Defense Research, Surveys, and Statistics Center (RSSC)

2014 Post-Election Voting Survey of the Active Duty Military

Nonresponse Bias Analysis

DMDC Report No. 2015-032
June 2015

2014 POST-ELECTION VOTING SURVEY OF THE ACTIVE DUTY MILITARY: NONRESPONSE BIAS ANALYSIS

Dr. Fawzi Al Nassir, David McGrath, and Jeff McLeod
(DMDC/RSSC)

Defense Manpower Data Center (DMDC)
Defense Research, Surveys, and Statistics Center (RSSC)
4800 Mark Center Drive, Suite 04E25-01, Alexandria, VA 22350-4000

Table of Contents

Appendixes

List of Tables

2014 POST-ELECTION VOTING SURVEY OF THE ACTIVE DUTY MILITARY NONRESPONSE BIAS ANALYSIS

Introduction

In 2006 the Office of Management and Budget recommended that a nonresponse bias analysis be completed for any government survey with a response rate below 80%. The purpose of nonresponse bias analysis is to determine whether survey estimates are biased due to nonresponse of some sample members. To make this determination, a statistical analysis is conducted to conclude 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 Post-Election Voting Survey of the Active Duty Military* (*2014 PEV5*) was approximately 16% based on AAPOR RR3, a nonresponse bias analysis was initiated.

The *2014 PEV5* nonresponse bias study was conducted in two phases. Phase 1 identified influential demographic variables, defined as variables that significantly predict whether a sample member responds to the survey. Identification of influential variables was based on both experience with Military surveys and responses to the current survey. Phase 2 examined whether the influential variables identified in Phase 1 show significant differences in predicting response patterns to some important survey questions. If a demographic variable exhibits statistical significance in both phases, then potential nonresponse bias is indicated and further analysis is performed.

The terms characteristics, variables, and factors are used interchangeably throughout this report. The *2014 PEV5* return dataset was used for the study. The dataset consists of active duty members in Army, Navy, Marines Corps, Air Force, and Coast Guard. The analysis was performed on the *2014 PEV5* complete respondents and nonrespondents consisting of 94,699 active duty members, with 16,873 18 to 29 year olds in the phone group and 5,955 18 to 29 year olds in the short web group. Statistical analyses were performed using SAS® and SPSS®. For information about the sampling frame, sampling design and weighting procedure refer to the *2014 Post-Election Voting Survey of the Active Duty Military: Statistical Methodology Report* (DMDC, 2015b). For information about the survey design, survey instruments and administration refer to the *2014 Post-Election Voting Survey of the Active Duty Military: Administration, Datasets, and Codebook* (DMDC, 2015a). For tabulation of responses refer to the *2014 Post-Election Voting Survey of the Active Duty Military: Tabulation of Responses* (DMDC, 2015c).

Analysis

Studies of nonresponse bias can be accomplished either by conducting another survey of nonrespondents or by using the original survey respondents. The latter is the approach used in this study. Two survey outcomes are critical in assessing nonresponse bias: response rate and the expected difference between respondents and nonrespondents on the 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, and Presser, 2000). Where bias is found, adjusting survey weights for nonresponse and poststratification can significantly reduce that bias.

The *2014 PEV5* nonresponse bias was assessed based on response to the survey and response to specific survey questions of interest. Demographic characteristics that show significant differences for both the response to the survey and to the questions of interest are considered characteristics with potential nonresponse bias. As noted above, the study was performed in two phases: Phase 1—Modeling Survey Response and Phase 2—Modeling Response to Critical Survey Questions.

Phase 1 - Modeling Survey Response

Respondents and nonrespondents are characterized based on a set of demographic variables. These characteristics were identified based on the survey response and based on DMDC's experience in Military surveys. Experience shows that variables such as the member's age, paygrade, and Service are critical in predicting Military survey response. Ten demographic variables based on DMDC's *June 2014 Active Duty Edit Master File* (*ADMF*) were identified, statistically tested, and determined to have significant predictive power on the *2014 PEV5* survey response. These variables are member's age group, paygrade group, Service, duty location, sex, marital status, occupation code, education, deployment since September 11, 2001, and race/ ethnicity. The deployment variable was based on the *June 2014 Contingency Tracking System* (*CTS*) for deployments in Iraq and Afghanistan (OIF and OEF operations). Table 1 shows the ten variables along with their corresponding levels.

