Paradata within the Total Survey Error Framework
Successes, Challenges, and Gaps

Frauke Kreuter

Federal Computer-Assisted Survey Information Collection (FedCASIC) Workshop

March 27, 2012
We have seen successful uses of paradata to gain efficiency, and to alert for errors.

We face serious challenges to expand the concurrent analytic use of paradata, the tailored collection of paradata, and the transfer across modes, surveys, and survey organizations.

We might benefit from widening the scope to other error sources, through linkage with cost data and others, and from the use of paradata in modelling.
Outline . . .

1. Current Activities and Typical Applications

2. Challenges in Collection, Analysis, and Communication

3. Paradata inside the Total Survey Error Framework
Outline

1. Current Activities and Typical Applications

2. Challenges in Collection, Analysis, and Communication

3. Paradata inside the Total Survey Error Framework
Prevalent Paradata in TSE Framework

Groves et al. 2004

Activities & Applications

Contact Data & Observations: Day/Time; Result

Keys & Key Strokes: Missing data, Response Times

Measurement

Construct \( \mu_i \)

Measurement

\( Y_i \)

Response

\( Y_i \)

Edited Response

\( Y_{ip} \)

Survey Statistic

\( \bar{Y}_{prw} \)

Representation

Target Population

\( \bar{Y} \)

Sampling Frame

\( \bar{Y}_C \)

Sample

\( \bar{Y}_s \)

Respondents

\( \bar{Y}_r \)

Postsurvey Adjustment

\( \bar{Y}_{rw} \)

Validity - Measurement Error - Processing Error - Coverage Error - Sampling Error - Nonresponse Error - Adjustment Error

Kreuter (JPSM & IAB/LMU)
Response Time

- **Substantive Use**
  - Attitudes as object-evaluation model
    

- **Post-hoc Use - Focus on Error**
  - Characteristics of Instrument and Setting:
    (+) poor wording, poor layout, length, complexity
    (-) logical order, practice, correct answers, decreasing motivation
    

- **Interview administration**
  
    e.g. Olson & Peytchev 2007, Couper & Kreuter 2012, Schafer 2012

  - **Interview falsification**
    
    Clements 2001; Penne, Snodgrass & Baker 2002

- **Concurrent Use - Focus on Error**
  - **Intervention if respondents answer too fast** Conrad et al. 2009
    or too slow Conrad, Schober & Coiner 2007
Call record data

- **Post-hoc Use** - Focus on Efficiency
  - Optimal call schedules
    - Example
  - Predictors of response
    - Example

- **Concurrent Use** - Focus on Efficiency
  - Call scheduling (CATI)
  - Monitoring
    - Example

- **Post-hoc Use** - Focus on Error
  - Nonresponse bias analyses
    - Example
    - e.g. FedStat Surveys - since OMB Standard and Guidelines 2006
  - Nonresponse bias adjustment

- **Concurrent Use** - Focus on Error
  - Interventions
    - Example
Outline

1. Current Activities and Typical Applications
2. Challenges in Collection, Analysis, and Communication
3. Paradata inside the Total Survey Error Framework
Now one could ask:

- Why do we not see more research on key stroke data?
- Why do analysts struggle with f2f contact protocol data?
- Why do FRs shy away from call record protocols?
- Why are interviewer observations not used for adjustment?
- Why are adjusters disappointed about interviewer observations?
And one could suggest:

- Why do we not see more research on key stroke data?  
  ⇒ Code repository shared across surveys and organizations

- Why do analysts struggle with f2f contact protocol data?  
  ⇒ Statistical methods development

- Why do FRs shy away from call record protocols?  
  ⇒ “There’s an app for that”

- Why are interviewer observations not used for adjustment?  
  ⇒ Flexible software to allow tailored indicators

- Why are adjusters disappointed about interviewer observations?  
  ⇒ Use of auxiliary information
Outline

1. Current Activities and Typical Applications
2. Challenges in Collection, Analysis, and Communication
3. Paradata inside the Total Survey Error Framework
TSE Framework

