How Worried Should We Be?

The Implications of Fabricated Survey Data for Political Science*

Oscar Castorena
Vanderbilt University
oscar.castorena@vanderbilt.edu

Mollie J. Cohen
University of Georgia
mj.cohen@uga.edu

Noam Lupu
Vanderbilt University
noam.lupu@vanderbilt.edu

Elizabeth J. Zechmeister
Vanderbilt University
liz.zechmeister@vanderbilt.edu

June 2020

* For their comments and advice, we are grateful to audiences at GESIS, Universidad de Los Andes, and Vanderbilt. We presented a previous version of this paper at the 2019 Annual conference of the Midwest Association for Public Opinion Research.

Abstract

Surveys are ubiquitous in the study of politics, making fabrication a critical issue for political science. Recent studies have focused on fabrication by respondents and firms, but a largely unexamined concern is enumerator fabrication. A prevailing view is that faked interviews affect inferences drawn from compromised datasets. Researchers have generated theories about how fabrication might affect our inferences. Yet, speculation has outpaced systematic testing. We leverage a rare dataset to address this gap: a national face-to-face survey in Venezuela in which a uniquely high volume of falsified interviews was detected, canceled, and replaced. Comparing the verified and fraudulent datasets, we find that descriptive inference is sometimes affected, but that correlational results hold, even in a dataset with an unusually high proportion of fabricated cases. Enumerators largely seem to fabricate plausible data. Though still egregious, enumerator-fabricated interviews represent little more than a minor threat to political science research.

Surveys are ubiquitous in the study of politics, making survey errors a critical issue for political science (Heath, Fisher, and Smith 2005; Lupu and Michelitch 2018). Among the most egregious is wholesale fabrication. Recent studies have focused on fabrication by either respondents who provide bogus responses or office employees who duplicate interviews (Lopez and Hillygus 2018; Kuriakose and Robbins 2016; Pew Research Center 2020). A largely unexamined concern is when fieldworkers fabricate data instead of conducting actual interviews. This is of paramount importance for political science since the majority of surveys in the discipline are administered in person (Lupu and Michelitch 2018).

There is no doubt that interviewers sometimes fabricate data. In our own work, we have detected enumerators at home "interviewing" friends; discussing ways to avoid location monitoring; and posing as respondents by modulating voices. Similar anecdotes have long raised concerns about the quality of academic survey data (Menold et al. 2013), and could affect public perceptions of scientific integrity. For example, in a recent public debate on this topic, one shocking headline read, "Many Surveys, About One in Five, May Contain Fraudulent Data" (Bohannon 2016).

The research community has addressed fake data through revisions to best practices and innovations in fraud detection (e.g., AAPOR 2003; Bredl, Storfinger, and Menold 2013; Cohen and Warner forthcoming; Crespi 1945; Gomila et al. 2017; Montalvo, Seligson, and Zechmeister 2018; Robbins 2018; Schraepler and Wagner 2003; Slomczynski, Powałko, and Krause 2017). But most approaches to detecting fake data are not well-suited to finding enumerator fabrication (e.g., Finn and Ranchhod 2017; Judge and Schechter 2009; Kuriakose and Robbins 2016; Schäfer et al. 2004). Consequently, we know little about the implications of enumerator fabrication for research on politics.

A prevailing view is that faked interviews affect inferences drawn from compromised datasets. Although it appears that only a small number of interviews – perhaps less than 2-5 percent – are faked in typical large-scale surveys (Bredl et al. 2013; Cohen and Larrea 2018), some simulations suggest that even low levels of fabrication can bias inferences (Sarracino and Mikucka 2017; Schraepler and Wagner 2003). Scholarship suggests two reasons to think that cheating produces biased data. First, if fabrication is motivated by incentives to complete interviews quickly, enumerators should favor approaches that allow them to speed through the questionnaire (Blasius and Thiessen 2015; Schraepler and Wagner 2003). Second, cheating interviewers may be more likely to select middling responses on scale items to avoid drawing scrutiny for too many extreme answers (Bredl et al. 2012; Menold et al. 2013; Porras and English 2004). In either case, we would observe different means and lower variance in faked data (Gomila et al. 2017).

