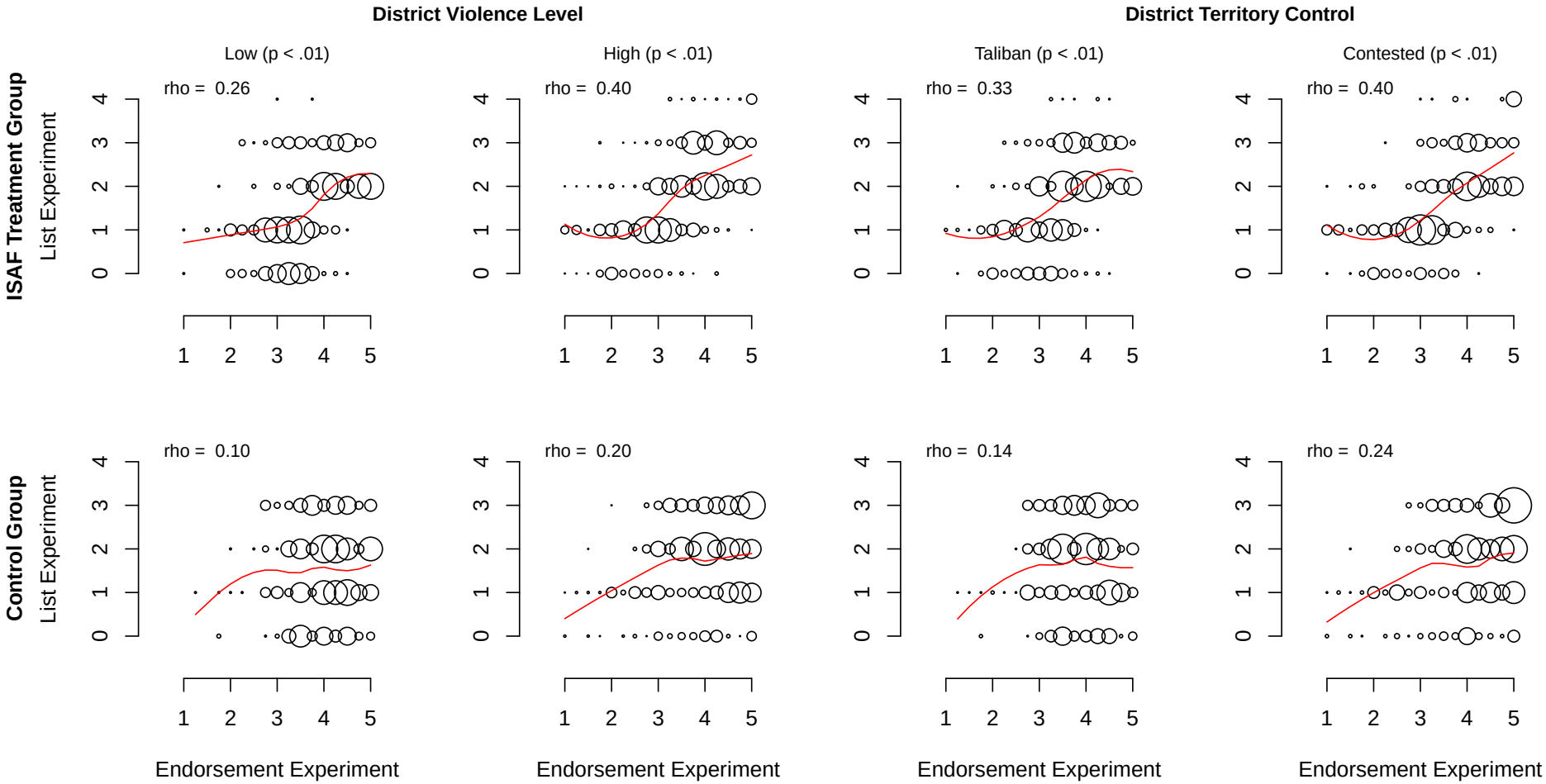


- A statistical test: $H_0 : \rho_0 = \rho_1$ and $H_1 : \rho_0 < \rho_1$ with bootstrap

Descriptive Comparison: Question by Question



Descriptive Comparison: Violence & Territorial Control



Models for List and Endorsement Experiments

- LIST EXPERIMENTS (Imai 2011, *JASA*; Blair & Imai 2012, *PA*):
  1. Likelihood framework with missing data
  2. Assumptions: no design effect, no liar
  3. Latent variable modeling for support

- ENDORSEMENT EXPERIMENTS (Bullock, Imai & Shapiro 2011, *PA*):
  1. Item response theory to combine multiple questions
  2. Assumptions: single policy dimension, no learning
  3. Latent variable modeling for support

- 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!

