OPA

OFFICE OF PEOPLE ANALYTICS

2019 Workplace and Gender Relations Survey of Reserve Component Members

Statistical Methodology Report

DATA DRIVEN SOLUTIONS FOR DECISION MAKERS

OPA

OFFICE OF PEOPLE ANALYTICS

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/dtic/order.html
Ask for report by DTIC# ADA1113657

OPA Report No. 2020-053
May 2020

2019 Workplace and Gender Relations Survey of Reserve Component Members

Statistical Methodology Report

DATA DRIVEN SOLUTIONS FOR DECISION MAKERS

OPA
OFFICE OF PEOPLE ANALYTICS

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:
https://discover.dtic.mil/products-services/
Ask for report by DTIC#

Acknowledgments

The Office of People Analytics (OPA) is grateful to numerous people for their assistance with the *2019 Workplace and Gender Relations Survey of Reserve Component Members* (*2019 WGRR*), which was conducted on behalf of Dr. Nathan Galbreath, Acting Director of the Department of Defense (DoD) Sexual Assault Prevention and Response Office (SAPRO). This survey was conducted under the leadership of Dr. Ashlea Klahr, Director of OPA's Health & Resilience (H&R) Research Division, and Ms. Lisa Davis, Deputy Director of OPA's H&R Research Division.

Individuals who contributed to the development of this survey include Dr. Aubrey Hilbert, Mr. Zachary Gitlin, Sarah Newman, Dr. Allison Greene-Sands, (DoD SAPRO), Ms. Shirley Raguindin (Office for Diversity, Equity, and Inclusion), Dr. Samantha Daniel, and Mr. Michael DiNicolantonio (OPA).

OPA's Statistical Methods Team, 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. Alex McMillan and Mr. Stephen Busselberg (Fors Marsh Group, LLC) implemented the weighting methods. Ms. Susan Reinhold provided the data processing support. Data Recognition Corporation (DRC) performed data collection and editing.

Table of Contents

Appendices

List of Tables

Table of Contents (Continued)

Page

2019 WORKPLACE AND GENDER RELATIONS SURVEY OF RESERVE COMPONENT MEMBERS STATISTICAL METHODOLOGY REPORT

Introduction

The Office of People Analytics' Center for Health and Resilience (OPA[H&R]) 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]). 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 on sensitive topics, which impact the health and well-being of military populations.

This report describes the statistical methodologies for the *2019 Workplace and Gender Relations Survey of Reserve Component Members* (*2019 WGRR*). The survey fielded from August 14, 2019 through November 12, 2019. Section 1 describes the sample design and selection of the sample. Section 2 describes the design of the survey communication and survey name experiments. Section 3 describes the weighting and variance estimation. Section 4 describes the statistical tests used for the *2019 WGRR.* Section 5 describes the calculation of contact, cooperation, and response rates for the full sample and population subgroups. Section 6 provides the results of the experiments. Section 7 is a nonresponse bias analysis. Survey estimates for select questions are found in the *2019 Workplace and Gender Relations Survey of Reserve Component Members: Results and Trends* (OPA, 2020a). Information about administration of the survey and detailed documentation of the survey dataset can be found in the *2019 Workplace and Gender Relations Survey of Reserve Component Members: Administration, Datasets, and Codebook* (OPA, 2020b).

Section 1: Sample Design and Selection

Target Population

The *2019 WGRR* was designed to represent individuals meeting the following criteria:

- Reserve component members from the Selected Reserve in Reserve Unit, Active Guard/Reserve (AGR/FTS/AR; Title 10 and Title 32), or Individual Mobilization Augmentee (IMA) programs from:
    - Army National Guard (ARNG),
    - U.S. Army Reserve (USAR),
    - U.S. Navy Reserve (USNR),
    - U.S. Marine Corps Reserve (USMCR),
    - Air National Guard (ANG), or

- U.S. Air Force Reserve (USAFR);

- Paygrades E1-O6

Sampling Frame

The sampling frame consisted of 793,216 Reserve component members who were not General/Flag officers or Coast Guard Reserve, using the March 2019 Reserve Components Common Personnel Data System (RCCPDS) Master File. Auxiliary frame data was obtained from the following files:

- March 2019 Reserve Duty Family File (contains the member's family information [e.g. marital status and children])

- March 2019 DoD Appropriated Fund Civilian Personnel Master File (identifies Military Technicians)

Sample Design

The sample for the *2019 WGRR* survey used a single-stage stratified design. Table 1 shows the four variables and associated variable levels that were used for stratification.

Table 1.
Variables for Stratification

Variable Description	Variable Name	Variable Levels
Reserve Component	RORG_CD	1. Army National Guard
		2. U.S. Army Reserve
		3. U.S. Navy Reserve
		4. U.S. Marine Corps Reserve
		5. Air National Guard
		6. U.S. Air Force Reserve
Gender	RSEX2	1. Male
		2. Female
Paygrade Grouping	RPAYGRP5	1. E1-E4
		2. E5-E9
		3. W1-W5/O1-O3
		4. O4-O6
Reserve Program	RPROGCIV	1. TPU
		2. AGR/TAR
		3. MilTech
		4. IMA

OPA partitioned the population frame into 123 strata that were initially determined by a full cross-classification of the aforementioned four stratification variables. Levels (specific levels from Table 1 such as "IMA") were collapsed when there were less than 200 in the stratum (e.g., collapsing "IMA" with "MilTech" to form a new stratification level). Reserve component and gender 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 probabilities varied among strata and individuals were not selected with equal probability overall. To achieve adequate sample sizes for all domains (reporting levels), OPA used a non-proportional allocation.

Sample Allocation

Unlike most OPA surveys where the sample size and design are determined by meeting precision requirements for required estimation domains (e.g., Army male), OPA decided to conduct a census of the Reserve forces in 2019 to reduce the survey burden on any individual member by ensuring that each Reserve component member received only one 2019 OPA survey. Therefore, OPA assigned all Reserve component members to either the *2019 Workplace Equal Opportunity (2019 WEOR), Workplace and Gender Relations Survey (2019 WGRR),* or *Status of Forces Survey (2019 SOFS-R).* OPA attempted to keep sample designs as similar as possible to prior administrations of these surveys, but this could not be completely achieved because each sample design was also influenced by requirements from the other two surveys. For instance, the WGRR surveys typically select almost all women within small Reserve components (e.g., Marine Corps Reserve), but for this administration many of these members needed to also be available for the other two Reserve surveys.

OPA designed the *2019 WGRR* sample to attempt to achieve estimates of percentages with associated precisions of less than 5% for 85 estimation domains (see Appendix A), but was unable to meet precision requirements for many of these domains for the reasons stated earlier. Note that the changes in sample designs for all three surveys do not affect the estimates[1] derived from these surveys, and comparisons with prior and future administrations are valid. 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 *2018 Status of Forces Survey of Reserve Component Members* (*2018 SOFS-R*), the *2017 Workplace and Gender Relations Survey of Reserve Component Members* (*2017 WGRR*), and the *2015 Workplace and Equal Opportunity Survey of Reserve Component Members (2015 WEOR)* were used to estimate these response 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

[1] While the expected value of any statistic (e.g., Navy female percent satisfied with job) is unaffected by the sample design, the margin of error of that statistic (e.g., +/- 3 percent) is greatly affected. Across the three Reserve surveys, many MOEs are smaller and many are larger than prior administrations because of the 2019 census of the Reserve forces.

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. OPA used a prevalence rate of 50% since it is most conservative and yields the largest estimated sample size.

The *2019 WGRR* total sample size was 269,475. Table 2 shows the sample sizes by stratification and experiment variables.