Table 1.
Independent Demographic Variables

Variable/Characteristic	Categories
Age Group	18 - 24 Years Old
	25 - 29 Years Old
	30 - 34 Years Old
	35 - 44 Years Old
	45 + Years Old
Paygrade Group	E1-E5
	E6-E9
	W1-W5
	O1-O3
	O4-O6
Service	Army
	Navy
	Marine Corps
	Air Force
	Coast Guard
Duty Location	US, Unknown
	Overseas
Sex	Male
	Female
Marital Status	Not Married
	Married
Occupation Code	Combat
	Combat Support
Education Level	No College
	Some College
	4-year Degree or Higher
Deployment	Never Deployed since 9/11/2001
	One or More Deployments since 9/11/2001
Race/Ethnicity	White
	Black
	Hispanic
	Other

Methodology: Analysis of Respondents and Nonrespondents

All 10 characteristics were examined individually using Logistic Regression. The dependent variable in the logistic model is a binary variable representing the response to the survey where the variable equals 0 for Nonresponse and 1 for Response. The predictors (or independent variables) are those demographic variables in Table 1. Ten separate logistic regression analyses were performed, one for each variable in Table 1. In other words, the

response to the survey was modeled using each of the 10 demographic variables one at a time. If the variable shows significant impact on predicting response to the survey then it is flagged as a potential driver of nonresponse bias and further analysis was performed.

Next, all variables with individual predictive power of survey response were tested simultaneously via a full logistic regression model. The full model is a main effect logistic model that includes all the variables exhibiting significant differences when tested individually. The purpose of testing the full model is to measure the effect of each variable controlling for the others (i.e., measuring the effect of one characteristic taking the other characteristics into consideration).

Logistic regression requires that one of the categories (levels) of the independent variable is set to be a reference category; typically either the first or the last level. We modeled using the first category as a reference. All other categories of the variable were compared to the reference category and the model parameters and odds ratios were derived and interpreted accordingly. If the characteristic significantly predicts response to the survey, the odds ratios are examined to determine the source of significance. To illustrate, modeling of the age variable is given as example. The other variables are similarly modeled and interpreted.

Modeling the age variable: The age group variable consists of five categories, 18-24, 25-29, 30-34, 35-44, and 45 years and older. The reference category is age group 18-24. Every other age category is compared to the reference category via the odds ratio. Table 2 shows the frequencies of each age category along with the number of respondents and nonrespondents and the reference assignment. Notice that zeros were assigned to the reference category (18-24). The first comparison to the reference will be for age 25-29, then for age 30-34, and so on.

Table 2.
Categorical Variable Coding

		Frequency	Nonresponse	Response	Parameter Coding			
					(1)	(2)	(3)	(4)
Age Group	18 - 24	25,708	22,129	3,579	0	0	0	0
	25 - 29	17,559	14,574	2,985	1	0	0	0
	30 - 34	7,587	6,127	1,460	0	1	0	0
	35 - 44	10,620	7,580	3,040	0	0	1	0
	45 +	3,929	2,373	1,556	0	0	0	1

Next, a Likelihood Ratio Chi-Square test was performed resulting in Chi-Square value for the model of 1,975 and p-value < .05 indicating that age group significantly predicts response to the survey (P-Value column, Table 3). Table 3 contains the significance testing results for the age variable and its categories. Notice that the reference category 18-24 years old is not displayed since the odds ratios of the other age categories are compared to the reference group and the odds ratio of the reference category to itself is 1.

Table 3.
Significance Testing of Age

	P-Value	Odds Ratio	95% C.I. for Odds Ratio	
			Lower	Upper
Age Group	0.000*			
25-29	0.000*	1.266	1.201	1.335
30-34	0.000*	1.473	1.378	1.576
35-44	0.000*	2.480	2.347	2.620
45 +	0.000*	4.054	3.769	4.361
Constant	0.000*	0.081		

* indicates statistical significance

Since age is a significant predictor of survey response, we next examined the odds ratio of each age group to determine the source of significant differences. The odds ratio (Odds Ratio column, Table 3) was calculated with respect to the reference; it is simply the odds of one category divided by the odds of the reference category. If the odds of the two groups are the same, then one would expect the odds ratio to be close to 1.0. An odds ratio greater than 1.0 indicates the comparison age group is more likely to respond to the survey than the reference group, while an odds ratio less than 1.0 implies the opposite. Consider the age group 25-29. The corresponding odds ratio is 1.266. This means that members of age 25-29 are about 1.3 times as likely to respond to the survey as members 18-24 years old, and their response is statistically different than members of age 18-24 (p-value <0.05). In other words, members of age group 25-29 respond at approximately a 30% higher rate than those in the reference group. Similarly, members 45 years old and older are about four times more likely to respond to the survey (odds ratio = 4.054) or equivalently, they respond at approximately 300% higher rate than the reference group and their response is significantly different than members age 18-24 (p-value <0.05). Notice that the odds ratio increases as the member age group increases suggesting that, in general, the older the member the higher the likelihood of responding to the survey. Moreover, not only is the overall age variable significant, but all age categories are statistically significant as well. The 95% Confidence Interval of the odds ratios is also given for further interpretation.

The other variables were similarly modeled and interpreted. All ten variables showed significant predictive power of survey response. Details of the analyses are given in Appendix A.