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

Y_i	Treatment group	Control group
4	(3,1)	
3	(2,1) (3,0)	(3,1) (3,0)
2	(1,1) (2,0)	(2,1) (2,0)
1	(0,1) (1,0)	(1,1) (1,0)
0	(0,0)	(0,1) (0,0)

List Experiments Framework

- N respondents
- J control items
- T_i: binary treatment indicator (1 = treatment, 0 = 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

Y_i	Treatment group	Control group
4	(3,1)	
3	(2,1) (3,0)	(3,1) (3,0)
2	(1,1) (2,0)	(2,1) (2,0)
1	~~(0,1)~~ ~~(1,0)~~	(1,1) ~~(1,0)~~
0	~~(0,0)~~	~~(0,1)~~ ~~(0,0)~~

  - *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
- For the Afghan experiment, an individual receives the same treatment across policies $T_i = T_{ij}$
- V_i: Covariates measured prior to the treatment

Statistical Modeling for Endorsement Experiments

- Multiple questions $\Longrightarrow$ item response theory

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

  - α_j: average popularity of policy j
  - β_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 $\frac{\partial}{\partial s_{ij}} \Pr(Y_{ij} = 1 \mid T_{ij} = 1) > 0$

- Hierarchical model of support:

$$s_{ij} \overset{\text{indep.}}{\sim} \mathcal{N}(V_i^\top \lambda, \omega^2)$$

Comparing and Combining the Two Models

- Key quantity: Probability of being a supporter
- List experiments:

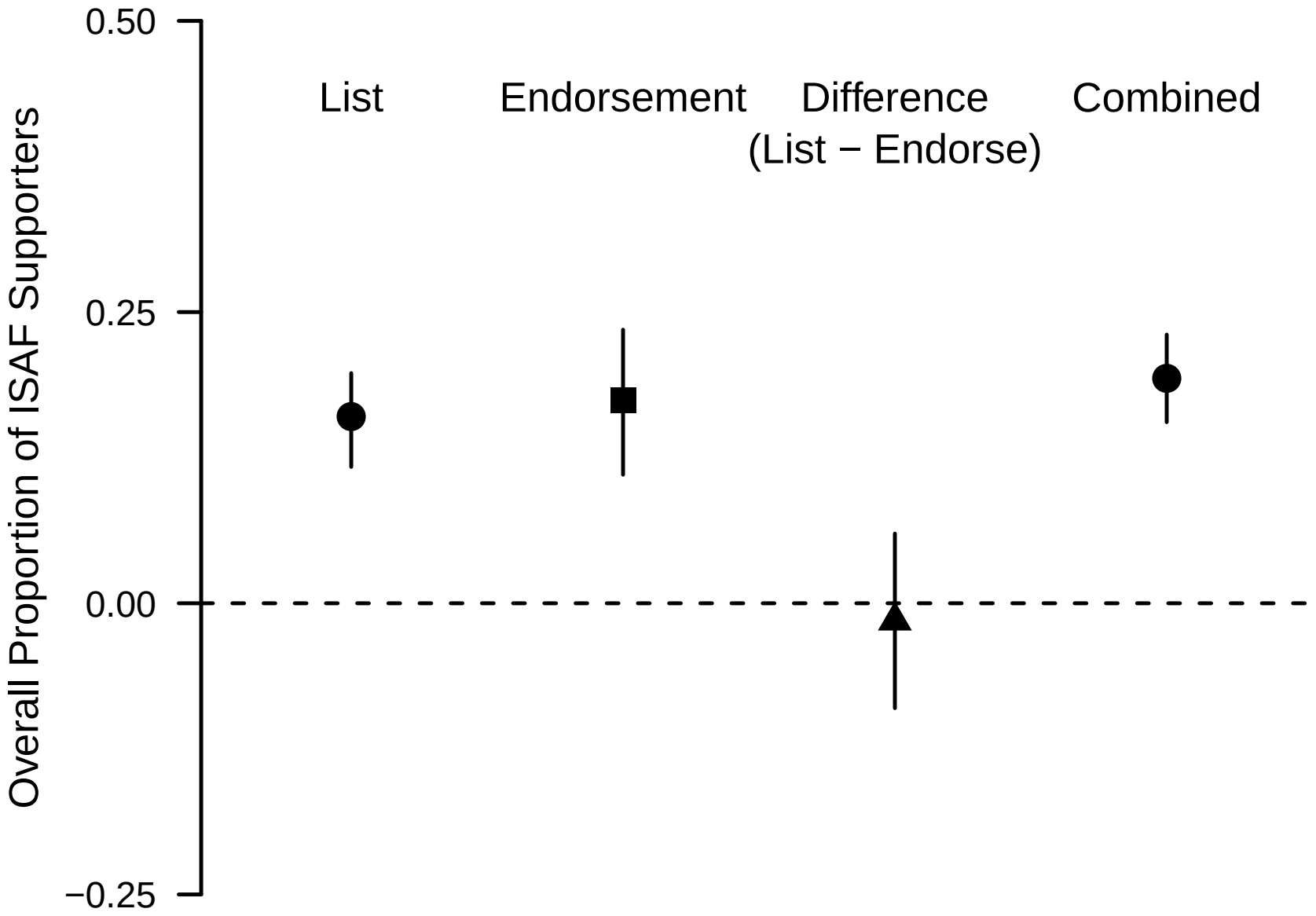
$$\Pr(Z_i^* = 1 \mid V_i) \quad = \quad \Phi(V_i^\top \gamma)$$

