Observed data
Imputed data—Missing completely at random
Imputed data—Fitted normal model
Imputed data—Various assumptions
Should the problems with polls make us worry about the quality of health surveys?

Andrew Gelman

Department of Statistics and Department of Political Science
Columbia University, New York

Centers for Disease Control and Prevention, Connecticut, 21 Feb 2017
There is no satisfactory statistical solution to deal with missing data that may not be at random. Assuming an extreme bias in the responders is one suggested method of dealing with low survey response rates. Getting a high response rate (>80%) from a small, random sample is considered preferable to a low response rate from a large sample. [9]
Trend data were not readily available from the large number of private organizations who conduct survey research and polls. These organizations claim to be experiencing problems as the following excerpt from the 1973 Conference on Surveys of Human Populations (American Statistical Association, 1974) shows:

"...spokesmen for a number of private survey organizations, large and small, who were queried by one of the conference participants, all report that their completion rates on general population samples now average approximately 60 to 65 percent, in spite of three or four callbacks. This recent experience is in contrast to a completion figure of 80 to 85 percent for the same firms in the decade of the sixties."
Survey Response Rates: Some Trends, Causes and Correlates.
Surveys Face Growing Difficulty Reaching, Persuading Potential Respondents

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Contact rate</strong> (percent of households in which an adult was reached)</td>
<td>90</td>
<td>77</td>
<td>79</td>
<td>73</td>
<td>72</td>
<td>62</td>
</tr>
<tr>
<td><strong>Cooperation rate</strong> (percent of households contacted that yielded an interview)</td>
<td>43</td>
<td>40</td>
<td>34</td>
<td>31</td>
<td>21</td>
<td>14</td>
</tr>
<tr>
<td><strong>Response rate</strong> (percent of households sampled that yielded an interview)</td>
<td>36</td>
<td>28</td>
<td>25</td>
<td>21</td>
<td>15</td>
<td>9</td>
</tr>
</tbody>
</table>

PEW RESEARCH CENTER 2012 Methodology Study. Rates computed according to American Association for Public Opinion Research (AAPOR) standard definitions for CON2, COOP3 and RR3. Rates are typical for surveys conducted in each year.
Survey response rates are going through the floor!
Survey response rates are going through the floor!
The poststratification identity

\[ \theta = \frac{\sum_{j=1}^{J} N_j \theta_j}{\sum_{j=1}^{J} N_j} \]
The poststratified estimate

\[ \hat{\theta} = \frac{\sum_{j=1}^{J} N_j \hat{\theta}_j}{\sum_{j=1}^{J} N_j} \]
Join the VP Debate
We’re polling LIVE, 9 PM ET
If the election were held today, who would you vote for?

- Barack Obama
- Mitt Romney
- Other
- Not sure
Xbox estimates, adjusting for demographics:
The experts speak!


- Nate Silver, *New York Times*, 6 Oct: “Mr. Romney has not only improved his own standing but also taken voters away from Mr. Obama’s column.”
Xbox estimates, adjusting for demographics and partisanship:
Why multilevel regression?
Nationally, Trump got 2% more of the vote than Romney.
Swing from 2012 to 2016: Lots of variation among states

(T rump vote) − (Romney vote)

Graph omits Utah, where T rump did 13% worse than Romney
Nationally, Trump got 2% more of the vote than predicted.
Trump did much better than predicted in states that Romney won in 2012.
Open problems in MRP

- Deep interactions
- Non-census variables
- Survey weights
- Cluster sampling
- Estimating regression coefficients
- Building trust in results