Since all ten characteristics differed significantly between the two groups (respondents and nonrespondents), all characteristics were then examined simultaneously to measure the impact of one variable in predicting response to the survey while controlling for the other nine variables. Logistic regression was again employed. As in the first step, the dependent variable represents the response to the survey and the independent variables are the demographic variables listed in Table 1. Likelihood Ratio Chi-Square tests with p-values < 0.05 (P-Value column, Table 4) indicate significant differences in response rates. The results of significance for each variable in the model and its corresponding categories are shown in Table 4. Notice that the reference category is not displayed in the table for the reason mentioned earlier. Column 1 shows the independent variables and their categories, columns 2 to 5 consist of the parameter

estimates (B), the standard errors of the estimate (S.E.), the Wald tests, and the degrees of freedom (df) associated with the variables and categories respectively.

Table 4.
Full Logistic Model with Ten Independent Variables for Phase 1

Variable	B	S.E.	Wald	df	P-Value	Odds Ratio	95% C.I. for Odds Ratio	
							Lower	Upper
Age Group			469.514	4	.000			
25-29	.102	.031	10.789	1	.001	1.107	1.042	1.176
30-34	.117	.044	7.013	1	.008	1.124	1.031	1.225
35-44	.614	.047	168.286	1	.000	1.848	1.684	2.028
45 +	1.054	.056	356.325	1	.000	2.868	2.571	3.199
Paygrade Group			156.008	4	.000			
E6-E9	.300	.036	68.315	1	.000	1.350	1.257	1.450
W1-W5	.359	.074	23.342	1	.000	1.432	1.238	1.656
O1-O3	.483	.045	115.172	1	.000	1.621	1.484	1.771
O4-O6	.272	.057	22.980	1	.000	1.313	1.175	1.468
Service			455.537	4	.000			
Navy	.194	.029	45.459	1	.000	1.214	1.148	1.285
Marine Corps	.199	.034	34.136	1	.000	1.220	1.141	1.304
Air Force	.463	.028	280.227	1	.000	1.588	1.504	1.676
Coast Guards	.956	.059	264.380	1	.000	2.601	2.318	2.919
Overseas	-.140	.024	34.679	1	.000	.869	.830	.911
Females	-.020	.030	.453	1	.501	.980	.925	1.039
Married	-.121	.024	25.555	1	.000	.886	.845	.929
Combat Support	.266	.028	90.656	1	.000	1.305	1.236	1.379
Education			90.620	2	.000			
Some College	.217	.033	42.842	1	.000	1.242	1.164	1.326
4 Year Degree or More	.318	.037	72.368	1	.000	1.375	1.278	1.480
Deployed 1 or more	-.016	.025	.399	1	.527	.984	.936	1.034
Race/Ethnicity			24.124	3	.000			
Black	-.073	.030	5.950	1	.015	.929	.876	.986
Hispanic	.031	.033	.855	1	.355	1.031	.966	1.101
Other	.130	.036	13.429	1	.000	1.139	1.062	1.221
Constant	-2.219	.036	3,833.899	1	.000	.109		

* indicates statistical significance

Most characteristics are still significant in the full model except for sex and deployment. The Hispanic category in the race variable was not significant; when the race variable was collapsed to two categories representing White and Other Minority, the race was not significance (p-value = .542, which is >.05). Accordingly, the race variable was dropped as well and new analyses were performed using the new full model of seven independent variables.

The Likelihood Ratio Chi-Square for this model was 3,100.797 and the corresponding p-value was <.05 suggesting that the model fits the data (i.e., the seven independent variables used in the model are significant in predicting survey response). Results of significance of the new full model with seven independent variables and their categories are given in Table 5. More details of the analysis are shown in Appendix B.

Table 5.
Final Full Logistic Model with Seven Independent Variables

	B	S.E.	Wald	df	P-Value	Odds Ratio	95% C.I. for Odds Ratio	
							Lower	Upper
Age Group			466.805	4	.000			
25-29	.095	.030	10.042	1	.002	1.099	1.037	1.166
30-34	.106	.043	6.195	1	.013	1.112	1.023	1.209
35-44	.602	.046	171.821	1	.000	1.826	1.669	1.998
45 +	1.039	.055	360.355	1	.000	2.826	2.539	3.146
Paygrade Group			155.926	4	.000			
E6-E9	.295	.036	66.983	1	.000	1.343	1.251	1.441
W1-W5	.355	.074	22.948	1	.000	1.426	1.233	1.649
O1-O3	.485	.045	117.358	1	.000	1.624	1.488	1.773
O4-O6	.270	.056	23.099	1	.000	1.310	1.174	1.463
Service			504.137	4	.000			
Navy	.216	.028	57.670	1	.000	1.241	1.174	1.312
Marine Corps	.210	.034	38.809	1	.000	1.233	1.154	1.317
Air Force	.470	.027	296.987	1	.000	1.600	1.517	1.688
Coast Guards	.981	.056	301.811	1	.000	2.667	2.387	2.979
Overseas	-.138	.024	33.825	1	.000	.871	.831	.912
Married	-.117	.024	24.472	1	.000	.890	.849	.932
Combat Support	.263	.028	90.335	1	.000	1.300	1.232	1.373
Education			91.891	2	.000			
Some College	.216	.033	42.650	1	.000	1.241	1.163	1.324
4 Year Degree or more	.320	.037	73.770	1	.000	1.378	1.280	1.482
Constant	-2.228	.034	4,376.910	1	.000	.108		

