Response to Notes Regarding the 2006 Survey of Active Duty Spouses: Reply

Willard C. Losinger

In "Response to Notes Regarding the 2006 Survey of Active Duty Spouses", Al Nassir *et al*. asserted that prior observations concerning the quality of Defense Manpower Data Center (DMDC) personnel surveys had been based upon a "preliminary data set with incorrect weighting variables", and consequently lacked "the necessary empirical support to be useful in improving the DMDC survey program." The observations had been made in close collaboration with the statistician who was responsible for the frame development, sampling, weighting and estimation work for the survey, and with the DMDC's supervisory survey statistician, and identified a large number of very basic mistakes with regard to issues that are well discussed in the statistics literature, and that could be corrected by reviewing basic texts. Since 2008, the DMDC's use of the Chi-squared Automated Interaction Detector (CHAID), to identify the best predictors for inclusion in the non-response logistic regression models, has been out-of-line with the standard application of CHAID to assist in non-response adjustments in surveys. Information from DMDC surveys should not be relied upon for formulating personnel policies and decisions until all of the statistical issues have been satisfactorily addressed. Information from DMDC surveys needs to be made more readily available to the public.

Key words: Statistical Methods, Data Quality, 2006 Survey of Active Duty Spouses, Defense Manpower Data Center

From September, 2006, until I retired in July, 2008, I held the position of Lead Survey Statistician with the Defense Manpower Data Center (**DMDC**) of the United States Department of Defense (**DOD**). My principal duties included providing advice on the quality of databases, and on appropriate methods of statistical analysis for sample surveys.[1] To that end, I identified a number of critical deficiencies with the

DMDC's surveys whilst tasked to write statistical methodology reports to describe the sample design, sample selection, weighting, and estimation procedures for the DMDC's 2006 Survey of Active Duty Spouses (**2006 ADSS**) and for the DMDC's the 2006 Survey of Reserve Component Spouses (**2006 RCSS**).[2] Briefly:

- stratification of the sample without proper consideration of the survey objectives rendered impossible the attainment of reportable information for many desired population groups;
- when problems were observed in the sampling, no efforts were made to re-stratify to correct the problems--instead, the decision was made to accept the fact that some of the wanted information would be unattainable;
- extreme over-stratification caused many of the sampling strata to have very small numbers of respondents, both expected and actual;
- logistic-regression models (used first to adjust sample weights for unknown eligibility and subsequently for survey completion among eligible respondents) contained an extremely large number of explanatory variables (plus two-way crossings), which led to absurd weight adjustments--most weights were adjusted very little (or not at all), while a few weights received enormous adjustments (sometimes more than 100-fold);
- no efforts were undertaken either to examine or to mitigate the effects of excessively variable weight adjustments, which can cause severely warped estimates;
- post-stratification cells (used in the final post-stratification adjustment to the weights) were inconsistent with the survey objectives;
- the post-stratification adjustment demonstrated that many survey estimates following the second logistic-regression-model adjustment were off by quite a lot (as much as 42 percent);
- sampling strata were collapsed together to form new “variance strata” for variance estimation, which effected a downward bias in the variance estimates, in order to make the survey results appear more precise and accurate than they actually were. The variance estimates that resulted

from the creation of the "variance strata" were inappropriate because the variance estimates did not reflect the actual sampling design. The margins of error presented with the *2006 ADSS* survey results grossly misrepresented the actual uncertainty associated with the estimates.

Al Nassir *et al.*stated that they had found that the problems that I had identified in the 2006 ADSS were based upon a "preliminary" data set dated July 2006, rather than a final "publicly-available" data set dated October 2006.[3] I worked very closely with the statistician who was responsible for creating the final weights, and with my supervisor, in composing the methodology report for the 2006 ADSS. If I had used the wrong data set, then I am certain that either the statistician or my supervisor would have told me. Neither Al Nassir, Lipari, nor Matos were involved in creating weights or estimates for the 2006 ADSS, although they did have the opportunity to review and comment upon the methodology report during 2007. The 2006 ADSS survey results were never widely shared. Reports and briefings that were presented to the Office of the Undersecretary of Defense for Personnel and Readiness during August and September, 2006, obviously were not based on the October data set. If the Codebook published for the 2006 ADSS is correct, then the final SAS data files were ADSS06B.7BDAT and ADSS06C.7BDAT, which were created in July, 2006.[4] During October, 2006, I noodled a bit with the 2006 ADSS weights, to see whether the weighting methods could be improved. If this is the data set that Al Nassir *et al.* found, and that eventually became the final "publicly-available" data set, then fine.

Al Nassir *et al*. claimed that "each iteration of the weighting process goes through vigorous quality control checks to assure accuracy of the weights.."[5] Prior to 2007, this just did not happen. During 2007, I introduced some basic quality-control checks. If these continue to be used, or if the DMDC has developed some different quality-control procedures since 2007, then fine. During 2006, I observed that survey quality control at the DMDC was abysmal.

