

Office of People Analytics (OPA)



2014 Status of Forces Survey of Active Duty Members

Nonresponse Bias Analysis 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/>

**2014 STATUS OF FORCES SURVEY OF ACTIVE
DUTY MEMBERS:
NONRESPONSE BIAS ANALYSIS REPORT**

**Ada Harris, Jeffrey Schneider, David McGrath, and Eric Falk
Defense Research, Surveys, and Statistics Center (RSSC)**

**Office of People Analytics (OPA)
Defense Research, Surveys, and Statistics Center
4800 Mark Center Drive, Suite 06E22, Alexandria, VA 22350-4000**

Table of Contents

	<u>Page</u>
Introduction.....	1
2014 SOFS-A Survey	2
Studies and Summary of Findings	3
Section 1: Comparing Population, Sample, and Respondent Composition.....	4
Summary of Sample Composition Compared With Survey Respondents.....	10
Section 2: Comparison of Weighted Sample Estimates with Frame Characteristics	10
Section 2 Conclusion	13
Section 3: Drop Off Analysis.....	14
Study 3 Conclusion.....	15
Section 4: Use of Prior SOFS-A respondents to estimate NRB	16
Study 4a: Effect of Prior Response on Later Response	17
Data.....	17
Analysis.....	19
Results.....	21
Study 4b: Prior Satisfaction	22
Data.....	23
Analysis.....	23
Results.....	25
Section 4 Conclusion	26
Overall Conclusions.....	26
References.....	29

Appendixes

A. Logistic Model for 4a Survey Response	31
B. Logistic Model for 4b Survey Response	35

List of Tables

1. 2014 SOFS-A Population, Sample Design, and Estimation for Service	6
---	---

Table of Contents (Continued)

	<u>Page</u>
2. 2014 SOFS-A Population, Sample Design, and Estimation for Paygrade	7
3. 2014 SOFS-A Population, Sample Design, and Estimation for Race/Ethnicity.....	7
4. 2014 SOFS-A Population, Sample Design, and Estimation for Gender.....	8
5. 2014 SOFS-A Population, Sample Design, and Estimation for Region.....	8
6. 2014 SOFS-A Population, Sample Design, and Estimation for Family Status with Final Weights	9
7. 2014 SOFS-A Population, Sample Design, and Estimation for Education with Final Weights	10
8. Selected Variables by Population Value and Weighted Estimates.....	11
9. Comparison of Population Values to Weighted Estimates for Years of Active Federal Military Service by Paygrade Group	12
10. Comparison of Population Values to Weighted Estimates for AFQT Percentile Score by Paygrade Group	13
11. Comparison of Population Values to Weighted Estimates for Age by Paygrade Group	13
12. SOFS-A1401 Drop-Offs (Last Question Answered) Frequency greater than 50.....	15
13. Missing Data Analysis for key Questions.....	15
14. Unique Unweighted Counts of Members Sampled in Multiple SOFS-A Surveys.....	17
15. Unweighted Number of SOFS-A(s) Sampled by Paygrade.....	18
16. Unweighted Number of SOFS-A(s) Responded by Paygrade.....	18
17. Distribution of Final Weights (n=16,503)	19
18. Predictor Variables for modelling survey response at T2 (for Studies 4a and 4b).....	20
19. Weighted Mean Satisfaction for Survey 'Avoiders' vs 'Takers' by Paygrade.....	22
20. Members completed survey and later sampled survey	23
21. Unweighted Response Rates for Later Survey (T2) by Reported Satisfaction Level from Prior Survey (T1)	24
22. Final Weight Moments (n=4,391)	24
23. Naive Bayes Confusion Matrix.....	26

2014 STATUS OF FORCES SURVEY OF ACTIVE DUTY MEMBERS NONRESPONSE BIAS ANALYSIS REPORT:

Introduction

In 2006 the Office of Management and Budget recommended that a nonresponse bias (NRB) analysis be completed for any government survey with a response rate below 80%. The purpose of NRB analysis is to determine whether survey estimates are biased due to nonresponse of some sample members. The Defense Research, Surveys, and Statistics (RSSC) Division within the Office of People Analytics (OPA)¹ conducted a series of statistical analyses to determine whether those who did not respond to the survey would have provided significantly different answers than those who did respond. Because the response rate for the *2014 Status of Forces Survey of Active Duty Military (2014 SOFS-A)* was 21.3% based on the American Association of Public Opinion Research response rate calculation (AAPOR Response Rate 4), a NRB analysis was initiated.

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 NRB associated with those observed variables if these variables are strongly associated with the behaviors. When all NRB can be eliminated in this manner, the absence is called *ignorable* or *missing at random* (Little & Rubin, 2002). The more observable demographic variables that were incorporated into the weights, the more plausible it is to assume that the weights eliminate any NRB.

The objective of this research was to assess the extent of NRB for the estimated percentage that were satisfied with the military way of life and the percentage that planned to stay (henceforth these rates will be referred to as satisfaction and retention) in the active duty. The percentages that were satisfied and planned to stay in the military were provided to the policy offices for planning purposes. The level of NRB can vary for every question on the survey, but OPA focused on the satisfaction and retention rate because these tended to be the more central questions on the survey. NRB occurs when survey respondents are systematically different from nonrespondents. Statistically, the bias in a respondent mean (e.g., satisfaction rate) is a function of the response rate and the relationship (covariance) between response propensities and the estimated statistics (i.e., satisfaction rate), and takes the following form:

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

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

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

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).

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 civilian studies in the past decade has resulted in a greater focus on potential NRB. OPA investigated the presence of NRB using many different methods, and this report summarizes the following methods and results:

1. Evaluate composition of sample compared with survey respondents,
2. Evaluate weighted point estimates of latent variables with the population,
3. Analyze questions on the survey instrument that have high item missing data due to respondent refusals (drop-offs), and
4. Analyze whether past survey respondents are more likely to respond to later surveys.

2014 SOFS-A Survey

The 2014 SOFS-A survey sample size was 65,097 active duty members selected from 1,348,423 active duty members on the May 2014 Active Duty Master File (ADMF). The frame included Army, Navy, Marine Corps, and Air Force active duty members who were ranked E1–O6 in September 2014 when the survey fielded. OPA selected a stratified random sample using the following five characteristics to define the stratification dimensions:

- Service (Army, Navy, Marine Corps, Air Force),
- Paygrade group (E1–E4, E5–E9, W1–W5, O1–O3, O4–O6),
- Race/ethnicity (Minority, Non-Minority),
- Location (US & US Territories, Europe, Asia & Pacific Islands), and
- Family status (Single with Children, Dual Service, other family).

Completed surveys were returned by 13,447 eligible sampled members, resulting in a 21% weighted response rate. These respondents were weighted to the full active population using standard weighting-class methods. The four-step weighting process is described below.

1. Assigning a base weight based on the inverse of the probability of selection,
2. Adjusting the base weight by distributing the weights from the cases of unknown eligibility to the cases of known eligibility,

3. Adjusting the weight from step 2 by distributing the weights from incomplete cases to the complete cases, and
4. Poststratifying the step 3 weight to known population totals for race/ethnicity, Service, gender, and paygrade.

SAS[®] was used for many of the statistical steps including sampling and weighting. For more information about the sampling frame, sampling design and weighting procedure refer to the 2014 Status of Forces Survey of Active Duty Military: Statistical Methodology Report (DMDC, 2015b). For information about the survey design, survey instruments and administration refer to the 2014 Status of Forces Survey of Active Duty Military: Administration, Datasets, and Codebook (DMDC, 2015a). For tabulation of responses refer to the 2014 Status of Forces Survey of Active Duty Military: Tabulation of Responses (DMDC, 2015c).

