Survey Methods for Sensitive Topics

by Graeme Blair
Columbia University

When comparativists rely on survey data, they implicitly invoke an important assumption: that respondents answered truthfully. This assumption may be violated when there are *incentives to conceal the truth*. Social pressures, fears of retaliation, and possible legal sanctions may drive respondents to answer questions in the least revealing way rather than entirely honestly. These incentives can operate when the topic is unmistakably sensitive — for example, in surveys about participation in protests or support for an unpopular policy — but they may also affect seemingly innocuous subjects like voter turnout. When there are incentives to conceal, our inferences about respondents (e.g. what proportion of them shared information with a militant) will be biased. Moreover, in some cases it will be difficult to determine the size or even the direction of the misreporting bias.

Incentives to conceal may even affect whether respondents agree to participate in a survey or to offer a response to a sensitive survey question. If the responses of participants — even truthful ones — differ from those of non-participants, then inferences about the population from the survey data (e.g. what proportion of Nigerians shared information with a militant) will suffer from nonresponse bias.

What can we do about these misreporting and nonresponse biases? In what follows, I review four survey techniques used by comparativists to address incentives to conceal truthful responses.[1] I first review survey administration practices designed to protect sensitive responses. For contexts in which these are insufficient, I review three experimental methods that can be used in addition that avoid soliciting exact answers to sensitive questions altogether. The experimental methods enable comparativists to ask survey questions that could not otherwise be asked due to ethical concerns and the risk of bias. However, these methods require additional assumptions that are often not testable, necessitating careful design and pilot testing. I conclude with a discussion of common critiques of the experimental techniques.

I. Survey Administration Protections

The first and most common approach is to build trust with respondents by implementing and communicating measures to protect sensitive answers through changes in how the responses are collected and stored. If respondents find these confidentiality measures convincing, they may be more likely to respond, and to respond truthfully. Specific approaches used by researchers include:

- Separating sensitive items from names, contact details, and questions that identify individuals such as exact age or family size (permanently, or via codes that can only be accessed by the researcher);
- Locking up paper surveys or encrypting electronic surveys;
- Ensuring that interviews take place in private locations without bystanders;
- Using interviewers who share the age/social group/gender/etc. of the respondent to address fears of outsiders asking sensitive questions;
- Self-administration on paper; via a recording (Chauchard, 2013); by touch-tone telephone; or on a smartphone, tablet, or computer.

In each case, the sensitive survey questions themselves are not modified. As a result, the great advantage to these measures is the simplicity of interpretation and analysis. The "yes" and "no" answers to a sensitive question can be directly analyzed with standard techniques such as means and logit or probit regressions. No additional assumptions beyond truthful responses are required.

II. Randomized Response Technique

The randomized response technique protects respondents by introducing random noise into their responses, so the responses could either reflect a truthful answer to the sensitive question of interest or an irrelevant response. To implement this technique, several design variants exist (Blair, Imai and Zhou, 2015). For example, in the forced response design, a randomization device such as a coin or a die directs each respondent to either automatically answer "yes" or "no," or to answer the question of interest truthfully. An example on estimating support for coalition forces in Afghanistan illustrates the technique (Blair, Imai and Zhou, 2015):

[1] I will discuss how they can enable researchers to solicit truthful answers to binary "yes" or "no" questions, though most of these techniques can be extended to other kinds of outcomes such as numerical responses.

For this question, I want you to answer yes or no. But I want you to consider the number of your dice throw. If 1 shows on the dice, tell me no. If 6 shows, tell me yes. But if another number, like 2 or 3 or 4 or 5 shows, tell me your own opinion about the question that I will ask you after you throw the dice. [TURN AWAY FROM THE RESPONDENT] Now you throw the dice so that I cannot see what comes out. Please do not forget the number that comes out. [WAIT TO TURN AROUND UNTIL RESPONDENT SAYS YES TO:] Have you thrown the dice? Have you picked it up? Now, during the height of the conflict in 2007 and 2008, did you know any militants, like a family member, a friend, or someone you talked to on a regular basis? Please, before you answer, take note of the number you rolled on the dice.