Table 2.
Sample Size by Stratification and Experiment Variables

Variable	Total	Army National Guard	US Army Reserve	US Navy Reserve	US Marine Corps Reserve	Air National Guard	US Air Force Reserve
Sample	269,475	114,579	63,746	17,995	13,160	34,602	25,393
Gender							
Male	167,106	77,374	34,318	10,600	12,536	19,807	12,471
Female	102,369	37,205	29,428	7,395	624	14,795	12,922
Paygrade Grouping							
E1-E4	134,810	69,478	31,479	4,625	9,250	11,778	8,200
E5-E9	92,295	31,636	18,620	9,805	2,178	17,834	12,222
W1-W5/O1-O3	24,412	10,363	8,311	1,319	759	2,096	1,564
O4-O6	17,958	3,102	5,336	2,246	973	2,894	3,407
Reserve Program							
TPU	229,150	101,524	56,909	14,861	11,683	24,465	19,708
AGR	18,803	6,520	3,974	3,095	556	3,810	848
MilTech	16,957	6,535	1,934	0	0	6,327	2,161
IMA	4,565	0	929	39	921	0	2,676
Survey Communication Experiment							
Postal, Email, and Paper Survey	219,572	93,361	51,941	14,662	10,723	28,195	20,690
Email Only	24,951	10,609	5,902	1,667	1,218	3,204	2,351
Survey Name Experiment							
Unnamed Survey	24,951	10,609	5,902	1,667	1,218	3,204	2,351
Project-specific Survey Name	24,952	10,609	5,903	1,666	1,219	3,203	2,352

After selecting the sample, OPA performed an additional check to verify the sample member was still eligible. OPA identified 3,307 (1.2% unweighted) sample members as record ineligible that became General/Flag officers or were no longer in the Selected Reserve in the April 2019 RCCPDS. Sample members who became ineligible during the field period were identified as self- or proxy-report ineligible. There were 735 (0.3%) sample members who were identified as being ineligible through either the survey instrument or other communications about the survey. OPA excluded ineligible sample members from further mailings and notifications (see Table 3).

Section 2: Design of Experiments

Prior OPA research has found evidence that survey communications and the name of the survey can have a substantial impact on both survey response rates and estimates. Because of these findings, OPA has continued to experiment with the way surveys are being communicated and publicized and implemented two embedded scientific experiments within the *2019 WGRR*. First, OPA designed a randomized experiment to determine the effect of changing the survey name from a project-specific (i.e. WGRR) to an unnamed survey that provides Reservists general information about possible survey questionnaire content. Second, OPA designed an experiment to determine the effect of multiple forms of communication (postal and email contact) versus a single form of contact (email only). For both experiments, OPA determines the effects on both survey response rates and survey estimates.

The goal of the experimental design was to maintain a control group that was as close as possible to prior survey administrations in order to measure the effect of changes. OPA randomly divided the sample into three treatment groups:

- Postal, Email, and Paper Survey: Unnamed Survey[2] – Received postal and email communications with no project-specific survey information (n=219,572)
- Email Only: Unnamed Survey – Received only email communications with no project-specific survey information (n=24,951)
- Email Only: Project-Specific Survey Name – Received only email notifications and used the WGRR survey name with discussion of project-specific survey topics (n=24,952)

This design allowed for two analyses. First, by comparing the Email Only: Unnamed Survey group to the Email Only: Project-Specific group, we can assess the impact of the survey name on response rates and key metrics. Second, by comparing the Postal, Email, and Paper Survey: Unnamed Survey group with the Email Only: Unnamed Survey group, we can assess the impact of including postal communications and a paper survey form.

[2] OPA emailed members of the 'unnamed survey' an advance letter where the first sentence said, 'You have been selected to participate in the Office of People Analytics' (OPA) only DoD-wide survey of the National Guard and Reserve.' This is contrasted with the 'project-specific' version of the communication which said, 'You have been selected to participate in the Office of People Analytics' (OPA) 2019 Workplace and Gender Relations Survey of Reserve Component Members.

Section 3: Weighting

OPA created analytical weights for the *2019 WGRR* 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 40 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 completion of the *2019 WGRR* critical items, defined as at least one of the six sexual assault items. Execution of the weighting process and computation of response rates both depend 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 dispositions; 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 a sample member refused the survey because it was "too long"; in the absence of any other information, the disposition would be "Active Refusal." However, if a family member of this same individual notified OPA that the sample member had left the military, the disposition of "Ineligible by self- or proxy-report" would override the later disposition, and OPA would code this individual as "ineligible" (SAMP_DC='2' in Table 3).

Case disposition counts for the *2019 WGRR* are shown in Table 3. There were 34,169 eligible, complete respondents (SAMP_DC = 4). Table 4 presents the number of eligible, complete respondents by stratification and experiment variables.

Table 3.
Case Dispositions for Weighting

Case Disposition (SAMP_DC)	Information Source	Conditions	Sample Size
1. Record ineligible	Personnel record	OPA used the following criteria to identify eligible members (all others are record ineligible): 1) Member had to be alive in the June 10, 2019 DBE (DEERS Database Extract) and 2) member had to be in the Selected Reserve and not a General/Flag officer in the April 2019 RCCPDS	3,307 (1.2%)
2. Ineligible by self- or proxy-report	Survey Control System (SCS)	Self or proxy reported that member was "retired," "no longer employed by DOD," or "deceased."	113 (0.04%)
3. Ineligible by survey self-report	Survey eligibility questions	The sampled member was determined to be ineligible based on their response to Question 1 of the survey: "Were you a member of a Reserve component on August 12, 2019?" Members who answered "No" were considered survey self-report ineligible."	622 (0.2%)
4. Eligible, complete response	Item response rate	Respondents needed to answer one of the six critical questions related to sexual assault.	34,169 (12.7%)
5. Eligible, incomplete response	Item response rate	Respondent answered some questions on the survey, but did not answer any of the critical sexual assault questions.	1,744 (0.7%)
8. Active refusal	SCS	Refused due to such reasons as "too long," "too intrusive," and "did not want additional communications," etc.	278 (0.1%)
9. Blank return	SCS	Blank questionnaire with no reason given.	617 (0.2%)
10. Postal Non-Deliverable (PND)	SCS	The final postal notification returned as postal non-deliverable. For 'email only' treatment group, OPA defines as PND sample cases with no email address.	24,315 (9.0%)
11. Nonrespondent	Remainder	Remaining sampled members who did not respond to survey.	204,310 (75.8%)
Total			**269,475 (100%)**

Table 4.
Complete Eligible Respondents by Stratification and Experiment Variables

Variable	Total	Army National Guard	US Army Reserve	US Navy Reserve	US Marine Corps Reserve	Air National Guard	US Air Force Reserve
Sample	34,169	10,728	8,081	2,725	1,002	7,363	4,270
Gender							
Male	19,654	6,920	4,167	1,482	927	4,116	2,042
Female	14,515	3,808	3,914	1,243	75	3,247	2,228
Paygrade Grouping							
E1-E4	7,309	2,795	1,527	250	467	1,467	803
E5-E9	15,891	4,652	3,022	1,373	209	4,450	2,185
W1-W5/O1-O3	5,113	2,158	1,687	347	112	506	303
O4-O6	5,856	1,123	1,845	755	214	940	979
Reserve Program							
TPU	23,560	7,277	6,103	2,294	760	4,196	2,930
AGR	5,003	1,903	1,201	417	82	1,181	219
MilTech	4,568	1,548	496	0	0	1,986	538
IMA	1,038	0	281	14	160	0	583
Survey Communication Experiment							
Postal, Email, and Paper Survey	29,281	9,336	6,959	2,369	853	6,132	3,632
Email Only	2,539	730	566	200	79	627	337
Survey Name Experiment							
Unnamed Survey	2,539	730	566	200	79	627	337
Project-Specific Survey Name	2,349	662	556	156	70	604	301

Nonresponse Adjustments and Final Weights

After case dispositions were resolved, OPA adjusted the sampling weights 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 complete 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) are set to 0.

The weighting adjustment factors for eligibility and completion were computed as the inverse of model-predicted probabilities. OPA used extreme gradient boosted (XGBoost[3]) decision trees to model the key outcomes separately for females and males.