Studies and Summary of Findings

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 level of nonresponse and bias (e.g., Keeter, Miller, Kohut, Groves, & Presser, 2000). Where bias is found, adjusting survey weights for nonresponse and poststratification using variables that are correlated with the response characteristics can significantly reduce that bias.

The 2014 SOFS-A NRB analysis was investigated by using several different methods, and this paper summarizes the following studies and results:

1. **Evaluate the composition of the survey respondents relative to the sample and population**—The 2014 SOFS-A sample composition demographically differs from the active duty population distribution due to intentional sampling strategies that allow OPA to make precise estimates for small subgroups. 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., Service, gender, region, family status) to survey nonrespondents. OPA found that the distribution of survey respondents was statistically significantly different from survey nonrespondents for six of the seven characteristics analyzed. Although poststratified weights reduce nonresponse bias for unobservable characteristics that are correlated with the poststratification variables, the differences on observable characteristics presents

NRB concerns. 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 *2014 SOFS-A*. Therefore, OPA interprets this study as presenting concerns that NRB may be present in estimates.

2. **Evaluate weighted point estimates of variables with known population values**—OPA assessed the robustness of survey weights by estimating three known population quantities not directly used in weighting: 1) Armed Forces Qualification Test (AFQT) scores, 2) years of active duty military service, and 3) age. An independent one sample t-test indicated that AFQT scores were systematically different for respondents than nonrespondents after controlling for potential demographic differences. For the other two variables there were no statistically significant differences between the weighted estimates and known population values. OPA concludes that this study provides some evidence of NRB because the weighted estimates for AFQT scores differ from the known population values. Members with high AFQT scores responded to *2014 SOFS-A* at slightly higher rates.
3. **Analyze questions on the survey instrument that have high item missing data due to respondent refusals (drop-offs)**—OPA determined the number of drop-offs for a question is driven more by the length of the questionnaire than the content of the questions, an effect OPA also observed when assessing NRB in the *2012 WGRA* survey (DMDC, 2013). The analysis of missing data on select sensitive questions provides no evidence of NRB.
4. **Analyze whether past survey respondents and respondents with certain characteristics (e.g., high job satisfaction) are more likely to respond to later surveys**—OPA found that past survey respondents and members who reported higher satisfaction on an earlier survey responded at higher rates to later surveys. This study provides evidence of NRB in estimates from *2014 SOFS-A* (and most likely any *SOFS-A*), with the effect being that estimates from most of OPA's active duty surveys are biased in a positive direction, meaning that survey estimates are slightly more favorable than reality.

Across these four studies, OPA finds evidence that even after weighting on large number of characteristics, NRB exists in *SOFS-A* estimates, and estimates are slightly more favorable than reality. However, weighting by a large number of characteristics likely keeps the NRB at a modest level.

Section 1: Comparing Population, Sample, and Respondent Composition

OPA considered whether, and how, survey nonresponse (unit nonresponse), affects NRB for this survey. In this section OPA evaluates the composition of the respondents and nonrespondents based on a set of demographic variables. Key demographics were identified based on the survey response and OPA's experience in military surveys. Previous studies

conducted by OPA indicate that variables such as member's age, paygrade, and Service are critical in predicting military survey response. 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 Army) to obtain enough expected responses to make statistically accurate estimates. While the sample and the number of responses might look out of alignment with the population, this is by design. OPA is able to use its military personnel data to weight the respondents in order to make survey estimates representative of the entire active duty population. The demographics considered in this analysis include Service, paygrade, race/ethnicity, gender, region, family status, and education. Service, paygrade, and race were directly controlled for in the poststratification stage and thus match the known population values exactly.

Table 1 (page 6) shows the breakdown of the population, sample, and respondent distributions by Service. Based on historically different response rates and the need to make estimates for each Service, OPA slightly oversampled the Marine Corps, and under sampled the Air Force (Table 1: columns b and d). For instance, Air Force makes up 24% of the active duty but since they respond better to survey only 20% of the *2014 SOFS-A* sample. There are fairly large differences between the unweighted sample size and unweighted respondents percentages, especially with Army and Air Force (Table 1: columns d and f). Army made up 39% of the sample and only 33% of the respondents. On the other hand, Air Force made up 20% of the sample, but 29% of the respondents. Finally, OPA uses post-survey weighting procedures (described in DMDC 2014-044) to adjust the 33% of Army respondents to make them representative of the Army's true 38% proportion of the overall active duty military population. The final weighting procedure aligns respondent proportions back with the military population for Service. Service was controlled in poststratification so the final weighted estimate is equal to the population (Table 1: columns b and h). Poststratified weights are likely to reduce nonresponse bias for outcomes that are correlated with the poststratification variables, but next we will explore if nonrespondents are different than respondents on observable characteristics in the next section.

OPA performed base weighted Chi-square tests of independence to examine the relationship between survey response and survey nonresponse (data not shown) on observable characteristics (i.e., Service, paygrade, race/ethnicity, gender, region, family status, and education). A survey respondent is complete eligible (n= 13,447) or self/proxy report ineligible (n= 190). A survey nonrespondent is defined as a postal non-deliverable, returned blank survey, active refusal, incomplete survey and a nonrespondent (n= 50,433). Record ineligibles (n= 1,027) are not included in the analysis. The relationship between survey response and Service was significant, χ^2 (df = 3, n= 1,326,518) = 840.9, $p < 0.0001$. The relationship between survey response and paygrade was significant, χ^2 (df = 4, n= 1,326,518) = 5,424.7, $p < 0.0001$. The relationship between survey response and race/ethnicity was significant, χ^2 (df = 1, n= 1,326,518) = 159.2, $p < 0.0001$. The relationship between survey response and gender was not significant, χ^2 (df = 1, n= 1,326,518) = 0.01, $p < 0.9$. The relationship between survey response and region was significant, χ^2 (df = 2, n= 1,326,518) = 66.4, $p < 0.0001$. The relationship between survey response and family status was significant, χ^2 (df = 4, n= 1,326,518) = 1,972.0, $p < 0.0001$. The relationship between survey response and education was significant, χ^2 (df = 3, n= 1,326,518)

=3,908.6, $p < 0.0001$. Unweighted Chi-square tests yielded the same conclusions as the weighted tests. Therefore, it is also possible that the respondents would also be different from the population on unobservable characteristics (e.g., satisfaction, retention, etc.). OPA interprets this as some presence of NRB.

Table 1.
2014 SOFS-A Population, Sample Design, and Estimation for Service

Service	Population		Sample		Respondents		Estimates (Final Weights)	
	Frequency (a)	Percent (b)	Frequency (c)	Percent (d)	Frequency (e)	Percent (f)	Frequency (g)	Percent (h)
Army	512,705	38	25,130	39	4,464	33	512,705	38
Navy	319,393	24	15,738	24	3,180	24	319,393	24
Marine Corps	190,625	14	10,920	17	1,896	14	190,625	14
Air Force	325,700	24	13,309	20	3,907	29	325,700	24
Total	1,348,423	100	65,097	100	13,447	100	1,348,423	100

Table 2 (page 7) shows the breakdown by paygrade grouping. Based on historically different response rates and the need to make estimates for each paygrade, OPA only slightly oversampled the junior enlisted members and under sampled senior enlisted members (Table 2: columns b and d). For instance, senior enlisted members make up 39% of the active duty but only 26% of the 2014 SOFS-A sample. On the other hand, junior enlisted are sampled roughly in proportion to their population (43% population, 47% sample). The basis for this approach is seen clearly in the differences between respondent percentages. The senior enlisted members, despite making up only 26% of the sample account for 32% of the respondents, while the junior enlisted members made up nearly half the sample (47%) of the sample, yet they represented only 18% of the respondents. These differences are adjusted based on known characteristics in post-survey weighting procedures, which aligned the respondent proportions equal to the military population for paygrade (Table 2: columns b and h).