Individual responses are protected because a "yes" or "no" answer may be a truthful answer or it may indicate that the respondent rolled a 1 or a 6. To identify how many respondents said "yes" to the sensitive question, the "yes" and "no" responses are combined with properties of the randomizing device (i.e. a standard die has approximately a 1/6 probability of landing on each side).

Noise can also be introduced in other ways. In the mirrored question design, a coin or die rolled by the respondent in private assigns her to answer either the sensitive question or its inverse ("yes" and "no" flipped). In the unrelated question design, respondents answer the sensitive question or an unrelated innocuous question.

The chief advantage of the randomized response technique is that no individual response can be exactly identified by anyone. In addition, the level of protection — how much can be learned about sensitive individual responses — is directly controlled by the researcher, who chooses the randomizing device (coin, dice, spinner, etc.). The strong protection and control do come at a cost: compliance with the instructions is assumed, and this may be a strong assumption with low-education respondents or in contexts in which saying "yes" is itself sensitive. There are, however, designs and models to adjust for non-compliance.

For design and analysis advice, see Gingerich (2010) and Blair, Imai and Zhou (2015).

III. List Experiments

The list experiment conceals individual responses to a sensitive survey item by aggregating those responses with responses to several other control questions. The respondent replies to a list experiment question with a count of the number of "yes" responses to a list of questions that includes the sensitive item. In this way, each individual's "yes" or "no" response to the sensitive question is hidden within the "yes" and "no" responses to other questions.

A second experimental group is used to estimate the proportion of respondents who said "yes" to the sensitive item (the quantity of interest). In this group, an identical question is asked, except that the list excludes the sensitive item. The average response in this control group is subtracted from the average response to the original question to identify the proportion of respondents who said "yes" to the sensitive item. An example on estimating support for coalition forces in Afghanistan illustrates the technique (Blair, Imai and Lyall, 2014):

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.

Sensitive Item Condition
Karzai Government
Foreign Forces
National Solidarity Program
Local Farmers

Control Item-Only Condition
Karzai Government
National Solidarity Program
Local Farmers

Now, please tell me how many of these groups or individuals do you broadly support?

Several list experiment applications have revealed increased respondent willingness to answer sensitive questions. However, a key disadvantage of the list experiment is that some individual responses are not protected. A respondent in the treatment group who answers that she supports all four groups is identified as a supporter of foreign forces in the example and a respondent who says no groups is identified as a non-supporter of foreign forces. Neither attitude is protected. As a result, some respondents may answer dishonestly to avoid having their views identified. There are methods to detect and adjust for this behavior (Blair and Imai, 2012), but the lack of complete protection may make all respondents more cautious.

For applications in comparative politics, see Corstange (2009), Gonzalez-Ocantos et al. (2012), and Meng, Pan and Yang (Forthcoming). For design and analysis advice, see Corstange (2009), Imai (2011), Blair and Imai (2012), Kramon and Weghorst (2012), Glynn (2013), Aronow et al. (2013), and Imai, Park and Greene (Forthcoming).

IV. Endorsement Experiments

The endorsement experiment is useful for measuring attitudes toward a political actor such as an elected official (see Rosenfeld, Imai and Shapiro (Forthcoming) for an inversion of the design to measure attitudes toward a policy). The method protects individual attitudes toward the sensitive actor by combining them with attitudes towards one or more public policies. In the endorsement condition, randomly assigned respondents are asked a question about a policy and told that the sensitive political actor endorses the policy. An individual's affect toward the actor is protected because a positive response could reflect affinity for either the policy or the actor.

