How Worried Should We Be? The Implications of Fabricated Survey Data for Political Science

Oscar Castorena
*Vanderbilt University*

Mollie Cohen
*University of Georgia*

Noam Lupu
*Vanderbilt University*

Elizabeth Zechmeister
*Vanderbilt University*

Follow this and additional works at: [https://ir.lib.uwo.ca/c-dem_conf_survey](https://ir.lib.uwo.ca/c-dem_conf_survey)

Part of the Political Science Commons

**Citation of this paper:**
[https://ir.lib.uwo.ca/c-dem_conf_survey/5](https://ir.lib.uwo.ca/c-dem_conf_survey/5)
How Worried Should We Be?
The Implications Of Fabricated Survey Data For Political Science

Dr. Oscar Castorena
Dr. Mollie J. Cohen
Dr. Noam Lupu
Dr. Elizabeth J. Zechmeister
We want high-quality survey data

• The fabrication of interviews by interviewers is a classic problem in survey research
  Crespi 1945; Gomila et al. 2017

• The field has developed best practices for minimizing, identifying fabrication
  AAPOR 2003; Cohen & Warner 2021; Kuriakose & Robbins 2016; Montalvo et al., 2018; Robbins, 2018; Slomczynski et al., 2017

• Questions remain about how common fabrication is and its effects on inference
How does fabricated data affect our inferences?

• Wholesale fabrication is rare
  Bredl et al., 2013; Cohen & Larrea, 2018; Menold et al., 2013

• But even low rates of fabrication can bias estimates (e.g., due to lower variance, middle responding)
  DeMatteis et al., 2020; Gomila et al., 2017

• Evidence is limited: it is hard to observe the counterfactual
Enter Venezuela, 2016-7

• 460 fabricated interviews (“fakes”) were identified and replaced during AmericasBarometer fieldwork

• We matched 420 fabricated interviews to validated replacements
Does fake data bias inferences?

We examine differences in datasets including fabricated interviews or matched valid interviews

1. Differences in averages, distributions, nonresponse
2. Differences in multivariate regression models
Some differences in means and distributions across data sets

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Fake vs. Matched</th>
<th>Compromised vs. Clean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference in means</td>
<td>11.5 – 48.7%</td>
<td>0.0 – 13.3%</td>
</tr>
<tr>
<td>Average magnitude (in SD)</td>
<td>0.20 – 0.31</td>
<td>0.08 – 0.11</td>
</tr>
<tr>
<td>Difference in variances</td>
<td>8.9 – 46.0%</td>
<td>0.9 – 4.4%</td>
</tr>
<tr>
<td>Item nonresponse</td>
<td>0.0 – 20.4%</td>
<td>0.0 – 0.9%</td>
</tr>
</tbody>
</table>

*Note:* Values result from tests of 115 items, comparing the fabricated interviews and the matched real data ($N = 420$) as well as the compromised data and the clean data ($N = 1,489$).
But few differences in regression models
Why don’t we see more differences?

• Interviewers fabricate plausible responses in response to fieldwork constraints

• Use a “mixed strategy” – fabricating some, not all

• Different implications from fastest path and middle response expectations
Why don’t we see more differences?

To assess, we compare the “true-fake” data to three simulated datasets:

1. The fastest path (identified by undergraduate RAs)
2. Data with imputed middle responses
3. Data with imputed random responses
Larger differences in simulated falsifications

**Item-Level Effects of Fabricated Data**

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Faked</th>
<th>Random</th>
<th>Speeding</th>
<th>Middling Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference in means</td>
<td>11.5 – 48.7%</td>
<td>68.0 – 81.4%</td>
<td>75.2 – 86.7%</td>
<td>71.7 – 85.8%</td>
</tr>
<tr>
<td>Average magnitude (in SD)</td>
<td>0.13 – 0.30</td>
<td>0.65 – 0.75</td>
<td>0.73 – 0.82</td>
<td>0.63 – 0.72</td>
</tr>
<tr>
<td>Difference in variances</td>
<td>8.9 – 46.0%</td>
<td>26.6 – 66.4%</td>
<td>77.9 – 92.9%</td>
<td>72.6 – 98.2%</td>
</tr>
<tr>
<td>Item nonresponse</td>
<td>0.0 – 20.4%</td>
<td>–</td>
<td>50.4 – 78.8%</td>
<td>–</td>
</tr>
</tbody>
</table>

*Note.* Values result from tests of 113 items. In each case, we generate results using either no adjustment to the standard errors as well as the Bonferroni correction, Hochberg’s step-up procedure, and Holm’s step-down procedure – and we report the range of values. We use a baseline cutoff of $p < .1$ for statistical significance.
In closing

• Data fabrication is egregious – not recommended!

• However, even very high rates of fake interviews may not bias inferences

• Fabricators may use their knowledge of the population to create plausible responses
Thanks!

mj.cohen@uga.edu