[3] XGBoost is an R package function and stands for Extreme Gradient Boosting which is a machine-learning algorithm used to determine the best model fit.

Weighting the *2017* and *2015 WGRR* was similar, but OPA reduced the number of key outcome variables in 2017 due to the smaller Reserve sample size (241,426 in 2017 and 485,774 in 2015). This reduction was continued in 2019 as the sample size remained smaller than 2015 (269,475). Table 5 shows the key outcome variables used in the XGBoost models for the *2015*, *2017*, and *2019 WGRR* surveys.

Table 5.
Key Outcome Variables Modeled in Stage One

Variable	2015	2017	2019
Female			
Gender Discrimination	X	X	X
Sexual Harassment	X	X	X
Sexual Assault Rate	X	X	X
Quid Pro Quo	X		
Non-Penetrative Sexual Assault	X		
Penetrative Sexual Assault	X		
Male			
Gender Discrimination	X	X	X
Sexual Harassment	X	X	X
Sexual Assault Rate	X	X	X

The *2019 WGRR* 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. Predicted values of the three outcomes from Table 5 were computed for both respondents and nonrespondents.[4] Two second stage models (eligibility and completion) were fit 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. The reciprocals of the predicted values from both of the second-stage models were used as nonresponse adjustment factors and applied first to cases with known eligibility status (SAMP_DC in (2-5)) and then to complete, eligible respondents (SAMP_DC=4). OPA weighted the eligibility model by the sampling weight, and the completion model by the eligibility-adjusted weight resulting from multiplying the sampling weight by the eligibility status adjustment factor. The weight prior to calibration through raking was equivalent to the sampling weight times the reciprocal of the predicted probability of response (providing eligibility status) times the reciprocal of the predicated probability of survey completion. Table 6 provides a list of

[4] OPA fit separate models for males and females so there are 6 first-stage models to predict key outcomes from Table 5.

the auxiliary variables included in the XGBoost models. Variables denoted with an asterisk (*) were included in both the first stage and second stage adjustment.

Table 6.
Variables Used to Model Key Outcome Variables

Variable	Variable Name	Variable Notes	Categories
Military Accession Program	ACC_SRC_CD2		See Appendix B
Armed Forces Qualification Test score	AFQT_SCRR	Officers set to missing	0-99
Member Age at Field Open Date	CAGE	Ages 17-68	1=20 and under 2=21-24 3=25-30 4=31-35 5=36 and above
Assigned Unit Navy Ashore/Afloat Code	ASSGN_UIC_NV_ASHR_AFLT_CD		2=Sea Duty-CONUS Ships 4=Non-rotated Sea Duty-Ships Homeported Overseas; 9=Unknown or not applicable
Email address purchase flag	BUYEMAIL		0=Do not buy email address, 1=Buy email address
Total Number of Children	CHILDCNT		0-10;
Organization Component Code	COMP_CD		G=Guard; V=Reserve
Contacted	CONTACTED	3 are missing	0=Not Contacted 1=Contacted
Reserve Forces Initial Entry Date (RCCPDS)	DIERF_DT2	6,310are missing	Range from 5-21
Duty Service Occupation Code	DTY_DOD_OCC_CD		100000-290500
Education level	EDC_LVLR		11=High school diploma or less 32=Completed High School-- No Diploma 41=Some college, no degree 45=Associate degree/Professional nursing diploma 51=Baccalaureate degree 61=Master's degree 62=Post master's degree 63=First professional degree 64=Doctorate degree 65=Post doctorate degree 99=Unknown
Email at Time of Sampling	EMAIL_FLD		Y=Have an e-mail N= No email
Email address flag	EMAILFLG		0=No email address 1=At least one email address

Variable	Variable Name	Variable Notes	Categories
Email status	EMAILSTAT		1=No email or all attempted email addresses invalid 2=At least one attempted email address not invalid
Ethnic affinity code	ETHNICR	Recoded from ETHNIC combining small categories into Other	AB=Chinese AC=Filipino AG=Korean AJ=Other Asian descent AK=Mexican AL=Puerto Rican AN=Latin American with Hispanic descent AO=Other Hispanic descent AR=US or Canadian Indian tribes BG=Other BH=None ZZ=N/A or Unknown
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	HOMFLG		N=No home address; Y=Address available
Mailing address available at the end of fielding	MAIL_FLD		N=No Y=Yes
Marital Status Code	MARITALR	Recoded from MARITAL	D=Divorced M=Married N=Never married O=Other
Home Address of Marine Corps Member is Midway	MIDWAYFLG		0=Not a Midway Home Address 1=Midway Home Address
Number of members assigned in UIC	N_AUIC		1-3,288
Number of people within members' specific occupation code	N_OCC		1-43,922
Percent male within members' specific occupation	P_OCCMALE		0-100%
Occupation Grouping	PDODOCCR	PDODOCC was recoded; There were 297 levels and this was formed by taking the first 2 characters	10-29
Military Longevity Pay Service Base Calendar Date	PEBD_DT2		1973-2019
Postal Non-deliverable	POSTAL_ND		N=No Y=Yes
Prior Regular Component Service	PRIOR_ASVC_INDC_CD		N=No W=NA

Variable	Variable Name	Variable Notes	Categories
Indicator Code (RCCPDS)			Y=Yes Z=Unknown
Race/Ethnic Category	RACE_ETH		A=AIAN B=Asian C=Black D=White E=Hispanic F=NHPI M=Multi Race Z=Unknown
Ready Reserve Service Projected End Calendar Date	RDYV_SVC_PE_DT	43,902 are missing	21-32
Numeric Organizational Code	RORG_CD*		1=Army National Guard 2=Army Reserve 3=Navy Reserve 4=Marine Corps Reserve 5=Air National Guard 6=Air Force Reserve
Paygrade Grouping	RPAYGRP9*		1=E1-E4 2=E5-E9 3=W1-W5 4=O1-O3 5=O4-O6
Reserve Category Programs	RPROGCIV*		1=TPU/Unknown 2=AGR/TAR 3=Military Technicians 4=IMA
Numeric Service Code	RSERVICE		1=Army 2=Navy 3=Marine Corps 4=Air Force
Sex	RSEX2*		1=Male 2=Female
Reserve Category Group Code	RSV_CATG		1=Selected Reserve (not including AGR) 2=Active Guard/ Reserve (AGR)
Reserve Subcategory Code	RSV_SCAT		A=Drilling Unit Member B=Individual Mobilization Augmentees (IMA) F=On Initial Active Duty For Training (IADT) G=Active Guard Reserve P=Person awaiting IADT Q=Awaiting Second Part of IADT T=Simultaneous Membership Program (SMP) V=FT members performing AD on FTNGD for >180, but exempt from X=SEL RES - Other Training Programs
Reserve Category Code	RSVCAT		S=Selected Reserve – Trained in Units T=Selected Reserve – Trained Individuals (non-unit) U=Selected Reserve – Training Pipeline

Variable	Variable Name	Variable Notes	Categories
All communications undelivered	UNDELIVERED		N=No Y=Yes NA=Not Applicable
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
Occupation was Closed to Females	WASCLOSED		0=Was not closed 1=Was closed
Active Federal Military Service	YOSR	13 are missing	0-45
Gender Discrimination	SDISC*		Predicted Propensity
Sexual Harassment	SEXHAR*		Predicted Propensity
Sexual Assault Rate	SA_RATE*		Predicted Propensity

* Variable used in both first-stage and second-stage adjustments

To further detail the nonresponse adjustments used in the *2019 WGRR*, recall from Table 3 that 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. The eligibility adjustment increased the weights for case disposition 2, 3, 4, and 5 to represent case dispositions 8, 9, 10, and 11. The second adjustment increased the weights of complete eligible cases (SAMP_DC=4) to compensate for incomplete eligible cases (SAMP_DC=5).

