The Department of Defense Innovations in Sexual Assault Prevention Pilot Program (ISAPPP)

Project 4: Evaluation of the U.S. Air Force's Wingman Intervention Training Program (WIT)

EVALUATION PLAN

Version: Option Year 1, Q2

Presented to:

GSA Federal Supply Service Specifications Section

Suite 8100

470 L’Enfant Plaza, SW

Washington, DC 20407

Presented by:

NORC at the University of Chicago

55 East Monroe Street, 30th Floor

Chicago, IL 60603

January 21, 2022

Table of Contents

Table of Contents....2
Statement of Need....3
Wingman Training Program (WIT) Overview....4
Program facilitators....6
Program participants....6
Exposure to WIT Program for Enlisted Personnel....8
*Logic Model and Theory of Change....9
*Evaluation Stakeholder Workgroup....11
Evaluation Plan....12
Research Questions (RQs)....12
Measurement....13
Process Assessments....13
Evaluability Assessments....13
Process Measures....14
Key Participant Outcomes....14
Contextual Measures....15
*Evaluation Research Design Plan....16
Timeline....19
Human Subjects Protection....19
Office of Management and Budget (OMB)....21
Data Sources....21
Data Collection Procedures....21
Power Calculations....24
*Analysis Plan....24
Data cleaning....24
Nonresponse and Missing Data Bias....24
Baseline Equivalence....25
Attrition....25
Nonresponse Adjustment....25
Process/Qualitative Analysis Plan....26
Descriptive/Exploratory Analyses of Survey Data....26
Multivariable Models....26
Exploratory Analysis of Subgroup Variation....27
Exploratory Analysis of Mediation and Moderation....27
Anticipated Challenges & Solutions....27
*Dissemination Plan....28
References....29

Appendix A. Implementer & Airmen/Guardians Feedback Forms.....33
Appendix B. VPI Survey Results.....37
Appendix C. Data Security.....47

***SAPRO Deliverables**

Statement of Need

Sexual harassment (SH) and sexual assault (SA) continue to be prevalent public health problems nationwide and U.S. military environments are not immune.[1-7] SH and SA incidents, which are part of a continuum of harm,[8] disrupt the sense of safety and cohesion within units and among military personnel, which hamper activities such as recruitment, training, and operational missions.[9] Service member survivors often battle emotional, physical, and mental trauma related to these incidents for the rest of their lives.[8] SA within the military is linked to a variety of negative health outcomes.[9-12] Furthermore, as a work environment, survivors of military sexual trauma often have to reside and work alongside their assailants, increasing the distress associated with the assault experience.[13] Given how problematic these incidents are both individually and organizationally, addressing SH and SA in the military is of utmost importance. Yet, doing so in military settings can be challenging.[14] Despite the high prevalence of SH and SA, there is a "critical gap" in military research on SA prevention strategies.[15,16] There is a lack of formal evaluations of whether existing SH and SA prevention program efforts being implemented in the military, have elicited attitude or behavior change or reduced rates of SH and SA.[16] As noted by the 2021 Independent Review Commission (IRC) on Sexual Assault in the Military, the Department of Defense (DoD) lacks sufficient data to make evidence-based decisions on the impact of prevention activities in military communities, particularly activities aimed at reducing perpetration.[8] The IRC calls for the removal of policy barriers and restrictions preventing research on sexual assault perpetration, for existing legal concerns within the DoD have limited the types of questions and inquiries available for research.[8] The IRC noted that there are distinct causal processes driving victimization versus perpetration and that without complementary research on perpetration—and the unique risk and protective factors that lead to perpetration—the military only has half of the total information needed to paint the full picture of how and why sexual assault occurs in the military.[8] As a result, the impact of prevention activities in military communities, particularly activities aimed at reducing SH and SA perpetration, remains relatively unknown.

Left unaddressed or under-addressed, SH and SA undermine personal safety and unit cohesion, possibly degrading military functioning

The United States Air Force (DAF) has made some important advances in prevention by introducing the evidence-based *Green Dot Bystander Intervention* training (later renamed the *Wingman Intervention Training - WIT)* program in 2015. *Green Dot*, and now *WIT*, is a primary prevention educational intervention designed to reduce power-based interpersonal violence affecting military and civilian members and families. To date, no formal evaluation has been conducted of the WIT program nor of the earlier Green Dot version of the program in DAF. To develop and implement a formal evaluation of the DAF WIT program, the Department of Defense SAPRO partnered with NORC at the University of Chicago (NORC) to conduct an evaluability assessment, develop an evaluation plan, and if feasible, implement this plan.

