Defense Equal Opportunity Management Institute's Weighting Strategy

DEFENSE EQUAL OPPORTUNITY MANAGEMENT INSTITUTE
DIRECTORATE OF RESEARCH DEVELOPMENT AND STRATEGIC INITIATIVES

Dr. Richard Oliver Hope Human Relations Research Center
Directed by Dr. Daniel P. McDonald, Executive Director
366 Tuskegee Airmen Drive Patrick AFB, FL 32925
321-494-2747

Prepared by
Hope Research Center
DEOMI R&D
Dr. Daniel P. McDonald

DEOMI PRESS

Tech Report No. 19–03

Executive Summary

The purpose of this report is to detail Defense Equal Opportunity Management Institute's (DEOMI) consideration, development, and evaluation of a weighting strategy to enhance generalizability of Defense Organizational Climate Survey (DEOCS) data. In considering the utility of a weighting strategy, the strengths and limitations of the DEOCS administration are assessed in conjunction with the potential benefit of a weighting strategy. An initial weighting strategy was developed to adjust estimates for nonresponse in the sample and the population. The preliminary evaluation of this procedure revealed: (1) similar ratios of individual-level demographics to the DoD population, (2) no evidence of nonresponse bias when employing sample weights, and (3) no significant difference in results between weighted and unweighted data. If improvement of estimates is minor, and the strength of the DEOCS data is in its representativeness, further consideration must be taken regarding whether weighting is necessary. Currently, the Office of People Analytics (OPA) is formally evaluating this process before providing unit climate results to DoD or Services and will be a resource for future direction in this domain.

Defense Equal Opportunity Management Institute's DEOCS Weighting Strategy

Defense Equal Opportunity Management Institute (DEOMI) developed a weighting strategy in order to leverage command climate data and extend inferences beyond the unit to the Service. This document outlines this process by (1) providing a background on the Defense Organizational Climate Survey (DEOCS) history, methodology, and resulting data; (2) providing a general overview of weighting; (3) detailing DEOMI's weighting strategy; and (4) providing an initial evaluation of this strategy and way forward. Currently, the Office of People Analytics (OPA) is formally evaluating the uses of DEOCS data.

1. Defense Organizational Climate Survey

This section provides background information regarding the DEOCS. It then describes the methodology of the survey administration, along with limitations and strengths of the current approach. The section concludes with a discussion of DEOCS data utility for commanders and the Department of Defense (DoD) community.

1.1 Background.

Approximately 30 years ago, DoD Directive 1350.2, "The Department of Defense Military Equal Opportunity Program," established the requirement for commanders/directors to monitor their organizations' climates, and take steps to remedy identified issues and concerns. Many command climate assessments were developed by the Services, and DEOMI developed its initial DoD-wide survey called the Military Equal Opportunity Climate Survey (MEOCS). The MEOCS evolved over the years into the DEOCS and usage grew. In 2012, over 1.8 million surveys were administered. Periodic climate assessment is mandated by the National Defense Authorization Act for FY13 (NDAA13), and the DEOCS is the DoD's exclusively authorized climate assessment instrument for commanders/directors to meet this requirement, per the Undersecretary of Defense for Personnel and Readiness [USD (P&R)] memo dated 20 Nov

2014. Still, unit member participation is voluntary and anonymous. The DEOCS is a command-requested organizational development survey used to assess members' shared perceptions of various human relations conditions, including organizational effectiveness, equal opportunity, and sexual assault prevention and response. It has been reviewed and approved through OPA (formerly Defense Manpower Data Center, DMDC), DoD Privacy Office, DoD Human Research Protection Program (HRPP), and has received licensure through the Washington Headquarters Services (WHS).

The DEOCS, previously known as the Military Equal Opportunity Climate Survey (MEOCS), was first released in 1990. In 2005, the MEOCS was modified and retitled DEOCS and, for the first time, was available as an online survey. Over the years, usage of the DEOCS grew substantially, from 154,381 surveys administered in FY05 to 3,104,702 surveys administered in FY18. The DEOCS is currently using version 4.1 and averages a 44% response rate across the DoD, making it a far-reaching and impactful tool for commanders to assess climate and readiness.

1.2 Methodology.

This section describes the procedure units use to administer the DEOCS, as well as how this procedure contributes to the sampling methodology. This section then highlights limitations and strengths to this approach.