- Endorsement experiments:

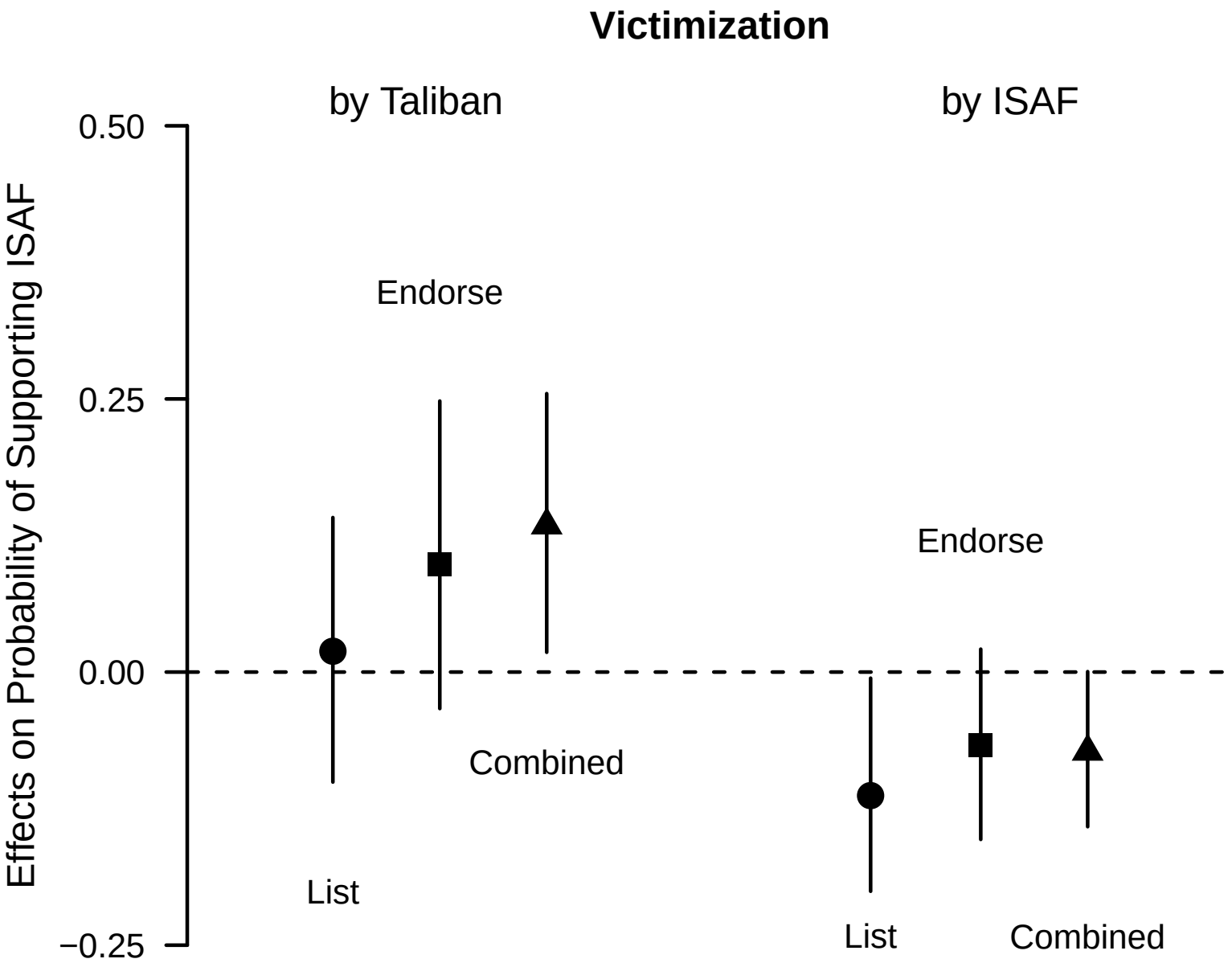
$$\Pr(s_{ij} > 0 \mid V_i) \quad = \quad \Phi(V_i^\top \lambda/\omega)$$