Paradata inside TSE Framework  Groves et al. 2004

Listing Information: Flags (USPS and phone), Geocode Precision, Edits

Target Population $\overline{Y}$

Key Strokes: Response Times; Back-Ups; Edits; Mouse Movements

Sampling Frame $\overline{Y}_C$

Vocal Characteristics: Pitch, Disfluencies, Pauses

Sample $\overline{Y}_s$

Contact Data & Observations: Day/Time; HU Characteristics; Proxy-Y

Respondents $\overline{Y}_r$

Key strokes: Back-Ups; Edits

Postsurvey Adjustment $\overline{Y}_{rw}$

Survey Statistic $\overline{Y}_{prw}$

Construct $\mu_i$

Measurement $Y_i$

Edited Response $y_{ip}$

Response $y_i$

Measurement Error

Validity

Processing Error

Measurement Error

Coverage Error

Sampling Error

Nonresponse Error

Adjustment Error
You . . .
Thank You!

fkreuter@survey.umd.edu
# Contact Rates by Hour and Day in NSFG

The set of calls included in each model was reduced from the total set of all calls for various reasons. Any calls that were after eight calls, for example, was not the goal. The goal was to provide household specific estimates. For example, if we appointment. The call number did not enter the models as a predictor. Estimating the average probability of being at home set as appointments were deleted since the purpose is to predict the probability of contact for a "cold" call, not an

### Table 3. Call Window Definitions

<table>
<thead>
<tr>
<th>Definition</th>
<th>Mon 9am-4pm, Tues-Fri 9am-4pm</th>
<th>Sat-Sun 9am-4pm</th>
<th>Sat-Sun-Mon 9am-4pm</th>
<th>Tues-Fri 4pm-9pm</th>
<th>Mon-Thurs 4pm-9pm</th>
<th>Sept 2008</th>
<th>Feb 2009</th>
<th>Sept 2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCA</td>
<td>SAT-SUN-MON 9AM-4PM</td>
<td>SAT-SUN 9AM-4PM</td>
<td>SAT-SUN-MON 9AM-4PM</td>
<td>TUES-FRI 4PM-9PM</td>
<td>MON-THURS 4PM-9PM</td>
<td>SEP 2008</td>
<td>FEB 2009</td>
<td>SEP 2009</td>
</tr>
<tr>
<td></td>
<td>SAT-SUN-MON 9AM-4PM</td>
<td>SAT-SUN 9AM-4PM</td>
<td>SAT-SUN-MON 9AM-4PM</td>
<td>TUES-FRI 4PM-9PM</td>
<td>MON-THURS 4PM-9PM</td>
<td>SEP 2008</td>
<td>FEB 2009</td>
<td>SEP 2009</td>
</tr>
<tr>
<td></td>
<td>SAT-SUN-MON 9AM-4PM</td>
<td>SAT-SUN 9AM-4PM</td>
<td>SAT-SUN-MON 9AM-4PM</td>
<td>TUES-FRI 4PM-9PM</td>
<td>MON-THURS 4PM-9PM</td>
<td>SEP 2008</td>
<td>FEB 2009</td>
<td>SEP 2009</td>
</tr>
<tr>
<td></td>
<td>SAT-SUN-MON 9AM-4PM</td>
<td>SAT-SUN 9AM-4PM</td>
<td>SAT-SUN-MON 9AM-4PM</td>
<td>TUES-FRI 4PM-9PM</td>
<td>MON-THURS 4PM-9PM</td>
<td>SEP 2008</td>
<td>FEB 2009</td>
<td>SEP 2009</td>
</tr>
<tr>
<td></td>
<td>SAT-SUN-MON 9AM-4PM</td>
<td>SAT-SUN 9AM-4PM</td>
<td>SAT-SUN-MON 9AM-4PM</td>
<td>TUES-FRI 4PM-9PM</td>
<td>MON-THURS 4PM-9PM</td>
<td>SEP 2008</td>
<td>FEB 2009</td>
<td>SEP 2009</td>
</tr>
<tr>
<td></td>
<td>SAT-SUN-MON 9AM-4PM</td>
<td>SAT-SUN 9AM-4PM</td>
<td>SAT-SUN-MON 9AM-4PM</td>
<td>TUES-FRI 4PM-9PM</td>
<td>MON-THURS 4PM-9PM</td>
<td>SEP 2008</td>
<td>FEB 2009</td>
<td>SEP 2009</td>
</tr>
<tr>
<td></td>
<td>SAT-SUN-MON 9AM-4PM</td>
<td>SAT-SUN 9AM-4PM</td>
<td>SAT-SUN-MON 9AM-4PM</td>
<td>TUES-FRI 4PM-9PM</td>
<td>MON-THURS 4PM-9PM</td>
<td>SEP 2008</td>
<td>FEB 2009</td>
<td>SEP 2009</td>
</tr>
</tbody>
</table>