To separate affect toward the actor from policy preferences, randomly assigned respondents are asked an identical question, but without the endorsement. This identifies policy preferences alone. Affect toward the actor is identified by subtracting average policy preferences from the average response to the endorsement question. This is the "endorsement effect." An example from Lyall, Blair and Imai (2013) that solicits attitudes toward the Taliban illustrates:

Policy-Only Condition
A recent proposal …

Endorsement Condition
A recent proposal **by the Taliban** …

In both conditions underlined above:

…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? Do you strongly agree, somewhat agree, are you indifferent, do you disagree, or do you strongly disagree with this policy?

The endorsement technique is perhaps the most protective of the methods because the direct sensitive question is not asked. Endorsement questions are also easy for enumerators and respondents to understand. There are two main downsides. First, the level of protection for respondents depends on the policies. In the example above, if most respondents disagree with the prison policy, an answer of "strongly agree" would indicate Taliban support. Second, the indirect nature of the question means that the magnitudes of endorsement effects are not directly interpretable without a behavioral assumption.

Examples of applications in comparative politics include Blair et al. (2013) and Lyall, Blair and Imai (2013). For design and analysis advice, see Bullock, Imai and Shapiro (2011) and Blair, Imai and Lyall (2014).

> By mitigating the biases from incentives to conceal truthful responses, these four methods enable comparativists to ask questions that could typically only be asked before by building trust over long periods of time with small pools of respondents.

V. Analysis of the Experimental Methods

Regression and other standard analyses can easily be conducted for the randomized response, list, and endorsement techniques using free software in the **R** statistical environment with the "*rr*," "*list*," and "*endorse*"

packages (Blair, Imai and Zhou, 2015; Blair, Zhou and Imai, 2015*b*; Blair and Imai, 2010; Shiraito and Imai, 2013). The *rr* and *list* packages also enable researchers to use randomized response and list experiment questions as predictors in a regression. Analysis is no more complicated than running a regression in STATA or **R**. For example, the list experiment regression command is:

```
ictreg( y.variable ~ x.variable, treat =
"treatment.variable", data = my.data )
```

VI. Critiques of the Experimental Methods

I now discuss five critiques that have been leveled at the experimental methods described above.

1. *There is still no incentive for respondents to answer truthfully*. This is an empirical question, and one for which there is not yet conclusive evidence. Nevertheless, in the small number of validation studies that have taken place, respondents have been more forthcoming with sensitive information and the estimates of known population parameters were closer to the truth using some of these methods compared to questions without protections (Rosenfeld, Imai and Shapiro, Forthcoming).

2. *The list experiment and randomized response technique are confusing to enumerators and respondents*. This is undoubtedly a problem. Carefully developing and pretesting instructions for respondents, directly training all enumerators, and identifying points of confusion by asking respondents a practice question can help.

3. *Respondents see through the designs and do not comply with them to avoid any risk*. An example illustrates this issue: in a list experiment in Afghanistan, zero of 2,754 respondents reported that they supported none or all of the groups mentioned — and these are the two responses that are not protected by the design (Blair, Imai and Lyall, 2014). This problem may be avoided by piloting different designs to find one that provides sufficient protection to encourage participation and compliance.

4. *Low power*. Each method requires a larger sample size than do direct questions. To improve power, researchers can use the "double list experiment" (Glynn, 2013), ask each control question separately in the list experiment control group (Corstange, 2009), ask multiple policy questions for the endorsement experiment, or combine responses from multiple measurements (see Aronow et al. (2013) on direct questions and list experiments, and Blair, Imai and Lyall (2014) on list and endorsement experiments).

5. *Difficult to design*. Intensive fieldwork is needed to identify appropriate question designs and instructions. Multiple pilot tests are also often needed to design the control items for list experiments and the policies for the endorsement experiment.

VII. Discussion

By mitigating the biases from incentives to conceal truthful responses, these four methods enable comparativists to ask questions that could typically only be asked before by building trust over long periods of time with small pools of respondents. The choice of which method to use will depend on the context, but the choice can be informed by the theoretical differences discussed here and by careful pilot testing. Moreover, the methods are not mutually exclusive. Survey administration protections can be combined with the experimental methods, and multiple experimental methods can be used for measurement (Blair, Imai and Lyall, 2014). Future research should validate these methods in contexts studied by comparativists in order to increase confidence in them and provide more concrete advice on the choice of techniques.

