Office of People Analytics (OPA)

2016 Workplace and Gender Relations Survey of Active Duty Members

Statistical Methodology Report

Additional copies of this report may be obtained from:

Defense Technical Information Center

ATTN: DTIC-BRR

8725 John J. Kingman Rd., Suite #0944

Ft. Belvoir, VA 22060-6218

Or from:

http://www.dtic.mil/

Ask for report by Report ID

OPA Report No. 2016-004
March 2017

2016 WORKPLACE AND GENDER RELATIONS SURVEY OF ACTIVE DUTY MEMBERS: STATISTICAL METHODOLOGY REPORT

Office of People Analytics (OPA)
Defense Research, Surveys, and Statistics Center
4800 Mark Center Drive, Suite 04E25-01, Alexandria, VA 22350-4000

Acknowledgments

The Office of People Analytics (OPA) is indebted to numerous people for their assistance with the 2016 Workplace and Gender Relations Survey of Active Duty Members (*2016 WGRA*), which was conducted on behalf of Major General Camille Nichols, Director, DoD Sexual Assault Prevention and Response Office (SAPRO).

Policy officials contributing to the development of this survey include Dr. Nathan Galbreath (Office of the Under Secretary of Defense, Personnel and Readiness, Sexual Assault Prevention and Response Office) and Ms. Shirley Raguindin (Office of the Under Secretary of Defense, Personnel and Readiness, Office of Diversity Management and Equal Opportunity). Service officials contributing to the development and administration of this assessment include Ms. Jessica Gallus (Army), Dr. Paul Garst (Department of Navy, SAPRO), Mr. Paul Rosen and Ms. Kimberly Lahm (Navy), Ms. Melissa Cohen and Dr. Jessica Zabecki (Marine Corps), Mr. James Thompson and Ms. Aileen Richards (Air Force).

RSSC's Statistical Methods Branch, under the guidance of Mr. David McGrath, Branch Chief, is responsible for all statistical aspects of this survey, including, sampling, weighting, nonresponse bias analysis, and the implementation of statistical hypothesis testing used in the survey program. Mr. Eric Falk, Team Lead of the Statistical Methods Branch, was responsible for managing the *2016 WGRA*. Mr. Jeff Schneider, Mathematical Statistician, used the OPA Sampling Tool to design the sample and implemented the weighting methods. Ms. Sue Reinhold provided the data processing support. Data Recognition Corporation (DRC) performed data collection and editing. Dr. Bob Fay and Dr. Minsun Riddles, Westat, consulted on statistical weighting methods.

Table of Contents

Page

Introduction....1

Sample Design and Selection....1
Target Population....1
Sampling Frame....2
Sample Design....2
Sample Allocation....4
Weighting....5
Case Dispositions....5
Nonresponse Adjustments and Final Weights....8
Comparison to the 2014 RAND RMWS Study and 2015 DMDC WGRR....15
Variance Estimation....15
Multiple Comparison Section....16
Contact, Cooperation, and Response Rates....17
Ineligibility Rate....19
Estimated Ineligible Postal Non-Deliverable/Not Contacted Rate....19
Estimated Ineligible Nonresponse....19
Adjusted Contact Rate....19
Adjusted Cooperation Rate....19
Adjusted Response Rate....19
Nonresponse Bias Analysis....20

References....23

List of Tables

1. Variables for Stratification and Key Reporting Domains....3
2. Sample Size by Stratification Variables....5
3. Case Dispositions for Weighting....7
4. Complete Eligible Respondents by Stratification Variables....8
5. Key Outcome Variables....9
6. Variables Used to model Key Outcome Variables....9
7. Variables used for Raking....13
8. Distribution of Weights and Adjustment Factors....14
9. Sum of Weights by Eligibility Status....15
10. Disposition Codes for Response Rates....18
11. Contacted, Cooperation, and Response Rates....20
12. Rates for Full Sample and Stratification Categories a....20

2016 WORKPLACE AND GENDER RELATIONS SURVEY OF ACTIVE DUTY MEMBERS: STATISTICAL METHODOLOGY REPORT

Introduction

The *Defense Research, Surveys, and Statistics Center*, Office of People Analytics (OPA), conducts both web-based and paper-and-pen surveys to support the personnel information needs of the Under Secretary of Defense for Personnel and Readiness (USD[P&R]).[1] These surveys assess the attitudes and opinions of the entire Department of Defense (DoD) community on a wide range of personnel issues. Health and Resilience (H&R) Surveys are in-depth studies of topics, which impact the health and well-being of military populations.

This report describes the statistical methodologies for the *2016 Workplace and Gender Relations Survey of Active Duty Members* (*2016 WGRA*). The first section describes the sample design and selection of the sample. The second section describes weighting and variance estimation, as well as a comparison to the *2014 RAND Military Workplace Study* (*2014 RMWS*) and *2015 Workplace and Gender Relations Survey of Reserve Component Members* (*2015 WGRR*). The third section describes the statistical tests used for the *2016 WGRA*. The fourth section describes the calculation of cooperation, completion, and response rates for the full sample and population subgroups. The fifth section provides an overview of the nonresponse bias analysis that will be done at a later date. Estimates for all survey questions are found in the *2016 Workplace and Gender Relations Survey of Active Duty Members: Tabulation Volume* (OPA, 2017a).

Sample Design and Selection

Target Population

The *2016 WGRA* was designed to represent individuals meeting the following criteria:

- Active duty members of the Army, Navy, Marine Corps, Air Force and Coast Guard excluding Public Health and NOAA members
- Including paygrades E1 to O6
- Be on the March 2016 Active Duty Master File (ADMF)
- Valid Personnel status (Not a prisoner, deserter, or unknown)

National Guard and Reserve members in active duty programs were excluded. Data were collected between July 22 and October 14, 2016.

[1] Prior to 2016, the Defense Research Surveys, and Statistics Center (RSSC) resided within the Defense Manpower Data Center (DMDC). In 2016, the Defense Human Resource Activity (DHRA) reorganized and moved RSSC under the newly established Office of People Analytics (OPA).