* indicates statistical significance

Wald's test and the corresponding p-values for all independent variables at all levels are significant (p-value < 0.05) suggesting that these variables exhibit significant predictive power of the *2014 PEV5* survey response.

The odds ratios (Odds Ratio column, Table 5) for each variable in the full model are derived taking the impact of the other variables in the model into consideration (i.e., controlling for the other variables). To illustrate the practical importance of the difference between results

from this model, where all the variables are examined simultaneously, and the results in the previous step, where each variable is examined independently, consider the odds ratio of the age variable in both cases. The odds ratio for age 45 and over without taking into account the effects of the other variables is 4.054 (Table 3). This indicates that 45 year olds and over are about four times as likely as 18-24 year olds (the reference group) to respond to the survey. However, in the final full model the odds ratio for 45 years old and older when the effect of the other demographic characteristics is taken into considerations is noticeably lower, it is 2.826 (Table 5). This indicates that if the impact of all the demographic variables is considered at once, members of age 45 years and older are about 2.8 times as likely as 18-24 year olds to respond to the survey.

Table 5 shows that even when the effects of the other variables are accounted for, the odds ratios of the demographic characteristics categories are all still statistically significant, and most odds ratios are substantially different from 1.0. For example, older members tend to respond at a significantly higher rate than young members; members age 45 and older are nearly three times or 182% more likely to respond than members age 18-24. Similarly, members with some college education are about 1.241 times or 24% more likely to respond than members who have never been to college; and those with 4 years degree and higher are 1.378 times or 38% more likely to respond.

Additionally, in accordance with industry practice, the *2014 PEV5* response data were weighted for nonresponse prior to analysis for the standard survey products. Studies of nonresponse bias support that adjusting survey weights for nonresponse and poststratifying to known control totals can significantly reduce that bias (e.g., Brick and Bose, 2001). Weighting for the *2014 PEV5* included two levels of nonresponse adjustments followed by poststratification to known frame variables. All seven variables that showed predictive power of nonresponse in this study; age, paygrade, Service, education, duty location, marital status, and occupation code were used in the weighing process to derive nonresponse adjustments. Other variables including race and sex were also included in the weighting process.

Having identified a set of variables impacting the response to the survey, next we measured the impact of these variables on survey questions. Variables exhibiting significant predictive power on survey response and on response to questions are considered drivers of potential nonresponse bias.

Phase 2 - Modeling Response to Critical Survey Questions

In this phase, the impact of the seven variables identified in Phase 1 to have significant predictive power on survey response were investigated in terms of their impact on the response to important survey questions. If a variable significantly predicted both the response to the survey (Phase 1) and the response to the survey question then we concluded that the estimates produced by these questions are at risk of nonresponse bias.

Of the 63 questions in the *2014 PEV5*, ten were chosen for the Phase 2 analysis. These questions were chosen based on their central importance to survey goals. Ongoing discussions with Federal Voter Assistance Program (FVAP) staff and requests for further analyses indicated

that these questions are representative of the most critical issues in the survey. The questions represent core understanding of five content areas:

- 2014 Voter Registration – Voter registration status (Q11).
- Absentee Ballots – Absentee ballot requests; use, receipt, completion, and return of Federal Post Card Application (FPCA); and receipt, completion, and return of regular absentee ballot (Q17, Q20, Q24, Q27).
- 2014 Election – Voting history, voting interest in 2014 election, and voting behavior in 2014 election (Q32, Q34).
- Voting Assistance – Need for and resources used for information or assistance in preparation for 2014 election (Q44, Q45).
- Federal Write-In Absentee Ballot (FWAB) – Use and awareness of FWAB (Q38).

The ten core and skip pattern questions under study are shown in Table 6.

Logistic regression models were used to assess the impact of the demographic variables on each question. The response to the question represents the dependent variable and the seven variables identified in Phase 1 are the predictors (independent variables).