1.2.1 Procedure.

The DEOCS 4.1 is available both online and in paper-and-pencil format, and takes approximately 22 minutes to complete. The core survey contains 56 items for civilians, and 57 items for military respondents. In addition to the core survey items, there are up to seven demographics questions; this number varies depending on demographic selections in the beginning of the survey. The survey has four major areas: (1) Demographics, (2) Organizational Effectiveness (OE), (3) Equal Opportunity/Equal Employment Opportunity (EO/EEO)/Fair Treatment, and (4) Sexual Assault Prevention and Response (SAPR). Commanders can also elect to add up to ten locally developed questions and five short answer questions. The DEOCS also provides respondents the opportunity to provide written comments at the end the OE, EO, and SAPR sections. All comments (verbatim) are provided to commanders in unit reports.

In most cases, the survey administrator requests the DEOCS on behalf of a unit commander. The survey administrator may be a DEOMI-trained equal opportunity advisor (EOA) or someone who is serving as a trusted agent to the commander and is responsible for overseeing the assessment process to include requesting, receiving, and distributing survey passwords.

The survey administrator has access to the DEOCS Survey Administrator Assistance Site (SAAS) management system. The SAAS enables survey administrators to manage DEOCS administration from beginning to end. They specify the start and stop dates for the survey, which typically remains open for 21 to 30 days. The survey administrator will send a memo from the commander to unit personnel, informing them of the upcoming voluntary survey and encouraging them to complete it. Survey administrators can track completion rates in real time

and can extend the stop date at any time during administration. This is typically done if participation rates are lower than what leadership strives to achieve, and extending the survey stop date avails more members the opportunity to complete it. In addition, the survey administrator or commander may opt to send interim reminders to personnel to complete the survey during the administration period. After the survey stop date, the survey administrator must log onto the SAAS account and request the DEOCS report. A DEOCS report is generated within 4–6 business days after being requested.

1.2.2 Sampling strategy.

The DEOCS is unique in that the survey is continuously available to DoD organizations to meet timelines specific to each commander, as directed by the NDAA13. NDAA13 requires leaders of DoD organizations to conduct a climate assessment that—per USD (P&R memo dated 20 Nov 2014)—includes administering the DEOCS as a component of that assessment within 120 days of taking command and annually, thereafter, while in command. This means that, unlike virtually all other DoD surveys, the DEOCS does not have an explicitly defined survey administration timeframe; instead, the survey is administered according to the life-cycle of the commander's tenure. This ties the results more directly to commander and leadership influence on the respective organization.

The NDAA13 requirement explicitly seeks to ensure leaders comprehensively characterize their organizations' sexual harassment and sexual assault prevention climate early after assuming command then at regular intervals while in command. This requirement helps support the assumption that all units will be sampled during any 12 month period. In other words, the survey is initiated by each commander in the DoD, across Services, and retained over time. Thus, the DEOCS administration resembles a census sample when data are collected and presented on an annual basis. This implies the DEOCS target population is all Service members, including all military and government civilians.

While a year of DEOCS data resembles a census sample when data are presented on a less than annual basis, the sample is best characterized as a non-probability or non-random sample, because units self-select into the sample according to their position in the NDAA13-defined climate assessment schedule. At even shorter sampling intervals, the sampling strategy closely resembles the single stage cluster sampling method, where entire clusters (or units) are included in the sample and every member within the cluster is provided an opportunity to participate; thus, individuals may self-select into the sample. This means that the administration method, on an annual basis, yields a complete census of units.

1.2.2.1 Limitations.

With every survey methodology comes trade-offs between academic rigor and field fidelity. The current methodology emphasizes the field's need for a practical climate survey; however, the non-probability sampling approach is noted as one of the primary limitations impacting DEOCS

data.[1] In discussing non-probability sampling, considerations are posed for the potential impact, if any, to DEOCS data quality.

Some constituents within the DoD have preferred a stratified random sample for survey administration when generalizing estimates to a larger population. Stratified random sampling is not as appropriate for DEOCS administration because the DEOCS is a commander's tool for internal climate assessment. In this setting, a non-probability sampling approach is preferred, where commanders ask every member within their unit to complete the survey, not just some members. While this approach works well at the unit level, we see the value of the DEOCS data beyond each individual unit. The remainder of the section will discuss obstacles to inference of non-probability sampling when generalizing results to larger populations. These include: 1) exclusion bias, 2) reliance on volunteers, and 3) participation rates (Baker et al., 2013).