### Example Table

<table>
<thead>
<tr>
<th>Hour</th>
<th>Sunday All</th>
<th>Sunday Elig</th>
<th>Monday All</th>
<th>Monday Elig</th>
<th>Tuesday All</th>
<th>Tuesday Elig</th>
<th>Wednesday All</th>
<th>Wednesday Elig</th>
<th>Thursday All</th>
<th>Thursday Elig</th>
<th>Friday All</th>
<th>Friday Elig</th>
<th>Saturday All</th>
<th>Saturday Elig</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>0.30</td>
<td>0.34</td>
<td>0.25</td>
<td>0.35</td>
<td>0.30</td>
<td>0.32</td>
<td>0.24</td>
<td>0.31</td>
<td>0.23</td>
<td>0.30</td>
<td>0.27</td>
<td>0.33</td>
<td>0.30</td>
<td>0.35</td>
</tr>
<tr>
<td>10</td>
<td>0.32</td>
<td>0.40</td>
<td>0.31</td>
<td>0.38</td>
<td>0.28</td>
<td>0.33</td>
<td>0.29</td>
<td>0.34</td>
<td>0.30</td>
<td>0.36</td>
<td>0.27</td>
<td>0.34</td>
<td>0.31</td>
<td>0.39</td>
</tr>
<tr>
<td>11</td>
<td>0.36</td>
<td>0.43</td>
<td>0.30</td>
<td>0.38</td>
<td>0.31</td>
<td>0.38</td>
<td>0.31</td>
<td>0.39</td>
<td>0.31</td>
<td>0.39</td>
<td>0.32</td>
<td>0.40</td>
<td>0.35</td>
<td>0.43</td>
</tr>
<tr>
<td>12</td>
<td>0.37</td>
<td>0.44</td>
<td>0.32</td>
<td>0.42</td>
<td>0.32</td>
<td>0.38</td>
<td>0.32</td>
<td>0.40</td>
<td>0.30</td>
<td>0.37</td>
<td>0.31</td>
<td>0.38</td>
<td>0.34</td>
<td>0.42</td>
</tr>
<tr>
<td>13</td>
<td>0.37</td>
<td>0.45</td>
<td>0.32</td>
<td>0.42</td>
<td>0.24</td>
<td>0.31</td>
<td>0.29</td>
<td>0.38</td>
<td>0.30</td>
<td>0.38</td>
<td>0.32</td>
<td>0.39</td>
<td>0.34</td>
<td>0.43</td>
</tr>
<tr>
<td>14</td>
<td>0.38</td>
<td>0.46</td>
<td>0.34</td>
<td>0.43</td>
<td>0.33</td>
<td>0.40</td>
<td>0.32</td>
<td>0.40</td>
<td>0.32</td>
<td>0.39</td>
<td>0.33</td>
<td>0.40</td>
<td>0.35</td>
<td>0.43</td>
</tr>
<tr>
<td>15</td>
<td>0.39</td>
<td>0.48</td>
<td>0.35</td>
<td>0.44</td>
<td>0.32</td>
<td>0.40</td>
<td>0.33</td>
<td>0.42</td>
<td>0.33</td>
<td>0.41</td>
<td>0.33</td>
<td>0.41</td>
<td>0.36</td>
<td>0.46</td>
</tr>
<tr>
<td>16</td>
<td>0.39</td>
<td>0.49</td>
<td>0.36</td>
<td>0.45</td>
<td>0.37</td>
<td>0.46</td>
<td>0.36</td>
<td>0.45</td>
<td>0.35</td>
<td>0.43</td>
<td>0.34</td>
<td>0.42</td>
<td>0.35</td>
<td>0.45</td>
</tr>
<tr>
<td>17</td>
<td>0.39</td>
<td>0.49</td>
<td>0.40</td>
<td>0.49</td>
<td>0.38</td>
<td>0.46</td>
<td>0.38</td>
<td>0.47</td>
<td>0.36</td>
<td>0.46</td>
<td>0.34</td>
<td>0.43</td>
<td>0.33</td>
<td>0.43</td>
</tr>
<tr>
<td>18</td>
<td>0.37</td>
<td>0.44</td>
<td>0.38</td>
<td>0.47</td>
<td>0.39</td>
<td>0.48</td>
<td>0.37</td>
<td>0.47</td>
<td>0.36</td>
<td>0.45</td>
<td>0.33</td>
<td>0.42</td>
<td>0.35</td>
<td>0.44</td>
</tr>
<tr>
<td>19</td>
<td>0.37</td>
<td>0.44</td>
<td>0.39</td>
<td>0.47</td>
<td>0.37</td>
<td>0.45</td>
<td>0.37</td>
<td>0.46</td>
<td>0.35</td>
<td>0.44</td>
<td>0.31</td>
<td>0.42</td>
<td>0.35</td>
<td>0.43</td>
</tr>
<tr>
<td>20</td>
<td>0.40</td>
<td>0.44</td>
<td>0.38</td>
<td>0.45</td>
<td>0.39</td>
<td>0.45</td>
<td>0.38</td>
<td>0.46</td>
<td>0.37</td>
<td>0.45</td>
<td>0.32</td>
<td>0.40</td>
<td>0.36</td>
<td>0.44</td>
</tr>
</tbody>
</table>
Nonresponse Bias in PASS  
Kreuter, Mueller, Trappmann 2010