Yet, it is also plausible that fabricated data barely differ from real data. Cheating interviewers may know how authentic responses tend to look based on prior experience conducting similar surveys (Waller 2013) – and use this knowledge to effectively mimic real data (Landrock 2017; Menold et al. 2013).

Assessing the effect of fabricated survey data on political science research is complicated by the absence of a counterfactual – the authentic data cheaters would have gathered from real respondents. As a result, most scholarship on interviewer fabrication relies on small datasets (e.g., Schraepler and Wagner 2003), simulations (e.g., Sarracino and Mikucka 2017), or data created by research assistants directed to fabricate responses (e.g., Landrock 2017; but see Finn and Ranchhod 2017). Researchers have theories about how large-scale fabrication might affect inferences, but speculation has outpaced systematic testing.

We leverage a rare opportunity to address this empirical gap. In their 2016-17 AmericasBarometer survey in Venezuela, Vanderbilt University's LAPOP Lab detected an unusually high volume of falsified interviews, and canceled and replaced them while fieldwork was in progress. By gaining access to this dataset, we are able to compare a clean dataset to the compromised one that would have resulted had the faked interviews not been replaced. We find that descriptive inference is sometimes affected, but that correlational results essentially hold, even in a dataset with an unusually high proportion of fabricated cases. Replication with a second dataset, the 2017 Peru AmericasBarometer survey, yields similar results. Enumerators largely seem to fabricate plausible data, which tamps down on the likelihood that faked interviews severely threaten political science research.

The Venezuela Dataset

Venezuela experienced acute crises in 2016, marked by efforts to recall the president, civil unrest (including frequent protests and looting), and deteriorating economic conditions (McCarthy 2017). This context of scarcity, unrest, and insecurity provided ample motivation for interviewers to skirt survey protocols.

LAPOP fielded a national face-to-face survey from November 2016 to February 2017. The survey used e-devices for data collection supplemented by extensive quality control, based on audio recordings, interviewer identity verification, timing features, and geographic coordinates, among other checks (Cohen and Larrea 2018; Montalvo et al. 2018). This extensive auditing allowed LAPOP to identify, cancel, and replace interviews while enumerators were still in the field. More than 650 interviews out of 1,500 were canceled and replaced due to quality concerns – a rate far higher than in any other LAPOP survey (Cohen and Larrea 2018).

We were granted access to all quality control data and auditors' notes. We used these to identify 460 of the canceled interviews as likely fraudulent.[2] The remaining 190 interviews were canceled for quality reasons – because an interviewer misread questions, for instance – but were likely real interviews. To compare fraudulent and authentic cases, we paired each fraudulent case to an authentic interview from the final, published dataset with an exact match on gender, age group, and primary sampling unit.[3] The result is a set of 420 fraudulent interviews matched to 420 authentic interviews, which we can compare directly.

Does Fraud Make A Difference?

Do falsified interviews differ from valid interviews? We focus on the means and distributions of questions with ordinal response scales, which constitute 86% of the attitudinal items on the survey. For each of these 113 items, we examine differences in means, variances, and item nonresponse rates across the fraudulent and matched authentic interviews, using a Bonferroni correction to account for multiple tests.

Table 1. *Item-Level Effects of Fabricated Data*

Comparison	Result
Difference in means	11.5%
Average magnitude (in SD)	0.13
Difference in variances	8.9%
Item nonresponse	0.0%

Note: Values result from tests of 113 items, comparing the fabricated interviews (N=420) and the matched real data (N=420). Full results in Figure A1 and Table A2.

[2] See appendix for coding procedures.

[3] See appendix for matching procedures.

