

Great Expectations: Changing Mode of Survey Data Collection in Military Populations¹

Ronald Z. Szoc

ICF International
530 Gaither Road, Rockville, MD, 20850

Jacqueline Pflieger

SRA International

Frances M. Barlas

Randall K. Thomas

GfK Custom Research, LLC

Proceedings of the 2013 Federal Committee on Statistical Methodology (FCSM) Research Conference

Abstract

As difficulties and costs of "in person" surveys have escalated, many studies have migrated to new modes of data collection. Most research concerning the differences in prevalence rates for health-related behaviors, such as alcohol use, as a function of data collection mode, has been done on civilian populations. We report results of a study of the effects of mode on such prevalence rates for a military population comprised of National Guard and Reserve units from all four military Department of Defense branches. Since 1980, the DoD Health Related Behaviors (HRB) surveys have been conducted onsite at military installations worldwide, with respondents completing paper questionnaires in group-administered, in-person sessions. For the most recent HRB National Guard and Reserve survey, we conducted a parallel study to assess the feasibility of conducting the HRB as a web-based, individually administered survey - 16,079 completed onsite and 18,895 completed online (response rates of 64% and 33% respectively). Controlling for sample differences, this afforded us the ability to compare self-reported prevalence estimates for 15 key risk and protective measures, such as alcohol use and BMI, between onsite and online modes. Two primary research questions were: (1) What differences were observed in behavioral health estimates as a result of mode of administration? and (2) Will changing modes retain the ability to make comparisons across services and still maintain trends over time? The results from hierarchical regression analyses were different from what might be expected from similar studies with civilian populations - online respondents reported significantly lower prevalence rates for risky behaviors than their onsite counterparts, while prevalence rate rank order was similar across modes. Thus, contrary to civilians, military members may be less likely to report engaging in risky behaviors in an online mode, and may require additional assurances of data anonymity for online surveys.

Introduction

The Department of Defense Survey of Health Related Behaviors (HRB) is designed to collect, analyze, and report on the lifestyle choices and behaviors that affect the mental and physical health of service members and that ultimately impact their readiness and retention. The survey, which began as an assessment of substance abuse in the Active Duty military in 1980, has evolved into a survey that covers a range of topics including current issues, such as prescription drug misuse. While the survey has been fielded with the Active Duty personnel since 1980, the first Survey of Health Related Behaviors Among the Guard and Reserve Force was conducted in 2006.² The 2009-2010 Department of

¹ Paper prepared for presentation at the Federal Committee on Statistical Methodology Annual Meeting in Washington, D.C. 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.

² Throughout this report we use the terms "Reserve" or "Reserve force" to mean the six components of the Guard and Reserve force that serve as the target population for this study - the Army Reserve, the Army National Guard, the Navy Reserve, the Marine Corps Reserve, the Air Force Reserve, and the Air National Guard.

Defense (DoD) Survey of Health Related Behaviors Among Reserve Component Personnel marks the second population-based examination of the state of health related behaviors in the Reserve and Guard force.

The HRB surveys have traditionally been conducted onsite at military installations widely dispersed around the world, with respondents completing paper questionnaires in group-administered in-person sessions. This traditional method of data collection is very time consuming and poses significant logistical challenges. Due to the dispersed nature of the Reserve and Guard force, conducting a representative survey using an in-person administration is logistically difficult and burdensome, as well as costly. Sending teams of data collectors to anywhere from 60 to 150 military installations worldwide requires extensive coordination with each branch of service and each installation, including the identification and assignment of Service Liaison Officers (SLOs) for each of the Services being surveyed and Installation Liaison Officers (ILOs) at each of the installations being visited. Coordinating data collection at each installation places a great deal of burden on the ILO for locating and scheduling all selected respondents at the installation as well as locating replacement respondents. Gathering respondents for survey sessions also disrupts the units to which the respondents belong. National Guard and Reserve units only meet one weekend a month and two weeks over the summer. The scheduling of weekend training activities occurs well in advance, and notification of the survey requirement can require a reorganization of planned training activities. In addition, allocating an hour or two of this limited training time to administer a survey can be frustrating for unit leaders. Finally, due to the high cost of travel, conducting in-person data collection requires that the selected sample be geographically clustered to minimize the number of site visits, which complicates statistical analysis of the data.

For the 2009-2010 Survey, TRICARE Management Activity (TMA) decided to undertake a pilot study primarily to assess the feasibility of conducting the HRB online. In addition to assessing the feasibility of conducting the HRB online, the pilot study was designed to conduct a mode test that would assess comparability of survey estimates between onsite and online modes of administration. Online data collection is an appealing alternative to in-person data collection as it alleviates many of these logistical and financial burdens. The onsite data collection teams are no longer required as there is no need for travel. While SLOs are important to help support the online data collection, ILOs are not needed. Respondents can complete an online survey at their convenience, eliminating the need to coordinate respondents' schedules and disrupt unit activities. In addition, online data collection eliminates the need for clustered sampling, allowing for a more random sample of respondents without regard to geographic location, which may provide a more accurate reflection of the diversity found among members of the Guard and Reserve force. Another important advantage of online survey administration is that it does not create an atmosphere of coercion, whether real or imagined; respondents can independently choose whether or not to participate without input from peers or superiors and without peer pressure.

Studying the Effects of Administrative Mode on Study Outcomes

In recent decades there has not only been an increase in the number of ways in which surveys are administered to respondents but also numerous variations in the way each survey mode can be administered to respondents. With the increase in the variety with which researchers administer surveys, a concern has arisen in the survey research community over the effects of the mode of administration on survey results. As a result of these possible differences by mode and the effects that they can have on survey results, changing the mode of administration on a survey like the HRB requires careful consideration and scientific investigation. In addition to considering the feasibility of such a change, a number of important methodological and research questions need to be investigated, including:

- Is it feasible to conduct the Health Related Behaviors Survey online?
- What effect, if any, will a change in mode of administration have on the prevalence estimates and incidence rates that the survey typically measures?
- Will the ability to compare trends across different iterations of the survey over time be lost or can it be maintained through the change in survey mode?
- How will the survey response rates be affected by a change in mode?

ICF designed and conducted an online pilot study to answer these questions, and to assess the feasibility and any ramifications of conducting the HRB online.

Online Data Collection Pilot Study

To conduct the online pilot study, ICF attempted to replicate the methodology used by the contractor performing the onsite survey administration. Minimizing differences in methodology was important to facilitate comparison between the two resulting datasets. The target population for both modes of administration included service members in the following six Reserve and Guard components: the Army Reserve, the Army National Guard, the Navy Reserve, the Marine Corps Reserve, the Air Force Reserve, and the Air National Guard.

The online sample was selected in two stages. In the first stage, a sample of Reserve and Guard units was randomly selected, with probability proportional to size, from the population of units, exclusive of those units selected by the onsite survey contractor. In the second stage, individual personnel were randomly selected from the sampled units. The final usable online sample included 18,895 Reserve and Guard members across the six components; the final usable onsite sample included 16,507 Reserve and Guard members across the six components.