Sampling Frame

The sampling frame consisted of 1,330,357 active duty members (1,291,357 DoD and 39,000 Coast Guard) determined from using the March 2016 ADMF. Auxiliary frame data were obtained from the following files:

- February 2016 Active Duty Family Database (ADFD)
- March 2016 Basic Allowance for Housing (BAH) File
- March 2016 Contingency Tracking System (CTS) Deployment File
- April 2016 Defense Enrollment Eligibility Reporting System (DEERS) Medical Point-in-Time Extract (PITE)
- April 2016 Unit Identification Code (UIC) Address File
- June 2016 Database Extract (DBE) File
- March 2016 Reserve Components Common Personnel Data System (RCCPDS) Master File (Dual Spouse Variable)

In addition, after selecting the sample, OPA performed additional checks to verify the member was still eligible. Any ineligible member in the sample was excluded from any mailings and notifications. This saved additional costs associated with the survey process. Individuals were included on the frame based on membership in both the March 2016 ADMF and the April 2016 PITE; sample members no longer in the April 2016 DEERS Medical PITE were dropped from the sample and recorded as record ineligibles. There were 9,247 (1.3%) members determined to be record ineligible from this process. Sample members who became ineligible during the field period, were identified as self- or proxy-report ineligible. There were 1,278 (0.2%) members who were identified as being ineligible through either the survey instrument or some other means.

Sample Design

The sample for the *2016 WGRA* survey used a single-stage stratified design. Design parameters from the DoD Sexual Assault and Prevention Office (SAPRO) specified an agreed upon fifteen installations that RSSC would consider when designing the sample to ensure that there were a sufficient number of respondents to make accurate estimates by base and gender.

RSSC implemented the stratification in two steps. First, the SAPRO installations were considered based on their size and expected number of respondents. From this analysis RSSC stratified eleven of the fifteen bases by gender. The remaining four bases were determined to be of sufficient size that no additional stratification was necessary. Then, the remaining population was stratified using the following five characteristics:

- Service (Army, Navy, Marine Corps, Air Force, Coast Guard),

- Gender (Male, Female),
- Paygrade grouping (E1-E4, E5-E9, W1-W5/O1-O3, O4-O6),
- Race/Ethnicity (non-minority, minority), and
- Family Status (Single with Children, Dual Spouse[2], all other).

Table 1 shows these five variables and associated variable levels.

Table 1.
Variables for Stratification and Key Reporting Domains

Stratification Variable	Variable Name	Categories
Service	CSERVICE	1. Army
		2. Navy
		3. Marine Corps
		4. Air Force
Gender	CSEX	1. Male
		2. Female
Paygrade group	CPAYGRP5	1. E1-E4
		2. E5-E9
		3. W1-W5/O1-O3
		4. O4-O6
Race	CRACECAT	1. Non-Minority
		2. Minority
Family Status	FAMSTAT4	1. Single w/ Children
		2. Dual Service Spouse
		3. All Others

OPA partitioned the population frame into 207 strata that were initially determined by the aforementioned eleven bases and a full cross-classification of the five stratification variables. Categories (specific categories from Table 1 such as "Single with Children") were collapsed when there were less than 200 in the stratum (collapsing "Minority" with "Non-Minority" to form a new stratification level "All Races"); occasionally, stratification variables were collapsed, in reverse order as listed. Service, gender, and paygrade group boundaries were always preserved.

OPA selected individuals with equal probability and without replacement within each stratum. However, because allocation was not proportional to the size of the strata, selection

[2] Members that have a spouse in the Active or Reserve Military are considered to have a "Dual Service Spouse"

probabilities varied among strata, and individuals were not selected with equal probability overall. To achieve adequate sample sizes for all domains (reporting categories), OPA used a nonproportional allocation.

Sample Allocation

OPA designed the sample to achieve the goal of reliable precision on estimates for outcomes associated with reporting a sexual assault (e.g., retaliation) and other measures that were only asked of a very small subset of members, especially for males. Given estimated variable survey costs and anticipated eligibility and response rates, OPA used an optimization algorithm to determine the minimum-cost allocation that simultaneously satisfied the domain precision requirements. Response rates from previous surveys were used to estimate eligibility and response rates for all strata. The *2014 Status of Forces Survey of Active Duty Members (SOFS-A)*, the *2013 SOFS-A*, the *2012 SOFS-A*, and the *2012 WGRA* were used to estimate these rates.

OPA determined the sample allocation by means of the OPA Sample Planning Tool (SPT), Version 2.1 (Dever & Mason, 2003). This application is based on the method originally developed by J. R. Chromy (1987) and described in Mason, Wheeless, George, Dever, Riemer, and Elig (1995). The SPT defines domain variance equations in terms of unknown stratum sample sizes and user-specified precision constraints. A cost function is defined in terms of the unknown stratum sample sizes and the per-unit cost of data collection, editing, and processing. The variance equations are solved simultaneously, subject to the constraints imposed, for the sample size that minimizes the cost function. Estimated eligibility rates are used and they modify the estimated prevalence rates used in the variance equations, thus affecting the allocation; response rates inflate the allocation, thus affecting the final sample size. Prevalence rates refer to a percentage that is used in determining the estimated variance used for the calculation of the sample size. For example, OPA used 50 percent since it is the most conservative and yields the largest estimated sample size.

There were 92 reporting domains defined for the *2016 WGRA* and the initial goal was to achieve below 5 percent precision on estimates. There was no administrative data associated with 16 of these domains since they associated with either sexual assault or sexual harassment. The precision requirement for each domain is typically based on an estimated prevalence rate of 50 percent with a 95 percent confidence interval half-width no greater than ± 5. However, given the rarity of events covered by many of the *2016 WGRA* questions, OPA ensured that a much tighter precision would be met for questions seen by all respondents, while making it likely that confidence interval half-widths of ± 5could be met for questions that are relevant to only a small portion of respondents. Therefore, OPA tightened the precision constraints accordingly. The overall sample for DoD was agreed to be approximately 75 percent of all women and 50 percent of all males. All Coast Guard members were selected for the survey. During the development of the sample design, another survey, the National Intimate Partner and Sexual Violence Survey (NISVS) was to be fielded at a similar time. OPA ensured that no overlap would occur with DoD active duty members and therefore set a limit of sampling a maximum of 85% from any stratum to leave necessary sample for the NISVS.