Logistic regression modeling requires the dependent variable to be binary; however, some of the dependent variables (response to questions) of interest are not binary. Accordingly, all survey questions in Table 6 that are not binary were dichotomized by collapsing groups of similar categories together to form a binary variable with values of 0 or 1. The response with the higher number (1 in this case) is the modeled category. The original and the collapsed levels are shown in Table 6. Table 6 consists of five columns; the variable name in the database, question number in the survey, question text, response levels and the collapsed levels/categories with the assigned value of 0 or 1. To clarify, consider the variable VOTED. VOTED has seven levels, levels 1 to 5 were collapsed to form the category "Definitely Voted" taking on value of 1, levels 6 and 7 formed the other category "Definitely did not vote/not sure" taking on value of 0, hence the new constructed dependent variable is a binary variable.

Table 6.
Selected 10 Questions for Phase 2

Variable	Q#	Question Text	Response Levels	Collapsed for Modeling
REGVOTER	11	Were you registered to vote in the United States?	1 Yes	No Collapsing or recoding
			0 No	
VOTED	34	In the election held on November 4, 2014, did you definitely vote in person on election day, definitely complete an absentee ballot by mail, e-mail, fax, or online on or before November 4, 2014, definitely not vote, or are you not completely sure whether you voted in that election? Mark one.	1 Definitely voted in person	1 Definitely voted
			2 Definitely voted by mail	
			3 Definitely voted by e-mail	
			4 Definitely voted at an online website	
			5 Definitely voted by fax	
			6 Definitely did not vote	0 Definitely did not vote/Not sure
			7 Not Sure	
REQABSBAL	17	Did you request an absentee ballot?	1 Yes	1 Yes
			2 No, but I automatically received an absentee ballot from a local election official.	0 No
			3 No, and I never received an absentee ballot.	
			4 No, I did not need an absentee ballot.	
REQFPCA	20	Did you use the Federal Post Card Application (FPCA) to request your absentee ballot or did you use another method?	1 Yes, I used an FPCA to request an absentee ballot.	1 Yes
			2 No, I used a State or local form to request an absentee ballot.	0 No
			3 No, I used a non-government website (e.g., Rock the Vote [RTV], Overseas Vote Foundation [OVF]) to request an absentee ballot.	
			4 No, I used another method.	
RECABS	24	Did you receive your regular absentee ballot?	1 Yes	No Collapsing or recoding
			0 No	
RETABS	27	Did you complete and return your regular absentee ballot?	1 Yes	No Collapsing or recoding
			0 No	

Table 6. (continued)

Variable	Q#	Question Text	Response Levels	Collapsed for Modeling
FWABUSE	38	Did you use the Federal Write-In Absentee Ballot (FWAB)?	1 Yes	1 Yes
			2 No, but I was aware of it	0 No
			3 No, and I was not aware of it	
INTEREST	32	How interested or uninterested were you in the election?	5 Very interested	1 Interested
			4 Somewhat interested	
			3 Neither interested nor uninterested	0 Uninterested or neither
			2 Somewhat uninterested	
			1 Very uninterested	
ASSIST	44	In preparation for the November 4, 2014, election, did you need any information or assistance (e.g., information on deadlines, how to request an absentee ballot)?	1 Yes	No Collapsing or recoding
			0 No	
SEEKA	45	Did you seek voting information or assistance from any of the following? FVAP	1 Yes	1 Yes
			2 No, but I was aware of it	0 No
			3 No, and I was not aware of it	
			3 No, and I was not aware of it	

For each of the variables listed in Table 6, logistic regression analysis was performed to determine whether the variables that predicted response to the survey (Phase 1) also predict the response pattern to the questions. Each model included all seven variables that were significant in predicting survey response from Phase 1. If the overall model fits the data (i.e., if the Likelihood Ratio Chi-Square test is significant [p-value <.05]) then that means that at least one of the characteristics is significant in predicting response to the questions. Accordingly, further investigation of the odds ratio is performed to determine which characteristics are significant (i.e., to determine which variables are potential drivers of nonresponse bias).

Table 7 consists of the significance testing results along with the most influential variables for each model, as well as the total number of respondents and number of respondents by category. All 10 models were significant with p-value < 0.05

Significant fit of the model to the data indicates that at least one of the variables in the model significantly predicts responses to the question. For example, for the model with the variable VOTED as dependent variable, the overall model with seven predictors from Phase 1 fits the data, meaning at least one of the seven variables is significant in predicting the response pattern to the voting question. For this model, four of the seven variables were significant (age, paygrade, duty location, and education). Since these four variables are significant in both phases then we say that the estimates derived from this question exhibit potential nonresponse bias.

Different sets of variables appear to have predictive power in different models (i.e., for different survey questions). Some variables appeared to be significant more than others:

- Education appeared in nine of the 10 models.
- Age Group, Pay Grade, and Duty Location appeared in eight models.
- Service appeared in four models.
- Occupation in two models.
- Marital status was not significant in any model, suggesting that marital status has negligible impact on nonresponse bias in this study.