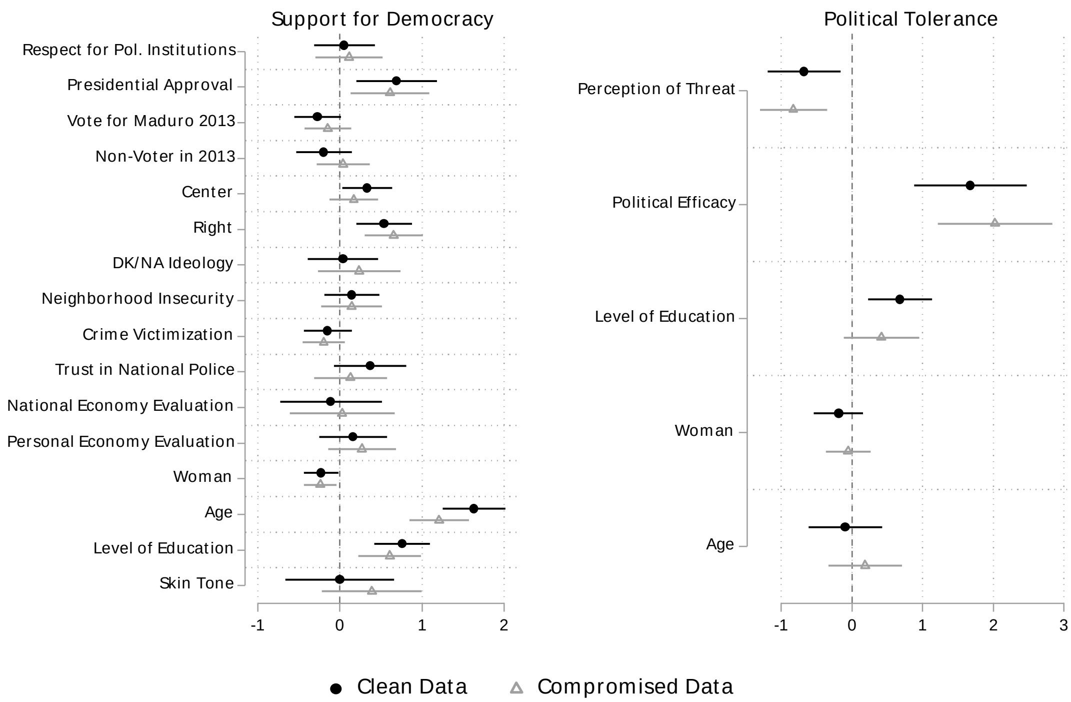
Table 1 reports the proportion of significantly different mean values for the 113 items. To minimize the potential for false negatives, we use a generous cutoff of $p<0.10$. Only thirteen items (11.5%) show significant mean differences between the fraudulent and clean interviews. On average, the magnitude of all differences is 0.13 standard deviations. Variance ratio tests assessing differences in the standard deviations indicate significant differences for ten variables (8.9%), and there are no significant differences in item nonresponse rates. On these metrics, enumerator-faked data differ from real data only rarely and not substantially.

Does Fraud Affect Inferences?

Most research draws inferences from the patterns observed in data. One approach is to compare means over time. We examined differences across time for the 49 items that were also included in LAPOP's 2014 Venezuela survey. Would we have reached different conclusions had LAPOP not identified and replaced the fake data? The answer is yes, but only 14.3% of the time (7 of 49 variables) and not because the cross-time differences are statistically distinct from each other, but because they split between falling just inside and just outside the 95% threshold (see Figure A5).

The Venezuela dataset also allows us to assess how a substantial amount of fake data affects standard political science regression models. We compare the results from identical regression analyses that use the final, clean dataset and a compromised dataset that replaces genuine responses with the matched fraudulent responses. We selected two models common to political behavior research. The first predicts support for democracy, an attitude of central concern in public opinion scholarship (e.g., Evans and Whitefield 1995; Mishler and Rose 1996). The second model predicts political tolerance, following seminal work by Gibson (1992; among others, see Golebiowska 1999; Shamir and Sullivan 1983). The support for democracy model includes independent variables whose means significantly differ between matched genuine and fabricated data, while the political tolerance model uses a dependent variable whose mean differed significantly (see Tables 1, A1, and A2).

Figure 1. *Comparing Regression Results*



Notes: Values represent coefficients estimated from OLS models using both the clean and compromised datasets with 95% confidence intervals. Standard errors account for sampling design effects.