Table 2.
2014 SOFS-A Population, Sample Design, and Estimation for Paygrade

Paygrade Grouping	Population		Sample		Respondents		Estimates (Final Weights)	
	Frequency (a)	Percent (b)	Frequency (c)	Percent (d)	Frequency (e)	Percent (f)	Frequency (g)	Percent (h)
E1-E4	579,984	43	30,278	47	2,442	18	579,984	43
E5-E9	529,886	39	16,733	26	4,328	32	529,886	39
W1-W5	19,535	1	1,306	2	465	3	19,535	1
O1-O3	131,939	10	10,551	16	3,338	25	131,939	10
O4-O6	87,079	6	6,229	10	2,874	21	87,079	6
Total	1,348,423	100	65,097	100	13,447	100	1,348,423	100

Table 3 shows the breakdown by minority status. Based on historically different response rates and the need to make estimates for minority and non-minority groups, OPA oversampled the minority group from their true proportion of 37% of the population to 41% of the sample. Subsequently, the oversampled minority group made up only 35% of the respondents. Minority status was the final variable used in the poststratification process, thus the weighted estimates are equal to the military population (Table 3: columns b and h).

Table 3.
2014 SOFS-A Population, Sample Design, and Estimation for Race/Ethnicity

Race/Ethnicity	Population		Sample		Respondents		Estimates (Final Weights)	
	Frequency (a)	Percent (b)	Frequency (c)	Percent (d)	Frequency (e)	Percent (f)	Frequency (g)	Percent (h)
Non-Hispanic White	849,764	63	38,373	59	8,790	65	849,764	63
Total Minority	498,659	37	26,724	41	4,657	35	498,659	37
Total	1,348,423	100	65,097	100	13,447	100	1,348,423	100

Table 4 (page 8) shows the breakdown by gender. Based on historical response rates, male and female members with similar characteristics respond at similar rates. Based on this information, OPA did not explicitly over or under sample either group, and the sample closely reflects their true population values. Additionally, weighted estimates are equal to their population percent without any additional controls. This result indicates that post-survey adjustment techniques may have reduced potential bias that might have existed from gender.

Table 4.
2014 SOFS-A Population, Sample Design, and Estimation for Gender

Gender	Population		Sample		Respondents		Estimates (Final Weights)	
	Frequency (a)	Percent (b)	Frequency (c)	Percent (d)	Frequency (e)	Percent (f)	Frequency (g)	Percent (h)
Male	1,148,168	85	53,880	83	11,092	82	1,140,762	85
Female	202,255	15	11,217	17	2,355	18	207,661	15
Total	1,348,423	100	65,097	100	13,447	100	1,348,423	100

Table 5 shows the breakdown by region. Based on historically similar response rates OPA sampled the different region categories in line with their known population percentages, taking care to ensure that enough members would respond in each category to make precise estimates for each region. Region was not controlled in poststratification, but the weighted estimates were around one to two percent of the population (Table 5: columns b and h). OPA determined that the magnitude of these differences was small and thus not likely to cause significant NRB in estimates.

Table 5.
2014 SOFS-A Population, Sample Design, and Estimation for Region

Region	Population		Sample		Respondents		Estimates (Final Weights)	
	Frequency (a)	Percent (b)	Frequency (c)	Percent (d)	Frequency (e)	Percent (f)	Frequency (g)	Percent (h)
U.S. (Incl. Territories)	1,193,955	89	57,832	89	11,776	88	1,178,734	87
Europe	68,700	5	3,325	5	885	7	83,617	6
Asia	85,768	6	3,940	6	786	6	86,072	6
Total	1,348,423	100	65,097	100	13,447	100	1,348,423	100

Table 6 (page 9) shows the breakdown by family status. OPA made final estimates by family status but did not significantly oversample any subgroup. Married with children have historically shown higher response rates and these were slightly under sampled in relation to other groups (34% in sample and 38% in population). Members with different family status responded at significantly different rates. For instance, “single without children” members made up 40% of the sample, but were only 24% of the respondents. In contrast, “married with children” comprised 34% of the sample, but were 49% of the respondents. Family status was not directly controlled in poststratification but the post-survey estimation procedures were able to

move the proportions based in alignment with the population. Again, it is unlikely not controlling for family status has any large effects on NRB in estimates.

Table 6.
2014 SOFS-A Population, Sample Design, and Estimation for Family Status with Final Weights

Family Status	Population		Sample		Respondents		Estimates (Final Weights)	
	Frequency (a)	Percent (b)	Frequency (c)	Percent (d)	Frequency (e)	Percent (f)	Frequency (g)	Percent (h)
Single With Children	65,095	5	3,970	6	872	6	58,680	4
Single Without Children	530,361	39	26,352	40	3,257	24	513,923	38
Married with Children	513,880	38	22,297	34	6,593	49	518,121	38
Married Without Children	238,229	18	12,453	19	2,721	20	256,937	19
Total	1,347,564	100	65,072	100	13,443	100	1,347,661	100

Note: Unknown family status frequencies are not shown. The population frequency of unknown is 859. The sample size of unknown family status is 25. The respondent size of unknown family status was 4. The final weighted frequency of unknown family status was 762.

Table 7 (page 10) shows the breakdown by education. OPA did not control for education during sampling, nor was it accounted for in the post-survey adjustment process. While active duty members of different educational groups do respond at different rates, these differences were only controlled for during the eligibility and completion adjustments in weighting. Members with “no college” were over half of the sample (62%), but only comprised 38% of the respondents, indicating that these members respond at significantly worse rates than other groups. In contrast, members with “graduate degrees or higher” accounted for only 11% of the sample, but were almost a quarter (24%) of the respondents. If these phenomena were not controlled during weighting, then the survey estimates in some instances could be biased. However, the eligibility and completion adjustments in weighting controlled for these differences, allowing OPA to bring the differing educational groups back into proper alignment (Table 7: columns b and h).

Table 7.
2014 SOFS-A Population, Sample Design, and Estimation for Education with Final Weights

Education	Population		Sample		Respondents		Estimates (Final Weights)	
	Frequency (a)	Percent (b)	Frequency (c)	Percent (d)	Frequency (e)	Percent (f)	Frequency (g)	Percent (h)
No College	908,584	67	40,477	62	5,107	38	874,785	65
Some College	165,668	12	6,475	10	1,736	13	179,041	13
4-year Degree	165,425	12	10,884	17	3,388	25	180,176	13
Grad/Prof Degree	108,746	8	7,261	11	3,216	24	114,420	8
Total	1,348,423	100	65,097	100	13,447	100	1,348,423	100

Summary of Sample Composition Compared With Survey Respondents

The purpose of this section of the NRB analysis was to determine whether there were differences between respondents and nonrespondents for seven observable characteristics (e.g., service, family status, and gender). OPA found that the distribution of survey respondents was statistically significantly different from survey nonrespondents for six of the seven characteristics: Service, paygrade, family status, region, race, and education.