The 2009-2010 survey instrument was adapted from the 2006 DoD Survey of Health Related Behaviors Among the Guard and Reserve Force and the 2008 DoD Survey of Health Related Behaviors Among Active Duty Military Personnel. The survey format, a self-administered questionnaire designed for scanning by optical-mark reader, was well-suited for adaptation to an online format in which the respondents were asked to click on one or more responses or, for some questions, to enter numbers into the appropriate cell. Like the onsite administration, the online survey was anonymous.

Selecting the final sample for this study consisted of a multi-step process necessitated by the fact that two different samples needed to be drawn: one for the onsite administration of the HRB Survey and a second separate sample for the online administration. The sections below provide the details of the process.

Survey Population

The survey population consisted of all non-deployed members of the Reserve and National Guard of the Army, Navy, Air Force, and Marine Corps, exclusive of Individual Mobilization Augmentees (IMA) and Individual Ready Reserve (IRR). The sampling frame consisted of 21,257 units comprising 795,676 individual members. The number of units by Reserve component is presented in Table 1 below:

	UIC* Sampling Frame		Units for Onsite Administration		Units for Online Administration	
<i>Guard and Reserve Component</i>	<i>Number of Units</i>	<i>Percent of Total</i>	<i>Number of Units</i>	<i>Percent of Total</i>	<i>Number of Units</i>	<i>Percent of Total</i>
Army Reserve	7,237	34.0	934	30.1	767	22.9
Army National Guard	6,372	30.0	482	15.5	836	25.0
Navy Reserve	3,161	14.9	848	27.3	489	14.6
Marine Corps Reserve	357	1.7	95	3.1	262	7.8
Air Force Reserve	2,449	11.5	433	14.0	430	12.9
Air National Guard	1,681	7.9	311	10.0	562	16.8
Total	21,257	100.0	3,103	100.0	3,346	100.0

*UIC is Unit Identification Code

The Army Reserve and Army National Guard had the most units in the Sampling Frame (64 percent of the total), followed by the Air Force (19.4 percent of the total), the Navy Reserve (14.9 percent), and the Marine Corps Reserve with the smallest number of units (1.7 percent). The proportion of units by Guard and Reserve component selected for the onsite and online samples differed from the sampling frame for several reasons: units were selected using Probability Proportional to Size (PPS) random sampling; units selected for the onsite administration were clustered by location for efficiency in survey administration; online unit selection was based on the remaining, non-

selected units from the original sample frame, and the final sample (the second-stage sampling unit) was randomized by strata for selection of individuals. Thus, there was no attempt to mirror the proportion of units by Guard and Reserve component for the samples chosen for onsite and online administration.

Sample Design and Selection

As noted above, the sample design and procedure was a multiple-step process, complicated by the need to draw two separate samples sequentially from which the point estimates were derived. The steps consisted of the following:

- The First-stage Sampling Unit (FSU) was the military unit, denoted by the Unit Identification Code (UIC); the Second-stage Sampling Unit (SSU) was the individual for the online and the onsite administration samples.
- The contractor for the onsite administration selected 3,103 FSUs from the list of UICs provided by the Defense Manpower Data Center. The remaining total of 18,154 FSUs constituted the new sample frame from which the online sample was selected.
- For the online selection, 3,346 UICs were selected from the 18,154 FSUs using Probability Proportional to Size (PPS) sampling. These 3,346 units comprised 310,217 individuals.
- This group of 310,217 individuals was stratified by Guard and Reserve component (six levels), pay grade (six levels), and gender (2 levels) prior to final selection. The six Guard and Reserve components are presented in Table 1 above. The pay grades were categorized into six levels: E1-E3, E4-E6, E7-E9, W1-W5, O1-O3, and O4 and higher.
- Individuals were randomly selected within each stratum, over-sampling the Officer pay grades and women. Over-sampling was employed in order to gain sufficient statistical power for the individual stratum

The sample sizes of most strata were large enough to yield sufficient power. The exceptions were the strata associated with female Officers, and the strata associated with all the Warrant Officer pay grades (i.e., pay grades W1 to W5).

Survey Instrument

ICF participated in the revision of the survey instrument with the onsite contractor and with TMA. This process incorporated feedback from subject matter experts and committees throughout the DoD. The questionnaire had 175 numbered questions with more than 600 possible responses. All responses were based on close-ended or numeric fill-in responses. The instrument covered the topics listed below:

- | | |
|--------------------|---|
| ▪ Demographics | ▪ Healthy lifestyle and disease prevention |
| ▪ Alcohol use | ▪ Safety and other health related behaviors |
| ▪ Tobacco use | ▪ Stress and mental health |
| ▪ Illicit drug use | ▪ Deployment and job satisfaction |

Once the paper-based version of the 2009-2010 Survey of Health Related Behaviors Among Reserve Component Personnel was finalized and approved by TMA, ICF transferred the questionnaire form to a web-based environment. ICF developed the web-based version of the instrument using IBM/SPSS's Data Collection (formerly Dimensions) software. The web-based instrument was programmed so that survey questions were presented to respondents in as visually and functionally similar manner as possible to the paper-based instrument. Toward that end, the survey was programmed with no skip or branching logic and without making any questions mandatory. Any other differences between the paper- and web-versions of the instrument were based on format changes necessitated by online presentation

The web-version of the questionnaire presented between one and three numbered questions per HTML web page, with a progress bar showing approximate percentage of survey completed; some of the longer grid questions were split and presented across two or three HTML pages. The items were presented on separate web pages so that

respondents did not have to vertically scroll down a single web page. A progress bar was added so that respondents could judge how much of the questionnaire remained because they could not see the entire survey as could those individuals taking the paper-version of the instrument. The final web-based survey was hosted on ICF's IBM/SPSS's Data Collection server.

Survey Administration - Invitations and Reminders

ICF worked with TMA to develop a communication strategy that included sending emails using six TMA listservs with a TMA utility (.mil) email account and a corporate utility email address specifically created for this project. We used the listserv and military email address to bolster the credibility of the emails that would be coming from the ICF corporate ".com" address. Using the listserv does not allow for any personalization of emails to individuals; the text of the outgoing message is the same for all recipients. Instead, we used the ICF email system to send tailored emails to respondents that included their survey access code (which was specific to their unit identification code or UIC). We used the listserv to alert the selected individuals that an email was coming from the ICF email domain. All notices from the listserv also included contact information for the survey administrator (ICF International) and a link to a letter of support from a ranking officer in the respondent's component. These letters of support were hosted on the TMA HPA&E organization website, providing further credibility for the study.

ICF sent invitations and reminders on a staggered start schedule in order to reduce email and web-survey server loads and because different components provided the supporting documentation (e.g., letter of support from command, whitelisting from each component's IT department) at different times.

ICF fielded the web-based survey first to the Air Force and Marine Reservists because their components were first to provide the letter of support and approval to commence from their Service Liaison Officer (SLO). Invitations were sent to the members of the remaining components as their supporting documentation and approvals were obtained.