Figure 1 plots the regression coefficients.[4] The dependent variable in the left-hand panel is based on a question that asks level of agreement with the statement, "Democracy may have problems, but it is better than any other form of government." Following previous research (e.g., Seligson 2007; Singer 2018), our model includes measures of respect for political institutions,

[4] Full results in Tables A4 and A6.

presidential approval, past vote choice, left-right self-placement, perceived neighborhood insecurity, crime victimization, trust in police, evaluation of the national economy, evaluation of one's personal economic situation, gender, age, education, and skin tone.[5] The results are broadly similar across the clean and compromised datasets. The coefficients for three variables change signs across the analyses (nonvoter in the 2013 elections, national economic evaluation, and skin tone), but none of these coefficients are significantly different from zero. Overall, both analyses support similar conclusions that are consistent with prior studies of support for democracy in the region.

For the political tolerance model (right-hand panel of Figure 1), the dependent variable is measured using an item that asks, "How strongly do you approve or disapprove that [regime critics] be allowed to conduct peaceful demonstrations in order to express their views?" Following Gibson (1992) and Duch and Gibson (1992), our models include measures of perception of threat, political efficacy, level of education, age, and gender. Again, the results are broadly similar across the clean and compromised datasets. The coefficient sign is different for only one variable, and that estimated coefficient is essentially zero. Overall, both results are in line with prior research: threat, lower efficacy, and lower education are associated with lower tolerance.

There is also similarity across coefficient parameters estimated using the clean versus compromised datasets. Chow tests of whether a coefficient estimated from the clean dataset is equal to the corresponding coefficient estimated from the compromised dataset yield no statistically significant differences (see Tables A5 and A7).

[5] All independent variables are scaled from zero to one. For wording and coding, see Table A3.

There are instances where slight differences could affect conclusions. For example, the coefficient for education in the tolerance analysis using the compromised dataset is outside the standard cutoff, while the corresponding coefficient using the clean dataset is within it. Still, the general picture is reassuring. Even unusually large amounts of fraud seem to make little difference to the inferences we draw. Those that rely on small effect sizes or estimates that are close to the threshold for statistical significance – results we should already treat with some skepticism – are most likely to be affected by compromised data.

How Do Interviewers Fake Data?

Concerns about the ill effects of fake survey data derive from the expectation that enumerators fabricate interviews in ways that substantially bias the data. Our results suggest that, in fact, interviewers create fake data that is internally consistent, such that the correlations between variables are realistic. Enumerators who fabricate interviews appear to mimic reality.

How do faking interviewers know what reality to mimic? The data suggest that they employ a mixed strategy, which allows them to learn about “true” public opinion before falsifying interviews. The Venezuela survey employed 79 interviewers; of these, 46 fabricated data. Prior to their termination, these enumerators averaged 10 fake interviews but also 21.5 real interviews. Most fabricators did so only after conducting real interviews: 50% of interviewers who eventually recorded a falsified interview had conducted at least five real interviews before their first fraudulent interview (only 25% of all cheaters falsified their first recorded interview). That is, most enumerators learned about response patterns prior to fabricating interviews.

Given such a strategy, we might expect that the faked interviews replace the kinds of interviews that are hardest to obtain. Given that LAPOP uses gender and age-based frequency matching for respondent selection,[6] a key challenge is recruiting working-age men, since they are frequently not home and less likely to participate (Silver et al. 2019). If interviewers are faking interviews when it becomes especially difficult to conduct real ones, the fabricated data should be disproportionately male and working-age.

This is precisely the pattern we observe. Table 2 compares the demographic composition of fabricated interviews with that of the full clean data. The fraudulent interviews contain

Table 2. *Demographic Composition of Clean and Fake Data*

	Male			Female		
Age group	**Fraud**	**Clean**	**Difference**	**Fraud**	**Clean**	**Difference**
18-25 years old	12.6%	11.0%	+1.6	10.0%	10.1%	-0.1
26-35 years old	11.7%	11.2%	+0.5	10.9%	11.0%	-0.1
36-45 years old	11.7%	9.8%	+1.9	6.7%	10.7%	-4.0
46-55 years old	13.6%	8.5%	+5.0	7.9%	8.4%	-0.5
56-65 years old	5.0%	5.9%	-0.9	5.2%	6.0%	-0.8
66+ years old	1.9%	3.9%	-2.0	2.9%	3.5%	-0.6

Notes: Values compare the fraudulent data to the full clean dataset.

slightly more young and middle-aged men and fewer older men and women of all age-groups. Interviewers appear to fabricate data to complete difficult fieldwork assignments.