Table 7.
Logistic Models for the 10 Questions

Variable	Question Text	Total Number of Respondents	Likelihood Ratio Chi-Square Test	P-Value	Variables with Significant Predictive Value
REGVOTER	Were you registered to vote in the United States?	12,551 Definitely Voted = 3,443 Definitely not Voted = 9,108	1,005.02	0.000*	Age Group Paygrade Group Duty location Education Service Combat
VOTED	In the election held on November 4, 2014, did you definitely vote in person on election day, definitely complete an absentee ballot by mail, e-mail, fax, or online on or before November 4, 2014, definitely not vote, or are you not completely sure whether you voted in that election? Mark one.	12,562 Yes = 3,045 No = 9,517	1,372.89	0.000*	Age Group Paygrade Group Duty location Education
REQABSBAL	Did you request an absentee ballot?	12,463 Yes = 2,558 No = 9,905	892.07	0.000*	Age Group Paygrade Group Education
REQFPCA	Did you use the Federal Post Card Application (FPCA) to request your absentee ballot or did you use another method?	2,538 Yes = 844 No = 1,694	50.43	0.000*	Duty location Combat Education
RECABS	Did you receive your regular absentee ballot?	2,548 Yes = 1,978 No = 570	54.82	0.000*	Paygrade Group Duty location
RETABS	Did you complete and return your regular absentee ballot?	3,428 Yes = 2,092 No = 1,336	293.42	0.000*	Age Group Paygrade Group Education
FWABUSE	Did you use the Federal Write-In Absentee Ballot (FWAB)?	12,502 Yes = 420 No = 12,082	147.23	0.000*	Age Group Paygrade Group Service Duty location Education

Table 7. (continued)

Variable	Question Text	Total Number of Respondents	Likelihood Ratio Chi-Square Test	P-Value	Variables with Significant Predictive Value
INTEREST	How interested or uninterested were you in the election?	12,569 Yes = 6,365 No = 6,204	1,262.94	0.000*	Age Group Paygrade Group Duty location Education
ASSIST	In preparation for the November 4, 2014, election, did you need any information or assistance (e.g., information on deadlines, how to request an absentee ballot)?	12,399 Yes = 2,432 No = 9,967	132.81	0.000*	Age Group Paygrade Group Service Duty location Education
SEEKA	Did you seek voting information or assistance from any of the following? FVAP	2,347 Yes = 471 No = 1,876	160.65	0.000*	Age Group Duty location Service Education

* indicates statistical significance

In conclusion, six of the seven predictors of survey response in Phase 1 are also significant in one way or another in Phase 2, suggesting that the six characteristics (age, paygrade, Service, duty location, education, and occupation (Combat, Combat Support) are indicative of potential nonresponse bias. However, as mentioned above, the *2014 PEV5* response data was weighted using all six variables in addition to marital status. Because the data were already adjusted to account for these variables, most of the nonresponse bias has likely been removed.

One of the benefits of nonresponse bias studies is to identify issues researchers should think of in future studies and to improve future analyses. In previous *PEV5* surveys, the variable "education" was not used in the weighting process. In *2014 PEV5* the variable "education" was used in the weighting process based on the recommendation of the *2012 PEV5* nonresponse bias study conducted by RSSC as "education" was identified as a powerful variable in predicting potential nonresponse bias in *PEV5* surveys. As shown, this variable continues to be powerful and using it in weighting further reduces the likelihood of nonresponse bias in the *PEV5* survey.

Conclusion

Potential nonresponse bias 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 nonresponse bias 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 PEV5*, a two-phase study was conducted.

In Phase 1, ten demographic variables were identified as having potential nonresponse bias impact on survey results. DMDC's experience with military surveys coupled with the responses of the *2014 PEV5* suggested these variables are critical in predicting survey response. Analyses in Phase 1 supported that seven of the ten variables did indeed show statistically significant and practically substantial differences in likelihood to respond to the survey.

Further investigation in Phase 2 indicated that six of these seven variables are significant predictors of response pattern to survey questions. Statistical significance of the six variables in both phases of the analysis indicated potential nonresponse bias of the *2012 PEV5* survey estimates.

However, all six characteristics were accounted for in the *2014 PEV5* response data weighting process including nonresponse and poststratification adjustments, so contribution of these variables to nonresponse bias is of little concern since such adjustments can significantly reduce that bias (e.g., Brick and Bose, 2001).

RSSC decision to use the variable "education" in the weighting process based on *2012 PEV5* nonresponse bias study was correct as the analysis showed that education is one of the most important factors in predicting potential nonresponse bias.

Finally, the impact of sophisticated weighting in reducing nonresponse bias is also related to the level of response. For the weighting process to be more effective in reducing nonresponse bias, steps should be taken to increase the response rate on future *PEV5* surveys, particularly among the younger sample members and those with less than a college degree.

References

Brick, J., and Bose, J. (2001). Analysis of Potential Nonresponse Bias. Proceedings of the Survey Research Methods Section of the American Statistical Association.