- Compare the coefficients: γ and λ/ω
- Combine the two models: $\gamma = \lambda/\omega$

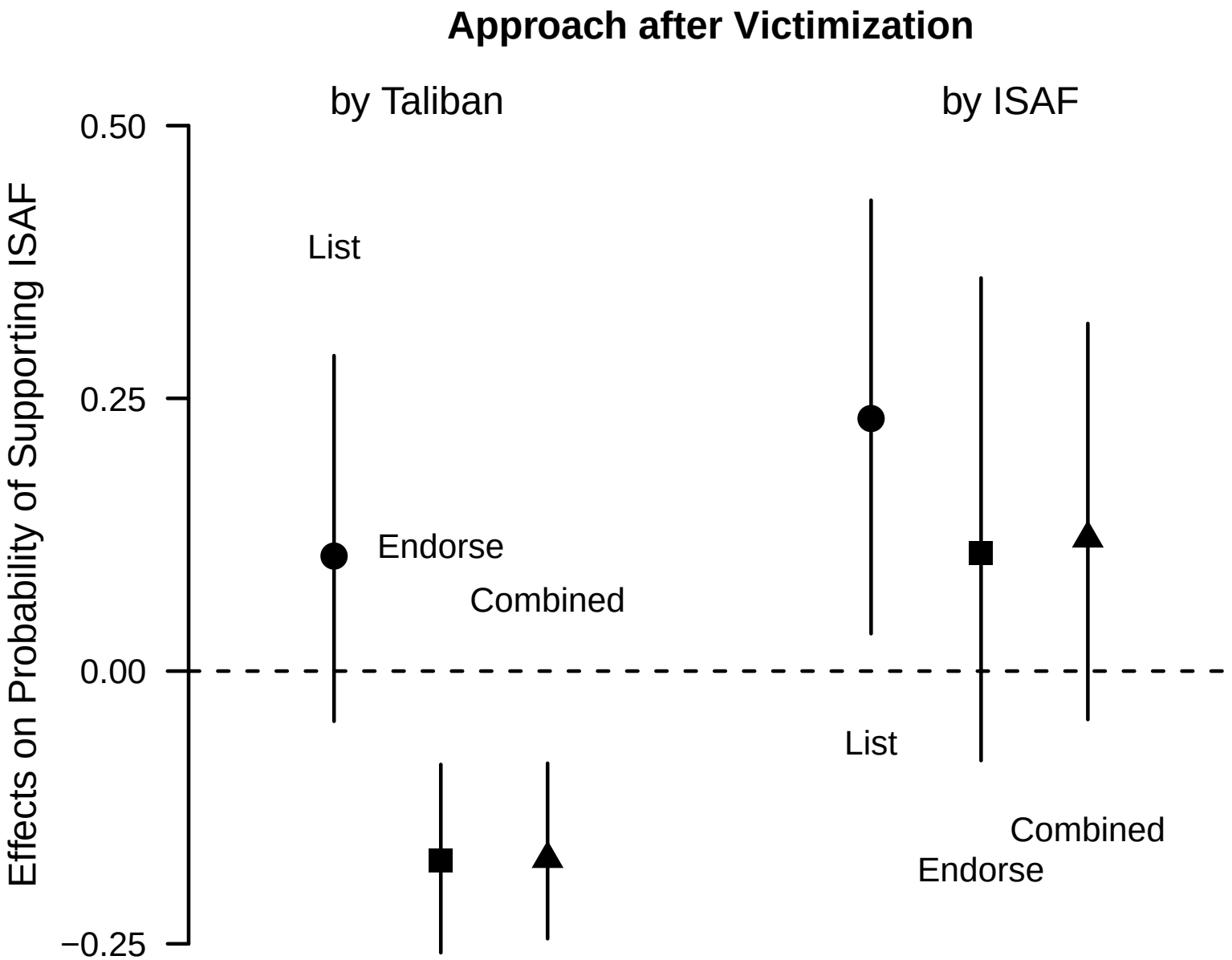