Differences between respondents and nonrespondents may suggest NRB. However, survey weighting effectively adjusts for these observable characteristics. Survey weighting also reduces any biases associated with unobservable characteristics (e.g., tobacco use) 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 or eliminates 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 SOFS-A. **Therefore, OPA interprets this study as presenting concerns that NRB may exist in SOFS-A estimates.**

Section 2: Comparison of Weighted Sample Estimates with Frame Characteristics

Survey researchers have observed that a potential key component of NRB analysis is to compare weighted estimates from the survey respondents with known population values. If there is a significant difference between these, this could be an indication of potential NRB. OPA methods make extensive use of demographic variables in every stage of weighting to reduce NRB given the differences in response rates. This analysis focuses on studying how effectively

OPA post-survey adjustments perform by studying three relevant continuous variables that were not used in weighting. The variables selected for study were:

1. Years of active federal military service,
2. Armed Forces Qualification Test (AFQT) percentile score, and
3. Age.

Table 8 shows the unweighted estimates, weighted estimates, and known population values for these three variables. Overall we see that on the aggregate level, OPA’s weighting methods move the estimate closer to the true parameter for years of active federal military service and age. Additionally the unweighted estimates highlight just how different the unadjusted respondent sample is from the population, and how effective OPA methods adjust to true values. Furthermore, the margin of errors around the estimates includes the true population values for active federal military service. However, the weighted values of AFQT Percentile Score are about three percentage points higher than the population values. OPA conducted a one sample t-test to determine whether weighted estimates of the sample on age and AFQT percentile are significantly different from known population values. The mean value of the weighted estimates for age (M= 28.8, SD= 76.1) was significantly different from the known population value, $t(13,446)= 2.5, p < 0.01$. The mean value of the weighted estimates for AFQT percentile (M= 66.0, SD= 238.9) was significantly different from the known population value, $t(6,675)= 11.9, p < .0001$. The mean value of the weighted estimates for active federal military service (M= 7.1, SD= 66.8) was not significantly different from the known population value, $t(13,425)= -0.49, p < 0.6$. Each of the variables presented in Table 8 are studied in more depth in Table 9 through Table 11. Although the estimated age was shown to be statistically significant, it is not practically significant. Based on this research, OPA has been using age as a variable for their weighting adjustments in all surveys.

Table 8.
Selected Variables by Population Value and Weighted Estimates

Population Variable (a)	Variable Name (b)	Unweighted Estimate (c)	Weighted Estimate (d)	Population Value (e)	Difference (percentage points) (e-d)
Active Federal Military Service Years *	AFMS_YR_QY	10.6	7.1 ± 0.1	7.2	0
AFQT Percentile Score *	AFQT_SCR_QY	64.6	66.0 ± 0.6	63.3	-2.7
Age	AGE	33.3	28.8 ± 0.1	28.7	-0.1

Note: Margins of error for unweighted and weighted estimates are done at the 95% Confidence Level. For active federal military service years; members with a value of 99 were removed from analysis. For AFQT percentile scores, only enlisted members and those with percentiles over 0 were used in estimates.

Table 9 displays years of active federal military service by paygrade group (separated into seven levels). Years of active federal military service indicates the length of time in years that the member has currently served in the active duty military. This is a relevant variable to study as members with more time on active duty have been shown to respond in higher rates. Thus if the weighted estimates did not effectively estimate the population, OPA weights might be biased based on years of service. However, overall the weighted estimates are consistent with the population values across each paygrade. The weighted estimates show low NRB because they are similar to the population. The weighted estimates effectively reduce bias because the unweighted estimates are closer to the true known population values.

Table 9.
Comparison of Population Values to Weighted Estimates for Years of Active Federal Military Service by Paygrade Group

Paygrade (a)	Unweighted Estimates (b)	Weighted Estimate (c)	Population Value (d)	Difference (percentage points) (d-c)
E1-E3	0.9	0.8± 0.06	0.9	0.1
E4	3.2	3.0± 0.1	3.1	0.1
E5-E6	9.8	8.8± 0.2	8.9	0.1
E7-E9	18.3	17.9± 0.2	17.9	0
W1-W5	17.6	16.7± 0.6	16.0	-0.7
O1-O3	7.3	6.8± 0.4	6.6	-0.2
O4-O6	17.7	17.4± 0.2	17.5	0.1

Note: Margins of error for unweighted and weighted estimates are done at the 95% confidence level

Table 10 displays the AFQT percentile score by enlisted paygrade grouping (officers do not take the AFQT). The AFQT is designed to determine enlistment eligibility and assign applications to military jobs, as well as aid students in career exploration. The AFQT percentile score indicates the percentile that enlisted members received on the test. While the AFQT score is only one measure of aptitude, it is designed to be an unbiased test that seeks to minimize adverse impact. Test items are specifically created so that they perform similarly across race and gender groupings. Thus, AFQT percentile score is relevant to study as the characteristics of members that perform very high on the test could be different from those that score lower.

As indicated in the introduction to Section 2, results of the weighted estimates compared to the population values indicate that in every enlisted paygrade the estimated AFQT score is higher than the population value. The weighted estimate and population values are significantly different and the result could indicate that members on the higher aptitude end of the spectrum are more likely to respond to an OPA survey. The implications of this result are discussed in the Section 2 conclusion.

Table 10.
Comparison of Population Values to Weighted Estimates for AFQT Percentile Score by Paygrade Group

Paygrade (a)	Unweighted Estimates (b)	Weighted Estimate (c)	Population Value (d)	Difference (percentage points) (d-e)
E1–E3	68.1	68.1± 1.3	63.5	-4.6
E4	67.8	67.5± 1.2	64.5	-3.0
E5–E6	63.1	64.3± 0.8	62.7	-1.6
E7–E9	61.8	62.8± 0.9	62.2	-0.6

Note: Margins of error for unweighted and weighted estimates are done at the 95% confidence level

The final continuous variable studied was the age of the military member. Paygrade is correlated with age, but since age was not directly used in weighting OPA was interested in understanding the potential differences between the weighted estimates and population values. Table 11 shows the results of the weighting and in this case indicates that OPA weighting is effective. The weighted estimates and population values are not significant and are evidence that not controlling for age had little if any overall effect on estimates.

Table 11.
Comparison of Population Values to Weighted Estimates for Age by Paygrade Group

Paygrade (a)	Unweighted Estimates (b)	Weighted Estimate (c)	Population Value (d)	Difference (percentage points) (d-e)
E1–E3	22.3	21.5± 0.2	21.4	-0.1
E4	25.8	24.8± 0.2	24.6	-0.2
E5–E6	31.5	30.4± 0.2	30.2	-0.2
E7–E9	39.0	38.7± 0.2	38.6	-0.1
W1–W5	38.9	37.8± 0.6	37.6	-0.2
O1–O3	30.5	29.7± 0.2	29.6	-0.1
O4–O6	42.1	41.9± 0.2	41.8	-0.1

Note: Margins of error for unweighted and weighted estimates are done at the 95% confidence level

Section 2 Conclusion

OPA’s weighted estimates are close to the true population values for two of the three characteristics in this study. OPA found that enlisted members with more aptitude, as measured by AFQT percentile score, are overrepresented in the weighted set of respondents, and consequently we should also use AFQT in future SOFS-A weighting. Because the estimates were close for age and years of service, the impact of AFQT on final estimates is likely modest.

Section 3: Drop Off Analysis

Section 3 analyzed item missing data for all questions on the *2014 SOFS-A* to investigate whether some respondents refuse to answer questions (item missing data) or quit the survey (i.e., drop-off) because of the design or sensitivity of the questions. If the decision to refuse to answer the question is not random (i.e., those who skip the “current deployment” question have different experiences than complete respondents), then a source of NRB may exist. OPA cannot directly test this because the data is missing for questions skipped by the respondent; however, unusual item nonresponse patterns or high item nonresponse may indicate possible NRB.

OPA analyzed missing-data patterns in the *2014 SOFS-A* instrument using a drop-off analysis technique. Drop-off analysis shows the last question a survey respondent answered on the survey. This analysis includes all respondents who started the survey. For example, if a respondent answered Q1 to Q5 and quit, the drop-off analysis would place the respondent in the frequency count at Q5. Drop-off analysis does not count for standard item missing data (e.g., a respondent skips one question, accidentally or on purpose, but returns to answer further questions). For example, if a respondent answered Q1 to Q5, skipped to Q7, answered Q7 to Q10, and then answered no further questions, the drop-off analysis would include the respondent in the frequency count for Q10.

Table 12 (page 15) presents the *2014 SOFS-A* drop-offs for questions where more than 50 respondents exited the survey. Most respondents (83%) complete the survey from start to finish, dropping off at the final question, Q222. Prior to the analysis, OPA subject matter experts identified questions 25 (satisfaction) and 27 (retention) as the two most important questions on the survey. These two questions showed no systematic increase in drop-offs. For question 25, eight people dropped off and for question 27 ten people dropped off. These two questions are defined here.