DMDC. (2015a). *2014 Post-Election Voting Survey of the Active Duty Military: Administration, datasets, and codebook* (Report No. 2015-009). Alexandria, VA: Author.

DMDC. (2015b). *2014 Post-Election Voting Survey of the Active Duty Military: Statistical methodology report* (Report No. 2015-008). Alexandria, VA: Author.

DMDC. (2015c). *2014 Post-Election Voting Survey of the Active Duty Military: Tabulation of responses*. (Report No. 2015-007). Alexandria, VA.

Keeter, S., Miller, C., Kohut, A., Groves, R. M., and Presser, S. (2000). "Consequences of Reducing Nonresponse in a National Telephone Survey," Public Opinion Quarterly, 2, 125–148. [1, 2].

Levy, Paul S., and Lemeshow Stanley. (1999). *Sampling of Populations: Methods and applications*. New York: J. Wiley and Sons.

Appendix A.
Simple Logistic Regression Results for Phase 1

Simple Logistic Regression Phase 1

Table A-1.
Variables in the Equation: Age Group

	B	S.E.	Wald	df	P-Value	Odds Ratio	95% C.I. for Odds Ratio	
							Lower	Upper
Step 1[a]: **Age Group**			2,043.836	4	.000			
25-29	.236	.027	76.590	1	.000	1.266	1.201	1.335
30-34	.388	.034	128.063	1	.000	1.473	1.378	1.576
35-44	.908	.028	1,049.989	1	.000	2.480	2.347	2.620
45 +	1.400	.037	1,410.948	1	.000	4.054	3.769	4.361
Constant	-1.822	.018	10,224.903	1	.000	.162		

[a]Variable(s) entered on step 1: Age Group.

Table A-2.
Variables in the Equation: Paygrade

	B	S.E.	Wald	df	P-Value	Odds Ratio	95% C.I. for Odds Ratio	
							Lower	**Upper**
Step 1[a]: **Paygrade**			1,743.759	4	.000			
E6-E9	.690	.024	810.468	1	.000	1.993	1.901	2.090
W1-W5	.829	.065	164.176	1	.000	2.291	2.018	2.601
O1-O3	.777	.034	514.825	1	.000	2.175	2.034	2.326
O4-O6	1.129	.035	1,054.241	1	.000	3.092	2.888	3.310
Constant	-1.776	.014	16,005.233	1	.000	.169		

[a]Variable(s) entered on step 1: Paygrade.

Table A-3.
Variables in the Equation: Service

	B	S.E.	Wald	df	P-Value	Odds Ratio	95% C.I. for Odds Ratio	
							Lower	Upper
Step 1[a]: **Service**			621.933	4	.000			
Navy	.121	.027	20.087	1	.000	1.128	1.070	1.190
Marine Corps	-.095	.032	8.709	1	.003	.910	.854	.969
Air Force	.426	.026	269.509	1	.000	1.530	1.455	1.610
Coast Guard	.986	.054	334.488	1	.000	2.681	2.412	2.980
Constant	-1.575	.017	8,816.795	1	.000	.207		

[a]Variable(s) entered on step 1: Cservice.

Table A-4.
Variables in the Equation: Region

	B	S.E.	Wald	df	P-Value	Odds Ratio	95% C.I. for Odds Ratio	
							Lower	Upper
Step 1[a]: **Europe, Asia, Pacific Islands**	.109	.021	26.965	1	.000	1.115	1.070	1.162
Constant	-1.466	.012	14,607.566	1	.000	.231		

[a]Variable(s) entered on step 1: Cregion2.

Table A-5.
Variables in the Equation: Sex

	B	S.E.	Wald	df	P-Value	Odds Ratio	95% C.I. for Odds Ratio	
							Lower	Upper
Step 1[a]: **Females**	.058	.028	4.273	1	.039	1.059	1.003	1.119
Constant	-1.439	.011	17,957.629	1	.000	.237		

[a]Variable(s) entered on step 1: Csex.

Table A-6.
Variables in the Equation: Marital Status

	B	S.E.	Wald	df	P-Value	Odds Ratio	95% C.I. for Odds Ratio	
							Lower	Upper
Step 1[a]: **Married**	.230	.020	128.323	1	.000	1.259	1.210	1.310
Constant	-1.567	.016	9,738.016	1	.000	.209		

[a]Variable(s) entered on step 1: Cmarital.

Table A-7.
Variables in the Equation: Occupation Category (Combat, Combat Support)

	B	S.E.	Wald	df	P-Value	Odds Ratio	95% C.I. for Odds Ratio	
							Lower	Upper
Step 1[a]: **Combat Support**	.332	.026	161.758	1	.000	1.394	1.324	1.467
Constant	-1.699	.024	5,134.893	1	.000	.183		

[a]Variable(s) entered on step 1: Combat_C.