Survey invitations and reminders were scheduled for three time periods: 1) the planned reminder period, 2) the extension of survey fielding, and 3) the booster time period for increasing the number of junior enlisted personnel participating. The planned reminder period encompassed the time period required to send the 10 planned types of correspondence for the first components that started the survey process. Initially all respondents received three authorization alert emails from the listserv and seven from the ICF corporate address. Because the onsite contractor's survey was delayed, TMA asked ICF to continue to solicit responses while the onsite contractor was in the field. During this extension of survey fielding, ICF sent three reminder emails scheduled three weeks apart followed by three more reminders scheduled four weeks apart. Thus, those components that had completed the original reminder schedule prior to the extension of the fielding period received an additional six reminders during the extension. Components that fielded later and had not completed their planned reminder schedule prior to the extension were incorporated into the extended reminder schedule once their planned reminder correspondence was complete. These components received fewer additional reminders during the extension period. Finally, towards the end of data collection, preliminary analyses showed that responses from junior personnel were low. To boost their participation rates, four additional reminders were sent specifically to individuals with pay grades of E1 to E4.

Email invitations were initially sent to the first group of respondents (Air Force and Marine Reservists) on August 3, 2010. The last email reminder was sent out May 11, 2011. The online system was left open for the selected participants to respond. Approximately 65 percent of usable surveys were completed within the first month of data collection, and 80 percent were completed within the first 3 months. As such, over 15,000 usable surveys were collected over three months of fielding. The last usable survey was provided on October 3, 2011.

All selected individuals were also sent paper invitations mailed to their unit address. These letters provided the same information as the email invitation with the survey's URL and their survey access code. This step was taken to include respondents who did not have a current email address registered with DMDC or who did not regularly check their email.

Ensuring Respondent Anonymity. The survey asked a number of questions on topics that are normally considered very private; in some cases, an indication of engaging in a specific behavior would constitute acknowledgement of engaging in illegal activity (e.g., taking recreational drugs). Therefore, it was important that the identity of

individuals could not be connected to their responses. The survey access codes that were sent to the respondents in email or via letter only identified the unit to which the individual was assigned and were not unique to the individual. Once the survey access code was entered, respondents were presented with a privacy statement. If respondents agreed to take the survey, they were asked to create a unique password that would allow them to re-access the survey if they were not able to complete it in one session. The respondent-generated password was not stored with the survey access code or any other identifying information (e.g., name, address, or email address); it was therefore not possible to connect survey responses with any individual.

Survey Support and Helpdesk. Technical support to address respondent concerns was provided by ICF's Survey Operations Center helpdesk. The helpdesk could be reached by email or toll-free telephone number. The helpdesk responded to email and phone calls between 8 AM – 6 PM Eastern Standard Time (EST); calls were automatically routed to voicemail outside of these hours. ICF responded to most messages or contacts within 1 to 2 business days. Respondents who wanted to verify the legitimacy of the request or authority of the survey or who had other questions needing to be addressed by a military source were forwarded to TMA's Contracting Office's Technical Representative or the components' SLO.

Preparing the Analytic Data File

At the close of the data entry period, the data were downloaded from the web environment to a raw SPSS data file. That file was subject to further cleaning, data transformations, and weighting procedures. The total number of respondent data records in the raw online data file was 22,083. As stated above, to assess the extent to which there was a "mode effect," the web survey software did not employ skip patterns, thus mimicking the experience a respondent would have with the paper survey. More data cleaning was therefore needed than if logical skip patterns had been implemented. First, rules were developed to define the characteristics of a "usable respondent record," or one that could be used for analysis. The rules for rejecting a respondent record were:

- If the respondent did not indicate being a member of any of the Guard and Reserve components (Question 1 in the survey), **or**
- If the respondent did not indicate their gender on the questionnaire (Question 5), **or**
- If responses for Pay Grade (Question 3) **and** for Education (Question 4) **and** for Age (Question 6) were **all** missing, **or**
- If five or fewer questions were answered above and beyond the demographic questions listed above, then the record was deleted from the working analytical dataset.

Application of these rules resulted in 3,188 cases being classified as not usable. The resulting analytic dataset for the online survey consisted of data from 18,895 respondents.

We applied several global data editing actions to the dataset:

- Creating a copy of all the variables so that the original values were maintained and editing could be done on a set of the "working variables." For example, the variable containing responses to Question 1 were copied to a new variable labeled EDQ001; Question 2 to a new variable labeled EDQ002, and so on. All further transformations and edits were done on these new variables.
- Calculation of summary or indicator variables for the various alcohol and drug items to facilitate the tabular display of data. For example, summary variables were created to indicate whether the respondent had imbibed any alcohol during the past 30 days and during the past 12 months.
- The values of various scales embedded within the questionnaire were calculated.

Finally, the data were weighted to reflect the two-stage sampling, with a post-stratification adjustment made based on known population figures.

Merging the Online and Onsite Data Files

To conduct the mode analysis comparing onsite and online collection modes, the two sources of data (i.e., onsite and online) needed to be merged into one dataset. The onsite file was converted to be consistent with the rules applied to the online file. It was important that identical variable recodes and transformations were applied across both datasets to eliminate all effects due to differences in variable coding between the two data files, and thus to accurately detect if a true mode effect was present.

The onsite dataset contained 16,507 usable respondent records, applying the "usable" definition above. The online dataset contained 18,895 usable respondent records. Combined, the dataset for the mode analysis contained 35,402 respondent records. ICF received the final onsite data files in SAS format. ICF converted the onsite SAS file to an SPSS file to be compatible with the online data file. After this conversion, ICF examined the onsite file to confirm that the data were raw and had not been previously cleaned or transformed in any way.

The first step in making the files compatible was to rename the onsite variables to the online variable names. The next step was to transform the values for each variable to ensure that all variables had identical values and value labels across datasets. The last step in making the onsite dataset comparable to the online dataset was recoding missing values (that were initially blanks in the data file) to '-8,' which indicated blank or missing responses. Once the files were the same for the onsite and online datasets, a 'Mode' variable was created in each respective file with the values of 0 (zero) signifying "onsite", and 1 (one) signifying online. The final step was to merge the onsite data file with the online data file, matched on identical variable names. After the onsite and online files were merged into one file, the online recode and transformation syntax was run in SPSS to recode variables and to create the scales. Variables were examined by mode to confirm that the recodes were accurate and consistent with the original meaning in the respective data files.

Analytical Variables and Key Survey Domains

The questionnaire covered eight topical domains including: demographic characteristics of the sample, alcohol use, tobacco use, illicit drug use, stress and mental health, healthy lifestyle and disease prevention, safety and other health related behaviors, and deployment and job satisfaction. Here, we provide an overview of the variables and scales measured in each domain.

Demographics. The demographic variables of interest in the survey were Guard and Reserve component, activation status, pay grade, education, gender, age, age at joining the military, marital status, race/ethnicity, the presence of children living at home, and income.