Finding that cheating interviewers create plausible fake interviews runs counter to what some have theorized. Some suggest that enumerators frequently provide responses in the middle of a scale or take the shortest path through the instrument when faking interviews (e.g., see discussions in Bredl, Winker, and Kötschau 2012; Bredl et al. 2013; Menold et al. 2013; Schraepler and Wagner 2003). These approaches – or mere random selection – would result in

[6] https://www.vanderbilt.edu/lapop/ab2016/AmericasBarometer_2016-17_Sample_Design.pdf.

far more differences. To demonstrate, we compare the observed differences in the fraudulent data to what would have been produced under other strategies. First, we generated a version of fraudulent data by programming a random selection of responses. Next, we hired 10 students and incentivized them to speed through the questionnaire as quickly as possible with devices similar to those used in the fieldwork, each 10 times. Finally, we created a version of the fraudulent data based on selecting middling responses.[7]

Table 3. *Comparing Actual and Simulated Fake Data*

Comparison	Faked	Random	Speeding	Middling Responses
Difference in means	11.5%	68%	75.2%	71.7%
Average magnitude (in SD)	0.13	0.55	0.64	0.54
Difference in variances	8.9%	26.6%	85.0%	98.2%
Item nonresponse	0.0%	--	50.4%	--

Notes: Values result from tests of 113 items comparing the listed data-generating process to the real data. Full results in Figures A1-A4.

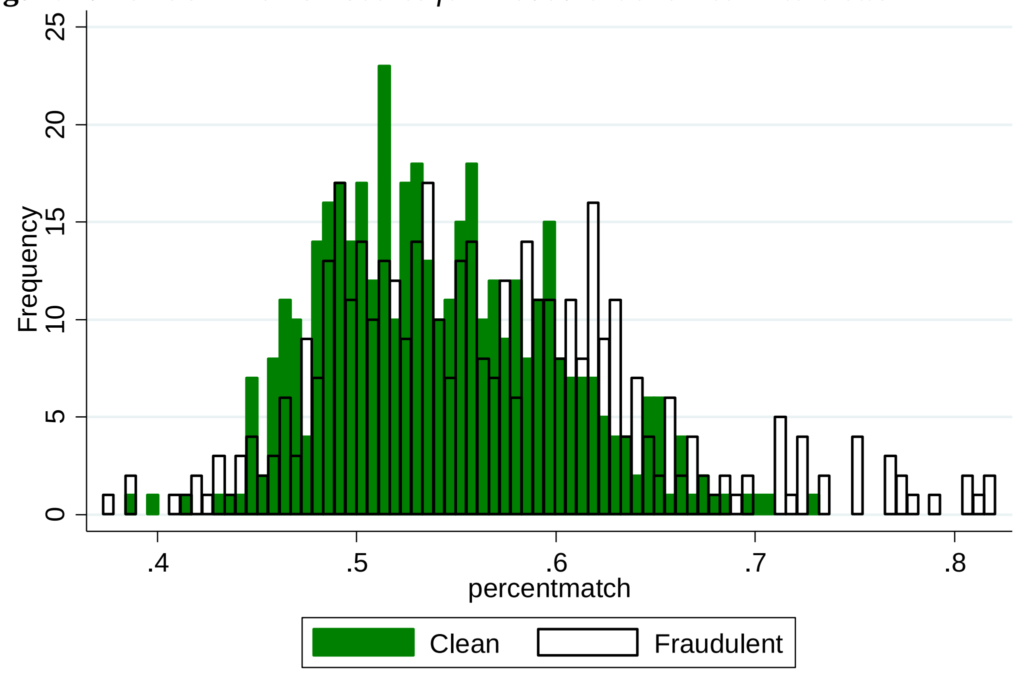
Table 3 compares the results from each of these simulations to the matched authentic interviews. In every case, we find far higher mean differences, average difference magnitudes, variance ratios, and item nonresponse. In other words, interviewers seem not to hew very closely to any of these hypothesized strategies when they fabricate interviews.[8]

[7] These were normally distributed with a mean at the center of a scale and standard deviation of 1/10th of its range.