25. *Overall, how satisfied are you with the military way of life?*

- *Very satisfied (5)*
- *Satisfied (4)*
- *Neither satisfied nor dissatisfied (3)*
- *Dissatisfied (2)*
- *Very dissatisfied (1)*

27. *Suppose that you have to decide whether to stay on active duty. Assuming you could stay, how likely is it that you would choose to do so?*

- *Very likely (5)*
- *Likely (4)*
- *Neither likely nor unlikely (3)*
- *Unlikely (2)*
- *Very unlikely (1)*

Table 12.
SOFS-A1401 Drop-Offs (Last Question Answered) Frequency greater than 50.

Last Question Answered	Number of Drop-Offs	Percent of Drop-Offs	Comment
Q29	118	0.81	The following question (Q30) is a very large matrix grid with 15 decision points. Respondents may have dropped-off due to question design.
Q46	55	0.38	The following question (Q47) again is a large matrix and is related to personal stress. Respondents may have caused respondent to drop-off due to sensitivity.
Q52	61	0.42	Question 53 asks ideas for problem solving and it is unclear why this question would cause a relatively larger drop-off
Q80	74	0.51	The following question (Q81) is a very large matrix grid with 14 decision points. Respondents may have dropped-off due to question design.
Q145	54	0.37	Question 146 is another matrix grid
Q148	54	0.37	Question 149 asks what the member plans to do after retirement and unclear what would cause these members to stop
Q152	53	0.36	The following question, In the past six months, did you talk to a counselor about... (Q153). Respondents may have caused respondent to drop-off due to item sensitivity.
Q214	113	0.77	The following question (Q215) seen by all respondents asks statements about attitudes about sexual assault. Respondents may have caused respondent to drop-off due to sensitivity.

Table 13 shows the missing data rates for complete eligible and partial respondents only (n=14,593). The table shows the percent missing for these key questions is less than 2 percent. Given the very high item completion rate, OPA’s analysis of missing data shows no evidence of NRB.

Table 13.
Missing Data Analysis for key Questions

Question	Variable	Potential Responses	Actual Responses	Percent Missing
Q25. Overall, how satisfied are you with the military way of life?	SATOVER	14,593	14,320	1.9%
Q27. Suppose that you have to decide whether to stay on active duty. Assuming you could stay, how likely is it that you would choose to do so?	RETINT1	14,593	14,302	2.0%

Study 3 Conclusion

Overall we see that the members may drop-off due to task burden and question sensitivity. Key questions exhibit typical missing data patterns, and OPA’s assessment of item

missing data and survey drop-offs doesn't present NRB concerns. Most respondents that who start the survey complete the survey.

Section 4: Use of Prior SOFS-A respondents to estimate NRB

The SOFS-A design is a repeated, cross sectional survey that OPA makes no attempt to increase nor decrease the percentage of members that overlap across samples. For this study, OPA matches members across SOFS-A iterations and determines whether prior survey response, or survey answers influence completion of a later survey. OPA conducted a similar analysis on the *2012 Workplace and Gender Relations Survey* (DMDC Report No. 2013–059) and the *2013 Workplace and Equal Opportunity Survey of Active Duty Members* (DMDC Report No. 2014–042) and an important conclusion can be drawn across both studies. These studies showed that prior survey respondents responded to later surveys at higher rates. This implies that even after conditioning on Service, paygrade, race, gender, and other demographic variables, there are a subset of Service members that are extremely cooperative (i.e. survey takers). NRB occurs if these cooperative members have attitudes/opinions that differ from the nonrespondents. Based on prior OPA research, OPA states our first hypothesis:

Hypothesis 1 (Study 4a): Active duty members that respond to a prior SOFS-A (time T1) will be more likely to respond to a later SOFS-A (time T2), even after conditioning on covariates related to survey response.

In addition to determining whether prior survey response predicts later survey response, OPA investigated whether 'reported satisfaction'² on a prior survey predicts later survey response. SOFS-A survey respondents and nonrespondents differ on observable³ characteristics (e.g., paygrade), but they may also differ on unobservable characteristics (e.g., job satisfaction). OPA can properly handle differences on observable characteristics through statistical weighting, but differences in unobservable characteristics can only be accounted for to the extent they are correlated with observable characteristics. Usually, OPA doesn't know whether respondents and nonrespondents differ on unobservable characteristics, however, we used prior survey data on satisfaction as a proxy for current satisfaction. If OPA found that 'satisfied' or 'dissatisfied' members have different response rates after controlling for covariates, survey estimates could be biased. Based on prior OPA research, we state our second hypothesis.

Hypothesis 2 (Study 4b): Active Duty members that report being 'Satisfied' with the military on a prior SOFS-A (T1) will be more likely to respond to a later SOFS-A (T2), even after conditioning on covariates related to survey response.

² Satisfaction is capture by question 25 "Overall, how satisfied are you with the military way of life?" The options are very satisfied (5), satisfied (4), neither satisfied nor dissatisfied (3), dissatisfied (2), and very dissatisfied (1).

³ In this report, OPA defines an observable characteristic as factually known for all members of the population (e.g., Service, paygrade, gender).

Study 4a: Effect of Prior Response on Later Response

Data

OPA selected and matched members from four SOFS-A (1201 SOFS-A, 1202 SOFS-A, 1301 SOFS-A, and 1401 SOFS-A). The sample size across the four survey administrations was 254,374 active duty members. Using the SOFS-A samples and returned data for the four recent surveys, OPA determined the number of members that were sampled in one, two, three, or all four surveys. Table 14 shows that most (93%) of the members surveyed for the SOFS-A program have been sampled in only one of the four SOFS-As.

Table 14.
Unique Unweighted Counts of Members Sampled in Multiple SOFS-A Surveys

Number of SOFS-A(s) Sampled in	Sampled Count
One	237,316 (93.3%)
Two	16,173 (6.4%)
Three	860 (0.3%)
Four	25 (0.01%)
Total	254,374

As OPA planned to assess whether members sampled (or responded) at different rates affected later survey response, it is important to show that these subgroups are not random samples from the active duty military. For instance, Table 15 (page 18) shows the paygrade breakdown for how many times each member has been sampled for the last four SOFS-A. Officers are more likely to be repetitively sampled than enlisted. This occurs because our survey tabulations require separate estimates by paygrade, and this requires similar numbers of officer and enlisted survey respondents. Because there are fewer military officers, OPA sampled officers more frequently. For instance, 39% of members sampled three times are O4–O6 while only 18% are E1–E4.

Table 15.
Unweighted Number of SOFS-A(s) Sampled by Paygrade

Paygrade	Percent Sampled in One SOFS-A	Percent Sampled in Two SOFS-A	Percent Sampled in Three SOFS-A	Percent Sampled in Four SOFS-A
Total Sampled	237,316	16,173	860	25
E1–E4	48	35	18	4
E5–E9	28	22	10	4
W1–W5	2	5	7	4
O1–O3	13	19	26	24
O4–O6	9	18	39	64
Total	100	100	100	100

The disproportionate paygrade distribution is exacerbated by differential response rates across paygrades (e.g., officers respond to RSSC surveys at much higher rates than enlisted). Table 16 shows that O4–O6 make up only 19% of responders for those sampled only once but this percentage increases the more they are sampled (63% for those sampled four times). The opposite effect is shown for E1–E4, whereas they make up 21% for those sampled only once but decrease to 0% as we reach sampled four times.