Alcohol Use. The alcohol use items on the survey covered a number of topics. The first section of alcohol-related questions on the survey specifically asked about the frequency and quantity of the consumption of beer, wine, and liquor over a 30 day timeframe as well as a 12 month timeframe. In addition, there were single-item indicators used in the analyses to represent heavy drinking, binge drinking, and drinking enough to feel drunk. Other alcohol indicators included multi-item measures such as reasons for limiting drinking, reasons for drinking, alcohol risk behaviors, and productivity loss at work due to drinking. The alcohol section contained items that were used in constructing two scales: the Ethanol Index and the Alcohol Use Disorders Identification Test (AUDIT) score. Additionally, we computed a categorical drinking level classification indicator.

Tobacco Use. The first section of tobacco-related questions on the survey specifically asked about the timing (age at first cigarette, smoking initiation since joining the military) and frequency and intensity of smoking based on 30 day (current smoker), 12 month (past smoker), and lifetime timeframes. In addition, there were single-item indicators used in the analyses to represent heavy smoking, regular smoking, and smoking cessation attempts. Other smoking indicators in the survey included multi-item measures such as perceived cigarette availability and acceptability and reasons for starting smoking regularly. Finally, there were measures of smokeless tobacco use in the past 30 days and in the past 12 months. The tobacco section contained one scale, the Fagerström Test for Nicotine Dependence.

Illicit Drug Use. The first few questions on drug use asked about the timing and intensity of drug use for 10 different drugs: marijuana, cocaine, LSD, PCP, MDMA, other hallucinogens, methamphetamines, heroin, GHB/GBL, and

inhalants, in the past 30 days, past 12 months, and lifetime use. There were additional items that asked about non-medical use of prescription stimulants other than methamphetamine, tranquilizers or muscle relaxers, sedatives or barbiturates, pain relievers, anabolic steroids, and erectile dysfunction (ED) drugs. An expanded question on prescription drug use for non-medical reasons was included in the current survey; it asked about the use of 24 different pain relievers in the past year. The section concluded with items on drug testing.

Stress and Mental Health. The stress and mental health questions contained items on stress (work-family conflict; work productivity loss; stress at civilian job, military job, and in the family; sources of stress; coping with stress), as well as mental health issues such as depression, anxiety, post-traumatic stress disorder, suicide ideation, suicide attempts, traumatic brain injury, the receipt of prescribed medication for mental health issues and whether professional counseling was sought for such issues. Physical and sexual abuse history were also covered. There were four major scales in this section – Depression, Generalized Anxiety Disorder (GAD), Post-Traumatic Stress Disorder (PTSD), and Traumatic Brain Injury (TBI).

Healthy Lifestyle and Disease Prevention. The healthy lifestyle items asked about moderate and vigorous physical activity and the number of days engaged in such activity in the past 30 days. Moderate activity was defined as “some exertion, but able to carry on a conversation comfortably during the activity.” Vigorous physical activity was defined as “finding it difficult to carry on a conversation during the activity.” The section also contained seven items asking about food intake in an average week, including fruit, vegetables, whole grains, dairy, lean protein, snack foods and sweets, and beverages.

Safety and Other Health Related Behaviors. The safety questions were single items that focused on seat belt use when driving or riding in a car, helmet use while driving or riding on a motorcycle, and hearing protection use when exposed to a loud noise at work. The questions in this section also focused on dental hygiene and sexually transmitted infections. There were two scales in the section: impulsivity/sensation-seeking and gambling.

Deployment and Job Satisfaction. Items in the deployment and job satisfaction portion of the survey asked about length of time in the Reserves, number and length of combat and noncombat deployments, activation status, theater of operations served in, combat exposure, change in substance use during deployment, job satisfaction, and satisfaction and change in relationship status with significant other after returning from deployment.

Response and Participation Rates

As noted earlier, the size of the sample across the six Guard and Reserve components was 79,913 individuals. It is necessary to compute the “eligible sample size” after taking into account the number of individuals in the original sample considered to be ineligible for the online administration.

Table 2 below shows the results of these calculations. The results are displayed for each of the six Guard and Reserve components. The left-most column in the table indicates the response element that is being taken into consideration in that row of the table. The right-most column contains totals, or averages in the case of percentages. The rows include:

- **Initial sample** is the number of individuals randomly selected from each Reserve component using the methods described earlier in the volume.
- **Number in sample missing email addresses** is the number of individuals supplied by DMDC but without email addresses. One of the requirements for the online survey was the possession of an email address by the persons in the sample so they could be invited to participate.
- **Number of emails not reaching the intended recipient** but with no notification of that event. This number is estimated from data from prior Status of Forces electronic surveys undertaken by DMDC. The estimated proportion of such emails was assumed to be four percent.
- **Number of emails returned by the various email servers as being "undeliverable,"** also known as "bouncebacks." Emails can be rejected for a number of reasons. When notification was received of a bounced

email, the information was recorded. However, the reasons for the bouncebacks were not investigated, nor were there any attempts to correct the email and retry sending it.

- **Net or effective sample size** is the initial sample size with the next three quantities subtracted from it. This number represents the number of eligible individuals in the online sample.
- **Number of participants** represents the number of individuals who clicked on the hyperlink in the email invitation or reminder and logged into the survey.
- **Number of usable respondents** is the number of respondent records who met all the criteria for being usable for analysis.
- **Response rate** is the ratio of the number of participants to the effective sample size.
- **Cooperation rate** is the ratio of usable respondent records to the number of participants.

The proportion of missing email addresses was relatively constant across all the Reserve components. The number of bouncebacks varied by component with the Air National Guard having many more than other Reserve components. The original sample was reduced by approximately 15 percent due to the missing addresses and non-delivery of emails.

Table 2. Survey Response Data and Performance Rates

Reserve Component								
Response Data		Army National Guard	Army Reserve	Navy Reserve	Air National Guard	Air Force Reserve	Marine Corps Reserve	Total Reserve Component
A	Initial Sample	11,174	16,699	16,678	11,119	16,676	7,567	79,913
B	Missing Email Addresses	81	117	80	58	48	43	427
C	Estimated mails not delivered	410	603	593	322	604	275	2,807
D	Email Bouncebacks	826	1,518	1,745	2,981	1,506	613	9,189
E	Net Sample (A) - (B) - (C) - (D)	9,857	14,461	14,260	7,758	14,518	6,636	67,490
F	Participants	3,622	3,760	3,029	5,043	5,566	1,063	22,083
G	Usable respondents	3,514	3,651	2,881	3,521	4,312	1,016	18,895
H	Response Rate (F) / (E)	36.7%	26.0%	21.2%	65.0%	38.3%	16.0%	32.7%
I	Cooperation Rate (G) / (F)	97.0%	97.1%	95.1%	69.8%	77.5%	95.6%	85.6%

The response rate was highest for the Air National Guard and lowest for the Marine Corps Reserve. This was offset by Air National Guard having the lowest cooperation rate, and the Marine Corps Reserve having a higher cooperation rate.