One issue with non-probability samples is *exclusion bias*, where "the vast majority of the target population is likely to have no chance of inclusion in the sample" (Baker et al., 2013, p. 21). Given that every individual within the DoD should have the opportunity to participate in the DEOCS, we posit that *exclusion bias* is likely not a major concern under the current administration process of the DEOCS. Each Service is charged with tracking compliance of the NDAA requirement; therefore, one exception could be the lack of consistency across Services for how compliance is monitored. One approach to improve the confidence in the representativeness of DEOCS data would be to centralize compliance tracking of meeting this administrative requirement. This will help ensure appropriate coverage of the target population. Ultimately, we can empirically examine whether exclusion bias is problematic for the data by engaging the Services, to determine whether commanders are complying with the NDAA requirement and, therefore, reasonable coverage for each Service and the DoD is achieved. The DoD is moving toward monitoring compliance.

Another limitation of non-probability sampling is the reliance on volunteers to complete the study, because the probability of participating is unknown (Baker et al., 2013). This limitation is not a valid concern for the coverage of units within the DoD, since policy requires the administration of the DEOCS on an annual basis. Additionally, the collection of auxiliary information (see Section II for further discussion of this stage) during the administration process allows for estimation of participation probability. With that said, the DEOCS—just as every other survey administrated within the DoD—is at risk for selection bias, whereas individuals who choose to take the DEOCS may differ from individuals who choose not to take the survey.

While the DEOCS is subject to selection bias, the response rate is much higher than other DoD surveys, which helps to lessen this concern. Baker et al. specifies that "[t]he issue of low response rates to probability-based samples and related concerns about nonresponse bias are at the heart of the arguments in favor of alternative approaches" (2013, p. 12). While initial efforts have been taken to study nonresponse bias in DEOCS data, we recommend further investigating this matter to determine the extent to which selection bias may influence the results.

[1] This section will focus on the limitations of non-probability sampling in relation to the sampling strategy of the DEOCS; DEOCS administration does have other limitations (e.g., relying on survey administrators to provide accurate auxiliary variables or demographics of the unit being surveyed); therefore, this is not an exhaustive account of all limitations regarding the current DEOCS administration process.

Although the non-probability sampling approach has been cited as one of the greatest limitations to the DEOCS, it is also best suited for a field survey. Additional benefits from this approach that offset any potential limitations to the data, in the way of generalizability, are discussed below.

1.2.2.2 Strengths

The current methodology of DEOCS administration serves commanders and, by extension, the DoD well. The primary objective of any sampling strategy is to make valid and reliable inferences to the target population. One of the main concerns when developing a sampling strategy is to ensure representativeness of your sample matches the target population. At the unit level, the target population is the entire unit, and all members are invited to complete the survey. When data is aggregated beyond the unit, DEOCS administration can be characterized as a non-probability or non-random sample. There are benefits of the current administration process that contribute to the representativeness of the sample data, both for the unit and aggregated beyond the unit.

First and foremost, the DEOCS is a commander's tool to assess that organization's human relations environment. The commander-initiated design of the DEOCS administration enhances the representativeness of the sample for a few reasons. First, commanders, while adhering to the NDAA13-defined climate assessment schedule, can choose a time period to initiate the survey that is optimal for the unit. Second, the DEOCS is available online and/or in paper-and-pencil form, making it more accessible to all units. Third, the value of the assessment is explicitly tied to the response rate of the survey; therefore, the commander—and EOA—are more likely to encourage participation because they understand that a higher response rate is key for establishing whether their results are representative of their unit. Fourth, commanders have strong buy-in in the process, as well as ownership over the results, to motivate change and improvement. This commander initiation also impacts respondent belief that their responses are more immediately relevant and potentially actionable—a unique quality of this DoD survey. Commander feedback of the results to their units further increases relevance and face validity of this climate assessment. This flexible, commander-focused administration is directly tied to the response rate of the survey, and, therefore, helps to improve the representativeness of unit-level results as well as provide reasonable coverage of the DoD population (for comparisons between quarterly samples and population, refer to any FY17 DEOMI quarterly reports titled: DEOMI Organizational Climate Survey Rollup Report Executive Summary).