The *2016 WGRA* total sample size was 735,329 (696,329 DoD and 39,000 Coast Guard); Table 2 provides the sample sizes by stratification variables.

Table 2.
Sample Size by Stratification Variables

Stratification Variable	Total	Army	Navy	USMC	Air Force	Coast Guard
Sample	735,329	282,584	173,326	108,936	131,483	39,000
Gender						
Male	576,436	228,527	126,255	97,216	91,199	33,239
Female	158,893	54,057	47,071	11,720	40,284	5,761
Paygrade Grouping						
E1-E4	407,334	166,457	90,565	79,082	58,891	12,339
E5-E9	232,688	78,444	62,362	22,462	50,894	18,526
W1-W5/O1-O3	65,496	27,509	13,837	5,386	13,337	5,427
O4-O6	29,811	10,174	6,562	2,006	8,361	2,708
Race						
Non-Minority	437,455	158,015	88,702	70,186	91,973	28,579
Minority	297,874	124,569	84,624	38,750	39,510	10,421
Family Status						
Single w/ Children	33,877	16,198	7,708	2,063	6,296	1,612
Dual Service Spouse	48,194	15,620	10,382	4,138	16,022	2,032
All Others	653,258	250,766	155,236	102,735	109,165	35,356

Weighting

Analytical weights for the *2016 WGRA* were created to account for unequal probabilities of selection and varying response rates among population subgroups. Sampling weights were computed as the inverse of the selection probabilities. The sampling weights were then adjusted for nonresponse using models that considered over 50 possible correlates of nonresponse. The adjusted weights were raked to match population totals and to reduce bias unaccounted for by the previous weighting steps. More details about the weighting process can be found later in this document.

Case Dispositions

As the first step in the weighting process, case dispositions were assigned based on eligibility for the survey and on completion of the questionnaire. Execution of the weighting process and computation of response rates both depended on this classification.

Final case dispositions for weighting were determined using information from personnel records, field operations (as recorded in the Survey Control System [SCS]), and returned

questionnaires. No single source of information is entirely complete and correct for determining the case disposition; inconsistencies among sources were resolved according to the order of precedence shown in Table 3. This order of execution is critical to resolving case dispositions. For example, suppose an individual in the sample refused the survey, with the reason that it was too long; in the absence of any other information, the disposition would be "eligible nonrespondent." Another example would be if we were provided a proxy report that the sample member had been left the military; in this instance the disposition would be "ineligible."

Case disposition counts for the *2016 WGRA* are shown in Table 3. Table 4 presents the number of complete eligible respondents (SAMP_DC = 4) by stratification variables: Service, gender, paygrade grouping, race, and family status.

Table 3.
Case Dispositions for Weighting

Case Disposition (SAMP_DC)	Information Source	Conditions	Sample Size
1. Record ineligible	Personnel record	OPA determined whether sampled members had a record in the DEERS point-in-time extract (PITE) prior to fielding the survey. No record in DEERS indicated the member either separated from the military, passed away, etc.	9,247 (1.3%)
2. Ineligible by self- or proxy-report	Survey Control System (SCS)	The sampled member or a proxy reported that member was ineligible due to such reasons as "Retired," "Ill," "Incarcerated," "No longer employed by DoD," or "Deceased."	296 (0.04%)
3. Ineligible by survey self-report	Survey eligibility questions	The sampled member was determined to be ineligible based on their response to Q1 of the survey questionnaire "Were you on active duty on [OPEN DATE]?"	982 (0.1%)
4. Eligible, complete response	Item response rate	Respondents needed to answer one of the six critical questions related to sexual assault.	151,010 (20.5%)
5. Eligible, incomplete response	Item response rate	Survey is not blank but none of the critical sexual assault questions were answered.	5,603 (0.8%)
8. Active refusal	SCS	Survey is returned blank due to such reasons as "Refused-too long," "Refused-inappropriate/intrusive," "Refused-other," "Unreachable at this address," "Refused by current resident," "Refused additional e-mails," or "Concerned about security/confidentiality."	1,654 (0.2%)
9. Blank return	SCS	Blank questionnaire returned with no reason given.	929 (0.1%)
10. PND	SCS	Postal non-deliverable or original address is non-locatable.	170,382 (23.2%)
11. Nonrespondent	Remainder	Remaining sampled members did not respond to survey.	395,226 (53.7%)
Total			**735,329**

Table 4.
Complete Eligible Respondents by Stratification Variables

Stratification Variable	Total	Army	Navy	USMC	Air Force	Coast Guard
Sample	151,010	44,782	28,594	14,362	44,691	18,581
Gender						
Male	108,547	32,587	19,478	11,915	29,061	15,506
Female	42,463	12,195	9,116	2,447	15,630	3,075
Paygrade Grouping						
E1-E4	42,493	11,753	6,147	5,901	14,840	3,852
E5-E9	70,752	19,843	14,837	5,903	20,389	9,780
W1-W5/O1-O3	22,979	8,512	4,331	1,695	5,279	3,162
O4-O6	14,786	4,674	3,279	863	4,183	1,787
Race						
Non-Minority	95,235	25,378	15,432	9,060	31,283	14,082
Minority	55,775	19,404	13,162	5,302	13,408	4,499
Family Status						
Single w/ Children	8,325	3,102	1,801	470	2,218	734
Dual Service Spouse	13,460	3,315	2,203	840	5,985	1,117
All Others	129,225	38,365	24,590	13,052	36,488	16,730

Nonresponse Adjustments and Final Weights

After case dispositions were resolved, the sampling weights were adjusted for nonresponse. First, the sampling weights for cases of known eligibility (SAMP_DC = 2, 3, 4, or 5) were adjusted to account for cases of unknown eligibility (SAMP_DC = 8, 9, 10, or 11). Next, the eligibility-adjusted weights for eligible respondents with completed questionnaires (SAMP_DC = 4) were adjusted to account for eligible sample members who returned an incomplete questionnaire (SAMP_DC = 5). All weights for the record ineligibles (SAMP_DC=1) were set to 0 and this weight was transferred to the other cases during post-stratification.