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 mitigate potential nonresponse bias of survey estimates. This information consists of totals for different levels of variables (such as demographic characteristics). For example, the variable RSEX2 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. Processing 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 (within a specified tolerance). Control totals were computed using information from the sampling frame. Table 7 shows the nine raking dimensions and associated levels.

Table 7.
Variables and Levels (Raking Dimensions) Used for Raking

Variable	Variable Name	Variable Levels
Reserve Component	RORG_CD	1. Army National Guard
		2. Army Reserve
		3. Navy Reserve
		4. Marine Corps Reserve
		5. Air National Guard
		6. Air Force Reserve
Reserve Program	RPROGCIV	1. TPU/Unknown
		2. AGR/TAR
		3. Military Technicians
		4. IMA
Paygrade Grouping	RPAYGRP9	1. E1-E4
		2. E5-E9
		3. W1-W5
		4. O1-O3
		5. O4-O6
Race/Ethnicity	RETHC4	1. Non-minority/Unknown
		2. Minority
Gender	RSEX2	1. Male/Unknown
		2. Female
Gender by Paygrade	GENPAY	1. Male E1–E4
		2. Male E5–E9
		3. Male W1–W5
		4. Male O1–O3
		5. Male O4–O6
		6. Female E1–E4
		7. Female E5–E9
		8. Female W1–W5
		9 Female O1–O3
		10. Female O4–O6
Gender by Program	GENPROG	1. Male TPU/Unknown
		2. Male AGR/TAR
		3. Male Military Technicians
		4. Male IMA
		5. Female TPU/Unknown
		6. Female AGR/TAR
		7. Female Military Technicians
		8. Female IMA
Gender by Race	GENRACE	1. Male Non-minority
		2. Male Minority
		3. Female Non-minority
		4. Female Minority

Variable	Variable Name	Variable Levels
Gender by Service by Paygrade	GENORGPAY	1. Male ARNG Enlisted
		2. Male ARNG Officer
		3. Male USAR Enlisted
		4. Male USAR Officer
		5. Male USNR Enlisted
		6. Male USNR Officer
		7. Male USMCR Enlisted
		8. Male USMCR Officer
		9. Male ANG Enlisted
		10. Male ANG Officer
		11. Male USAFR Enlisted
		12. Male USAFR Officer
		13. Female ARNG Enlisted
		14. Female ARNG Officer
		15. Female USAR Enlisted
		16. Female USAR Officer
		17. Female USNR Enlisted
		18. Female USNR Officer
		19. Female USMCR Enlisted
		20. Female USMCR Officer
		21. Female ANG Enlisted
		22. Female ANG Officer
		23. Female USAFR Enlisted
		24. Female USAFR Officer

Table 8 summarizes the distributions of the sampling, eligibility, completion, and final weights, and the corresponding adjustment factors for the complete eligible respondents. As described earlier in the report, eligible respondents are those individuals who were eligible to participate in the survey and completed at least one of the critical sexual assault questions.

The mean sampling weight for the entire sample was 2.9 (data not shown) and the mean for the eligible respondents was 3.3. The nonresponse adjustment for eligibility status makes the largest adjustment to the weights (mean is 7.0), in terms of increasing both the mean and the coefficient of variation (CV) of the weights. The two remaining adjustments for nonresponse, completion and raking (mean is 1.1 and 1.0, respectively), have a modest effect on increasing the mean weight. The final weights, after raking, have the largest difference between the minimum and maximum values (weights range from 2.1 to 201.4)

Table 8.
Distribution of Weights and Adjustment Factors for Complete, Eligible Respondents

Eligibility Status	Statistic	Sampling Weight	Eligibility Status Adjustment	Eligibility Status Adjusted Weight	Complete Eligible Response Adjustment	Complete Eligible Response Adjusted Weight	Raking Adjustment	Final Weight With Non-response and Raking Adjustment
Eligible Respondents	N	34,169	34,169	34,169	34,169	34,169	34,169	34,169
	MIN	1.3	1.1	2.3	1.0	2.6	0.8	2.1
	MAX	8.8	40.5	156.3	1.3	160.5	1.5	201.4
	MEAN	3.3	7.0	20.8	1.1	21.7	1.0	22.6
	STD	1.8	6.2	18.2	0.0	18.7	0.2	21.0
	CV	0.54	0.89	0.87	0.02	0.86	0.15	0.93

Under simplifying assumptions, Kish (1965) approximates the relative increase in variance due to weight variation as 1 plus the coefficient of variation (CV) squared ($1+ [CV]^2$). Because the CV of the weights is less than 1 (0.93), the increase in variance due to weighting is less than 2 (1.86). 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 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 With Nonresponse and Raking Adjustments
1. Eligible respondent	111,163	711,560	742,565	772,945
2. Ineligible	2,378	19,582	19,582	20,271
3. Non-respondent	670,386	31,066	0	0
4. Record ineligible	9,290	9,290	9,290	0
Total	**793,216**	**771,499**	**771,437**	**793,216**

Note. Rows may not add up to total due to rounding.

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 *2019 WGRR* 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. The *2019 WGRR* variance estimation strata corresponded closely to the design strata; however, it was necessary to collapse some sampling strata containing fewer than 50 complete eligible responses with non-zero final weights with similar strata. There were 98 variance strata defined for the *2019 WGRR*.

Section 4: Multiple Comparisons

To support the WGRR reports and briefings, OPA conducts a large number of statistical tests to identify significant differences across demographic groups or compare estimates with prior years. 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. To protect against erroneous statistically significant results during the *2019 WGRR*, OPA used a p-value of 0.01 for its statistical tests. OPA chose this cut-off after empirically testing a statistical method called False Discovery Rate correction (FDR) developed by Benjamini and Hochberg (1995) in several prior OPA population-based surveys.

When comparing groups, a hypothesis whether there are no statistically significant differences (null hypothesis) versus there are statistically significant differences (alternative hypothesis) is tested. OPA mainly uses independent two sample t-tests and 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 0.01 means there is a one percent chance of finding a difference as large as the observed result if the null hypothesis were true.

Section 5: 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).[5]

NC = Non-contact (SAMP_DC =10)

O = Other (SAMP_DC = 11)[6]

e(O) = Estimated ineligible nonrespondents

e(NC) = Estimated ineligible PND

N_C = Adjusted contacted sample

N_E = Adjusted eligible sample

N_R = Complete eligibles[7]

[5] OPA considers these all cases of known eligibility.
[6] These are all nonrespondents which OPA considers cases of unknown eligibility.
[7] Complete eligible is an OPA term that applies to self-administered surveys, which relates to the terms complete and partial interviews used by AAPOR.

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

Ineligibility Rate

The ineligibility rate (IR) is defined as the following and needs to be calculated 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 contacted rate (ACR) is defined as:

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

Adjusted Cooperation Rate

The adjusted cooperation rate (ACOR) is defined as:

ACOR = (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 contact rate and the cooperation rate. Table 11 shows both weighted and unweighted contact, cooperation, and response rates for the *2019 WGRR*.

Finally, Table 12 shows weighted contact, cooperation, and response rates for the full sample by the stratification and experiment variables. The final weighted response rate for the survey was 14.5%.

Table 11.
Contacted, Cooperation, and Response Rates

Type of Rate	Computation	Unweighted (percent)	Weighted (percent)
Contacted	Adjusted contacted sample/Adjusted eligible sample	90.9	91.4
Cooperation	Usable responses/Adjusted contacted sample	14.4	15.8
Response	Usable responses/Adjusted eligible sample	13.1	14.5

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, Stratification Level, and Experiment Variables

Variables	Variable Levels	Contact Rate (percent)	Cooperation Rate (percent)	Weighted Response Rate (percent)
Sample	Sample	91	16	14
Component	Army National Guard	91	13	11
	Army Reserve	92	15	14
	Navy Reserve	89	19	17
	Marine Corp Reserve	83	10	8
	Air National Guard	96	24	23
	Air Force Reserve	94	19	18
Gender	Male	92	16	14
	Female	91	16	15
Paygrade Grouping	E1-E4	87	6	6
	E5-E9	94	19	18
	W1-W5/O1-O3	94	23	22
	O4-O6	97	35	34
Reserve Program	TPU	91	12	11
	AGR/TAR	93	30	28
	Military Technicians	96	29	28
	IMA	96	24	23
Survey Communication Experiment	Email, Postal, and Paper survey	90	17	15
	Email Only	96	12	12
Survey Name Experiment	Unnamed survey	96	12	12
	Project-Specific name	96	11	11

Note. Reported rates are weighted. Unweighted rates can be influenced by the sample design. This table was rounded for clarity.