A primary purpose of the online pilot study was to evaluate the **feasibility** of conducting an online survey among Reserve and Guard units. The ICF data collection established that more completes can be gathered more quickly and for a lower cost than the traditional onsite methodology. A secondary purpose of the pilot study was to **compare the results obtained using an onsite methodology with an online methodology** of data collection. In this section we summarize selected analyses of mode effects.

The method and composition of the sample drawn for onsite administration was different from that drawn for online administration. The onsite sample was drawn first from the available sample frame of units with probability

proportional to size, with geographic clustering of units so as to enable more efficient onsite data collection. Respondents were initially randomly selected to represent the six levels of pay grade and two levels of gender stratification, with oversampling of females and officers. However, a non-random replacement strategy was employed onsite to replace respondents who did not show up at their appointed time. By contrast, the online sample was drawn second, from the units remaining after onsite sample selection. The online sample was drawn in two stages. First, units were randomly drawn proportional to size, and then respondents within units were randomly drawn within strata, with a similar oversample of females and officers as was done in the onsite administration. However, unlike the onsite methodology, there was no replacement strategy used for respondents who failed to respond to the web survey.

Selecting the samples from two different pools of units, each having different non-response biases and probabilities of selection,³ could have created major sample differences between the two modes of administration even before survey administration. Rather than being constant, sample differences could be covarying with mode of administration.

Summary of Analytical Approach

Since the sampling design, the obtained sample, and the methods used to calculate the weights for the online and onsite samples were different and are therefore not comparable, we used the raw, unweighted data to conduct hierarchical regression analyses to determine mode effects. The mode regression analyses examined the extent to which differences in the dependent variable of interest could be statistically attributed to the mode of data collection, after controlling for demographic and other differences between samples.

After merging the two datasets, we investigated sample differences between the modes. To do so, we first examined correlations between mode of administration and sample variables (i.e., measured variables that may reflect sample characteristics, such as demographics and psychological characteristics). We then entered the variables that were correlated with mode into a step-wise regression model, entering variables that predicted mode of administration one at a time until the increment in explanatory power was minimized. Using this procedure, we identified 10 variables, shown in Table 3, that were most predictive of mode of administration. These sample-dependent variables served as covariates that allowed us to remove some of the sample differences from other dependent variables of interest to more clearly determine the mode effect. All non-dichotomous categorical variables were “dummy-coded” to fully represent each variable among the set of covariates and are noted in Table 3. While this process allowed us to adjust for differences in the samples using measured variables, many unobserved differences likely remain between the samples, limiting the comparability of the samples.

Variable	Description
▪ Age	Age in 4 ordinal categories (24 or younger, 25-34, 35-44, 45 or older)
▪ Impulsivity	Average of 9 items, scale ranging from 1-4
▪ Activation Status	Activated full-time status; non-activated is the reference group
▪ Gender	Female; males are the reference group
▪ Service	5 dummy codes; Air Force is the reference group
▪ Pay grade	5 dummy codes; O4-O10 is the reference group
▪ Depression	Sum of 6 dichotomized items; scale ranging from 0-6
▪ Marital status	Married; unmarried is the reference group
▪ Race/ethnicity	3 dummy codes; African American is the reference group
▪ Education	Education in 4 ordinal categories (high school or less, some college, college graduate/some grad school, graduate degree)

We focused our mode analysis on selected risk behaviors representing the most important outcomes. We selected key variables from each of the variable domains to assess a diversity of variables across the survey. The variables chosen for the mode analysis included:

³ The PPS sample of units selected for the onsite administration was selected first and purged from the frame for the (PPS) selection of units for the online administration. Therefore, the first sample likely included units that tended to be larger than those in the second sample overall.

- Heavy alcohol use, defined as consuming five or more drinks per typical drinking occasion at least once a week in the past 30 days;
- Binge drinking, defined as consuming five or more drinks on one occasion at least once in the past 30 days;
- Hazardous drinking, defined as a score of eight or higher on the Alcohol Use Disorders Identification Test (AUDIT);
- Current smoking status, defined as smoking at least one cigarette in the past 30 days;
- Heavy smoking, defined as smoking one or more packs of cigarettes per day in the past 30 days;
- Nicotine dependence, measured by six items from the Fagerström Test for Nicotine Dependence;
- Smokeless tobacco use, defined as use of chewing tobacco, snuff, or other smokeless tobacco in the past 12 months;
- Marijuana use in the past 12 months;
- Illicit drug use except marijuana in the past 12 months, including cocaine, hallucinogens, amphetamines/stimulants, tranquilizers, barbiturates/sedatives, heroin/opiates, analgesics, and inhalants;
- Pain reliever misuse in the past 12 months, defined as any non-medical use of 24 different pain relievers;
- Lifetime suicide ideation, defined as ever seriously considering suicide;
- Being overweight, defined by a Body Mass Index (BMI) of 25 or greater; and
- Incidence of Sexually Transmitted Infections (STIs) in the past 12 months, including gonorrhea, syphilis, chlamydia, and genital herpes.

To reach comparable results, we employed hierarchical logistic regression for all outcome measures; as such, we dichotomized the dependent variables where necessary. In the first stage of the hierarchical analysis, we controlled for demographic and other differences in the composition of the two samples using the covariates we identified in Table 3. This effectively removed variance between the samples due to these factors. In the second stage, we analyzed the ‘mode’ dichotomous variable which served as the primary independent variable to detect mode differences for the dependent variable/outcome of interest, such as, for example, heavy alcohol use.

Results from Regression Analysis of Mode Effects

Table 4 provides a summary of the 13 outcome variables.

Dependent Variable	Onsite Paper Survey^a	Online Web Survey^a	Model Pseudo R²	Pseudo R² due to Mode	Mode Odds Ratio^b
Alcohol					
1. Heavy Alcohol Use	22.1%	8.9%	.343	.010	.51**
2. Binge Drinking	37.8%	23.7%	.201	.001	.80**
3. Hazardous Drinking (AUDIT)	17.7%	7.5%	.260	.007	.57**
Tobacco					
4. Current Smoker	23.4%	14.7%	.164	.000	.86**
5. Heavy Smoker	10.4%	7.5%	.105	.001	.81*
6. Nicotine Dependence	13.1%	12.2%	.071	.001	.81
7. Smokeless Tobacco	17.7%	8.8%	.240	.000	1.00
Illicit Drugs					
8. Marijuana Use	3.7%	1.2%	.207	.014	.35**
9. Illicit Drug Use	15.7%	13.4%	.076	.007	.64**
10. Pain Reliever Misuse	5.7%	3.5%	.106	.011	.48**
Mental Health, Healthy Lifestyle and Disease Prevention					
11. Suicide Ideation	6.8%	4.3%	.207	.002	.72**
12. Overweight/Obesity (BMI)	61.5%	65.1%	.177	.001	1.14**
13. Sexually Transmitted Infections	1.8%	1.0%	.111	.000	.88

^a Estimated prevalence rates after controlling for covariates.