The weighting adjustment factors for eligibility and completion were computed as the inverse of model-predicted probabilities. The *2016 WGRA* models paralleled those developed by RAND for 1) the *2014 RAND Military Workplace Study* (*2014 RMWS*) (Morral, Gore, & Schell, 2014, 2015), which surveyed both the active duty and Reserve members and 2) the *2015 Workplace and Gender Relations Survey of the Reserve Component* (*2015 WGRR*). As in the *2014 RMWS* and *2015 WGRR* surveys, RSSC modeled the following six outcome variables separately for females and males: sexual harassment, gender discrimination, sexual quid pro quo, attempted sexual assault, non-penetrative sexual assault, and penetrative sexual assault. Table 5 provides a list of the key outcome variables used in the gradient boosted decision tree models (GBM) models.

Table 5.
Key Outcome Variables

Variable	Variable Name	Question Type
Hostile Work Environment	HWE	Military Equal Opportunity
Gender Discrimination	SDISC	Military Equal Opportunity
Sexual quid pro quo	QPQ	Military Equal Opportunity
Attempted Sexual Assault	SA_A_ADJ	Sexual Assault
Non-penetrative sexual assault	SA_T_ADJ	Sexual Assault
Penetrative sexual assault	SA_P_ADJ	Sexual Assault

The *2016 WGRA* nonresponse adjustment involved two steps, each of which produced a set of models. The first step used data from the eligible, complete respondents to develop stage one models for the key outcome variables. The models were fitted separately by gender. Predicted values of the six outcomes from Table 5 were computed for both respondents and nonrespondents. Two second stage models (eligibility and completion) were fitted separately by gender to predict the probability of response, using the results from the stage one models along with a limited number of other predictors: Service, paygrade, race. In addition survey form type (paper vs. web) was used for the second stage completion model. The reciprocals of the predicted values from the second model were used as nonresponse adjustments and applied to the respondents. The GBM models were weighted; first by the sampling weight, and second by the eligibility-adjusted weight resulting from multiplying the sampling weight by the eligibility status adjustment. Then, the models were adjusted by multiplying the eligibility status weight by the completion status adjustment. Table 6 provides a list of the candidate auxiliary variables considered for the GBM models.

Table 6.
Variables Used to Model Key Outcome Variables

Variable	Variable Name	Categories
Military Accession Program	ACC_SRC_CD2	ACC_SRC_CD was recoded. Any accession code that had less than 50 respondents were put into the category '0'
Mailing Address Match Flag	ADDMATCH	0=Address is different; 1=Address is the same
Armed Forces Qualification Test score	AFQT_CAT_CD2	AFQT_CAT_CD was recoded; Groups with less than 100 respondents were combined into '4Z';
Member Age	AGE	17-71
Basic Allowance for Housing Indicator	BAHREC	N=Not receiving BAH, Y=receiving BAH, Z=Unknown, .=Missing
Number of People that are Female/Male at Base	BASEMALE_PCT	BASEMALE and BASESIZE were used to create percentage that were male

Variable	Variable Name	Categories
Base name of Member	BASENAME_CD	BASENAME was recoded; Any base with less than 50 complete eligible responses were combined into an "*** All Small Bases' group
Number of People at Base	BASESIZE_CD	BASESIZE was recoded into subgroups
Email address purchase flag	BUYEMAIL	0=Do not buy email address, 1=Buy email address
Total Number of Children	CHILDCNT	0-12; 99's were coded as missing
Duty Location in the World Regions	CREGION1	1='US & US territories, Other, Unknown', 2='Europe', 3='Asia & Pacific Islands'
Service of Member	CSERVICE	1=Army, 2= Navy, 3= Marine Corps, 4= Air Force, 5= Coast Guard
Gender of Member	CSEX	1=Male, 2= Female
Current deployment status	CUR_DEPLOY	1=Yes; 0=No
Number of Deployments	DCOUNT	1-27
Deployment flag in the last 12 months	DEPLOY12	1=Yes; 0= No
Deployment flag in the last 24 months	DEPLOY24	1=Yes; 0= No
Dual Spouse Flag	DUAL_FLAG	Dual="Dual Spouse'; OTHR="Not a dual spouse'
Duty UIC Match Flag; Address is the Same	DUICMATCH	0=Duty UIC is different; 1=Duty UIC is the same
Education level	EDUC_CD	EDUC was recoded; Less than 100 respondents were put into similar education levels
E-mail at Time of Sampling	EMAIL	1=Have an e-mail ; 0= no email
Email address flag	EMAILSTAT_CD	EMAILSTAT was recoded: '1=No email or all attempted email addresses invalid, 2=At least one attempted email address not invalid
Ethnic affinity code	ETH_CD	ETH was recoded; Less than 100 respondents were put into other ethnicity group (OTH)
Family Status	FAMSTAT	0= Unknown marital status and/or child status, 1= Single with child(ren), 2= Single without child(ren), 3= Married with child(ren), 4=Married without child(ren)
Home Address Flag	FLG_H	N=No home address; Y=Home address
Retired or Separated from Service Flag	LEFTSERV	0=No; 1=Yes
Marital Status Code	MRTL_STA_CD	MRTL_STA was recoded; Less than 100 respondents were put into 'O'
Number of members in member's duty UIC	N_DUIC	1-6,084
Number of males in member's duty UIC	N_DUICMALE	1-4,562
Number of people within members' specific occupation code	N_OCC	1-85,772
Number of males in member's primary occupation	N_OCCMALE	1-85,772