1.3 Value of the DEOCS.

Organizational assessment is a key primary prevention strategy that EO professionals and commanders rely on to detect issues before they rise to a level of more egregious behavior. In addition to preventing negative behaviors, organizational assessments can be harnessed to promote positive behaviors. This is because command climate is an important driver of members' attitudes, behaviors and, ultimately, mission effectiveness. In order to leverage and maximize the positive influence of command climate, leadership must establish and maintain the right climate. The DEOCS is a commander's management tool that provides insight into the

positive and negative attributes of the command climate. This knowledge can be used to identify potential areas of concern and organizational strengths.

The DEOCS provides a comprehensive report that diagnoses potential organizational issues. Commanders use this information to assess negative and positive indicators of their commands, and use this information to modify policies, programs, and practices, and reallocate resources. DEOCS findings help to drive follow-on assessment actions, namely focus groups and/or individual interviews, systematic observations, and/or review of records and reports. Because the DEOCS is just one part of a larger organizational assessment, triangulating DEOCS results with information collected using other methods can help commanders gain a holistic picture of their command's climate.

The value of the DEOCS extends well beyond the commander's utility. The current administration process allows for a large and robust climate dataset that can be used as a prescriptive tool. This dataset can be paired, by Unit Identification Code (UIC), with other data sources to identify protective and problematic unit-level behaviors. DEOCS data can be examined to improve the prescriptive information provided to commanders, thereby enhancing readiness across the DoD. Put another way, the resulting data from all of the DEOCS administrations, across units, can potentially be examined together to better understand the state of climates within the DoD.

Historically, DEOMI provided reports containing estimates of climate-related constructs aggregated to the Services and DoD agencies. From FY12 to FY17, DEOMI provided unweighted estimates of these constructs in quarterly and ad hoc reports tailored to each constituency. DEOMI was cautious not to draw conclusions based on the data, but, instead, provided descriptive statistics of the climate constructs. Additionally, DEOMI incorporated appropriate caveats into all reports that the results applied to the particular sample of data analyzed and cautioned against generalizing. To enhance the ability to draw conclusions about DEOCS data to extend inferences beyond the unit to the Services, DEOMI partnered with SRA/Westat/DMDC to develop a weighting plan and strategy. These entities were the same experts who were consulted for other DoD surveys that are generalized and reported to congress. In FY18, DEOMI began applying sample weights to DEOCS 4.1 data for higher-level quarterly and ad hoc requests/reports. The remainder of this paper will discuss the use of statistical weights to improve estimation, DEOMI's approach to weighting DEOCS data, and DEOMI's initial efforts to evaluate the weights.

2. Improving Estimation with Weighting

This section defines weighting, defines criteria for utilizing a weighting strategy, and describes two main types of weighting adjustments. A discussion of weighting in survey research follows, reviewing the uses and utility of weighting.

2.1 Weighting defined.

Weighting refers to the statistical adjustment of survey data after it has been collected to improve the accuracy of the inferences made (Lavrakas, 2008). In survey research, there are two broad

categories to consider when discussing the sample: responders and non-responders. Responders are defined as those who completed the essential questions required for the report of interest. Non-responders are defined as individuals who chose not to respond (refusal), were not eligible to respond (ineligible), or were eligible but were not contacted for a response (non-contact). The decision to utilize a statistical adjustment is made when there is concern about survey nonresponse. Pike (2008) notes two ways in which nonresponse can become a concern. First, nonresponse bias may occur if non-respondents differ from respondents on the survey variables of interest. Second, there is concern if response rates differ among demographic subgroups and these subgroups have differing responses on the survey variables of interest (Pike, 2008). However, researchers have found that, for many surveys, this bias can be relatively minor (Pike, 2008; e.g., see Groves et al., 2004; Keeter et al., 2000; Kuh, 2001).

Use of weighting adjustments tends to occur when concerns about the response rate and demographic subgroups within the sample exist. There are three criteria that should be considered prior to weighting: (1) differences in response rates of subgroups, (2) significant differences between subgroups on the survey variables, and (3) no meaningful differences between responders and non-responders (Pike, 2008). When these criteria are present, weighting may be appropriate, and there are two main types of adjustments to choose from: population weighting and sample weighting.

