Great Expectations: Changing Mode of Survey Data Collection in Military Populations

Ronald Z. Szoc, PhD
Jacqueline Pflieger, PhD
Frances M. Barlas, PhD
Randall K. Thomas

Federal Committee on Statistical Methodology
Research Conference
November 4-6, 2013

The opinions and conclusions presented herein are those of the authors and do not necessarily reflect the opinions of the Department of Defense.
WHAT WE FOUND

- Significant differences in prevalence of self-reported behaviors between onsite and online respondents, with online mostly lower
  - Still true, even after controlling for demographic differences between onsite and online samples
- Significant but small mode effects with the most significant being:
  - Marijuana use in past 12 months
  - Pain reliever misuse in past 12 months
  - Heavy alcohol use in past 30 days
- Non-Significant mode effects:
  - Smokeless tobacco use in past 12 months
  - STIs in past 12 months
- While the absolute values for many measures were significantly different by mode, the results for both modes retained the essential ordering of information
WHAT DOES IT MEAN?

- Military survey respondents “act differently” compared to civilians with online surveys when self-reporting risky or undesirable or non-healthy behaviors
- The differences may be due to a variety of factors:
  - Trust or lack of trust with regard to military online systems
  - Peer pressure
- More research is needed to identify those factors affecting responses by military vs. civilian populations
- Do these results generalize to workplace settings for online data collection?
WHAT follows

- Brief description of Health Risk Behavior Survey methodology
  - Assumptions
  - Analytical approach
- Variables
  - Dependent variables (i.e., self-reported behaviors)
  - Covariates
- Results
- Lessons Learned
HEALTH RELATED BEHAVIORS SURVEY PILOT STUDY

- HPA&E/TMA contracted with Acentia and ICF to conduct a pilot study of Web administration of the survey
- The pilot study was to be done in parallel with the onsite administration conducted by another contractor so that analysis of mode differences could be conducted
In order to examine a “pure” mode effect, all aspects of the online survey mirrored the onsite mode as much as possible, except for the actual mode of administration:

- The methodology for the sampling frame and selecting individuals to be part of the sample was similar.
- The wording of the questionnaire items and response choices was identical.
- The order of presentation of the questions was identical to the paper instrument.
- Logic skip patterns for the questions were not be used, to better reflect the paper survey experience.
- Data collection period for online mode was the same as the onsite mode in duration.
- Online survey was anonymous as was the onsite survey.
**METHODOLOGICAL APPROACH: SAMPLING**

- Two stage sampling process
  - First-stage Sampling Unit frame was comprised of 21,257 UICs across six Guard and Reserve components
    - Onsite contractor selected 3,103 UICs for the onsite mode
    - ICF selected 3,346 UICs for the online mode from the remaining 18,154 units using Probability Proportional to Size sampling
  - Second-stage Sampling Unit for online was the individual
    - 310,257 individuals from 3,346 UICs stratified by Guard/Reserve component, pay grade category (6 levels), and gender
    - 79,913 individuals randomly selected using the strata
    - Officers and women were oversampled
Mode was made a dichotomous variable: Onsite = 0, Online = 1

Correlations examined between mode and covariates hypothesized to account for differences between samples

Regression analyses conducted on selected risk and protective behaviors as key outcomes to determine effect of mode on dependent variables of interest, using unweighted data

Correlational analysis conducted on risk and protective measures to determine if the results obtained from each mode were ordered similarly across key groups of interest, using weighted data
MODE ANALYSIS: DEPENDENT VARIABLES

- Heavy alcohol use
- Binge drinking
- Hazardous drinking
- Being current smoker
- Heavy smoking
- Nicotine dependence
- Smokeless tobacco use
- Marijuana use
- Illicit drug use (except marijuana)
- Pain reliever misuse
- Suicide ideation
- Being overweight/obese
- Sexually transmitted infections
MODE ANALYSIS: COVARIATES

- Age category (17-24, 25-34, 35-44, 45+)
- Impulsivity (5 item scale)
- Activation/full-time status (non-activated reference)
- Gender (males reference)
- Service component (5 ‘dummy’ variables, AF reference)
- Pay grade (5 ‘dummy’ variables, O4-O10 reference)
- Depression (6 item scale)
- Marital status (non-married reference)
- Race/ethnicity (3 ‘dummy’ variables, African American reference)
- Education (high school or less, some college, college graduate/some graduate school, graduate degree)
<table>
<thead>
<tr>
<th></th>
<th>Army National Guard</th>
<th>Army Reserve</th>
<th>Navy Reserve</th>
<th>Air National Guard</th>
<th>Air Force Reserve</th>
<th>Marine Corps Reserve</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Sample</td>
<td>11,174</td>
<td>16,699</td>
<td>16,678</td>
<td>11,119</td>
<td>16,676</td>
<td>7,567</td>
<td>79,913</td>
</tr>
<tr>
<td>B. Ineligibles</td>
<td>1,317</td>
<td>2,238</td>
<td>2,418</td>
<td>3,361</td>
<td>2,158</td>
<td>931</td>
<td>12,423</td>
</tr>
<tr>
<td>C. Net sample</td>
<td>9,857</td>
<td>14,461</td>
<td>14,260</td>
<td>7,758</td>
<td>14,518</td>
<td>6,636</td>
<td>67,490</td>
</tr>
<tr>
<td>(A) - (B)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. Participants</td>
<td>3,622</td>
<td>3,760</td>
<td>3,029</td>
<td>5,043</td>
<td>5,566</td>
<td>1,063</td>
<td>22,083</td>
</tr>
<tr>
<td>E. Usable surveys</td>
<td>3,514</td>
<td>3,651</td>
<td>2,881</td>
<td>3,521</td>
<td>4,312</td>
<td>1,016</td>
<td>18,895</td>
</tr>
<tr>
<td>F. Response Rate (D) / (C)</td>
<td>36.7%</td>
<td>26.0%</td>
<td>21.2%</td>
<td>65.0%</td>
<td>38.3%</td>
<td>16.0%</td>
<td>32.7%</td>
</tr>
<tr>
<td>G. Cooperation Rate (E) / (D)</td>
<td>97.0%</td>
<td>97.1%</td>
<td>95.1%</td>
<td>69.8%</td>
<td>77.5%</td>
<td>95.6%</td>
<td>85.6%</td>
</tr>
</tbody>
</table>
## Summary of Regression Analysis