Variable	Variable Name	Categories
On or Off Base Status	OFFBASE	0=Unknown, 1=On Base (No BAH), 2=Off Base (receiving BAH)
Percent of males in member's duty UIC	P_DUICMALE	0-100%
Percent male within members' specific occupation	P_OCCMALE	0-100%
Paygrade of Member (20 level)	PAYGRADE	E1-E9, W1-W5, O1-O6
Occupation Grouping	PDODOCC_CD	PDODOCC was recoded; There were 298 levels and this was formed by taking the first 2 characters
Race/Ethnic Category	RACE_ETH	A=AIAN, B=Asian, C=Black, D=White, E=Hispanic, F=NHPI, M=Multi Race, Z=Unknown
Strength Accounting Codes	STR_ACCT_CD2	STR_ACCT_CD was recoded; the A20's were put with the A24
Active Federal Military Service Base Calendar Date	TAFMS_DT2	TAFMS_DT2 was recoded: Took the year and month
Years of service	TAFMS_YR_QY	1-42; 99's were coded to missing
US Citizen Citizenship Origin Code	US_CITZ_ORIG_CD	A='Born within the US, GU, PR or VI', B='US citizen, parent became a citizen by naturalization', C='Born outside US,GU,PR or VI to at least one citizen parent', D='US citizen by naturalization', Y='Not a US citizen', Z='Origin not determined'
US Citizenship Status Code	US_CITZ_STAT_CD	A=US national, C=US citizen, N=Non US citizen or national, Z=Unknown

To further detail the nonresponse adjustments used in the *2016 WGRA*, in Table 3, SAMP_DC (case disposition) 2, 3, 4, and 5 denote cases with known eligibility, whereas SAMP_DC 8, 9, 10, and 11 correspond to cases for which eligibility is unknown. Consequently, the first of the two nonresponse adjustments increased the weights for case dispositions 2, 3, 4, and 5 to represent dispositions 8, 9, 10, and 11. The second adjustment increased the weights of complete cases with disposition 4 to compensate for incomplete eligible cases with SAMP_DC = 5.

To increase response to the *2016 WGRA*, nonrespondents to the web version of the survey were sent a paper form of the questionnaire. The paper version included the key survey items, but it omitted many secondary items on the web questionnaire, presenting the recipient with approximately 100 questions instead of the approximately 225 on the web version. The primary set of weights was based on responses from the full data set including both the web and paper versions. To support analysis of items only on the web version, a second set of weights was produced, following the same steps as the full data set excluding the paper questionnaire. For this weighting, all paper questionnaire respondents were treated as nonrespondents, including in the fitting of the GBM models. This second set of weights is intended solely for analysis of web-only items. The primary set of weights provides the basis for estimating the key outcomes from the survey items collected on both the web and paper versions of the questionnaire.

Finally, the nonresponse-adjusted weights were modified through a process called raking. The purpose of raking is to use known information about the survey population to increase the precision of survey estimates. This information consists of totals for different levels of variables (such as demographic characteristics). For example, the variable CSEX has two levels: male and female. During the raking process, sampled individuals are first categorized into the cells of a table defined by two or more variables—called raking dimensions. The goal of raking is to adjust the weights so that they add up to the known totals—called control totals—for the different levels within each raking dimension. Preceding one dimension at a time, raking computes a proportional adjustment to the weights associated with each level of the raking dimension. After all dimensions are adjusted, the process is repeated until the totals for all levels of the raking dimensions are equal to the corresponding control totals (at least within a specified tolerance).

Control totals were computed from information from the sampling frame. There were four raking dimensions, defined below and shown in Table 7:

- DoD (2 level) crossed with paygroup (7 level),
- DoD (2 level) crossed with race (2 level),
- DoD (2 level) crossed with gender (2 level) and paygroup (5 level) and
- Service (5 level) crossed with gender (2 level), and enlisted/officer status (2 level).

Table 7.
Variables Used for Raking

Variable	Variable Name	Categories	
DoD x paygroup (CDOD x CPAYGRP7)	DODPAY7	1. DoD * E1-E3	8. CG * E1-E3
		2. DoD * E4	9. CG * E4
		3. DoD * E5-E6	10. CG * E5-E6
		4. DoD * E7-E9	11. CG * E7-E9
		5. DoD * W1-W5	12. CG * W1-W5
		6. DoD * O1-O3	13. CG * O1-O3
		7. DoD * O4-O6	14. CG * O4-O6
DoD x race (CDOD x CRACECAT)	DODRACE	1. DoD * Non-minority	3. CG * Non-minority
		2. DoD * Minority	4. CG * Minority
DoD x Gender x Pay (CDOD x GENDER x CPAYGRP5)	DODGENPAY	1. DOD * Male * E1-E4	11. CG * Male * E1-E4
		2. DOD * Male * E5-E9	12. CG * Male * E5-E9
		3. DOD * Male * W1-W5	13. CG * Male * W1-W5
		4. DOD * Male * O1-O3	14. CG * Male * O1-O3
		5. DOD * Male * O4-O6	15. CG * Male * O4-O6
		6. DOD * Female * E1-E4	16. CG * Female * E1-E4
		7. DOD * Female * E5-E9	17. CG * Female * E5-E9
		8. DOD * Female * W1-W5	18. CG * Female * W1-W5
		9. DOD * Female * O1-O3	19. CG * Female * O1-O3
		10. DOD * Female * O4-O6	20. CG * Female * O4-O6
DoD x Gender x Service x Officer (CDOD x CSEX x CSERVICE X CPAYGRP6)	DODGENSVCOFF	1. DOD * Army * Male * Enlisted	11. DOD * Navy * Female * Enlisted
		2. DOD * Army * Male * Officer	12. DOD * Navy * Female * Officer
		3. DOD * Navy * Male * Enlisted	13. DOD * USMC * Female * Enlisted
		4. DOD * Navy * Male * Officer	14. DOD * USMC * Female * Officer
		5. DOD * USMC * Male * Enlisted	15. DOD * AF * Female * Enlisted
		6. DOD * USMC * Male * Officer	16. DOD * AF * Female * Officer
		7. DOD * AF * Male * Enlisted	17. CG * Male * Enlisted
		8. DOD * AF * Male * Officer	18. CG * Male * Officer
		9. DOD * Army * Female * Enlisted	19. CG * Female * Enlisted
		10. DOD * Army * Female * Officer	20. CG * Female * Officer

Table 8 summarizes the distributions of the sampling weights, intermediate weights, final weights, and corresponding adjustment factors by eligibility status for the primary weights. Eligible respondents are those individuals who were not only eligible to participate in the survey but also completed at least one of the critical sexual assault questions. Record ineligible individuals are those who were not eligible to participate in the survey according to administrative records; no weights were computed for these cases.