2.2 Population versus sample weighting.

Population weighting requires that researchers have knowledge of the population distribution across a number of demographic characteristics, known as auxiliary variables (Kalton & Kasprzyk, 1986). For example, sex, age, and race are considered auxiliary variables that can be separated into combinations or classes (e.g., white junior enlisted male). Knowledge of class distribution within the population allows researchers to apply weights to the classes, in order to reduce nonresponse bias (Kalton & Kasprzyk, 1986).

If researchers do not have knowledge of the population distributions, then sample weighting may be appropriate. However, sample weighting requires that researchers subsequently have additional information about the respondents and non-respondents in their sample (Pike, 2008). That is, within the sample, researchers have knowledge of the demographic distribution and, by applying weights to the classes in which responders have been separated, the sample of responders becomes a closer representation of the complete sample of individuals (Pike, 2008).

2.3 Weighting in survey research: uses and utility.

The application of weighting to reduce nonresponse bias is considered useful if weighting can reduce the variance of survey estimates (Little & Vartivarian, 2005). For example, consider the case of a demographic sub-sample of voters that is not well-represented. If researchers have access to the population distribution, an estimation of the population's sample distribution, and knowledge that there is a correlation between the demographics used for weighting and the polling topic, then there is a greater likelihood that a weighting adjustment may improve the accuracy of the conclusion made from the poll (Alvarez & Nagler, 2005). Similarly, the 2012-2013 National Adult Tobacco Survey utilized weighting to better examine the prevalence of

tobacco use and tobacco-related indicators. Using data gathered from the U.S. Census and the American Community Survey, researchers noted they would be better able to remove nonresponse bias under the assumption that gender is correlated with their study outcome. Therefore, their weights were adjusted according to the gender totals within the population (Centers for Disease Control and Prevention, 2014). In relation to the DEOCS, this would mean attending to demographic variables such as rank and gender combinations that are known to be correlated with outcomes of interest, specifically the *sexual harassment* and *sexual assault* climate areas.

The general concern is that, regardless of weighting method, socio-demographics (e.g., sex, age, race) are often the only variables that are available to aid in reducing nonresponse bias; however, the actual influence on survey response can come from the survey questions themselves (van Goor & Stuiver, 1998). That is, regardless of demographic characteristics, if survey questions incite extreme reactions—specifically feelings that one is being threatened or feelings of disinterest—individuals may be less likely to respond. For example, researchers studying alcohol consumption found that individuals on both extreme ends of the spectrum (i.e., heavy and infrequent drinkers) were less likely to respond to the survey (Lemmens, Tan, & Knibbe, 1988). Another study found similar results for absenteeism: non-responders were individuals with a very high or very low number of absences (Bebbington, 1970). This suggests that weighting for nonresponse can be useful; however, the availability of resources out in the field for non-responders may not lead to a meaningful improvement in results (i.e., reducing nonresponse bias; van Goor & Stuiver, 1998). This is an additional consideration for weighting DEOCS data that will require attention to the specific research or project focus, based on which variables may be theorized to have significant relationships with our climate areas.

Keeping both the utility and constraints of weighting in mind, a weighting strategy was developed for the DEOCS that could be applied to a wide range of DEOCS estimates. The goal was to improve the accuracy of inferences with a focus on making adjustments based on demographic weighting classes. The following section will discuss DEOMI's approach to weighting DEOCS data and the initial efforts to evaluate the weights.

3. DEOMI's Weighting Strategy

Under collaboration with SRA/Westat/DMDC, a weighting plan and strategy was developed to include consideration of both non-responders in the unit sample and the DoD population. This section discusses DEOMI's weighting strategy, highlighting work taken to collect accurate unit demographics as well as up-to-date population data available by UIC. Broadly, this strategy involves first developing a sample weight, then developing a population weight, and finally combining the adjustments into one overall weight. The full strategy is discussed in this section, but current barriers limit DEOMI's ability to leverage the population weight portion of this strategy. This section will provide details on each of these steps and conclude with a practical discussion of DEOMI's current status on this strategy and current practices in weighting.