Figure 1a* Cumulative mean over quintiles of no. of contact attempts; Current Welfare Status, by age group

Figure 1b* RMSE over quintiles of no. of contact attempts; Current Welfare Status; by age group

\[ \text{Proportion (w/f = 1)} \]

\[ \text{root mean squared error (RMSE)}; \leq 41\text{yrs bias; } \leq 41\text{yrs std. err.; } \leq 41\text{yrs root mean squared error (RMSE); } >41\text{yrs bias; } >41\text{yrs std. err.; } >41\text{yrs root mean squared error (RMSE)} \]

\[ \text{... though often no comparison} \]
Monitoring Effort in PASS Wave 5  Mueller 2011

The image shows a line graph illustrating the monitoring effort over calendar weeks for different samples:

- W5 Recipient Sample
- W5 Recipient Sample [New PSUs]
- W5 Gen. Pop. Sample [New PSUs]
- Panel HH W4 dropout
- Panel HH W4 Interview
- Split-off households

The x-axis represents the calendar week, ranging from 6 to 40, and the y-axis represents the proportion of ever contacted individuals, ranging from 0 to 1.
Response Rates by Subgroup in NSFG  

Lepkowskiet al. 2012

![Graph showing response rates by subgroup in NSFG. The graph compares response rates for Black Males 15-19, Hispanic Males 15-19, Other Males 15-19, Black Males 20-44, Hispanic Males 20-44, and Other Males 20-44 across different days.](image-url)
Structurally, an audit trail is a comma delimited text file, where each line in the audit trail (except for header information) shows the date and time when a keystroke was entered. An example of an audit trail for a case appears in Figure 1.