The mean sampling weight is 2.0 for the complete eligibles. The nonresponse adjustment for eligibility status that follows next makes the biggest single adjustment to the weights, in terms of increasing both the mean and the coefficient of variation (C.V.) of the weights. The two remaining adjustments for nonresponse among the eligible population and the final raking have a modest effect on increasing the mean weight. The corresponding factors shown in the last two columns of Table 8 have small C.V.'s; in other words, the factors in each column differ from each other by relatively small amounts.

Table 8.
Distribution of Weights and Adjustment Factors

Statistic	Sampling Weight	Eligibility Status Adjusted Weight	Complete Eligible Response Adjusted Weight	Final Weight	Eligibility Status Factor	Complete Eligible Response Factor	Raking Factor
N	151,010	151,010	151,010	151,010	151,010	151,010	151,010
MIN	1.00	1.2	1.2	1.1	1.2	1.0	0.9
MAX	3.8	129.2	130.0	140.9	88.3	1.6	1.2
MEAN	2.0	8.1	8.4	8.6	4.4	1.0	1.0
STD	0.8	5.9	6.2	6.6	4.2	0.02	0.04
C.V.	0.4	0.73	0.74	0.77	0.95	0.02	0.04

Under simplifying assumptions, Kish (Kish, 1965) approximates the relative increase in variance due to weight variation as 1 plus the C.V. squared ($1+(C.V.)^2$). Because the C.V. of the weights is less than 1 (0.77), the increase in variance due to weighting is less than 2 (1.59). Given the task of the weighting adjustments is to compensate for differential nonresponse and its possible impact on the bias of key outcome variables, the increase in variance due to weighting appears reasonable.

Table 9 shows the sum of the weights at different stages of weighting. The weights adjusted for known eligibility status distribute the sampling weights for nonrespondents with unknown eligibility status among the remaining dispositions. The eligible response adjusted weights then compensate for eligible respondents providing incomplete surveys. By design, the final raking adjustments redistribute record ineligibles and other dispositions excluded from the final weights to match the total number in the original frame.

Table 9.
Sum of Weights by Eligibility Status

Eligibility Category	Sum of Sampling Weights	Sum of Eligibility Status Adjusted Weights	Sum of Complete Eligible Response Adjusted Weights	Sum of Final Weights
1. Eligible weighted	306,268	1,219,367	1,270,256	1,301,077
2. Ineligible weighted	2,424	28,489	28,489	29,280
3. Non-response unweighted	1,006,399	52,316	0	0
4. Record ineligible unweighted	15,266	15,266	15,266	0
Total	**1 ,330,357**	**1 ,315,438**	**1,314,012**	**1 ,330,357**

Comparison to the 2014 RAND RMWS Study and 2015 DMDC WGRR

RAND found that increasing the number of weighting variables and using GBM improved the *2014 RMWS* survey weights, therefore, for comparability purposes OPA decided to also use this approach for the *2015 WGRR*. The description of the *2016 WGRA* weighting was set in the context of the methodologies used for *2014 RMWS* and the *2015 WGRR* and was described in the preceding section. The comparison is further elaborated here.

The software used for the *2015 WGRR* was built on the approach used by RAND in the *2014 RMWS*. Both weighting methodologies used the statistical computing software R and specifically functions from the packages "gbm" (Ridgeway, 2009) and "TWANG" (Ridgeway, 2004). RAND researchers provided the specific R scripts they used for their final production runs of the *2014 RMWS* weighting. For the *2016 WGRA* improvements were made by using a newer state of the art package for gradient boosted decision trees, "xgboost" (Chen, 2016). In addition, OPA rewrote the necessary TWANG functions to leverage "xgboost" in both stages of weighting. Initial results on the test cases provided in the TWANG documentation show results at least as good, with faster runtimes in comparison to "gbm."

The weighting for the *2016 WGRA* and the *2015 WGRR* also differed in some respects from the *2014 RMWS*. The *2016 WGRA* and the *2015 WGRR* weighting incorporated the two nonresponse steps (eligibility and completion), necessitating use of weights throughout the analysis. Some of the modeling in the *2014 RMWS* had been unweighted.

Variance Estimation

Sampling error is the uncertainty associated with an estimate that is based on data gathered from a sample of the population rather than the full population. Note that sample-based estimates will vary depending on the particular sample selected from the population. Measures of the magnitude of sampling error, such as the variance and the standard error (the square root of the variance), reflect the variation in the estimates over all possible samples that could have been selected from the population using the same sampling methodology. Analysis of the *2016 WGRA* data required a variance estimation procedure that accounted for the weighting

procedures. The final step of the weighting process was to define strata for variance estimation by Taylor series linearization. For each strata/variance strata, OPA ensured that there were at least 25 complete eligible responses with non-zero final weights. The variance strata closely mirrored the original strata and collapsing only occurred in four strata.

Multiple Comparison Section

When statistically comparing groups (e.g., Army vs. Navy estimates of the effectiveness of the sexual assault training), a statistical hypothesis whether there are no differences (null hypothesis) versus there are differences (alternative hypothesis) is tested. OPA mainly uses independent two sample t-tests for its statistical tests. The conclusions are usually based on the p-value associated with the test-statistic. If the p-value is less than the critical value then the null hypothesis is rejected. Any time a null hypothesis is rejected (a conclusion that estimates are significantly different), it is possible this conclusion is incorrect. In reality, the null hypothesis may have been true, and the significant result may have been due to chance. A p-value of 0.05 means there is a five percent chance of finding a difference as large as the observed result if the null hypothesis were true.