3.1 Developing sample weights.

DEOMI uses a sample weight to adjust the sample for non-responders in units that have participated in the DEOCS. In this adjustment, a non-responder in a participating unit is a member who elected to not participate in the survey. This procedure involves (1) collecting auxiliary information from survey administrators to form weighting classes (rank and gender combinations), (2) assessing this auxiliary data for quality, (3) using the auxiliary data to inform a Python (Python Software Foundation) code that simulates non-responder rows in the dataset, and (4) conducting a weighting class method using a chi-square approach (via IBM SPSS CHAID) to establish a sample weight that adjusts for nonresponse.

Unit administrators are first asked to provide the unit demographic data for assigned unit members in 16 weighting classes (e.g., junior enlisted females, junior enlisted males). Analysts evaluate the quality of the administrators' data by comparing each unit's proportions of demographic subgroups to other units that are similar in Service and size. After the quality check, Python code uses this information to generate "dummy" non-responder rows in the sample dataset to create a structured dataset that is compatible with IBM SPSS CHAID. Finally, the chi-square approach is used to assess the relationship between the rank×gender weighting classes and unit membership, predicting response. The final sample weight is created by taking the reciprocal of the predicted probability of being a responder.

3.2 Developing population weights.

Next, DEOMI proposes a population weight to adjust the sample for non-responders who are in units that have *not* participated in the DEOCS. In this adjustment, a non-responder is a member within a unit that did *not* participate in the DEOCS during the time period of the sample and, therefore, were never given an opportunity to respond to the survey. DEOMI has two potential methodologies for creating this weight. The first procedure, adopted by DEOMI, involves (1) coding population data into weighting classes (e.g., rank, sex, race, ethnicity, and Service combinations), (2) merging the weighting classes into the sample data, and (3) utilizing a ratio method to establish population weights. The second procedure omits the steps for coding into weighting classes by employing a raking ratio strategy.

DEOMI obtains population data from DMDC on a quarterly basis. The first procedure requires that this data is categorized into 7,277 weighting classes and merged into the sample data. Next, each unit's proportions are compared to the population, and a ratio method is used to establish a population weight. The final step is to create a combined adjustment of the population and sample weights through multiplication.

3.3 Current status of weighting strategy.

Currently, DEOMI has completed syntax for all facets of this strategy; however, we have only executed the sample weighting portion. Practical issues arise from the format of the DMDC-provided population data. Currently, the Army does not provide DMDC with population data using the same coding structure of "race" as the other Services. Specifically, Army includes

Hispanic as a race and not an ethnicity. This impacts the ability to include race in the population weighting classes and stalls out the ability to create final population weights. Without the population adjustments, DEOMI provided sample-adjusted estimates in roll-up reports each fiscal quarter, from FY18 Q1 to FY18 Q3, to the DoD Sexual Assault Prevention and Response Office and DoD Suicide Prevention Office. Then, on 1 October 2018, a DHRA memo instructed DEOMI to discontinue providing roll-up reports of DEOCS data, pending further OPA evaluation.

4. Initial Evaluation of DEOMI Weighting Strategy

In June 2018, OPA provided DEOMI with technical recommendations for weighting during the evaluation of DEOCS version 4.1. The technical recommendations included the following four analysis methods to evaluate the validity and reliability of the weights:

1. Check whether all eligible respondents have final weights and whether all other sample members do not have weights.

2. Check whether the summation of final weights equals the population totals.

3. Check the trend line of the key survey weighted estimates, overall and in key subpopulations, to see whether the changes across waves are reasonable.

4. Check the unequal weighting effect of the final weights and the design effects of the weighted estimates to see whether there is large weight variability and whether any extremely large weights have an impact on the weighted estimates.

DEOMI conducted the recommended analyses as part of the evaluation recommendations. Results from the analyses demonstrated:

1. All eligible respondents have final weights.

2. DEOCS unit sample totals size could be inflated due to having anonymous vs confidential data. DEOCS data is anonymous for individuals (confidential for units). Because of this, DEOMI's weighting strategy uses Python code to add "dummy" rows in the data, based on the auxiliary fields provided by the survey administrator. The Python code can only add rows for individuals who are not represented in the data, but cannot *remove* rows based on respondents misrepresenting their demographics. However, inflated totals are scaled back through a process of "normalizing" the data.