**Figure 1: Excerpt from an audit trail file**

```
"1/11/2004 9:15:50 AM","Enter Form:1","Key:XXXXXXXX 
"1/11/2004 9:15:50 AM","Metafile name:C:\WINCM\DATA\STUDIES\CEQ_BA01\e-inst\inst.bmi"
"1/11/2004 9:15:50 AM","WinUserName:FR"
```

```

```
"2/11/2004 5:52:21 PM","Enter Field:Sect03.ANYRENT","Status:Normal","Value:"
"2/11/2004 5:52:24 PM","Leave Field:Sect03.ANYRENT","Cause:Next Field","Status:Normal","Value:"
```

```
"1/11/2004 6:16:44 AM","Leave Form:1","Key:XXXXXXXX 
```

```
```

```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
ESS ROC data – manual entries

... data are incomplete and error prone

<table>
<thead>
<tr>
<th>CntryC-e</th>
<th>visit</th>
<th>result fromESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>31544</td>
<td>30101118</td>
<td>1 NO contact at all</td>
</tr>
<tr>
<td>31545</td>
<td>30101118</td>
<td>2 NO contact at all</td>
</tr>
<tr>
<td>31546</td>
<td>30101118</td>
<td>3 NO contact at all</td>
</tr>
<tr>
<td>31547</td>
<td>30101118</td>
<td>4 NO contact at all</td>
</tr>
<tr>
<td>31548</td>
<td>30101118</td>
<td>5 NO contact at all</td>
</tr>
<tr>
<td>31549</td>
<td>30101118</td>
<td>6 NO contact at all</td>
</tr>
<tr>
<td>31550</td>
<td>30101118</td>
<td>7 Contact with R NO interview</td>
</tr>
<tr>
<td>31551</td>
<td>30101118</td>
<td>8 NO contact at all</td>
</tr>
<tr>
<td>31552</td>
<td>30101118</td>
<td>9 NO contact at all</td>
</tr>
<tr>
<td>31553</td>
<td>30101118</td>
<td>10 NO contact at all</td>
</tr>
<tr>
<td>31554</td>
<td>30101118</td>
<td>. Interview 1</td>
</tr>
</tbody>
</table>

... times are all but randomized

... interviewer have their own preferences
### Contact record

<table>
<thead>
<tr>
<th>Number of calls</th>
<th>S0350 CALL #1</th>
<th>S0351 CALL #2</th>
<th>S0352 CALL #3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Date (day / month / year)</td>
<td>_ / _ / _</td>
<td>_ / _ / _</td>
<td>_ / _ / _</td>
</tr>
<tr>
<td>B. Day of week</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Exact time began</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. Interviewer I.D.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. Contact with</td>
<td>Respondent</td>
<td>Informant</td>
<td>No One</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>F. Mode of contact</td>
<td>Personal</td>
<td>Telephone</td>
<td>Personal</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>G. Tel. Number if obtained</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H. Household Unit listing obtained</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>I. Detailed description of contact or attempt to contact</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J. Result code</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Contact record (continued)

<table>
<thead>
<tr>
<th>Number of calls</th>
<th>S0355 CALL #6</th>
<th>S0356 CALL #7</th>
<th>S0357 CALL #8</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Date (day / month / year)</td>
<td>_ / _ / _</td>
<td>_ / _ / _</td>
<td>_ / _ / _</td>
</tr>
<tr>
<td>B. Day of week</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Exact time began</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. Interviewer I.D.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. Contact with</td>
<td>Respondent</td>
<td>Informant</td>
<td>No One</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>F. Mode of contact</td>
<td>Personal</td>
<td>Telephone</td>
<td>Personal</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
**CONTACT STRATEGIES ATTEMPTED**

- Select the categories that describe the strategies used on this contact attempt.
- Enter all that apply, separate with commas.