[8] Faking interviewers do shy away slightly from extreme response options (see Figures A6 and A7).

A final way to examine the fabrication production process is to consider their similarity to real interviews. If interviewers are mimicking true responses, fake data should look similar to real data – and may even underestimate population heterogeneity. We can see this using Kuriakose and Robbins' (2016) `percentmatch` program, which assigns a score to each interview based on the extent to which responses match those of the most similar interview.

Figure 2. `Percentmatch` *Scores for Fraudulent and Real Interviews*



Notes: Histogram of `percentmatch` scores for matched clean and fraudulent interviews. A value of one indicates complete duplication between an interview and its most similar interview.

Figure 2 summarizes these scores for the fraudulent interviews and the clean interviews with which they are matched. Indeed, the fraudulent cases do exhibit greater similarity in their responses to other interviews.[9]

[9] Figure 2 also shows that `percentmatch` does not detect enumerator-generated fabrication: all of the scores are below the 0.85 threshold for likely fabrication.

All of this suggests that when enumerators fake interviews, they do so with some knowledge about the distribution and consistencies in real interviews – and they try to mimic real data. This informed fabrication process largely maintains the patterns and relationships found in genuine data.

How Worried Should We Be?

Interview fabrication is egregious, but the consequences may be less severe than prior studies suggest. Widespread fabrication is rare. But our analyses reveal that, even in unusual instances where fabrication is extensive, its effects on the kinds of analyses that political scientists conduct are likely minimal. Fabrication affects descriptive statistics for a small proportion of variables. Correlations, the basis of most political science applications of survey data, are remarkably similar across both clean and fake data.

This is because interviewers use their knowledge of real data when fabricating interviews. In contrast to expectations that cheating enumerators might choose responses at random, to maximize speed, or toward the middle of the response range, we find they tend to mimic real data, basing their false responses on real interviews they conduct prior to fabricating others. As a result, even a large proportion of fake interviews has little effect on researchers' inferences.

We rely, however, on data from just one study, so how generalizable are these results? The political and socioeconomic context in Venezuela in 2016-17 was troubled, creating incentives for enumerators to shirk. But it was not unlike the other developing contexts in which researchers regularly field face-to-face surveys. LAPOP's monitoring protocols are more extensive than those employed by most researchers, but nearly all survey researchers employ some type of monitoring (Lupu and Michelitch 2018), dating back to Crespi's (1945) important contribution on enumerator cheating.

We can assess the question of generalizability empirically by using data from the only other case in which LAPOP identified, canceled, and replaced a sufficient number of interviews to replicate the analyses in this paper. LAPOP's 2017 Peru dataset contains 116 likely fabricated interviews (4.4% of the sample) – a rate that is still higher than the estimated level of fabrication in the average scientific public opinion survey (Bredl et al. 2013; Cohen and Larrea 2018). We matched these data to authentic interviews and the results affirm our conclusions (see appendix). The average difference in means is of a similar magnitude, though fewer of these differences are significant given the smaller sample size.

These findings are reassuring for scholarship that relies on face-to-face surveys. Still, we do not advocate being cavalier about fabricated data. Ensuring the accuracy and quality of the data we use is paramount. Our results do highlight that fabricated data could change inferences about fine-grained comparisons or small effects near standard statistical thresholds. But our findings also suggest that recent skepticism about the reliability of interviewer-administered survey data – especially from international surveys – is exaggerated. Scientific research should always strive for accuracy, but our research reveals that we can still learn a great deal from survey data even when some of it is fake.