References

Aronow, Peter M., Alexander Coppock, Forrest W. Crawford and Donald P. Green. 2013. "Combining list experiment and direct question estimates of sensitive behavior prevalence." *arXiv preprint arXiv:1312.1268* .

Blair, Graeme, C. Christine Fair, Neil Malhotra and Jacob N. Shapiro. 2013. "Poverty and support for militant politics: Evidence from Pakistan." *American Journal of Political Science* 57(1):30–48.

Blair, Graeme and Kosuke Imai. 2010. "list: Statistical methods for the item count technique and list experiment." *Available at The Comprehensive R Archive Network (CRAN)* .

Blair, Graeme and Kosuke Imai. 2012. "Statistical analysis of list experiments." *Political Analysis* 20(1):47–77.

Blair, Graeme, Kosuke Imai and Jason Lyall. 2014. "Comparing and combining list and endorsement experiments: Evidence from Afghanistan." *American Journal of Political Science* 58(4):1043–1063.

Blair, Graeme, Yang-Yang Zhou and Kosuke Imai. 2015*a*. Design and analysis of randomized response technique. Technical report.

Blair, Graeme, Yang-Yang Zhou and Kosuke Imai. 2015*b*. "rr: Statistical Methods for the Randomized Response Technique." *Available at The Comprehensive R Archive Network (CRAN)* .

Bullock, Will, Kosuke Imai and Jacob N. Shapiro. 2011. "Statistical analysis of endorsement experiments: Measuring support for militant groups in Pakistan." *Political Analysis* 19(4):363–384.

Chauchard, Simon. 2013. "Using MP3 players in surveys: The impact of a low-tech self-administration mode on reporting of sensitive attitudes." *Public Opinion Quarterly* 77(S1):220–231.

Corstange, Daniel. 2009. "Sensitive questions, truthful answers? Modeling the list experiment with LISTIT." *Political Analysis* 17(1):45–63.

Gingerich, Daniel W. 2010. "Understanding off-the-books politics: Conducting inference on the determinants of sensitive behavior with randomized response surveys." *Political Analysis* 18(3):349–380.

Glynn, Adam N. 2013. "What can we learn with statistical truth serum? Design and analysis of the list experiment." *Public Opinion Quarterly* 77(S1):159–172.

Gonzalez-Ocantos, Ezequiel, Chad Kiewiet De Jonge, Carlos Meléndez, Javier Osorio and David W Nickerson. 2012. "Vote buying and social desirability bias: Experimental evidence from Nicaragua." *American Journal of Political Science* 56(1):202–217.

Imai, Kosuke. 2011. "Multivariate regression analysis for the item count technique." *Journal of the American Statistical Association* 106(494):407–416.

Imai, Kosuke, Bethany Park and Kenneth F. Greene. Forthcoming. "Using the predicted responses from list experiments as explanatory variables in regression models." *Political Analysis* .

Kramon, Eric and Keith R. Weghorst. 2012. "Measuring sensitive attitudes in developing countries: Lessons from implementing the list experiment." *Newsletter of the APSA Experimental Section* 3(2):14–22.

Lyall, Jason, Graeme Blair and Kosuke Imai. 2013. "Explaining support for combatants during wartime: A survey experiment in Afghanistan." *American Political Science Review* 107(04):679–705.

Meng, Tianguang, Jennifer Pan and Ping Yang. Forthcoming. "Conditional receptivity to citizen participation evidence from a survey experiment in China." *Comparative Political Studies* .

Rosenfeld, Bryn, Kosuke Imai and Jacob Shapiro. Forthcoming. "An empirical validation study of popular survey methodologies for sensitive questions." *American Journal of Political Science* .

Shiraito, Yuki and Kosuke Imai. 2013. "Endorse: R package for analyzing endorsement experiments." *Available at The Comprehensive R Archive Network (CRAN)* .