<table>
<thead>
<tr>
<th></th>
<th>Onsite Survey</th>
<th>Online Survey</th>
<th>Mode Difference (Onsite-Online)</th>
<th>Model Pseudo R²</th>
<th>Incremental Pseudo R²</th>
<th>Mode Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Heavy Alcohol</strong></td>
<td>22.1%</td>
<td>8.9%</td>
<td>13.2%</td>
<td>.343</td>
<td>.010</td>
<td>.511***</td>
</tr>
<tr>
<td><strong>Binge Drinking</strong></td>
<td>37.8%</td>
<td>23.7%</td>
<td>14.1%</td>
<td>.201</td>
<td>.001</td>
<td>.809***</td>
</tr>
<tr>
<td><strong>Hazardous Drinking</strong></td>
<td>17.7%</td>
<td>7.5%</td>
<td>10.2%</td>
<td>.260</td>
<td>.007</td>
<td>.575***</td>
</tr>
<tr>
<td><strong>Current Smoker</strong></td>
<td>23.4%</td>
<td>14.7%</td>
<td>8.7%</td>
<td>.164</td>
<td>.000</td>
<td>.862***</td>
</tr>
<tr>
<td><strong>Heavy Smoker</strong></td>
<td>10.4%</td>
<td>7.5%</td>
<td>2.9%</td>
<td>.105</td>
<td>.001</td>
<td>.812**</td>
</tr>
<tr>
<td><strong>Nicotine Dependence</strong></td>
<td>13.1%</td>
<td>12.2%</td>
<td>0.9%</td>
<td>.071</td>
<td>.001</td>
<td>.810*</td>
</tr>
<tr>
<td><strong>Smokeless Tobacco</strong></td>
<td>17.7%</td>
<td>8.8%</td>
<td>8.9%</td>
<td>.240</td>
<td>.000</td>
<td>1.006</td>
</tr>
<tr>
<td><strong>Marijuana</strong></td>
<td>3.7%</td>
<td>1.2%</td>
<td>2.5%</td>
<td>.207</td>
<td>.014</td>
<td>.354***</td>
</tr>
<tr>
<td><strong>Illicit Drugs</strong></td>
<td>15.7%</td>
<td>13.4%</td>
<td>2.3%</td>
<td>.076</td>
<td>.007</td>
<td>.645***</td>
</tr>
<tr>
<td><strong>Pain Reliever Misuse</strong></td>
<td>5.7%</td>
<td>3.5%</td>
<td>2.2%</td>
<td>.106</td>
<td>.011</td>
<td>.487***</td>
</tr>
<tr>
<td><strong>Suicide Ideation</strong></td>
<td>6.8%</td>
<td>4.3%</td>
<td>2.5%</td>
<td>.207</td>
<td>.002</td>
<td>.728***</td>
</tr>
<tr>
<td><strong>Overweight/Obese</strong></td>
<td>61.5%</td>
<td>65.1%</td>
<td>-3.6%</td>
<td>.177</td>
<td>.001</td>
<td>1.148***</td>
</tr>
<tr>
<td><strong>STIs</strong></td>
<td>1.8%</td>
<td>1.0%</td>
<td>0.8%</td>
<td>.111</td>
<td>.000</td>
<td>.882</td>
</tr>
</tbody>
</table>

* * p<.05; ** p<.01; *** p<.001
MODE CORRELATIONAL ANALYSIS VARIABLES: RISK AND PROTECTIVE FACTORS

- Non-hazardous AUDIT
- Heavy alcohol use past 30 days
- Alcohol binge episode in past month
- Non-smoker
- Heavy smoker
- Nicotine dependence
- Any smokeless tobacco use
- Past year marijuana use
- Past year illicit drug use, except marijuana
- Past year pain reliever misuse
- Vigorous exercise
- 7 or more hours of sleep per night
- Having 1 or more combat deployments
- BMI in overweight or obese range
- Work stress
While the absolute values of many measures differed significantly by mode, the results retained the essential ordering of information. This indicates that each mode measures information similarly and can usefully describe differences between groups of individuals, both as a cross-section and for data collected longitudinally.
WHAT WE FOUND REDUX

- Significant differences in prevalence of self-reported behaviors between onsite and online respondents, with online mostly lower
  + Still true, even after controlling for demographic differences between onsite and online samples

- Significant but small mode effects with the most significant being:
  - Marijuana use in past 12 months
  - Pain reliever misuse in past 12 months
  - Heavy alcohol use in past 30 days

- Non-Significant mode effects:
  - Smokeless tobacco use in past 12 months
  - STIs in past 12 months

- While the absolute values for many measures were significantly different by mode, the results for both modes retained the essential ordering of information
LESSONS LEARNED

- Encourage taking the survey on a computer owned by the individual, and not on a work-related computer, to reduce social desirability bias resulting from concerns about privacy.
- Online modes should obtain personal email addresses in addition to work-related email addresses to facilitate.
- Various methods for ensuring anonymity and respondent trust should be investigated so online modes achieve the same level of acceptance as paper administration.