Section 6: Results of Experiments

The survey communication and survey name experiments for *2019 WGRR* were first analyzed for their impact on response rates. The communication experiment compared response rates from the 'Postal, Email, and Paper Survey' treatment group with the 'Email-Only' group (to make fair comparisons this only uses data within the unnamed survey). Response rates for these groups were 14.9% and 11.6%, respectively, and as expected adding postal notifications and a paper survey form has a statistically significant positive effect on response rates, $\chi 2$ (df=1, n= 244,523) = 134.9, $p < 0.001$. Table 13 shows that the gain in response rate is fairly consistent across Reserve components, gender, and paygrade groupings, and most comparisons are statistically significant. The postal communications improved response rates the most for E1-E4 at 73% (2.5 percentage points). This increase for E1-E4 was much larger than either the *2019 WEOR* (16%) or the *2019 SOFS-R* (7%), likely because of the paper survey embedded within the WGRR postal notifications that were not present in the other two surveys. Because E1-E4 have the lowest response rates of any Reserve subgroup, the inclusion of paper surveys likely reduces nonresponse bias in OPA survey estimates.

Table 13.
Response Rates by Survey Communication Experiment

Variable	Postal, Email, and Paper Survey (Unnamed Survey)	Email Only (Unnamed Survey)	$\chi 2$[8]	p-value
Total	14.9	11.6	134.93	<.0001
Reserve Component				
Army National Guard	11.9	8.6	63.77	<.0001
Army Reserve	14.6	10.7	44.08	<.0001
Navy Reserve	17.7	12.6	19.55	<.0001
Marine Corps Reserve	8.2	6.6	3.68	0.055
Air National Guard	23.2	21.3	4.71	0.030
Air Force Reserve	18.3	15.0	11.04	0.001
Gender				
Male	14.8	11.4	98.23	<.0001
Female	14.9	12.0	54.37	<.0001
Paygrade Grouping				
E1-E4	6.0	3.5	111.84	<.0001
E5-E9	18.4	15.9	25.41	<.0001
W1-W5	31.1	21.6	10.80	0.001
O1-O3	21.0	15.2	28.23	<.0001
O4-O6	35.1	27.1	32.87	<.0001

Table 14 shows results from the survey name experiment, which compared response rates from the project-specific survey name (WGRR) to an unnamed survey[9] (note that this comparison is made only within the 'email only' part of the sample). Response rates for the unnamed survey were 11.6% compared with 10.8%, which was statistically significant, $\chi 2$ (df=1, n= 49,903) = 5.0, p = 0.025. The unnamed survey produced higher response rates for all Reserve components, both genders, and almost all paygrade groupings, although the effects are smaller than the communications experiment and most comparisons are not statistically significant. Response rate improvements ranged from about even for O1-O3 to 24% for Navy Reserve. OPA conducted the same survey name experiment in the *2019 WEOR* and *2019 SOFS-R* surveys, and a similar version in the 2016 and 2018 Post Election Voting (PEV) Surveys of active duty military. For *2019 WEOR*, results were similar to *2019 WGRR* where the unnamed survey (or generic survey title in PEV) [10] produced equal or higher response rates than OPA's traditional project-specific survey name. For *2019 SOFS-R*, Status of Forces is already a survey name that

[8] For Tables 13-16, the Wald Chi-square test was generated using the PROC SURVEYLOGISTIC with a weight statement within SAS 9.3 and SAS/STAT 12.1

[9] The unnamed survey did not mention a specific survey name on postal and email communications. It also discussed possible survey topics rather than the precise content of the survey.

[10] The Post-Election Voting Surveys (PEV) of active duty military in 2016 and 2018 also had a version of the survey name experiment where the 'voting name' was compared with a survey name of 'Quick Compass of Active Duty Military'.

is uninformative regarding survey topics, and in fact we do find that both survey name treatments produced similar response rates within *2019 SOFS-R.* The OPA statistical methods team recommends confirming this finding in future surveys, and if confirmed perhaps moving away from project-specific survey communications on OPA surveys.

Table 14.
Response Rates by Survey Name Experiment

Variable	Unnamed Survey (Email Only)	Project-Specific Survey Name (Email Only)	$\chi 2$	p-value
Total	11.6	10.8	5.05	0.025
Reserve Component				
Army National Guard	8.6	8.1	1.24	0.265
Army Reserve	10.7	10.2	0.52	0.469
Navy Reserve	12.6	10.2	3.56	0.059
Marine Corps Reserve	6.6	6.1	0.23	0.630
Air National Guard	21.3	20.5	0.41	0.523
Air Force Reserve	15.0	13.7	1.18	0.276
Gender				
Male	11.4	10.8	2.54	0.111
Female	12.0	10.8	6.46	0.011
Paygrade Grouping				
E1-E4	3.5	2.9	4.59	0.032
E5-E9	15.9	14.9	2.22	0.136
W1-W5	21.6	21.6	0.00	0.987
O1-O3	15.2	15.3	0.00	0.971
O4-O6	27.1	24.9	1.58	0.209

The second analysis was to determine whether respondents in different treatment groups reported experiencing different rates of sexual assault, sexual harassment, or sexual discrimination. The estimates may differ due to a change in measurement error or nonresponse error. For measurement error, a respondent to a 'Gender Relations' survey may have a heightened awareness of sexual harassment and assault, and this could raise their awareness and hence the likelihood of reporting an experience. While possible, OPA statisticians believe the more likely scenario is that the survey communications or survey name altered the composition of WGRR respondents, and therefore also potentially altered the magnitude and direction of nonresponse bias in the estimates.

Table 15 shows the estimates for males and females experiencing these key metrics, comparing respondents in the 'Postal, Email, and Paper Survey' treatment group with the 'Email-Only' group. The table shows that no key sexual assault/sexual harassment estimates were significantly different between communication methods. This fails to support a historical

OPA finding that sexual harassment and assault rates are generally slightly higher on paper surveys than web surveys.

Table 15.
Key Estimates by Gender by Survey Communication Experiment

Gender	Variable[11]	Postal, Email, and Paper Survey (Unnamed Survey)	Email Only (Unnamed Survey)	$\chi 2$	p-value
Males	Sexual Assault Rate	0.3%	0.2%	1.46	0.227
	Sexual Harassment	4.5%	3.6%	1.75	0.185
	Gender Discrimination	1.6%	1.8%	0.33	0.563
Females	Sexual Assault Rate	3.6%	3.8%	0.06	0.813
	Sexual Harassment	17.3%	18.6%	0.63	0.429
	Gender Discrimination	11.5%	11.5%	0.00	0.949

For the survey name experiment, Table 16 shows estimates for males and females experiencing these key metrics , comparing respondents in the project-specific survey name (WGRR) to an unnamed survey (recall that this comparison is within the 'email only' group. The table shows that no key sexual assault/sexual harassment estimates were significantly different between survey names. While no comparisons reached statistical significance, it is interesting to note that rates are higher for five of the six experiences from the project-specific survey name. This provides very weak support for the hypothesis that a survey name containing 'Gender Relations' may influence response from harassment/assault victims at slightly higher rates. However, no conclusions should be drawn here and OPA statisticians recommend repeating this experiment with larger sample sizes.