^b *Odds ratio is significant at $p < .05$. **Odds ratio is significant at $p < .01$. ***Odds ratio is significant at $p < .001$.

The table presents the list of dependent variables examined, along with the estimated prevalence rates after controlling for the differences between the onsite and online modes of survey administration. In addition, the table shows the overall model Pseudo R^2 (the effect of mode and covariates on the dependent variable), the Pseudo R^2 due to mode (i.e., Incremental Pseudo R^2), and the mode odds ratio for the effect of mode on the dependent variable and its significance. Overall, marijuana use in the past 12 months, pain reliever misuse in the past 12 months, and heavy alcohol use in the past 30 days demonstrated the most significant, albeit small, mode effects. The variables that had a non-significant difference between modes of survey administration were smokeless tobacco use in the past 12 months and contraction of STIs in the past 12 months. Results show that the online survey administration mode typically resulted in lower prevalence rates of health related behaviors, but differences were smaller after controlling for the differences between samples.

Correlational Analysis of Mode Effects

Although the analyses above found that the online administration results typically indicated lower prevalence estimates, another way to understand the results from the modes is to ask if the results obtained from each mode are ordered similarly across key groups of interest. For example, one can ask whether one component scores the lowest on an outcome measure in both modes, or whether one gender-pay grade cell in the strata has the lowest prevalence in both modes while another has the highest prevalence in both modes. To analyze the comparability of the order of the survey results, we split the file into smaller cells to obtain weighted prevalence estimates for 15 key risk and protective measures:

- Non-hazardous AUDIT (1=Audit<8; 0=Audit≥8)
- Heavy Alcohol Use Past 30 Days
- Alcohol Binge Episode in Past Month
- Non-Smoker (1=Non-smoker; 0=Smoker)
- Heavy Smoking
- Nicotine Dependence
- Any Smokeless Tobacco Use
- Past Year Marijuana Use for Mode
- Past Year Illicit Drug Use, Except Marijuana
- Past Year Pain Reliever Misuse
- Vigorous exercise (3+ days per week)
- 7 or more hours of sleep per night
- Having 1 or more combat deployments
- BMI in overweight or obese range
- Work stress ('some' or 'a lot')

To have a sufficient number of prevalence estimates within strata, we split the file by mode and gender-pay grade (2 x 12) for each variable and split by mode and component (2 X 6). This resulted in 270 prevalence estimates for each mode of administration ((12 + 6 cells) x 15 outcome measures = 270 estimates). We then correlated the prevalence estimates obtained for each mode. Results for the 270 point estimates calculated for each mode showed a correlation of .975 ($p<.0001$). Figure 1 portrays the correspondence. While the absolute values of the measures for many measures were significantly different 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.

After exploring the correlations between the prevalence estimates of key questionnaire items for the onsite and online modes, we examined the extent to which results from one mode predicted results in the other mode to assess the extent of bias. To predict the mean unit change for the online items, the online prevalence estimates for the key items were regressed on the onsite prevalence estimates. The result of this analysis showed that for each proportion unit change in the onsite survey (i.e., change of 1 for a key item), there was a corresponding unit increase of 1.045 (unstandardized) in the online survey's proportion for that item ($F_{(1,267)}=5115.819, p<.001$). Alternatively, if we wanted to predict the proportion unit change for the onsite survey using the online proportions, the results showed a .910 (unstandardized, $F_{(1,267)}=5115.819, p<.001$) unit change; in other words, for each one unit change in the online proportion, the onsite proportion changed by .910. Both of these unstandardized regression coefficients show that, even if there are significant differences between the prevalence estimates, changes in the key item means for one

administration mode predicted almost an identical change in the estimate for the other administration. These findings can be useful for converting results from one mode to another, as might occur when we want to revise prior results collected onsite to make them appear ‘as if’ they were collected online (also known as ‘backcasting’). Backcasting is a useful tool to facilitate the continuation of trends when switching modes.