Al Nassir *et al.* asserted that the DMDC "expends a great deal of effort to ensure that its methods are in line with current research standards" and "is a firm believer and practitioner of continuous process improvement, including enhancements to its statistical methods."[6] As evidence, Al Nassir *et al.* stated that, in 2008, the DMDC began using the Chi-squared Automated Interaction Detector (**CHAID**) to "identify the best predictors for inclusion in the non-response logistic regression models."[7] CHAID is type of decision-tree method (based upon adjusted significance testing) that has been available since 1980[8], and is an exploratory technique that is an *alternative* to multiple linear regression and logistic regression, particularly when the data set is not well-suited to regression analysis.[9] In adjusting survey weights for non-response, CHAID (and other branching algorithms that have been developed more recently) is typically used to form weighting classes *directly*, thus avoiding the need for logistic-regression models.[10] Using CHAID to identify predictors for inclusion in non-response logistic regression models is highly irregular, and Al Nassir *et al.* may wish to reflect upon their approach.

Al Nassir *et al.* concluded that the "DMDC takes its mission to collect, analyze and report data to support the military community very seriously. While there may be steps that DMDC can take to advance this mission, the suggestions made in Losinger (2010) are not among them."[11] The problems that I had pointed out in the DMDC surveys comprised very basic mistakes with regard to issues that are well discussed in the statistics literature, and constitute relatively simple concepts that could be corrected by reviewing basic texts.

On January 21, 2009, President Obama issued a memorandum, for the heads of executive departments and agencies, to establish a system of transparency, public participation, and collaboration in government, with the goals of strengthening democracy and of promoting efficiency and effectiveness in government.[12] Toxic leadership remains a vexing problem within the DOD, for both military personnel and civilian employees.[13] While many federal managers may feel overwhelmed and threatened by the

President's goals, my objective was to stimulate discussion that would lead to improved surveys and better management at the DOD. Every year, DMDC surveys consume millions of dollars, and the time of hundreds of thousands of survey participants. According to Al Nassir *et al.*, DMDC surveys "have had profound effects upon military spouses and their families, such as directing funding to valued support programs and shaping policy to better reflect the needs of military families."[14] The statistical methods applied to DMDC surveys are poor enough that information from DMDC surveys should not be relied upon for formulating personnel policies and decisions until all of the statistical issues have been satisfactorily addressed.

The DMDC maintains a highly aggressive schedule of surveys, such that individuals may be contacted for multiple surveys each year. The success of a survey is highly dependent upon the willful participation of the members of the sample. DMDC activities are financed by American taxpayers, and the results of the surveys are used as input in decisions that affect millions of Americans. Particularly in light of the President's commitment to unprecedented levels of openness and transparency in government, the Undersecretary of Defense for Personnel and Readiness should be neither afraid nor ashamed to make DMDC survey results fully available to the public. At the very least, the DMDC should make reports from all of its surveys freely accessible on its website.

Notes

1. Lead Survey Statistician Position Description, DOD position description number J827950.

2. Willard C. Losinger, "Notes regarding the 2006 Survey of Active Duty Spouses," *Armed Forces & Society,* 36(2010): 558-570.

3. Fawzi al Nassir, Rachel Lipari and Kenneth Matos, "Response to Notes Regarding the 2006 Survey of Active Duty Spouses", *Armed Forces & Society*, in press.

4. *2006 Survey of Active Duty Spouses: Administration, Datasets, and Codebook*. (Arlington, VA: Defense Manpower Data Center, Report No. 2006-034, 2007), 28.

5. Fawzi al Nassir, Rachel Lipari and Kenneth Matos, "Response to Notes Regarding the 2006 Survey of Active Duty Spouses", *Armed Forces & Society*, in press.

6. ibid.

7. ibid.

8. Gordon V. Kass, "An Exploratory Technique for Investigating Large Quantities of Categorical Data", *Applied Statistics*, 29(1980): 119–127.

9. Evgeny Antipov; & Elena Pokryshevskaya, "Applying CHAID for Logistic Regression Diagnostics and Classification Accuracy Improvement", *Journal of Targeting, Measurement and Analysis for Marketing,* 18(2010): 109-117.

10. J. Michael Brick & Graham Kalton, "Handling Missing Data in Survey Research", *Statistical Methods in Medical Research,* 5(1996): 215-238.

11. Fawzi al Nassir, Rachel Lipari and Kenneth Matos, “Response to Notes Regarding the 2006 Survey of Active Duty Spouses”, *Armed Forces & Society*, in press.

12. Barack Obama, "Transparency and Open Government", Memorandum for the Heads of Executive Departments and Agencies, *Federal Register* 74(15, January 26, 2009):4685-4686.

13. George E. Reed and R. Craig Bullis, “The impact of destructive leadership on senior military officers and civilian employees,” *Armed Forces & Society,* 36 (2010): 5-18.

14. Fawzi al Nassir, Rachel Lipari and Kenneth Matos, “Response to Notes Regarding the 2006 Survey of Active Duty Spouses”, *Armed Forces & Society*, in press.