

SUPPORTING STATEMENT – PART B

B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS

The collection of information will employ statistical methods, and thus the following information is provided in this Supporting Statement:

1. Description of the Activity

This is an outcome evaluation design to test the effectiveness of the Wingman Intervention Training (WIT) program at the Department of the Air Force (DAF), designed to prevent sexual harassment (SH) and sexual assault (SA). The respondent universe includes enlisted First-Term Airmen/Guardians in the DAF. Respondents will be recruited from First-Term Airmen Centers (FTAC) to assure that our sample includes Airmen/Guardians who are new to the Air Force and potentially vulnerable to SH and/or SA.

This is a new collection. Eligible participants are First-Term Airmen/Guardians. Airmen/Guardians in the treatment group will each be recruited to participate in a baseline survey prior to WIT exposure and a follow-up survey six months after WIT exposure. A comparison sample of First-Term Airmen/Guardians on bases not offering the WIT program will be recruited to take the baseline and follow-up surveys on a timeline parallel to the treatment group participants. The current data collection is designed to recruit the following sample:

| | Baseline survey, starting March 2022 | Follow-up survey, starting September 2022 |
|------------------|--|---|
| Treatment Sites | N~2,000 | N~2,000 |
| Comparison Sites | N~2,000 | N~2,000 |

2. Procedures for the Collection of Information

a. Statistical methodologies for stratification and sample selection.

The WIT program at DAF is a one-hour program. Thus, respondents will be First-Term Airmen/Guardians assigned to participate in the one-hour WIT program. These First-Term Airmen/Guardians will be surveyed (on a rolling basis based on entry date to their First-Term Airmen/Guardians Center and thus scheduled WIT training) starting March 2022 and once more (starting in September 2022) via a 6-month follow-up survey. The baseline and follow up surveys will assess whether the intended outcomes of reducing sexual harassment/assault were achieved, and the fidelity forms will be a check on whether all the key WIT curriculum components were implemented correctly. There is no stratification in the sample design; all eligible First-Term Airmen/Guardians are included in the sample. The Airmen/Guardians fidelity feedback forms will be administered to a random sample of five Airmen/Guardians receiving the Wingman Intervention Training. This sample will be selected by the interventionist (the DAF staff member implementing the training) picking a random point on the roster of WIT attendees

and selecting every fifth case until five surveys are completed. One fidelity form will be completed for each WIT session by the WIT interventionist (DAF staff member implementing the training).

b. Estimation procedures.

No sample estimation procedures are included in this program evaluation design.

c. Degree of accuracy needed for the Purpose discussed in the justification.

To ensure the credibility of evaluation findings, NORC has conducted statistical power calculations to determine the credibility of detecting a significant program effect at specific sample sizes. Statistical power provides an estimate of the probability of identifying a relationship through a significant statistical test when, in fact, such an impact exists. To calculate our power estimates we used formulas for computing the expected test statistic found in many power analysis texts in conjunction with Microsoft Excel's routines for evaluating the standard normal curve.

The primary analyses will compare the treatment group DAF bases against the comparison group DAF bases (bases not implementing WIT). Power estimates were computed for N= 4,000 for the first-term Airmen/Guardians for this quasi-experiment (2,000 DAF first-term Airmen/Guardians receiving the treatment and at least 2,000 DAF First-Term Airmen/Guardians in the comparison condition), for the various effect sizes. This calculation is based upon an alpha level of .05, a two-tailed statistical test, and covariates that explain 25% of outcome variation (i.e., at pre-test). An effect of .24 is considered a small effect size based on Cohen's formulation (1988).¹

For this main scenario, our power is over .80 for any effect size of .20 or above (still in the small effect size range). Our power is higher for larger effect sizes. This scenario of 2,000 treatment cases and 2,000 comparison cases will also provide ample power to explore subgroup differences (e.g., differences by sex).

d. Unusual problems requiring specialized sampling procedures; and

We do not anticipate a need for specialized sampling procedures given the study design.

e. Use of periodic or cyclical data collections to reduce respondent burden.

Two data points are necessary to identify change over time. Surveys are administered prior to the intervention and again six months after the baseline survey was completed. Respondents will be administered the web-based baseline survey starting in March 2022 on a rolling basis (based on each Airmen/Guardians entry date to their First-Term Airmen/Guardians Center and thus scheduled WIT training), with a six-month intake period until September 2022. A 6-month follow-up survey starting in September 2022 will follow, the data collection for which will close March 2023.