Figure 1. Correlation of online and onsite prevalence estimates

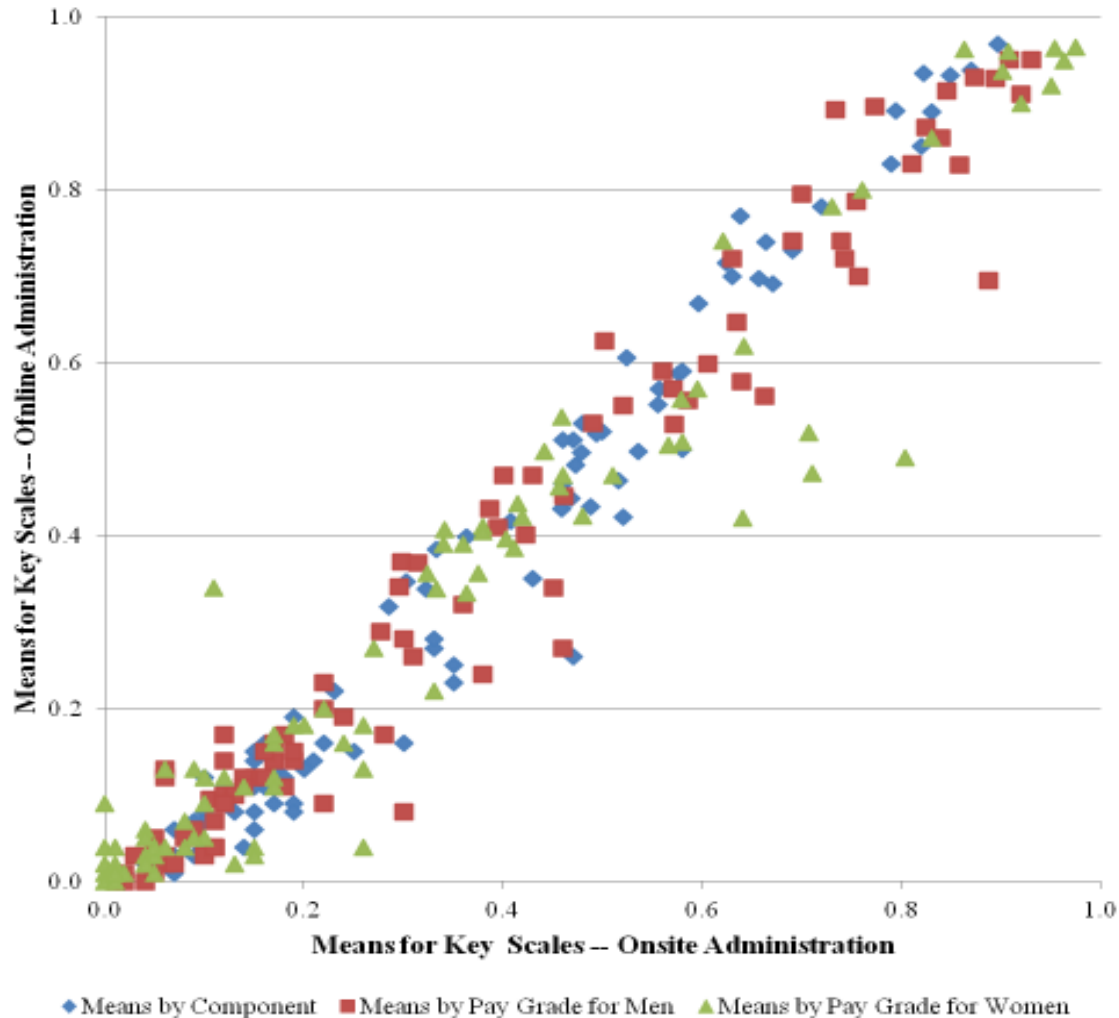
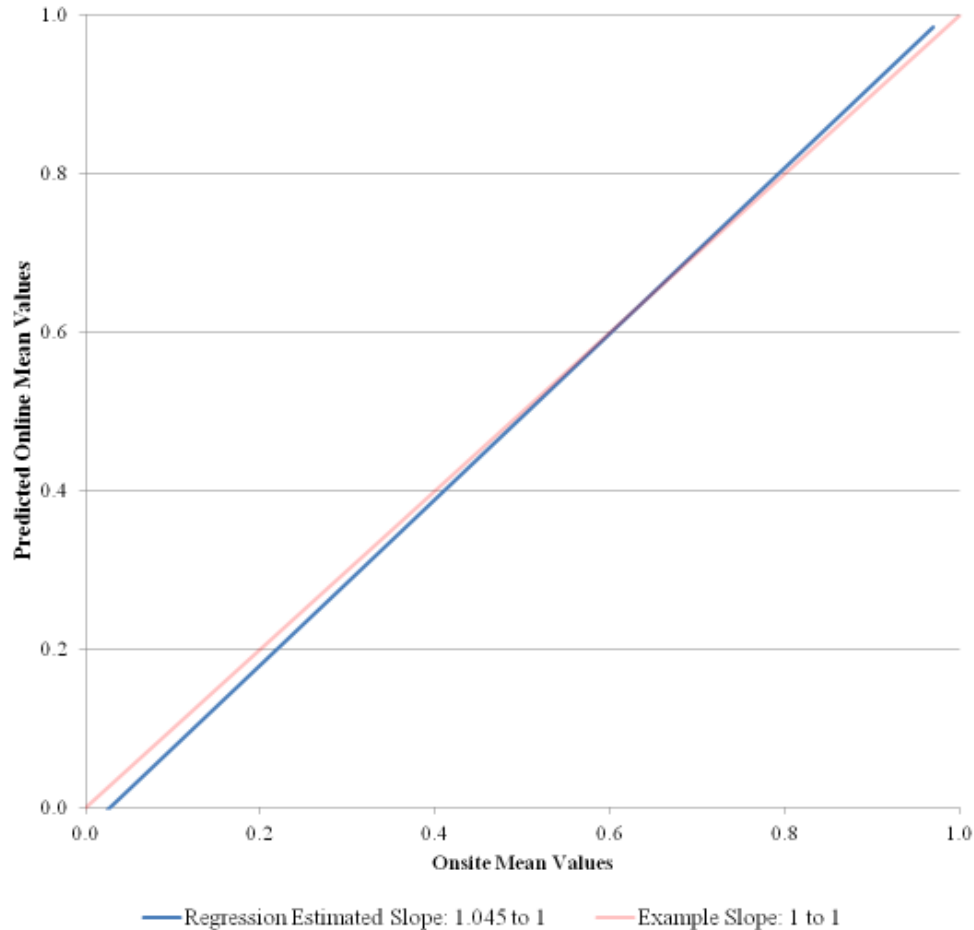


Figure 2, below, presents the slope of the regression equation obtained from regressing the online proportions for key items on the onsite proportions. In a perfect relationship, a change in the key items obtained from the onsite mode would correspond to an identical change in the key items in the estimate from the online mode. As shown in the graph below, the prediction of online values from the observed onsite values was close to a perfect relationship.

Figure 2. Extent of Bias of Online Mean Values Predicted by Onsite Mean Values Compared to a Perfect Relationship



Summary of Mode Analysis Results

In our initial analysis of the raw data from both modes, there were significant sample differences for a number of demographic characteristics, including age, gender, pay grade, and education. Even after controlling for many of these differences, we obtained significantly different results for many of the outcome measures, although many of the differences between outcome measures were smaller. There are a number of possible reasons why we might obtain different results from similar samples when using different modes of administration, which are presented below in Table 5.

Table 5. Comparability of Onsite and Online Methodology

Sources of Mode Differences	Examples
Sampling	Sample geographic clusters, strata, sample proportions, non-random replacement in onsite
Method of survey administration	In-person group self-administration with proctors , paper survey, web survey
Survey administration setting	Group setting in large room versus individual setting at one's computer
Differential non-response	People who respond in one mode may be different than those who respond in another mode
Social desirability pressures	Social situations, including group settings and human interviewers

Table 5. Comparability of Onsite and Online Methodology

Sources of Mode Differences	Examples
Item wording	Respondent instructions may differ based on mode of administration
Response wording and response choices	May not be the same due to mode, may need a 'Decline to answer' to be explicitly presented in one mode but not another
Item order may be different	Order may change based on efficiency for mode
Item grouping	Singly presented items versus presenting items as groups
Response order within items	Telephone surveys have auditory biases (recency) while paper and online surveys have visual biases (primacy)
Sample filtering/logic skips	Respondents may be asked follow-up questions differently in one mode versus another
Weighting algorithms	Procedures for accounting for design effects and computing post-stratification weights may be different
Weighting targets	Population targets may be derived from different sources
Methods of response imputation/backfilling	Data cleaning and filling of empty cells may have different rules
Season of survey administration	People do different behaviors at different times of the year, so they may report differently if times are different, especially if data collection occurs during a 12-month period

As a result of the differences we observed between the modes, it is important to identify more clearly the reasons that onsite and online prevalence estimates diverge. At a basic level, it may be that when respondents complete surveys in an onsite setting with peers they may be tempted to exaggerate their use of alcohol or illicit drugs. Respondents in an anonymous environment such as an online survey may not experience such pressures to exaggerate. Conversely, if in the online environment respondents feel that their responses are monitored online and therefore may be identifiable, they may under-report substance use and other behaviors that might put their jobs in jeopardy. Additional research would assist in understanding the true respondent reactions to both onsite and online modes of administration. While a number of DoD-wide surveys have gone to an all-online mode of administration, such as the Status of Forces surveys, further study of the psychological environment of the respondent is warranted to ensure that accurate information concerning sensitive behaviors can be collected online.

Discussion

This pilot study of the effects of mode on the data collected for the 2009-2010 Survey of Health Related Behaviors Among Reserve Component Personnel had four primary research questions. The pilot study sought to address these questions by conducting the HRB online with a stratified random sample of members of the Guard and Reserve Component, and by comparing results obtained via the online mode to the data obtained through the onsite mode. It should be noted that the implementation of the online mode tried to simulate as closely as possible the onsite experience with a paper instrument.