Table 16.
Unweighted Number of SOFS-A(s) Responded by Paygrade

Paygrade	Percent Sampled=1 and responded =1	Percent Sampled=2 and responded >=1	Percent Sampled=3 and responded >=1	Percent Sampled=4 and responded >=1
Total Responded	52,341	5,936	437	19
E1–E4	21	14	6	0
E5–E9	37	22	8	5
W1–W5	4	7	8	5
O1–O3	19	25	28	26
O4–O6	19	32	49	63
Total	100	100	100	100

Because the subset of sampled members in this analysis differ from the full set of respondents used in the production estimates for the individual surveys, OPA created new analytical weights for this nonresponse bias study. OPA raked the unweighted respondents

(n=16,503) to the 2014 active duty (*1401 SOFS-A*)⁴ population totals using the following six dimensions: 1) Service, 2) paygrade groups, 3) sex, 4) education, 5) race/ethnicity, and 6) marital status for. Table 17 shows the distribution of the final rake-adjusted weights.

Table 17.
Distribution of Final Weights (n=16,503)

Moment	Final Weight
Mean	81.7
Standard Deviation	57.9
Max	282.2
99%	262.9
95%	197.8
90%	167.4
75%	121.9
50%	70.6
25%	29.3
10%	18.4
5%	16.1
1%	12.1
Min	10.1

Analysis

To assess whether members that respond to an early survey (T1) have a higher propensity to respond to a later survey (T2), OPA fit a logistic regression model predicting response at T2.⁵ Table 22 (page 24) shows the predictor variables. The study 4a variable of interest is ‘response at T1, but OPA included eight other demographic variables⁶ based on their known effects on response rates. OPA fit both unweighted and weighted logistic regressions. Table 18 (page 20) shows the predictor variables, including the variable of interest for both Study 4a and 4b, along with eight demographic variables used in both models.

⁴ The population totals are slightly different for each of the four surveys, and OPA could have selected the frame count for any of them; we selected *1401 SOFS-A*.

⁵ The target variable is a binary variable depicting survey response to a later survey (i.e. if the member completed a later survey the observation has the value 1; otherwise 0).

⁶ Members can have different demographic characteristics at T1 and T2; OPA could have selected either set and we chose the demographics at T2.

Table 18.
Predictor Variables for modelling survey response at T2 (for Studies 4a and 4b)

Variable	Variable Label	Categories	Use of Variable
SAMP_T1	Response to prior survey	Nonresponse	Variable of Interest, Study 4a
		Response	
SATOVERR ^a	Q25 Satisfaction on prior survey	Dissatisfied	Variable of Interest, Study 4b
		Neither Satisfied nor dissatisfied	
		Satisfied	
CEDUC	Education Status	No College	Demographics used in study 4a and 4b
		Some College	
		4-year Degree	
		Grad/Prof Degree	
CMARITAL	Marital Status	Married	
		Not Married	
CYOS	Years of Service	1 to 2 years	
		3 to 5 years	
		6 to 9 years	
		10 or more years	
FAM	Family Status	Single with Children	
		Dual Service Spouse	
		All Others	
PAY5	Paygrade	E1–E4	
		E5–E9	
		W1–W5	
		O1–O3	
		O4–O6	
RACE	Race	Non-minority	
		Minority	
SERVICE	Service	Army	
		Navy	
		Marine Corps	
		Air Force	
SEX	Gender	Male	
		Female	

^a: OPA collapsed Q25 Job Satisfaction from the five categories to three categories due to the sparse levels for the logistic model in study 4b.

Results

Both the unweighted and weighted models⁷ show that response at T1 is a statistically significant predictor ($p < 0.05$) of response at T2 (Appendix A). After controlling for covariates known to impact survey response, these results show that prior SOFS-A respondents are 3 to 4 times more likely to respond to a later SOFS-A compared to a prior nonrespondent. Ideally, survey respondents would be missing at random (MAR, Little and Rubin, 2002) within classes, but these results show that subsets of members are ‘survey takers’ and then there necessarily exist other subsets that are ‘survey avoiders’. This casts doubt that SOFS-A data satisfies the MAR assumption, and estimates will be biased if survey avoiders and survey takers have different opinions, which we next assess using job satisfaction. OPA is assessing the bias at T2, and therefore analyzed satisfaction at T2 by several demographics. Although the effect is relatively small, Table 19 shows that survey takers report higher satisfaction than survey avoiders for variables that were significant predictors in the logistic model. For instance, Table 19 shows mean satisfaction for E1–E4 survey avoiders is 3.3 compared with 3.5 for survey takers.

At this point, OPA has shown evidence that prior SOFS-A survey respondents are more likely to take later SOFS-A, and these ‘survey takers’ appear to be more satisfied. Although Table 19 (page 22) shows margins of error and many of these differences would be statistically significant, OPA does not compute statistical significance here because survey takers and survey avoiders are non-random, demographically different subgroups, and all statistical tests should be examined with many covariates, which we assess in the next section

⁷ Because the results were similar, OPA only shows the weighted models in Appendix A and B.

Table 19.
Weighted Mean Satisfaction for Survey 'Avoiders' vs 'Takers' by Paygrade

Variables	Survey Avoiders		Survey Takers	
	Frequency	Mean Satisfaction	Frequency	Mean Satisfaction
Paygrade				
E1–E4	38,240	3.30 ±0.12	14,772	3.48 ±0.17
E5–E9	66,454	3.68 ±0.10	66,656	3.84 ±0.09
W1–W5	3,122	3.95 ±0.14	3,884	4.01 ±0.13
O1–O3	20,042	3.65 ±0.09	22,197	3.77 ±0.08
O4–O6	13,019	4.07 ±0.08	26,393	4.06 ±0.06
Service				
Army	42,601	3.57 ±0.10	35,974	3.81 ±0.09
Navy	32,620	3.45 ±0.13	26,402	3.82 ±0.11
Marine Corps	20,807	3.75 ±0.10	18,994	3.86 ±0.09
Air Force	44,849	3.71 ±0.12	52,531	3.86 ±0.09
Family Status				
Single with Children	9,398	3.77 ±0.22	8,806	3.86 ±0.22
Dual Service Spouse	17,678	3.59 ±0.19	15,747	3.64 ±0.20
All Others	113,800	3.61 ±0.06	109,348	3.86 ±0.05
Education				
4-year Degree	80,319	3.47 ±0.09	50,391	3.76 ±0.09
Grad/Prof Degree	18,897	3.76 ±0.16	27,131	3.79 ±0.14
4-year Degree	25,727	3.75 ±0.10	26,801	3.84 ±0.08
Grad/Prof Degree	15,933	3.94 ±0.11	29,578	4.01 ±0.07
Years of Service				
1 to 2 years	21,078	3.38 ±0.16	9,453	3.57 ±0.18
3 to 5 years	31,893	3.38 ±0.12	20,671	3.57 ±0.13
6 to 9 years	25,474	3.57 ±0.14	20,761	3.62 ±0.14
10 or more years	62,431	3.83 ±0.09	83,017	3.99 ±0.06
Total	140,876	3.61±0.06	133,901	3.84 ±0.05

Note: Survey avoiders did not responded at T1 and did respond at T2 for this analysis. Survey takers responded at T1 and T2 for this analysis.

Study 4b: Prior Satisfaction

This study investigates Hypothesis 2, assessing whether reported satisfaction on a prior survey predicts response to a later survey.

Data

This study uses a subset of the data from study 4a, where the same four surveys are used, but this analysis requires a prior survey completion⁸ at T1 rather than just being in sample at T1. Unlike study 4a where we looked at satisfaction at T2, this study examines how reported satisfaction at T1 affects response propensity at T2. Table 20 shows the 4,391 members that responded to a prior survey and OPA coincidentally sampled at a later iteration. OPA constructed the dataset using the sample member's last completed survey and the next survey the member was sampled in. For example, if a member responded to *1201 SOFS-A* and *1301 SOFS-A* and was sampled for *1401 SOFS-A*, the member is in the dataset only once, and we used satisfaction from *1301 SOFS-A* and response status from *1401 SOFS-A*. Although an imperfect measure of current job satisfaction, this variable construction places the member's proxy satisfaction as temporally close as possible to the survey outcome of interest, response status at T2. Table 20 shows the breakdown of the 4,391 members by survey completion and sample combination.