In survey research there is interest in conducting more than one comparison, i.e. multiple comparisons. For example, 1) testing whether the percentage of sexual assaults among senior officers is the same as the percentage of sexual assaults across all other enlisted members, and 2) testing that the percentage of sexual harassments for junior officers is the same as the percentage of sexual harassments with all enlisted members and so on. When performing multiple independent comparisons on the same data the question becomes: "Does the interpretation of the p-value for a single statistical test hold for multiple comparisons?" If 200 independent statistical (significance) tests were conducted at the 0.05 significance level, and the null hypothesis is supported for all, 10 of the tests would be expected to be significant at the p-value < 0.05 level due to chance. These 10 tests would have incorrectly assumed to be statistically significant—known as false positives or false discoveries. Holding the significance level constant, the more tests that are conducted the greater the number of false discoveries.

This is known in statistical hypothesis testing as the multiple comparisons problem. Numerous techniques have been developed to reduce the false positives associated with conducting multiple statistical tests. It should be noted that there is no universally accepted approach for dealing with the problem of multiple comparisons.

The method that OPA uses to control for false discoveries is known as the False Discovery Rate correction (FDR) developed by Benjamini and Hochberg (1995). FDR is defined as the expected percentage of erroneous rejections among all rejections. The idea is to control the false discovery rate which is the proportion of "discoveries" (significant results) that are actually false positives. The approach can be summarized as follows:

- Determine the number of comparisons (tests) of interest, call it m;
- Determine the tolerable False Discovery Rate (FDR Rate), call it α;
- Calculate the p-value for each statistical test;

- Sort the individual p-values from smallest to largest and rank them, call the rank k.
- For each ranked p-value calculate the FDR-adjusted alpha (threshold) which is defined as (k * ∝)/m
- Determine the cutoff delineating statistically significant results from non-significant results in the sorted file as follows: Look for the maximum rank (k) such that the ordered p-value is less than the FDR-adjusted alpha (i.e., look for the maximum k after which the p-value becomes greater than the threshold), call this maximum k the cutoff. Any comparison (p-value) with rank less than the cutoff is considered statistically significant.

OPA computed the FDR thresholds (FDR adjusted alpha) separately for the two types of comparisons—current year and trends. For both types of tests, OPA implemented the FDR Multiple Comparison corrections to control the expected rate of false discoveries (Type I errors) at ∝ = 0.05. For the current year estimates from the *2016 WGRA*, OPA performed 130,739 separate statistical tests (e.g., sexual harassment rates for men versus women). Of the 130,739 current year statistical tests, 62,447 were statistically significant. In addition, OPA performed another 12,002 separate statistical tests to compare estimates from the *2016 WGRA* to the *2014 RMWS* (i.e., trends). For trends, 3,456 of the 12,002 statistical tests were significant. For the current year, the FDR threshold was .02388 and for trends the FDR threshold was .01440.

Contact, Cooperation, and Response Rates

Contact, cooperation, and response rates were calculated in accordance with the recommendations of the American Association for Public Opinion Research (AAPOR, 2016 Standard Definitions), which estimates the proportion of eligible respondents among cases of unknown eligibility (SAMP_DC = 10 and 11).

The *contact rate* uses the concepts of AAPOR standard formula CON2 and is defined as

$$CON2 = \frac{(I+P)+R+O-e(O)}{(I+P)+R+O+NC-e(NC+O)} = \frac{\text{adjusted contacted sample}}{\text{adjusted eligible sample}} = \frac{N_C}{N_E}.$$

The *cooperation rate* uses the concepts of AAPOR standard formula COOP2 and is defined as

$$COOP2 = \frac{(I+P)}{(I+P)+R+O-e(O)} = \frac{\text{complete eligibles}}{\text{adjusted contacted sample}} = \frac{N_R}{N_C}.$$

The *response rate* uses the concepts of AAPOR standard formula RR4 and is defined as

$$RR4 = \frac{(I+P)}{(I+P)+R+O+NC-e(NC+O)} = \frac{\text{complete eligibles}}{\text{adjusted eligible sample}} = \frac{N_R}{N_E}.$$

Where:

I = Fully complete responses according to RR4 are greater than 80% complete (SAMP_DC=4)

P = Partially complete responses according to RR4 are between 50 – 80% complete (SAMP_DC=4)

R = Refusal and break-off according to RR4 are less than < 50% complete (SAMP_DC=5, 8, and 9)[3]

NC = Non-contact (SAMP_DC =10)

O = Other (SAMP_DC = 11)[4]

e(O) = Estimated ineligible nonrespondents

e(NC) = Estimated ineligible PND

N_C = Adjusted contacted sample

N_E = Adjusted eligible sample

N_R = Complete eligibles[5]

Table 10 shows the corresponding sample disposition codes associated with the response categories.

Table 10.
Disposition Codes for Response Rates

Response Category	SAMP_DC Values
Eligible Sample	4, 5, 8, 9, 10, 11
Contacted Sample	4, 5, 8, 9, 11
Complete Eligibles	4
Not Returned	11
Eligibility Determined	2, 3, 4, 5, 8, 9
Self Report Ineligible	2, 3

[3] OPA considers these all cases of known eligibility
[4] These are all nonrespondents which OPA considers cases of unknown eligibility
[5] Complete eligibles is an OPA term that applies to self-administered surveys in comparison to the terms complete and partial interviews used by AAPOR

Ineligibility Rate

The ineligibility rate (IR) is defined as the following and needs to be calculated for both weighted and unweighted to be applied to Table 10:

IR = Self Report Ineligible/Eligibility Determined.

Estimated Ineligible Postal Non-Deliverable/Not Contacted Rate

The estimated ineligible postal non-deliverable or not contacted (IPNDR) is defined as:

IPNDR = (Eligible Sample - Contacted Sample) * IR.

Estimated Ineligible Nonresponse

The estimated ineligible nonresponse (EINR) is defined as:

EINR = (Not Returned) * IR.

Adjusted Contact Rate

The adjusted contact rate (ACR) is defined as:

ACR = (Contacted Sample - EINR)/(Eligible Sample - IPNDR - EINR).

Adjusted Cooperation Rate

The adjusted cooperation rate (ACR) is defined as:

ACR = (Complete Eligible)/(Contacted Sample - EINR).

Adjusted Response Rate

The adjusted response rate (ARR) is defined as:

ARR = (Complete Eligible)/(Eligible Sample - IPNDR - EINR).