¹ Cohen J. *Statistical power analysis for the behavioral sciences*. Hillsdale, New Jersey: Lawrence Erlbaum Associates; 1988.

Without a follow-up data collection on all participating DAF bases, we will be unable to assess the outcomes associated with the WIT program. In other words, given the need to identify change over time, two surveys (baseline/follow-up) are the least number of surveys possible.

3. Maximization of Response Rates, Non-response, and Reliability

We propose to compare responders with non-responders in terms of basic aggregated demographic information and adjust for non-response bias with appropriate methods (e.g., non-response weights) if needed. To address item-level missing data (respondents skip some questions), we will first assess the amount of missing data and whether missingness is at random.

There are several facets of the research design that will contribute to a strong response rate for this data collection and thus the overall reliability and validity of the program evaluation effort. The research team has been engaged with DAF HAF/A1Z Resilience Office over the past year to fully understand the context of the sexual harassment and sexual assault prevention programming as well as approaches to survey implementation with first term Airmen/Guardians to ensure a strong design. DAF leadership is also fully informed of and approves the program evaluation design. The recruitment protocols and survey language have been carefully reviewed with DAF staff; with a small panel of consultants which includes experts in the fields of SH and SA prevention, both within and outside the military context; and with a small sample of volunteer DAF Airmen/Guardians to assure that the recruitment and survey language is understandable and acceptable to the target population of first term Airmen/Guardians. NORC has developed the data collection protocols to be consistent with best practices in the field on survey design and implementation. NORC has tested the online anonymous survey link on different web browser platforms using different NORC laptops and personal computing devices to ensure that recruited participants will encounter a user-friendly design without technical glitches, enabling easy survey participation. DAF personnel have also tested the links for connectivity.

Finally, respondents will be offered incentives for survey participation based upon the guidance of DAF HAF/A1Z leadership and following established practices in the field of survey research for reliable and valid data collection. Similar incentive strategies were offered to service members within military service branches to elicit high response rates. In the summer and fall of 2021, DoD Research Site One and DoD Research Site Two had two sets of service members participate in similar surveys as those proposed for Airmen/Guardians. The cohorts at each respective Research Site were administered with different incentive protocols, where the Research Site One cohorts received monetary incentives while the Research Site Two cohorts received one PMI (an opportunity for PM Inspection; i.e., to sleep in later in the morning). Despite Research Site Two's reportedly much higher response rates on prior surveys, the cohort response was about half that of the Research Site One's cohort response. Please see the corresponding table with response rates from the cohorts below.

| Cohort | Active Sample Size (N) | Preliminary Total Completes | Preliminary Percent Complete |
|-----------------------------------|-------------------------------|------------------------------------|-------------------------------------|
| Research Site One-cohort 1 | 1,174 | 1,042 | 88.8% |
| Research Site One-cohort 2 | 1067 | 648 | 60.7% |
| Research Site Two-cohort 1 | 1201 | 946 | 78.8% |
| Research Site Two-cohort 2 | 977 | 308 | 31.5% |

While attrition at follow-up is a common phenomenon in survey data collection, DAF Airmen/Guardians are not usually offered tokens of appreciation for survey participation at their respective Air Force bases and thus the current data collection is expected to achieve a strong response rate. Specifically:

For the baseline survey, participants from the DAF that complete survey will receive a \$10 gift code to Amazon.com. For the follow-up survey, participants that complete a survey will receive a \$15 gift code to Amazon.com. No incentives will be provided for the Airmen/Guardians fidelity feedback form nor the WIT interventionist fidelity form because each will take only five minutes or less to complete.

4. Tests of Procedures

The survey recruitment and survey language were tested through discussions with a small sample of Airmen/Guardians from outside the target study cohorts. The email recruitment connectivity and anonymous online survey connectivity (i.e., the technological aspects of the data collection) were tested with the respective DAF staff involved in planning the data collection. The Airmen/Guardians fidelity feedback form and the WIT interventionist fidelity form were reviewed with DAF staff with extensive knowledge of the operation of the WIT program.

5. Statistical Consultation and Information Analysis

a. Provide names and telephone number of individual(s) consulted on statistical aspects of the design.

Bruce Taylor, PhD
Elizabeth Mumford, PhD
Neha Trivedi, PhD
Lysa Vasquez, MPH

Joshua Lerner, PhD

b. Provide name and organization of person(s) who will actually collect and analyze the collected information.

Cynthia Simko, MA
Neha Trivedi, PhD
Lysa Vasquez, MPH
Mireya Dominguez
Rachel Lipofsky