Table 20.
Members completed survey and later sampled survey

Completed SOFS-A	Sampled in 1202 SOFS-A (n)	Sampled in 1301 SOFS-A (n)	Sampled in 1401 SOFS-A (n)	Total
1201 SOFS-A	625	686	418	1,729
1202 SOFS-A		1,027	614	1,641
1301 SOFS-A			1,021	1,021
Total	625	1,713	2,053	4,391

Analysis

Table 21 (page 24) shows the relationship between satisfaction at T1 and response rates at T2, by paygrade. Generally, members who reported higher satisfaction at T1 responded at higher rates at T2, but paygrade O4–O6 did not follow the pattern. This analysis suggests that although junior enlisted have low response rates there is a subset of junior enlisted members that are cooperative that may have attitudes/opinions that differ (e.g., higher satisfaction) from nonrespondents. For instance, 23% of E1–E4 dissatisfied members responded to the subsequent survey compared with 30% for satisfied members.

⁸ Note: this analysis required both a prior survey completion and answering the job satisfaction question.

Table 21.

Unweighted Response Rates for Later Survey (T2) by Reported Satisfaction Level from Prior Survey (T1)

Paygrade at Time 2	Dissatisfied at T1		Neither Satisfied nor dissatisfied at T1		Satisfied at T1	
	Frequency	Response Rate at T2	Frequency	Response Rate at T2	Frequency	Response Rate at T2
E1-E4	80	23%	118	23%	284	30%
E5-E9	96	38%	140	33%	675	50%
W1-W5	33	42%	32	59%	244	56%
O1-O3	98	43%	138	47%	855	49%
O4-O6	77	64%	168	52%	1,352	58%
Total	384	41%	596	41%	3,410	52%

Similar to study 4a, OPA created analytical weights for this study to make these 4,391 respondents match the active duty population for a set of subgroups. OPA raked sampled members (n=4,391) to population totals using the following six dimensions: 1) Service, 2) paygrade groups, 3) sex, 4) education, 5) race/ethnicity, and 6) marital status for *1401 SOFS-A*. Table 22 shows the final standard moments for the final rake adjusted weights.

Table 22.

Final Weight Moments (n=4,391)

Moment	Final Weight
Mean	307.1
Standard Deviation	388.9
Max	1,675.1
99%	1,675.1
95%	1,240.1
90%	999.6
75%	472.3
50%	99.0
25%	64.5
10%	41.7
5%	41.7
1%	33.6
Min	28.8

Results

OPA continued assessing the effect of prior satisfaction on survey response in a multivariate analysis. OPA fit a logistic model predicting response at T2 using the predictors in Table 18 (i.e., satisfaction at T1 and demographics). Demographic characteristics can vary from T1 to T2; OPA used the demographics from T2. The logistic model shows satisfaction at T1, paygrade, and years of service were significant predictors for survey response at T2. These results show that even after controlling for many demographics, a member's reported satisfaction on SOFS-A at T1 will be more likely to respond to a later SOFS-A. The odds ratio show prior satisfied respondents are approximately 1.6 times likely to respond to a later SOFS-A compared to neither satisfied nor dissatisfied respondents. Appendix B contains the estimated coefficients from the logistic model. Because OPA can only control for observable characteristics (i.e., all the variables except 'prior satisfaction' from Appendix B), OPA estimates are likely positively biased because 'satisfied' members have higher response rates, even after controlling for covariates.

To determine whether OPA's findings are robust to the chosen statistical method (i.e., logistic regression), OPA performed the same analysis using the naïve Bayes probabilistic classification method, which can be applied to data with categorical predictors. Naïve Bayes assumes independence of predictors; rarely is this assumption met in practice. The algorithm often competes very well with other classification techniques even if there are dependencies among variables. Naïve Bayes model is easy to build and runs quickly on large datasets because of its simplicity. Naïve Bayes is based on Bayes theorem which can be used to make predictions based on prior knowledge and current evidence. OPA is interested in the posterior probability of survey response given evidence (i.e., satisfaction, paygrade, Service, gender, family status, education, and years of service). Similar to the logistic model, the Bayes classifier predicts survey response at time T2 conditioning on covariates. OPA used the same independent and dependent variables from the logistic regression. The dataset was split into training and test set using 10 fold cross-validation. The original dataset is randomly divided into 10 equal groups. Nine of the groups are used as training data, and the remaining group is used as the validation data. This process is repeated 10 times with each group used once as the validation data. The 10 estimates are averaged to produce a single estimate. The weighted survey response rate is 53% (i.e., the a-priori probabilities are survey response: 0.53, survey nonresponse: 0.47). The naïve Bayes classifier can achieve an accuracy of 58%. Table 23 (page 26) shows the confusion matrix for the naïve Bayes classification. Overall, the classification method performed reasonably well.

Table 23.
Naive Bayes Confusion Matrix

Actual Class	Predicted Class		Total
	Survey Response=Yes	Survey Response=No	
Survey Response=Yes	1,165	799	1,964
Survey Response=No	1,056	1,371	2,427
Total	2,073	2,221	4,391

The two most current SOFS-A surveys used approximately 18–20 variables in weighting. Further analysis would consider all 18 variables and a ranking matrix for the naïve Bayes model to improve accuracy.

Section 4 Conclusion

OPA finds evidence that prior survey respondents are more likely to respond to later SOFS-A surveys. This violates the MAR assumption that members have the same response propensities within weighting classes. In addition, OPA finds evidence that members that respond to a prior survey and report high job satisfaction are more likely to respond to later SOFS-A surveys. OPA concludes there is a subset of active duty members who are satisfied with the military, and also tend to respond to military surveys (survey takers). This leads to a positive bias in estimates from OPA’s SOFS-A surveys. While OPA has continued to state that NRB is likely the largest source of error in military surveys, this analysis is the most conclusive yet that estimates from SOFS-A, and likely most active duty military surveys, show more positive results than reality.

Overall Conclusions

Potential NRB is expected to exist in any survey when some sample members do not respond to the survey. It is not possible to directly measure the effect of NRB on survey estimates without additional information from the sample members who did not respond. To indirectly estimate the impact of nonresponse on survey results for the 2014 SOFS-A, four separate investigations were conducted.

In Section 1, OPA showed that while sophisticated post-survey adjustment techniques bring weighted estimates into alignment with true population values, and thus help eliminate potential NRB, there is concern that the sample and respondents composition are very different for nearly every observable characteristic.

Further investigation in Section 2 indicated that OPA methods were robust enough to accurately estimate population values that were not directly used in weighting. While OPA noted that AFQT score is significantly higher in the weighted point estimates, the impact on NRB is likely modest. However, OPA will include AFQT percentiles scores for future

weighting to further reduce any NRB associated with academic aptitude not captured by educational level.

Section 3 indicated that members that did not respond to the survey are not dropping off as a function of sensitive questions, but more due to highly burdensome lists of questions. Key questions exhibit typical missing data patterns, and OPA does not find evidence of respondents skipping certain questions due to privacy or other sensitive concerns. OPA recommends discussing questions that cause the large drop-offs with those involved in questionnaire design for future surveys.

Section 4 showed that members who responded to a prior SOFS-A survey were more likely to respond to a later SOFS-A survey, even after controlling for demographics. In addition, members that reported high job satisfaction in a prior survey were more likely to respond to a later survey. Both of these Section 4 studies present NRB concerns, and OPA concludes that SOFS-A estimates are positively biased, meaning that survey estimates are slightly more favorable than reality.