Based on the analysis described in the preceding sections, we can address these research questions as follows.

- Is conducting the Health Related Behaviors Survey online feasible? Yes. Not only is doing so feasible, it may become preferable. The majority of online completed questionnaires were obtained during the first three months of data collection. This contrasts with onsite data collection which lasted approximately one year. In addition to being faster, online data collection was also more efficient, requiring fewer overall resources from the military, with minimal disruption to normal Guard and Reserve activities.
- What effect, if any, will a change in mode of administration have on the prevalence estimates and incidence rates? In comparison to past iterations of the HRB Survey, the prevalence estimates for unhealthy behaviors, such as excessive ingestion of alcohol, smoking, and excessive drug use and misuse tend to be

reported at lower rates by online respondents. This finding corresponds to findings that we have obtained in other studies of military populations, including the Air Force Community Assessment Survey that has been conducted for the past 12 years. While we can speculate as to the reasons behind this finding (e.g., peer pressure during onsite administration or lack of trust with regard to anonymity of online responses), it does seem to be consistent across a number of different behaviors. However, as noted in the analysis of the effects of mode, the overall order of prevalence rates (as contrasted with their absolute value) is almost identical between the two modes. Thus, from a mode perspective, both modes measure the same underlying constructs.

- Will the ability to compare trends across different iterations of the survey over time be lost or can it be maintained through the change in survey mode? Self-report questionnaires can only estimate the prevalence of behaviors. Such self-reporting is influenced by a variety of factors and is only an approximation to the true underlying prevalence rate. In our initial analysis of the raw data from both modes, there were significant differences for a number of demographic characteristics, including age, gender, pay grade, and education. Even after controlling for these sample differences (though many were unmeasured and therefore not able to be controlled for), we obtained significantly different results for many of the outcome measures. We generally found that we obtained lower rates for high risk behaviors and higher rates for protective behaviors when the survey was conducted online, which may reflect differential mode bias that can be quantified. As noted, the order of prevalence rates is very similar between the two modes. Thus, it may be useful to discuss the possibility of establishing a new trending baseline for online data collection that can be used in future trend analysis. However, based on the very high and significant degree of correspondence that we obtained between modes, we can use derived formulae predicting online from onsite data and apply these formulae to earlier iterations in order to 'backcast,' or adjust prior data to resemble data collected online. In this way, prior years' data can be adjusted so that trend lines can continue into the future, though based on the new mode of data collection. This type of analysis would involve additional cost, but would need to be done only once. Data collected online can detect differences between services as well as changes in risk and protective behaviors, and can be used to provide feedback on program effectiveness.
- How will the survey response rates be affected by a change in mode? The overall online response rate was 32.7 percent. This is consistent with response rates reported by other online surveys of military populations, such as recent Status of Forces surveys conducted by Defense Manpower Data Center (DMDC).⁴ We believe that the high response rates traditionally obtained with past iterations of the HRB Survey can be obtained in the online mode with suitable changes to administrative operations.

The economics and advantages of moving survey data collection to an all electronic environment - whether over the Internet or some other infrastructure - are many and compelling: greater efficiency, less disruption, logistical simplicity, more elegant and efficient sampling methodology, scalability to arbitrarily large samples, and faster analysis. The advantages of the online mode are illustrated by one startling fact obtained in this project:

Onsite administration of the HRB Survey at 153 military installations took approximately 13 months of calendar time, and involved over 20 teams of two to three people. For the online administration, 80% of all usable surveys, over 15,000 in number, were collected during the first three months of data collection for each component.

Conceivably, because of the scalability of online methods, surveys can be administered to an even larger sample or even the census of the Reserve components with only incremental additional costs. Thus, in a time of budget cuts and diminishing resources, online methods of data collection have a genuine advantage.

The wide-spread availability and use of computers in many forms, such as tablets, "smart phones," laptops, and desktops is fundamentally changing the ways in which we work, play, and communicate. Newer generations of Soldiers, Airmen, Sailors, and Marines are not only more comfortable interacting with computers and machines,

⁴ See for example the December 2009 Status of Forces Survey of Reserve Component Members, which had a response rate of 27 percent.

they often prefer it. The ubiquity of the Internet and its attendant means of access, including Wi-Fi and broadband, along with devices that use those electronic pathways, is slowly but surely revolutionizing the manner in which data are collected, processed, analyzed, and used. The survey industry is undergoing profound change in ways that will touch all private and public sector organizations.

The potential benefits of online data collection are formidable. The question, then, becomes one largely of "When", not "If". The answer to "If" was provided some time ago. The transition, however, must be done carefully and deliberately, and with eyes wide open, in order to reap the benefits of this transition.

Bibliography

- Babor, T. F., Higgins-Biddle, J. C., Saunders, J. B., & Monteiro, M. G. (2001). *The Alcohol Use Disorders Identification Test (AUDIT): Guidelines for Use in Primary Care* (2nd Ed.). World Health Organization, Department of Mental Health and Substance Dependence, Geneva, Switzerland.
- Cherpitel, C. J. (1999). Substance use, injury, and risk-taking dispositions in the general population. *Alcoholism: Clinical and Experimental Research*, 23, 121-126.
- Defense Manpower Data Center. (2009). November 2008 Status of Forces Survey of Active Duty Members: Administration, Datasets and Codebook. DMDC Report No. 2008-033.
- Defense Manpower Data Center. (2010). December 2009 Status of Forces Survey of Reserve Component Members: Tabulations of Responses. DMDC Report No. 2010-002.
- Heatherton, T. F., Kozlowski, L. T., & Frecker, R. C. (1991). The Fagerström test for nicotine dependence: A revision of the Fagerström tolerance questionnaire. *Br J Addict* 86, 1119–1127.
- Kuczmarski, R. J., & Flegal, K. M. (2000). Criteria for definition of overweight in transition: Background and recommendations for the United States. *American Journal of Clinical Nutrition*, 72, 1074-1081.
- Radloff, L. S. (1977). A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385-401.
- Robins, L. N., Helzer, J. E., Croughan, J., & Ratcliff, K. S. (1981). National Institute of Mental Health Diagnostic Interview Schedule: Its history, characteristics, and validity. *Archives of General Psychiatry*, 38, 381-389.
- Spitzer, R. L., Kroenke, K., & Williams, J. B. (1999). Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. Primary care evaluation of mental disorders. Patient Health Questionnaire. *Journal of the American Medical Association*, 282, 1737-1744.
- Weathers, F. W., Litz, B. T., Huska, J. A., & Keane, T. M. (1994). *The PTSD checklist-civilian version (PCL-C)*. Boston, MA: National Center for PTSD.
- Wolkowitz, O. M., Roy, A., & Doran, A. R. (1985). Pathological gambling and other risk-taking pursuits. *Psychiatric Clinics of North America*, 8, 311-322.