The final response rate is the product of the location rate and the completion rate. Table 11 shows both weighted and unweighted location, completion, and response rates for the *2016 WGRA*.

Finally, Table 12 shows weighted contact, completion, and response rates for the full sample by the stratification variables. The final weighted response rate for the survey was 23.5 percent.

Table 11.
Contacted, Cooperation, and Response Rates

Type of Rate	Computation	Unweighted	Weighted
Contacted	Adjusted contacted sample/Adjusted eligible sample	76.5%	79.9%
Cooperation	Usable responses/Adjusted contacted sample	27.4%	29.4%
Response	Usable responses/Adjusted eligible sample	21.0%	23.5%

Note: Weighted response rates are the official reported rates. Unweighted response rates can be influenced by the sample design.

Table 12.
Rates for Full Sample and Stratification Categories

Domain Variable	Domain	Contact Rate	Completion Rate	Response Rate
Sample	All	80%	29%	23%
Service	Army	78%	25%	19%
	Navy	77%	25%	19%
	Marine Corps	72%	23%	16%
	Air Force	89%	39%	35%
	Coast Guard	94%	52%	48%
Gender	Male	79%	28%	23%
	Female	82%	35%	28%
Paygroup	E1-E4	65%	17%	11%
	E5-E9	90%	33%	30%
	W1-W5	91%	37%	34%
	O4-O6	97%	50%	49%
Race	Non-minority	82%	30%	25%
	Minority	77%	27%	21%
Family Status	Single With Children	86%	30%	26%
	Dual Service Spouse	89%	33%	29%
	All Others	79%	29%	23%

Note: Reported rates are weighted. Unweighted rates can be influenced by the sample design.

Nonresponse Bias Analysis

Survey nonresponse has the potential to introduce bias in the survey estimates. To the extent that nonrespondents and respondents differ on observable characteristics (e.g., Service, paygrade, etc.), OPA uses weights to adjust the sample so the weighted respondents match the full population on key observable characteristics. This eliminates the portion of nonresponse bias (NRB) associated with those characteristics. When all NRB can be eliminated in this manner, the missingness is called *ignorable* or *missing at random* (Little & Rubin, 2002). Conditioning the weights on a very high number of observable demographics, like RSSC uses for military surveys, increases the likelihood that weighting effectively reduces NRB. OPA's complete assessment of NRB and the corresponding report were not ready at the time this report was finalized; however, the limited analysis conducted thus far that compares survey estimates of

reported sexual assaults to actual 'true' reports retained by DoD's Sexual Assault Prevention Office (SAPRO) showed no signs of NRB. OPA is in the process of evaluating NRB using the following four studies: 1) comparing the composition of the sample compared with survey respondents by key demographics, 2) comparing weighted survey estimates of sexual assaults to actual reports, 3) comparing estimates from the NRB follow-up survey (*2016 WGRA-N*[6]) to the *2016 WGRA*, and 4) evaluating the sensitivity of different post-survey adjustments (weighting methods) on survey estimates.

[6] After the production survey closed, OPA sampled a subset of about 2016 WGRA nonrespondents and conducted a short survey to assess NRB, as well as learn why members didn't complete the *2016 WGRA*

References

American Association for Public Opinion Research. (2016). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (9th Ed.). AAPOR. Retrieved from http://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*. 289–300. Retrieved from http://www.jstor.org/stable/2346101

Chen, T. (2016). xgboost: Extreme Gradient Boosting (Version 0.6-4) [Computer software]. Retrieved from http://lib.stat.cmu.edu/R/CRAN/

Chromy, J.R. (1987). Design optimization with multiple objectives. In *1987 proceedings of the Section on Survey Research Methods : papers presented at the Annual Meeting of the American Statistical Association, San Francisco, California, August 17–20, 1987* (pp. 194–199). Alexandria, VA: American Statistical Association. Retrieved from http://www.amstat.org/sections/srms/Proceedings/papers/1987_029.pdf

Dever, J.A., & Mason, R.E. (2003). DMDC sample planning tool (Version 2.1) [Computer program and software]. Arlington, VA: DMDC.

Kish, L. (1965). *Survey Sampling* (pp. 424–433). New York: John Wiley & Sons, Inc. doi:10.1002/sim.1513

Little, R.J., & Vartivarian, S. (2005). Does weighting for nonresponse increase the variance of survey means? Survey Methodology 31(2):161-168.

Mason, R.E., Wheeless, S.C., George, B.J., Dever, J.A., Riemer, R.A., & Elig, T.W. (1995). Sample allocation for the status of the armed forces surveys. In Proceedings of the Section on Survey Research Methods (Vol. II, pp. 769–774). Alexandria, VA: American Statistical Association. Retrieved from http://www.amstat.org/sections/srms/Proceedings/papers/1995_133.pdf

Morral, A.R., Gore, K.L., & Schell, T.L. (Eds.). (2014). *Sexual assault and sexual harassment in the U.S. military: Volume 1. Design of the 2014 RAND military workplace study* (No. RR-870/1-OSD). Santa Monica, CA: RAND Corporation.

Morral, A.R., Gore, K.L., & Schell, T.L. (Eds.). (2015). *Sexual assault and sexual harassment in the U.S. military: Volume 2. Estimates for Department of Defense service members from the 2014 RAND military workplace study* (No. RR-870/2-OSD). Santa Monica, CA: RAND Corporation.

Morral, A.R., Gore, K.L., & Schell, T.L. (Eds.). (2015). *Sexual assault and sexual harassment in the U.S. military: Volume 4. Investigations of potential bias in estimates from the 2014 RAND military workplace study* (No. RR-870/2-OSD). Santa Monica, CA: RAND Corporation.

Ridgeway, G. (2009). TWANG, Toolkit for weighting and analysis of nonequivalent groups (Version 1.4-9.3) [Computer software]. Retrieved from http://lib.stat.cmu.edu/R/CRAN/

This page is reserved for insertion of Standard Form 298, page 1 -- this is best accomplished by replacing this page after the document has been converted to PDF

Office of People Analytics (OPA)