- [ ] 1. Advance letter given
- [ ] 2. Scheduled appointment
- [ ] 3. Left note / appointment card
- [ ] 4. Left promotional packet / informational brochure
- [ ] 5. Called household
- [ ] 6. Left message on answering machine
- [ ] 7. FR will request No One Home Letter
- [ ] 8. FR will request Refusal Letter
- [ ] 9. FR will request Better Understanding Letter
- [ ] 10. Called contact persons
- [ ] 11. Stake out
- [ ] 12. Checked with neighbors
- [ ] 13. Contacted other family members
- [ ] 14. Contacted property manager
- [ ] 15. Visited county assessor / post office / permit office
- [ ] 16. On-line tracking database
- [ ] 17. Sought help from SFR / RO
- [ ] 18. Reassignment
- [ ] 19. Offered incentive
- [ ] 20. CED double placement
- [ ] 21. Used MAF or ALMI
- [ ] 22. None
- [ ] 23. Other - specify

---

Kreuter (JPSM & IAB/LMU)  
Paradata  
FedCASIC / 14
Interviewers forget to record variables or some of them might be difficult to observe. The factor analysis approach might help smooth out this measurement error by using multiple variables to measure a latent tendency for someone to express a certain type of concern. Table 4 shows the resulting correlations of the factor scores with survey participation and the survey variables. As expected, the factors based on the individual paradata variables are more strongly correlated with survey participation than the survey variables. The general resistance factor is the strongest correlate with participation. The average correlation between the factor scores and the survey variables is approximately .03. The largest correlation is .15.

Table 4. Correlation of factor scores with participation and survey variables (2006 data).

<table>
<thead>
<tr>
<th>Variable set</th>
<th>Correlation with participation</th>
<th>Correlation with survey variables (absolute values)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average</td>
<td>Maximum</td>
</tr>
<tr>
<td>Cooperation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor 1: Time concerns</td>
<td>-.25</td>
<td>.03</td>
</tr>
<tr>
<td>Factor 2: Privacy or content concerns</td>
<td>-.27</td>
<td>.02</td>
</tr>
<tr>
<td>Factor 3: General resistance</td>
<td>-.47</td>
<td>.02</td>
</tr>
<tr>
<td>Factor 4: Gatekeeper issues</td>
<td>-.30</td>
<td>.02</td>
</tr>
</tbody>
</table>

We also looked at the correlation of the individual paradata variables with participation. The top half of Table 5 shows that the cooperation variables are better correlates with participation than the contactability variables. The strongest correlates with participation are the indicators for the sample person indicating they are not interested or do not want to be bothered (-.69) and the sample person hangs up or slams the door on the interviewer (-.64). The bottom half of the table summarizes the correlations between the paradata variables and the vector of survey variables. The average correlations are only in the .02-.07 range. Only in a few cases are the correlations larger than .2. We also ran separate logistic regression models on the contactability and cooperation variables to obtain a response propensity for each case based on these variables. However, these propensities were not correlated any stronger with the survey variables than many of the individual paradata variables.
# Measurement Error: Gender

McCulloch & Kreuter 2012

Pooled data: 28 CATI surveys (n=25,635), Marist College (MIPO):

<table>
<thead>
<tr>
<th></th>
<th>Respondent male</th>
<th>Respondent female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guess - male</td>
<td>97.36</td>
<td>13.87</td>
<td>49.63</td>
</tr>
<tr>
<td>Guess - female</td>
<td>2.64</td>
<td>86.13</td>
<td>50.37</td>
</tr>
<tr>
<td>Total</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>White</th>
<th>Afr. Am.</th>
<th>Hispanic</th>
<th>Asian</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guessed Correct</td>
<td>92.1%</td>
<td>87.4%</td>
<td>92.1%</td>
<td>92.0%</td>
<td>92.7%</td>
</tr>
<tr>
<td>Guessed Incorrect</td>
<td>7.9%</td>
<td>12.6%</td>
<td>7.9%</td>
<td>8.0%</td>
<td>7.3%</td>
</tr>
</tbody>
</table>
Interviewer vs. Census Tract

Casas-Cordero et al. 2012

Fig. 3. Estimates of Intraclass Correlation from the Linear Unconditional Models

NG watch sign
Security signs
Tended yards
Graffiti
Cigarettes
Security gates
Litter, glass
Cond. bldgs.
Trash, junk
Aband. cars
Damaged walls
Bars on windows
Boarded up
Vacant lots
Painted graff
Dog sign
Empty bottles
Drug items

Kreuter (JPSM & IAB/LMU)
NSFG Interviewer Observations from 16 Quarters (n=15,044):