3. In order to check the trend line of the key survey weighted estimates overall, average favorable response percentages of the key survey variables were examined over a six-month period of time (using "month" variable). The means demonstrated small variation, with all remaining within two standard deviations (SD). Future efforts should examine the trend lines for key subgroups and compare weighted estimates to unweighted estimates. Efforts to further evaluate and improve the weighting process were paused

with the distribution of the DHRA memo, suspending aggregated reports. DEOMI effort in this area is pending OPA's evaluation.

4. The unequal weighting effect of the final weights was examined for large weight variability. Kish's (1965) formula for calculating UWE was used, which defines UWE as the ratio of the variances of the weighted mean to the variance of the unweighted mean. Results indicated that the UWE was less than two across all weights; therefore, trimming was not applied.

DEOMI also enlisted the assistance of outside academics through its Summer Research Faculty program to perform further evaluation on the weighting strategy prior to finalizing it. Dr. Wei Wan, an associate professor of mathematics at Claflin University, and Dr. Richard Harris, a demographer and professor of social work at University of Texas, San Antonio, provided advisement and feedback throughout development of the weighting strategy and during its evaluation. Dr. Wan specifically examined the strategy to evaluate it for nonresponse bias. Additionally, Dr. Harris compared results of both weighted and unweighted data to determine if there was a significant difference. Using a three-month sample for the evaluation, findings indicated (1) similar ratios of individual-level demographics to the DoD population, (2) no issue with nonresponse bias when weighting for nonresponse, and (3) no significant difference when comparing results of weighted and unweighted data. Each of these studies will be examined in further detail in the following paragraphs.

Dr. Wan's (2018) research examined whether making adjustments for nonresponse will lead to increased variability in the weights and lower precision of the survey estimates. To do this, he referenced methods that OPA has used to evaluate nonresponse bias for the Workplace Gender Relations Assessment (WGRA). Dr. Wan's results demonstrated that the absolute relative bias statistic (the absolute difference between the respondent estimate and the full sample estimate divided by the full sample estimate) was minor and that there is little nonresponse bias (Wan, 2018).

Dr. Harris (2018) examined whether weighting DEOCS data produced significantly different results than unweighted data. Specifically, Dr. Harris looked at the Unwanted Workplace Experience factor in his study. His results demonstrated that there are only small observed differences, and that these small differences suggest there would be little benefit to weighting associated with the nonresponse patterns. Additionally, no significant differences were found using a Rake Weight procedure to adjust for differences in proportions between the sample and population (Harris, 2018).

Taken together, DEOMI concluded that the weighting strategy was valid, based on the results of OPA's recommended analyses during the evaluation of DEOCS version 4.1. The results of Dr. Wan's research demonstrated that weighting DEOCS data produced little nonresponse bias, and the difference between the respondent estimate and full sample estimate was minor. However, Dr. Harris found that only small observed differences occurred between weighted and unweighted data, which calls into question whether the data should be weighted. Further discussion should be conducted by OPA regarding the utility of weighting DEOCS data as they continue their formal evaluation of this process.

5. Conclusion

DEOCS data represents an annual census of units and can be considered a representative cluster sample on a quarterly basis. Methodological strengths of this survey's administration include the flexible administration by unit commanders and the NDAA requirement that all units periodically participate. These strengths increase commander buy-in, elevating responses rates. Acknowledging that the measurement and validity of climate "estimates" are inherently fuzzy, there is always value in improving these estimates.

Employing a weighting strategy is one method that may enhance estimates. Although Pike (2008) discusses the relative costs and benefits of utilizing weighting, it is important to keep in mind that it is not a silver bullet solution to providing complete confidence in the estimates derived from any data source. In fact, adjusting for nonresponse can potentially decrease the precision of estimates and increase the variances and standard errors of estimators (Kalton, 1983; Kalton & Kasprzyk; Kish, 1965; Pike, 2008). DEOMI explored the utility of weighting DEOCS data, and developed a weighting strategy to adjust for nonresponse in both the sample and population. In an initial evaluation of this weighting process, DEOMI found that DEOCS data contains low nonresponse bias and no difference between weighted and unweighted results (likely due to a combination of high response rates and low nonresponse bias). With similar weighted and unweighted results, it may warrant revisiting whether weighting brings more costs than benefits in understanding estimates of command climate.