References

American Association for Public Opinion Research (AAPOR). 2003. "Interviewer Falsification in Survey Research: Current Best Methods for Prevention, Detection and Repair of Its Effects." http://www.aapor.org/AAPOR_Main/media/MainSiteFiles/falsification.pdf.

Bohannon, John. 2016. "Many Surveys, About One in Five, May Contain Fraudulent Data." *Science*, February 24.

Blasius, Jörg, and Victor Thiessen. 2015. "Should We Trust Survey Data? Assessing Response Simplification and Data Fabrication." *Social Science Research* 52: 479–93.

Bredl, Sebastian, Nina Storfinger, and Natalja Menold. 2013. "A Literature Review of Methods to Detect Fabricated Survey Data." In *Interviewers' Deviations in Surveys*, eds. Peter Winker, Natalja Menold, and Rolf Porst, pp. 3–24. Frankfurt, Germany: PL Academic Research.

Bredl, Sebastian, Peter Winker, and Kerstin Kötschau. 2012. "A Statistical Approach to Detect Interviewer Falsification of Survey Data." *Survey Methodology* 38(1): 1–10.

Cohen, Mollie J., and Sebastian Larrea. 2018. "Assessing and Improving Interview Quality in the 2016/17 AmericasBarometer." LAPOP Lab Methodological Note No. 2.

Cohen, Mollie J., and Zach Warner. Forthcoming. "How to Get Better Survey Data More Efficiently." *Political Analysis* 1–27.

Crespi, Leo P. 1945. "The Cheater Problem in Polling." *Public Opinion Quarterly* 9(4): 431–45.

Duch, Raymond M. and James L. Gibson. 1992. "'Putting up with' Fascists in Western Europe: A Comparative, Cross-Level Analysis of Political Tolerance." *Western Political Quarterly* 45(1): 237–73.

Evans, Geoffrey, and Stephen Whitefield. 1995. "The politics and economics of democratic commitment: Support for democracy in transition societies." *British Journal of Political Science* 25(4): 485–514.

Finn, Arden, and Vimal Ranchhod. 2017. "Genuine Fakes: The Prevalence and Implications of Data Fabrication in a Large South African Survey." *World Bank Economic Review* 31(1): 129–57.

Gibson, James L. 1992. "Alternative Measures of Political Tolerance: Must Tolerance be 'Least-Liked'?" *American Journal of Political Science* 36 (2): 560–77.

Gomila, Robin, Rebecca Littman, Graeme Blair, and Elizabeth Levy Paluck. 2017. "The Audio Check: A Method for Improving Data Quality and Detecting Data Fabrication." *Social Psychological and Personality Science* 8(4): 424–33.

Golebiowska, Ewa A. 1999. "Gender Gap in Political Tolerance." *Political Behavior* 21(1): 43–66.

Heath, Anthony, Stephen Fisher, and Shawna Smith. 2005. "The Globalization of Public Opinion Research." *Annual Review of Political Science* 8: 297–333.

Judge, George, and Laura Schechter. 2009. "Detecting Problems in Survey Data Using Benford's Law." *Journal of Human Resources* 44(1): 1–24.

Kuriakose, Noble, and Michael Robbins. 2016. "Don't Get Duped: Fraud through Duplication in Public Opinion Surveys." *Statistical Journal of the IAOS* 32(3): 283–91.

Landrock, Uta. 2017. "Explaining Political Participation: A Comparison of Real and Falsified Survey Data." *Statistical Journal of the IAOS* 33(2): 447–58.

Lopez, Jesse, and D. Sunshine Hillygus. 2018. "Why So Serious?: Survey Trolls and Misinformation." Unpublished manuscript. https://ssrn.com/abstract=3131087.

Lupu, Noam, and Kristin Michelitch. 2018. “Advances in Survey Methods for the Developing World.” *Annual Review of Political Science* 21: 195–214.

McCarthy, Michael M. 2017. “Venezuela’s Manmade Disaster.” *Current History* 116(787): 61–7.