Overall Proportion of ISAF Supporters



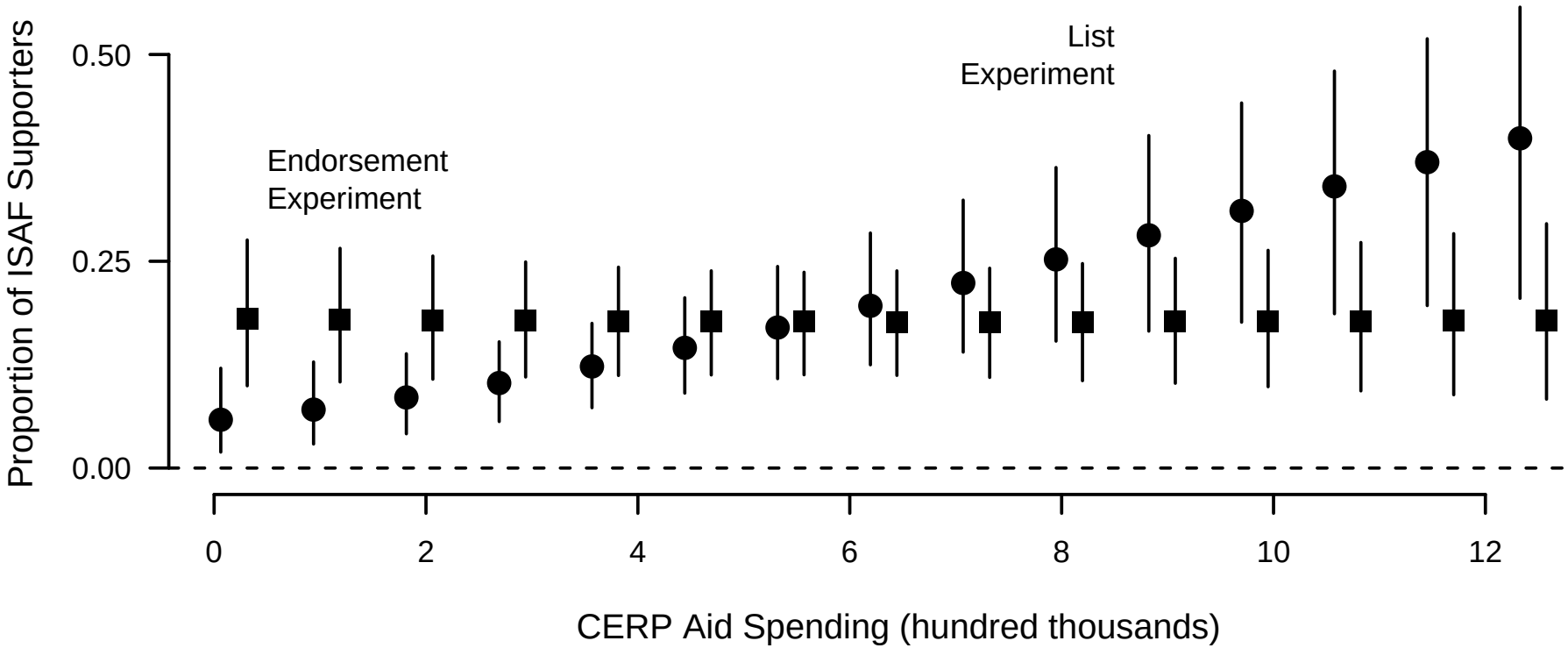
Effects of Taliban and ISAF Victimization