References

- American Association for Public Opinion Research. (2015). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (8th Ed.). AAPOR. Retrieved from https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions2015_8theditionwithchanges_April2015_logo.pdf
- Brick, J., & Bose, J. (2001). Analysis of Potential Nonresponse Bias. Proceedings of the Survey Research Methods Section of the American Statistical Association. August 5–9, 2001.
- DMDC. (2013). *2012 Workplace and Gender Relations Survey of Active Duty Members : Nonresponse Bias Analysis Report* (Report No. 2013–059). Alexandria, VA: Author.
- DMDC. (2015). *2014 Status of Forces Survey of Active Duty Members: Administration, datasets, and codebook* (Report No. 2014–043). Alexandria, VA: Author.
- DMDC. (2015). *2014 Status of Forces Survey of Active Duty Members: Statistical methodology* (Report No. 2014–044). Alexandria, VA: Author.
- DMDC. (2015). *2014 Status of Forces Survey of Active Duty Members: Tabulation volume* (Report No. 2014–042). Alexandria, VA: Author.
- Groves, R.M., & Couper, M.P. (1998). *Nonresponse in household interview survey*. New York: John Wiley & Sons, Inc. doi: 10.1002/9781118490082
- Groves, R.M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A Meta-Analysis. *Public Opinion Quarterly*, 72, 167–189. doi:10.1093/poq/nfn011
- Keeter, S., Miller, C., Kohut, A., Groves, R. M., & Presser, S. (2000). “Consequences of Reducing Nonresponse in a National Telephone Survey,” *Public Opinion Quarterly*, 2, 125–148. [1, 2].
- Little, R.J., & Rubin, D.B. (2002). *Statistical analysis with missing data* (2nd ed.). New York: John Wiley & Sons, Inc. doi: 10.1002/9781119013563

Appendix A.
Logistic Model for 4a Survey Response

Logistic Model for 4a Survey Response (n=16,503)

	Parameter Estimate	Standard Error	Wald Test Statistic	df	P-value	Odds Ratio	95 Percent C.I. for Odds Ratio	
							Lower C.I.	Upper C.I.
Response at T1			567.1597	1	<.0001			
Nonresponse	Reference Category							
Response	1.3468	0.0566	567.1597	1	<.0001	3.845	3.442	4.296
Paygrade			49.1461	4	<.0001			
E1–E4	Reference Category							
E5–E9	0.3994	0.0944	17.8891	1	<.0001	1.491	1.239	1.794
W1–W5	0.6549	0.1335	24.059	1	<.0001	1.925	1.482	2.501
O1–O3	0.7355	0.118	38.8711	1	<.0001	2.087	1.656	2.629
O4–O6	0.5945	0.1496	15.7822	1	<.0001	1.812	1.351	2.43
Service			24.7701	3	<.0001			
Army	Reference Category							
Navy	0.1665	0.0695	5.7469	1	0.0165	1.181	1.031	1.353
Marine Corps	0.0922	0.0652	2.0006	1	0.1572	1.097	0.965	1.246
Air Force	0.3729	0.0766	23.6744	1	<.0001	1.452	1.249	1.687
Family Status			7.6322	2	0.022			
Single with Children	Reference Category							
Dual Service Spouse	0.3842	0.1523	6.3647	1	0.0116	1.468	1.089	1.979
All Others	0.3191	0.1205	7.0165	1	0.0081	1.376	1.087	1.742
Race/Ethnicity			1.3266	1	0.2494			
Non-minority	0.0666	0.0578	1.3266	1	0.2494	1.069	0.954	1.197
Minority	Reference Category							
Gender			0.2487	1	0.618			
Males	0.0398	0.0798	0.2487	1	0.618	1.041	0.89	1.217
Females	Reference Category							
Education			8.6279	3	0.0347			
No College	Reference Category							
Some College	0.1778	0.0913	3.7943	1	0.0514	1.195	0.999	1.429
4-year Degree	0.2714	0.1055	6.6232	1	0.0101	1.312	1.067	1.613
Grad/Prof Degree	0.2804	0.1283	4.773	1	0.0289	1.324	1.029	1.702
Marital Status			0.2121	1	0.6452			
Not Married	Reference Category							
Married	0.0315	0.0685	0.2121	1	0.6452	1.032	0.902	1.18
Years of Service			72.8516	3	<.0001			
1 to 2 years	Reference Category							
3 to 5 years	0.0847	0.0886	0.9141	1	0.339	1.088	0.915	1.295
6 to 9 years	0.249	0.1102	5.1084	1	0.0238	1.283	1.034	1.592
10 or more years	0.7829	0.1138	47.3688	1	<.0001	2.188	1.751	2.734

Appendix B.
Logistic Model for 4b Survey Response

Logistic Model for 4b Survey Response(n=4,391)

	Parameter Estimate	Standard Error	Wald Test Statistic	df	P-value	Odds Ratio	95 Percent C.I. for Odds Ratio	
							Lower C.I.	Upper C.I.
Satisfaction			12.5864	2	0.0018			
Dissatisfied	0.0999	0.2033	0.2416	1	0.6231	1.105	0.742	1.646
Neither Satisfied nor dissatisfied	Reference Category							
Satisfied	0.4625	0.1467	9.9437	1	0.0016	1.588	1.191	2.117
Paygrade			19.5789	4	0.0006			
E1-49	Reference Category							
E5-E9	0.6003	0.1971	9.2782	1	0.0023	1.823	1.239	2.682
W1-W5	0.6303	0.2468	6.5227	1	0.0107	1.878	1.158	3.046
O1-O3	0.8164	0.1914	18.2005	1	<.0001	2.262	1.555	3.292
O4-O6	0.7503	0.2522	8.8537	1	0.0029	2.118	1.292	3.471
Service			4.9231	3	0.1775			
Army	0.2556	0.1435	3.1716	1	0.0749	1.291	0.975	1.711
Navy	Reference Category							
Marine Corps	0.0671	0.1438	0.2181	1	0.6405	1.069	0.807	1.418
Air Force	0.2758	0.154	3.2068	1	0.0733	1.318	0.974	1.782
Family Status			1.6969	2	0.4281			
Single with Children	Reference Category							
Dual Service Spouse	0.3106	0.2853	1.1855	1	0.2762	1.364	0.78	2.386
All Others	0.2752	0.2137	1.6576	1	0.1979	1.317	0.866	2.002
Race/Ethnicity			2.9596	1	0.0854			
Non-minority	0.1903	0.1106	2.9596	1	0.0854	1.21	0.974	1.502
Minority	Reference Category							
Gender			0.3112	1	0.5769			
Male	Reference Category							
Female	0.0834	0.1494	0.3112	1	0.5769	1.087	0.811	1.457
Education			4.6208	3	0.2018			
No College	Reference Category							
Some College	0.2499	0.1455	2.9509	1	0.0858	1.284	0.965	1.708
4-year Degree	0.0797	0.1565	0.2595	1	0.6105	1.083	0.797	1.472
Grad/Prof Degree	0.2595	0.1808	2.0586	1	0.1514	1.296	0.909	1.848
Marital Status			0.003	1	0.9564			
Not Married	0.00732	0.1339	0.003	1	0.9564	1.007	0.775	1.31
Married	Reference Category							
Years of Service			12.9973	3	0.0046			
1 to 2 years	0.2237	0.1984	1.2717	1	0.2595	1.251	0.848	1.845
3 to 5 years	Reference Category							
6 to 9 years	0.138	0.1929	0.5117	1	0.4744	1.148	0.787	1.675
10 or more years	0.5704	0.1848	9.5278	1	0.002	1.769	1.231	2.541

REPORT DOCUMENTATION PAGE

*Form Approved
OMB No. 0704-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY)		2. REPORT TYPE		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (Include area code)

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATES COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report, e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5d. PROJECT NUMBER. Enter all project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

**Office of People Analytics
(OPA)**