[11] Variable names are SA_RATE, SEXHAR, and SDISC

Table 16.
Key Estimates by Gender by Survey Name Experiment

Gender	Variable	Unnamed Survey (Email Only)	Project-Specific Survey Name (Email Only)	$\chi 2$	p-value
Males	Sexual Assault Rate	0.2%	0.4%	0.88	0.349
	Sexual Harassment	3.6%	4.5%	0.52	0.471
	Gender Discrimination	1.8%	1.9%	0.01	0.903
Females	Sexual Assault Rate	3.8%	4.1%	0.05	0.824
	Sexual Harassment	18.6%	17.2%	0.38	0.540
	Gender Discrimination	11.5%	11.8%	0.02	0.889

From these experiments, OPA concludes that the addition of postal communications and a paper survey significantly improves response rates over response rates obtained via an email only survey. These results match prior research and the identical experiments conducted in the 2019 SOFS-R and 2019 WEOR. The methods of survey communication shows no effect on sexual harassment, assault, or discrimination rates. The effect of the experimental manipulation of the survey name is smaller, but OPA concludes there is evidence that the communications with our historical survey names (project-specific) produce slightly lower response rates, or perhaps the inclusion of the phrase 'only DoD-wide survey of the National Guard and Reserve' assisted response in the unnamed communications. In addition, there may be very limited evidence that project-specific survey name produces slightly higher sexual harassment and assault rates, although this effect is small and no differences are statistically significant. This is an area for future survey research.

Section 7: Nonresponse Bias Analysis

Survey nonresponse has the potential to introduce bias in the estimates of key outcomes. To the extent that nonrespondents and respondents differ on observed characteristics, OPA can use weights to adjust the sample so the weighted respondents match the full population on the most critical characteristics. This eliminates the portion of nonresponse bias (NRB) associated with those observed variables if these variables are strongly associated with the behaviors being estimated. When all NRB can be eliminated in this manner, the missingness is called *ignorable* or *missing at random* (Little & Rubin, 2002). The more observable demographic variables that are incorporated into the weights, the more plausible it is to assume that the weights eliminate any NRB.

Nonresponse bias occurs when survey respondents are systematically different from nonrespondents. Statistically, the bias in a respondent mean (e.g., sexual assault rate) is a function of the response rate and the relationship (covariance) between response propensities and the estimated statistics (i.e., sexual assault rate), and takes the following form:

$$Bias\ (\bar{y}_r) = \frac{\sigma_{yp}}{\bar{p}} = \left(\frac{\rho_{yp}}{\bar{p}}\right)\sigma_y\sigma_p, \text{ where:}$$

$\bar{y}_r$ = estimated sexual assault rate

σ_{yp} = covariance between *y* and response propensity,

$\bar{p}$ = mean propensity over the sample,

ρ_{yp} = correlation between y and p,

σ_y = standard deviation of y,

σ_p = standard deviation of p.

NRB can occur with high or low survey response rates, but the decrease in overall survey response rates within the Department, as well as in civilian studies, in the past decade has resulted in a greater focus on potential NRB. OPA conducted an extensive NRB study on the *2015 WGRR*. When the essential survey conditions (i.e., survey mode, contacts, response rates [including subgroups]) remain mostly constant, the level and direction of NRB should remain similar. Therefore, OPA conducted an abbreviated NRB study on the *2017 WGRR in an attempt to confirm that the levels and direction of NRB was the same as 2015 WGRR* by comparing the sample composition with the survey respondents. This same analysis of the level and direction of NRB was conducted for the *2019 WGRR*. If these comparisons are the same across survey iterations, OPA asserts that the NRB is similar and the *2019 WGRR* requires no further assessments. That result is confirmed in the following section.

Studies of NRB can be accomplished either by 1) conducting a follow-up survey of nonrespondents or 2) by using the survey responses and characteristics of the respondents to assess NRB. The latter is the approach that was used in this report. Two survey outcomes are critical in assessing NRB: response rates and the expected difference between respondents and nonrespondents on survey estimates.

It is common that survey quality is judged by response rates; they are the most visible measure of survey quality. However, response rates do not necessarily provide an accurate measure of survey bias. Low response rates are only indicative of the possibility of survey bias. A number of research studies have found little relationship between the level of nonresponse and bias (e.g., Keeter, Miller, Kohut, Groves, & Presser, 2000). Where bias is found, adjusting survey weights for nonresponse and raking using variables that are correlated with the response characteristics can significantly reduce that bias.

Comparing Survey Respondents with Survey Nonrespondents

The *2019 WGRR* NRB analysis compared the sample composition with the survey respondent composition and assessed whether the patterns matched the *2017 WGRR* results. The *2019 WGRR* sample composition demographically differs from the Reserve component member population distribution due to intentional sampling strategies that allow OPA to make precise estimates for small subgroups and to sample constraints from simultaneously sampled surveys. The respondent composition differs from the sample distribution in predictable ways due to subgroups (e.g., junior enlisted members) responding at different rates. This analysis assesses

whether survey respondents possess similar observable characteristics (e.g., gender, component, and paygrade grouping) to survey non-respondents.

OPA draws optimized samples to reduce survey burden on members as well as produce high levels of precision for important domain estimates by using known information about the military population and their response propensity. It is important to note that OPA samples are often not proportional to their respective population. Depending on specific subgroups, OPA will over or under sample a specific group (e.g., E1-E4 US Army Reserve) to obtain enough expected responses to make statistically accurate estimates. Therefore, the sample composition is out of alignment with the population, and this is intentional. OPA is able to use its military personnel data to weight the respondents in order to make survey estimates representative of the entire Reserve component population. The demographics considered in this analysis include: gender, Reserve component, and paygrade grouping, which were directly controlled for in the raking stage and thus exactly match the known population values.

Table 17 shows the population, sample, and response breakdown by gender. OPA intentionally oversampled females in order to achieve reliable precision on estimates for outcomes conditional on reporting a sexual assault (i.e., retaliation) and other measures that were only asked of a very small subset of members (Table 17: columns b and d). For example, females make up 20% of the Reserve population but 38% of the *2019 WGRR* sample. The final weighting procedure (i.e., raking) pulls the respondents back into alignment with the gender composition in the Reserve components to ensure final weighted estimates do not over-represent females.

OPA performed a base-weighted Chi-square test of independence to examine the relationship between survey response and gender. Survey respondents are defined as complete eligible (n=34,169) or self/proxy report ineligible (n=735). OPA defines survey nonrespondents as SAMP_DC levels 5-11 (n=231,264; see Table 3). Record ineligibles (n=3,307) are not included in the analysis. The relationship between gender and survey response was not significant, χ^2 (df=1, n= 266,168) = 0.2[12], p = 0.655. The results indicate that different genders did not respond at significantly different rates. While males (moved from 62 to 58 percent) and females (38 to 42 percent) have different sample and respondent percentages, weighted response rates were similar between genders. Table 18 shows the response patterns in 2017 were similar, where males moved from 67 to 63 percent and females moved from 33 to 37 percent. Therefore, 2019 estimates are at similar risk of NRB as 2017 survey estimates due to only small differences in response rates by gender.

[12] The weighted Chi-square was generated using the PROC SURVEYFREQ with a weight statement within SAS 9.3 and SAS/STAT 12.1. The Rao-Scott correction to the Chi-square test was used since the data comes from a complex sample survey.