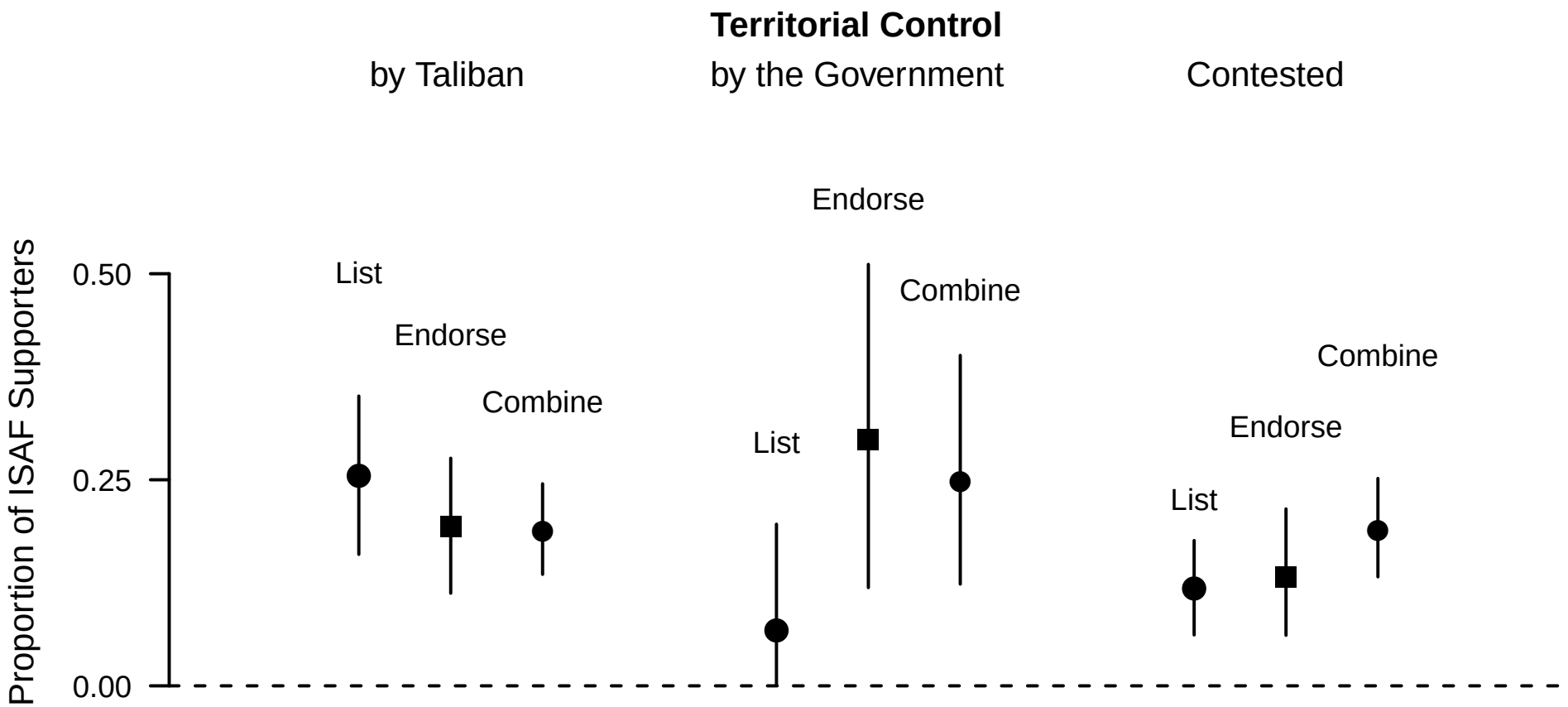
Effects of Taliban/ISAF Post-Harm Mitigation Efforts



Effects of CERP Aid Spending



Proportion of ISAF Supporters by Territorial Control



Concluding Remarks

- Challenges of eliciting truthful responses to sensitive questions
- List and endorsement experiments: indirect questioning methods

- Need for validation $\Longrightarrow$ multiple measurement strategy
- Statistical methods for comparing and combining list and endorsement experiments
- Open-source software `list` and `endorse` for implementation

- Practical suggestions:
  1. Randomize the treatment across, not within, respondents
  2. List experiments are more prone to social desirability bias than endorsement experiments
  3. Multiple pre-tests and focus groups

The project website for papers and software:

http://imai.princeton.edu/projects/sensitive.html

Email for comments and suggestions:

kimai@princeton.edu