Menold, Natalja, Peter Winker, Nina Storfinger, and Christoph J. Kemper. 2013 “A Method for Ex-Post Identification of Falsifications in Survey Data.” In *Interviewers’ Deviations in Surveys*, eds. Peter Winker, Natalja Menold, and Rolf Porst, pp. 25–47. Frankfurt, Germany: PL Academic Research.

Mishler, William, and Richard Rose.1996. "Trajectories of fear and hope: support for democracy in post-communist Europe." *Comparative political studies* 28(4): 553-581.

Montalvo, J. Daniel, Mitchell A. Seligson, and Elizabeth J. Zechmeister. 2018. “Improving Adherence to Area Probability Sample Designs: Using LAPOP’s Remote Interview Geo-Locating of Households in Real-Time (RIGHT) System.” LAPOP Lab Methodological Note No. 4.

Pew Research Center. 2020. “Assessing the Risks to Online Polls from Bogus Respondents.” https://www.pewresearch.org/methods/wp-content/uploads/sites/10/2020/02/PM_02.18.20_dataquality_FULL.REPORT.pdf.

Porras, Javier, and Ned English. 2004. “Data-Driven Approaches to Identifying Interviewer Data Falsification: The Case of Health Surveys.” Proceedings of the American Statistical Association, Section on Survey Research Methods, 4223–8.

Robbins, Michael. 2018. “New Frontiers in Detecting Fabrication.” In *Advances in Comparative Survey Methods*, eds. Timothy P. Johnson, Beth-Ellen Pennell, Ineke A.L. Stoop, and Brita Dorer, pp. 771–806. Hoboken, NJ: Wiley.

Shamir, Michal, and John Sullivan. 1983. The Political Context of Tolerance: The United States and Israel. *American Political Science Review* 77(4): 911–28.

Sarracino, Francesco, and Malgorzata Mikucka. 2017. “Bias and Efficiency Loss in Regression Estimates Due to Duplicated Observations: A Monte Carlo Simulation.” *Survey Research Methods* 11(1): 17–44.

Schäfer, Christin, Jörg-Peter Schräpler, Klaus-Robert Müller, and Gert G. Wagner. 2004. “Automatic Identification of Faked and Fraudulent Interviews in Surveys by Two Different Methods.” DIW Discussion Paper No. 441. https://www.diw.de/documents/publikationen/73/diw_01.c.42515.de/dp441.pdf.

Slomczynski, Kazimierz M., Przemek Powałko, and Tadeusz Krause. 2017. “Non-unique Records in International Survey Projects: The Need for Extending Data Quality Control.” *Survey Research Methods* 11(1): 1–16.

Schraepler, Joerg-Peter, and Gert G. Wagner. 2003. “Identification, Characteristics and Impact of Faked Interviews in Surveys: An Analysis by Means of Genuine Fakes in the Raw Data of SOEP.” IZA Discussion Paper No. 969. http://hdl.handle.net/10419/20205.

Seligson, Mitchell A. 2007. “The Rise of Populism and the Left in Latin America.” *Journal of Democracy* 18(3): 81–95.

Silver, Laura, Martha McRoy, Kat Devlin, and Patrick Moynihan. 2019. “Who’s Home and Who Isn’t? The Challenges of Conducting Face-to-Face Interviews in Jordan.” *Decoded* / Pew Research Center. https://medium.com/pew-research-center-decoded/whos-home-and-who-isn-t-the-challenges-of-conducting-face-to-face-interviews-in-jordan-95a2a7081ded

Singer, Matthew M. 2018. “Delegating Away Democracy: How Good Representation and Policy Success Can Undermine Democratic Legitimacy.” *Comparative Political Studies* 51(13): 1754–88.

Slomczynski, Kazimierz M., Przemek Powalko, and Tadeusz Krauze. 2017. “Non-Unique Records in International Survey Projects: The Need for Extending Data Quality Control.” *Survey Research Methods* 11(1): 1–16.

Waller, Lloyd George. 2013. “Interviewing the Surveyors: Factors which Contribute to Questionnaire Falsification (Curbstoning) Among Jamaican Field Surveyors.” *International Journal of Social Research Methodology* 16(2): 155–64.