Table 17.
2019 WGRR Population, Sample Design, and Response Composition for Gender

Gender	Population		Sample		Respondents		Weighted Estimates (Final Weights)	
	Frequency (a)	Percent (b)	Frequency (c)	Percent (d)	Frequency (e)	Percent (f)	Frequency (g)	Percent (h)
Male	631,734	80	167,106	62	20,127	58	631,734	80
Female	161,482	20	102,369	38	14,777	42	161,482	20
Total	**793,216**	**100**	**269,475**	**100**	**34,904**	**100**	**793,216**	**100**

Table 18.
2017 WGRR Population, Sample Design, and Response Composition for Gender

Gender	Population		Sample		Respondents		Weighted Estimates (Final Weights)	
	Frequency (a)	Percent (b)	Frequency (c)	Percent (d)	Frequency (e)	Percent (f)	Frequency (g)	Percent (h)
Male	650,440	80	162,554	67	26,546	63	650,440	80
Female	157,687	20	78,872	33	15,269	37	157,687	20
Total	**808,127**	**100**	**241,426**	**100**	**41,815**	**100**	**808,127**	**100**

Table 19 shows the breakdown of the population, sample, and respondent distributions by Reserve component. There are fairly large differences between the unweighted sample size and unweighted respondents percentages, especially with Army National Guard (43% of the sample and only 32% of the respondents; Table 19: columns d and f), US Marine Corps Reserve (5 to 3 percent), Air National Guard (13 to 21 percent), and US Air Force Reserves (9 to 12 percent). Similar results are found in *2017 WGRR* where Army National Guard moved from 27 to 20 percent, Air National Guard moved from 10 to 17 percent, and US Air Force Reserve moved from 12 to 17 percent (Table 20). The final weighting procedure aligns respondent proportions back with the Reserve population for the components (Table 19: columns b and h).

OPA performed base weighted Chi-square test of independence on respondents and nonrespondents by component. The relationship between component and survey response was significant, χ^2 (df=5, n= 266,168) = 2795.6, $p < 0.0001$. The results indicate that different components respond at different rates and unweighted respondents are prone to nonresponse bias if not adjusted. Response patterns (e.g., Air Force Reserve and Air National Guard respond at higher rates) are the same across the 2017 and 2019 surveys, and therefore OPA concludes that NRB levels and direction will also be similar.

Table 19.
2019 WGRR Population, Sample Design, and Response Composition for Component

Reserve Component	Population		Sample		Respondents		Weighted Estimates (Final Weights)	
	Frequency (a)	Percent (b)	Frequency (c)	Percent (d)	Frequency (e)	Percent (f)	Frequency (g)	Percent (h)
Army National Guard	330,976	42	114,579	43	10,997	32	330,976	42
US Army Reserve	190,213	24	63,746	24	8,231	24	190,213	24
US Naval Reserve	58,715	7	17,995	7	2,811	8	58,715	7
Marine Corps Reserve	38,185	5	13,160	5	1,032	3	38,185	5
Air National Guard	106,391	13	34,602	13	7,483	21	106,391	13
US Air Force Reserve	68,736	9	25,393	9	4,350	12	68,736	9
Total	**793,216**	**100**	**269,475**	**100**	**34,904**	**100**	**793,216**	**100**

Table 20.
2017 WGRR Population, Sample Design, and Response Composition for Component

Reserve Component	Population		Sample		Respondents		Weighted Estimates (Final Weights)	
	Frequency (a)	Percent (b)	Frequency (c)	Percent (d)	Frequency (e)	Percent (f)	Frequency (g)	Percent (h)
Army National Guard	341,374	42	64,581	27	8,562	20	341,374	42
US Army Reserve	198,250	25	52,753	22	9,390	22	198,250	25
US Naval Reserve	57,984	7	33,293	14	6,555	16	57,984	7
Marine Corps Reserve	38,202	5	37,669	16	2,998	7	38,202	5
Air National Guard	104,165	13	24,203	10	7,146	17	104,165	13
US Air Force Reserve	68,152	8	28,927	12	7,164	17	68,152	8
Total	**808,127**	**100**	**241,426**	**100**	**41,815**	**100**	**808,127**	**100**

Table 21 shows the breakdown of the population, sample, and respondent percentage distributions by paygrade grouping. Based on historically different response rates and the need to make estimates for each paygrade, OPA oversampled the junior enlisted members and under sampled senior enlisted members (Table 21: columns b and d). For instance, senior enlisted members make up 42% of the Reserve component but only 34% of the *2019 WGRR* sample. On the other hand, junior enlisted are oversampled in proportion to their population (42% population, 50% sample). The basis for this approach is seen clearly in the differences between respondent percentages. The senior enlisted members account for 47% of the respondents, despite making up only 34% of the sample, while the junior enlisted members made up approximately half the sample (50%), yet represented only 22% of the respondents. Similar

results are found in *2017 WGRR* where junior enlisted members moved from 48 to 20 percent, and senior enlisted members moved from 29 to 38 percent (Table 22). These differences are adjusted based on known characteristics in post-survey weighting procedures, which aligned the respondent proportions equal to the Reserve population for paygrade (Table 21: columns b and h).

OPA performed base weighted Chi-square test of independence for paygrade grouping. The relationship between paygrade grouping and survey response was significant, χ^2 (df=3, n= 266,168) = 12265.5, $p < 0.0001$. The results indicate that different paygrade groupings respond at different rates and unweighted respondents are prone to nonresponse bias if not adjusted. Response patterns (e.g., junior enlisted respond at the lowest rates) are the same across the 2017 and 2019 surveys, and therefore OPA concludes that NRB levels and direction will also be similar.

Table 21.
2019 WGRR Population, Sample Design, and Response Composition for Paygrade

Paygrade Grouping	Population		Sample		Respondents		Final Weighted Estimates	
	Frequency (a)	Percent (b)	Frequency (c)	Percent (d)	Frequency (e)	Percent (f)	Frequency (g)	Percent (h)
E1-E4	333,602	42	134,810	50	7,539	22	333,602	42
E5-E9	329,762	42	92,295	34	16,245	47	329,762	42
W1-W5/O1-O3	70,367	9	24,412	9	5,165	15	70,367	9
O4-O6	59,485	7	17,958	7	5,955	17	59,485	7
Total	**793,216**	**100**	**269,475**	**100**	**34,904**	**100**	**793,216**	**100**

Table 22.
2017 WGRR Population, Sample Design, and Response Composition for Paygrade

Paygrade Grouping	Population		Sample		Respondents		Final Weighted Estimates	
	Frequency (a)	Percent (b)	Frequency (c)	Percent (d)	Frequency (e)	Percent (f)	Frequency (g)	Percent (h)
E1-E4	341,450	42	115,693	48	8,209	20	341,450	42
E5-E9	336,824	42	69,846	29	15,761	38	336,824	42
W1-W5	12,371	2	3,529	1	1,351	3	12,373	2
O1-O3	60,627	8	26,854	11	6,675	16	60,625	8
O4-O6	56,855	7	25,504	11	9,819	23	56,855	7
Total	**808,127**	**100**	**241,426**	**100**	**41,815**	**100**	**808,127**	**100**

Summary

The purpose of this NRB analysis was to determine whether there were differences between respondents and nonrespondents for three observable characteristics (gender, Reserve component, and paygrade grouping). Similar to the *2017 WGRR*, OPA found that the distribution of survey respondents was statistically significantly different from survey nonrespondents for Reserve component and paygrade grouping and that while gender was not found to be significant in 2019, response patterns by gender were similar to 2017.

Differences between respondents and nonrespondents on observable characteristics may suggest NRB. However, survey weighting effectively adjusts for these observable characteristics. Survey weighting also reduces any biases associated with unobservable characteristics (e.g., sexual assault rate) that are correlated with the observable characteristics.

Comparing survey respondents with the survey sample cannot definitively detect NRB. For example, if the respondents and nonrespondents look similar on observable characteristics, there is no evidence of NRB. However, if the respondents and nonrespondents look different on observable characteristics, OPA reduces this source of NRB during survey weighting. Therefore, neither of these two outcomes has the capability of detecting NRB. The relationship between observable and unobservable characteristics is unknown, and therefore the most desirable outcome would be where respondents and nonrespondents match on observable characteristics, something OPA does not find in either the *2017 WGRR* or *2019 WGRR*.

In this analysis, OPA observes that response patterns for the *2019 WGRR* are very similar to patterns from the *2017 WGRR* and concludes that the level of NRB should essentially be the same in both surveys. In the NRB studies conducted in *2017 WGRR* and *2015 WGRR*, OPA found little evidence of NRB and OPA draws that same conclusion here.