If weighting is still a desired component of an analysis plan for DEOCS data, future work can continue to improve upon DEOMI's weighting strategy (e.g., include more weighting classes, improve population data source, and complete the population weighting process). With any improvements, we must consider the balance between incremental improvement and changes that may undermine the survey's primary purpose—serving the commander. Serving the commander with a unit-focused, flexible administration must continue to be the foundational goal of the DEOCS; this strategy will also secure its methodological strengths (i.e., unit participation and individual response rates). Changes to DEOCS survey methodology may jeopardize resulting data quality and the validity of subsequent statistical conclusions, not to mention the utility to the unit commander.

Our current understanding is that, using a three-month sample of data, we tend to see (1) similar ratios of individual-level demographics to the DoD population, (2) no evidence of nonresponse bias when employing sample weights, and (3) no significant difference in results between weighted and unweighted data. Currently, OPA is formally evaluating this process before providing unit climate results to DoD or Services and will be a resource for future direction in this domain.

References

Alvarez, R. M., & Nagler, J. (2005, July). *A method for weighting survey samples of low-incidence voters.* Paper prepared for the 2005 Political Methodology Summer Meeting, Tallahassee, FL.

Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., Gile, K. J., & Tourangeau, R. (2013, June). Report of the AAPOR task force on non-probability sampling. Retrieved from: https://www.aapor.org/AAPOR_Main/media/MainSiteFiles /NPS_TF_Report _Final_7_revised_FNL_6_22_13.pdf

Centers for Disease Control and Prevention (2014). *2012–2013 National Adult Tobacco Survey Weighting Specifications*. Paper prepared by Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. Atlanta, GA: U.S.

Bebbington, A. C. (1970). The effect of nonresponse in the sample survey with an example, *Human Relations, 23*(3), 169-180.

Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly, 68*(1), 2-31.

Harris, R. J. (2018). *Unwanted Workplace Experiences in the Military: Issues Affecting Measurement of Patterns and Prevalence of Sexual Harassment.* Unpublished manuscript, Defense Equal Opportunity Management Institute, Patrick Air Force Base, FL.

Kalton, G., & Kasprzyk, D. (1986). The treatment of missing survey data. *Survey Methodology, 12*(1), 1-16.

Keeter, S., Miller, C., Kohut, A., Groves, R. M., & Presser, S. (2000). Consequences of reducing nonresponse in a national telephone survey. *Public Opinion Quarterly, 64,* 125-148.

Kish, L. (1965). *Survey Sampling*. Oxford, England: Wiley

Kuh, G. D. (2001). The national survey of student engagement: Conceptual framework and overview of psychometric properties. Bloomington, IN: Indiana University Center for Postsecondary Research. Retrieved February 24, 2006 from the national Survey of Student Engagement Web site: http://nsse.iub.edu/html/psychometric_framework_2002.cfm

Lavrakas, P. J. (2008). Surveys by telephone. In W. Donsbach & M. W. Traugott (Eds.). *The SAGE Handbook of Public Opinion Research* (pp. 249-261). London: SAGE Publications Inc.

Lemmens, P. H. H. M., Tan, E. S., & Knibbe, R. A. (1988). Bias due to nonresponse in a Dutch survey on alcohol consumption. *British Journal of Addiction, 83*(9), 1069-1077.

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

Pike, G. R. (2008). Using weighting adjustments to compensate for survey nonresponse. *Research in Higher Education, 49*, 153-171.

van Goor, H., & Stuiver, B. (1998). Can weighting compensate for nonresponse bias in a dependent variable? An evaluation of weighting methods to correct for substantive bias in a mail survey among Dutch municipalities. *Social Science Research, 27*, 481-499.

van Hoeven, L.R., Janssen, M.P., Roes, K.C.B., & Koffijberg, H. (2015). Aiming for a representative sample: Simulating random versus purposive strategies for hospital selection. *BMC Medical Research Methodology, 15*, 90.

Wan, W. (2018). *Analysis of Data from DEOCS Survey: Weighting by Decision Tree, Evaluation of Weights, and Nonresponse Bias Analysis.* Unpublished manuscript, Defense Equal Opportunity Management Institute, Patrick Air Force Base, FL.