Table A-8.
Variables in the Equation: Education

	B	S.E.	Wald	df	P-Value	Odds Ratio	95% C.I. for Odds Ratio	
							Lower	Upper
Step 1[a]: **Education**			1356.180	2	.000			
Some College	.532	.029	332.440	1	.000	1.702	1.608	1.803
4 Year Degree or More	.834	.023	1,261.698	1	.000	2.303	2.199	2.411
Constant	-1.693	.013	16,843.466	1	.000	.184		

[a]Variable(s) entered on step 1: Ceduc4.

Table A-9.
Variables in the Equation: Number of Deployments

	B	S.E.	Wald	df	P-Value	Odds Ratio	95% C.I. for Odds Ratio	
							Lower	Upper
Step 1[a]: **Deployed 1 or more**	.243	.020	144.015	1	.000	1.275	1.225	1.326
Constant	-1.571	.016	10,132.059	1	.000	.208		

[a]Variable(s) entered on step 1: Deployment constructed from NDEPLOY variable.

Table A-10.
Variables in the Equation: Race/Ethnicity

	B	S.E.	Wald	df	P-Value	Odds Ratio	95% C.I. for Odds Ratio	
							Lower	Upper
Step 1[a]: **Race/Ethnicity**			37.388	3	.000			
Black	-.104	.028	13.604	1	.000	.901	.853	.952
Hispanic	-.111	.032	11.974	1	.001	.895	.841	.953
Other	.100	.034	8.633	1	.003	1.105	1.034	1.182
Constant	-1.411	.012	12,903.298	1	.000	.244		

[a]Variable(s) entered on step 1: Crace_Eth.

Appendix B.
Full Model for Phase 1

Logistic Regression: Full Model, Phase 1

Table B-1.
Categorical Variables Coding

		Frequency	Parameter Coding			
			(1)	(2)	(3)	(4)
Age Grouping 5 Constructed	18 - 24	25,708	.000	.000	.000	.000
	25 - 29	17,559	1.000	.000	.000	.000
	30 - 34	7,587	.000	1.000	.000	.000
	35 - 44	10,620	.000	.000	1.000	.000
	45 +	3,929	.000	.000	.000	1.000
Paygrade Group 7 Constructed	E1-E5, Enl Unknowns	40,981	.000	.000	.000	.000
	E6-E9	13,587	1.000	.000	.000	.000
	W1-W5, War Unknowns	1,245	.000	1.000	.000	.000
	O1-O3, Off Unknowns	5,208	.000	.000	1.000	.000
	O4-O6	4,382	.000	.000	.000	1.000
Numeric Service	Army	25,009	.000	.000	.000	.000
	Navy	14,643	1.000	.000	.000	.000
	Marine Corps	10,073	.000	1.000	.000	.000
	Air Force	14,019	.000	.000	1.000	.000
	Coast Guard	1,659	.000	.000	.000	1.000
Education Grouping 4 Constructed	No College/ Unknown	44,783	.000	.000		
	Some College	8,078	1.000	.000		
	4-year Degree or higher	12,542	.000	1.000		
Regional / Deployment Grouping Constructed	US & Territories, Other, Unknown	44,591	.000			
	Europe, Asia, Pacific Islands	20,812	1.000			
COMBAT_C	Combat	13,608	.000			
	Combat Support	51,795	1.000			
MARITAL Status Grouping Constructed	Not Married	27,756	.000			
	Married	37,647	1.000			

Table B-2.
Variables in the Equation

		B	S.E.	Wald	df	Sig.
Step 0	Constant	-1.431	.010	20,853.461	1	.000

Table B-3.
Variables not in the Equation

			Score	df	P-Value
Step 0	Variables	**Age Group**	2,170.259	4	.000
		25-29	81.251	1	.000
		30-34	.015	1	.902
		35-44	708.653	1	.000
		45 +	1,106.956	1	.000
		Paygrade	1,811.294	4	.000
		E6-E9	388.783	1	.000
		W1-W5	61.066	1	.000
		O1-O3	211.231	1	.000
		O4-O6	685.143	1	.000
		Service	642.411	4	.000
		Navy	1.616	1	.204
		Marine Corps	91.082	1	.000
		Air Force	259.330	1	.000
		Coast Guards	293.575	1	.000
		Overseas	26.981	1	.000
		Married	128.638	1	.000
		Combat Support	162.844	1	.000
		Education	1,395.708	2	.000
		Some College	123.018	1	.000
		4 Year Degree or More	1,091.990	1	.000
	Overall Statistics		3,314.283	17	

Table B-4.
Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	3,100.797	17	.000
	Block	3,100.797	17	.000
	Model	3,100.797	17	.000

Table B-5.
Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	61,057.332	.046	.074

[a]Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

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

This page is reserved for insertion of Standard Form 298, page 2 -- this is best accomplished by replacing this page after the document has been converted to PDF


DEPARTMENT OF DEFENSE
UNITED STATES OF AMERICA