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 *Proceedings of the Section on Survey Research Methods, presented at the annual meeting of the American Statistical Association, San Francisco, CA, August 17-20, 1987* (pp. 194-199). Alexandria, VA: The Association.

Dever, J. A., and Mason, R. E. (2003). *DMDC sample planning tool: Version 2.1*. Arlington, VA: DMDC.

Keeter S, Kohut A, Miller A, Groves R, Presser S. "Consequences of Reducing Non-response in a Large National Telephone Survey" Public Opinion Quarterly. 2000; 64(2):125–48.

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

Little, R.J., & Rubin, D.B. (2002). *Statistical analysis with missing data* (2nd ed.). New York: John Wiley & Sons, Inc. doi: 10.1002/9781119013563

Mason, R. E., Wheeless, S. C., George, B. J., Dever, J. A., Riemer, R. A., and Elig, T. W. (1995). Sample allocation for the Status of the Armed Forces Surveys. In *Proceedings of the Section on Survey Research Methods, Volume II, American Statistical Association* (pp. 769–774). Alexandria, VA: The Association.

OPA. (2020a). *2019 Workplace and Gender Relations Survey of Reserve Component Members: Results and Trends* (Report No. 2020-051). Alexandria, VA: DMDC.

OPA. (2020b). *2019 Workplace and Gender Relations Survey of Reserve Component Members: Administration, Datasets, and Codebook* (Report No. 2020-052). Alexandria, VA: DMDC.

Appendix A.
Reporting Domains

DATA DRIVEN SOLUTIONS FOR DECISION MAKERS

OPA
OFFICE OF PEOPLE ANALYTICS

Reporting Domains

Domain	Domain Level	Population Size	Percent Sampled	Expected Sample Size
1	All Domains	793,216	34	269,693
2	Reserve	355,849	34	120,277
3	Army National Guard	330,976	35	114,518
4	Air National Guard	106,391	33	34,577
5	National Guard	437,367	34	149,142
6	US Army Reserve	190,213	34	63,721
7	US Navy Reserve	58,715	31	17,967
8	US Marine Corps Reserve	38,185	35	13,174
9	US Air Force Reserve	68,736	37	25,364
10	Enlisted	663,364	34	226,870
11	E1-E4	333,602	40	134,775
12	E5-E9	329,762	28	92,333
13	Officers	129,852	33	42,332
14	W1-W5/O1-O3	70,367	35	24,417
15	O4-O6	59,485	30	17,964
16	TPU/Unknown	639,350	36	228,887
17	AGR/FTS/AR	74,020	25	18,801
18	Military Technician	67,392	25	16,983
19	IMA	12,454	37	4,558
20	Non-Hispanic White	499,457	32	161,325
21	Total Minority	293,759	37	108,103
22	Females	161,482	63	102,380
23	Females*Enlisted	135,838	62	84,763
24	Females*E1-E4	73,806	62	45,834
25	Females*E5-E9	62,032	63	38,956
26	Females*Officers	25,644	69	17,617
27	Females*W1-O3	14,586	68	9,904
28	Females*O4-O6	11,058	70	7,707
29	Females*TPU/Unknown	129,523	65	84,319
30	Females*AGR/FTS/AR	16,169	54	8,780
31	Females*Military Technician	12,517	57	7,135
32	Females*IMA	3,273	67	2,177
33	Females*Non-Hispanic White	82,306	64	52,676
34	Females*Total Minority	79,176	63	49,643
35	Females*Reserve	79,959	63	50,374
36	Females*Army National Guard	59,052	63	37,203
37	Females*Army National Guard*Enlisted	52,655	63	32,909
38	Females*Army National Guard*Officers	6,397	67	4,292
39	Females*Air National Guard	22,471	66	14,786
40	Females*Air National Guard*Enlisted	19,472	65	12,676

41	Females*Air National Guard*Officers	2,999	71	2,126
42	Females*National Guard	81,523	64	52,012
43	Females*US Army Reserve	45,660	65	29,451
44	Females*US Army Reserve*Enlisted	36,251	63	22,766
45	Females*US Army Reserve*Officers	9,409	71	6,652
46	Females*US Navy Reserve	14,022	53	7,390
47	Females*US Navy Reserve*Enlisted	11,234	50	5,583
48	Females*US Navy Reserve* Officers	2,788	65	1,812
49	Females*US Marine Corps Reserve	1,565	40	624
50	Females*US Air Force Reserve	18,712	69	12,930
51	Females*US Air Force Reserve*Enlisted	15,004	69	10,383
52	Females*US Air Force Reserve*Officers	3,708	69	2,544
53	Males	631,734	27	167,410
54	Males*Enlisted	527,526	27	142,432
55	Males*E1-E4	259,796	34	89,110
56	Males*E5-E9	267,730	20	53,278
57	Males*Officers	104,208	24	24,802
58	Males*W1-O3	55,781	26	14,503
59	Males*O4-O6	48,427	21	10,267
60	Males*TPU/Unknown	509,827	28	144,791
61	Males*AGR/FTS/AR	57,851	17	10,008
62	Males*Military Technician	54,875	18	9,823
63	Males*IMA	9,181	26	2,387
64	Males*Non-Hispanic White	417,151	26	108,459
65	Males*Total Minority	214,583	27	58,581
66	Males*Reserve	275,890	25	69,800
67	Males*Army National Guard	271,924	29	77,498
68	Males*Army National Guard*Enlisted	233,955	29	68,315
69	Males*Army National Guard*Officers	37,969	24	9,188
70	Males*Air National Guard	83,920	24	19,805
71	Males*Air National Guard*Enlisted	71,637	24	16,906
72	Males*Air National Guard*Officers	12,283	23	2,862
73	Males*National Guard	355,844	27	97,145
74	Males*US Army Reserve	144,553	24	34,259
75	Males*US Army Reserve*Enlisted	116,103	24	27,284
76	Males*US Army Reserve*Officers	28,450	25	6,999
77	Males*US Navy Reserve	44,693	24	10,592
78	Males*US Navy Reserve*Enlisted	33,320	27	8,830
79	Males*US Navy Reserve* Officers	11,373	15	1,751
80	Males*US Marine Corps Reserve	36,620	34	12,524
81	Males* US Marine Corps Reserve*Enlisted	32,480	34	10,978
82	Males*US Marine Corps Reserve*Officers	4,140	37	1,544
83	Males*US Air Force Reserve	50,024	25	12,456
84	Males*US Air Force Reserve*Enlisted	40,031	25	10,048
85	Males*US Air Force Reserve*Officers	9,993	24	2,428

Appendix B.
Military Accession Program

DATA DRIVEN SOLUTIONS FOR DECISION MAKERS

OPA
OFFICE OF PEOPLE ANALYTICS

Military Accession Program

Military Accession Program
1=Induction
2=Voluntary enlistment in a Regular Component
3=Vol enlist - Rsv Comp for Reg DEP - 10 USC 12103/10 USC 513
4=Voluntary enlistment - Rsv Comp, Sec 511, ref(b). Excl DEP
A=U.S. Military Academy
B=U.S. Naval Academy
C=U.S. Air Force Academy
D=U.S. Coast Guard Academy
F=Air National Guard Academy of Military Sciences
G=ROTC/NROTC scholarship program
H=ROTC/NROTC non-scholarship program
J=OCS, AOCS, OTS, or PLC
L=National Guard state OCS
M=Direct appointment authority, Commissioned Off, professional
N=Direct appointment authority, Commissioned Off, all other
P=Aviation training program other than OCS, AOCS, OTS, or PLC
Q=Limited Duty Officer Program
R=Direct appointment authority, warrant officer
S=Direct appointment authority, commissioned warrant officer
T=Warrant Officer Aviation Training Program
W=NA
X=Other
Z=Unknown or Not Applicable

OPA

OFFICE OF PEOPLE ANALYTICS