DoD and U.S. Air Force (DAF) Prevention Strategy

The Department of Defense Sexual Assault Prevention and Response Office (SAPRO) is

responsible for oversight of the Department's SA policy (see https://www.sapr.mil/). SAPRO works closely with the Services and the civilian community to develop and implement innovative and research-based prevention and response activities. The overarching goal is educating Service men and women in prevention efforts to reduce the prevalence of SH and SA within DoD. Each service has its own Sexual Assault Prevention and Response (SAPR) program to fulfill the mission of SAPRO. Within DAF, the Air Force SAPR works together with the Integrated Resilience Directorate (AF/A1Z) to plan, implement and evaluate interpersonal violence prevention programs within DAF.

Sexual Assault in the DAF and the WIT Program

The DAF is composed of 329,839 active duty individuals, including 64,025 officers and 265,814 enlisted Airmen/Guardians, as of 2018.[17] The average age of an officer is 35, the average age for enlisted personnel is 28, and 13% of DAF are officers below the age of 26.[17] The DAF is made up of mostly males, with only 21.1% of DAF being women. Among officers, 22.4% are women, and among the enlisted, 20.8% are women.[17]

An estimated 15.4% of women in the Air Force and 4.0% of men experienced SH.[18] For women, this was a statistically significant increase in SH from 2016 when the rate was 13.2%.[18] There was no change in SH from 2016 for men.[18] SH can lead to the loss of talented Airmen/Guardians—25% of women and 20% of men who were harassed reported that they had taken steps to separate from the military.[19] SA within DAF, like the other service branches, is a significant public health problem linked to negative health outcomes.[9-12] The 2018 Workplace and Gender Relations Survey of Active Duty Members found that in 2018, 4.3% of Air Force women and 0.5% of Air Force men experienced a SA in the past 12 months, representing a statistically significant increase from 2016 (2.8% for women and 0.3% for men).[18]

Sexual assaults are a major reason that some female service members leave the military, and military-related sexual trauma often leaves many female veterans struggling to transition back to civilian life, with some ending up homeless.[14, 20, 21] The DAF maintains the position that SA is a crime in stark opposition to DAF core values and culture of dignity and respect.[22] The DAF encourages the reporting of SAs and supports victims through the military justice system and victim services. In fiscal year 2019, 1,683 SAs were reported to the DAF SAPR Office, which was a 9% increase from fiscal year 2018.[22]

The *WIT program* evaluation plan is organized as follows. We first detail the prevention strategy including a description of the *WIT program* and logic model. Second, we describe the stakeholders who will support the *WIT program* evaluation design and implementation. Third, we discuss our evaluation plan, including: the research questions and hypotheses, our methods for measuring and answering those questions, and the feasibility of doing so. Fourth, we describe our evaluation design, the timeline to implement this design and human subject protections. Fifth, we discuss data collection methods, including data sources, measures, and the power necessary to detect effects. Lastly, we lay out our analysis plans, anticipated challenges and reporting plans

Wingman Training Program (WIT) Overview

In a review of past year exposure to SA training across the U.S. Military, based on responses from over 24,000 active duty personnel who completed the 2010 DoD Workplace and Gender Relations Survey, Holland and colleagues[23]

found that 9% of military personnel self-report having no training on SA during the last 12 months. They also report that 54% of military personnel report receiving comprehensive training on SA, exposing them to approaches to addressing SA actions, interventions, reporting mechanisms, and resources; 30% reported partial training exposure covering some important topics, but missing others; and 7% reported minimal training exposure, missing important topics. In addition, Air Force personnel reported the greatest access to comprehensive SA training and had the lowest rates of SA (compared to the other Services, including the Navy, Marines, and Army).[23]

The DAF *WIT Program*, the focus of this evaluation, is based on the *Green Dot* program. *Green Dot* is an evidence-based comprehensive approach to sexual and domestic violence prevention that capitalizes on the power of peer influence across all levels of the social ecology. It is a bystander intervention approach that teaches tools to address personal, peer, or organizational barriers that impede intervening in high risk situations related to SA.[24] *Green Dot* asks participants to consider two fundamental questions: (1) What kind of responsibility do I feel for the people and situations that cross my path? and, (2) What are realistic options for me when I decide I do want to get involved?[24] *Green Dot* helps participants define their own boundaries and then equips them with realistic ways to get involved or intervene.[24] *Green Dot* targets all members of the community as potential bystanders.[25] It seeks to engage bystanders through awareness, education and skills-practice, in (1) proactive behaviors that establish intolerance of violence as the norm and (2) reactive interventions in high-risk situations, which (hopefully) result in the reduction of violence.[25]

As a result of feedback from DAF leadership on the potential negative inference of the term 'bystander', DAF has adapted and rebranded the program as the Wingman and Leader Intervention.[26, 27] The DAF wanted to focus on encouraging leaders who foster a positive climate to encourage Airmen/Guardians to safely intervene in potential SH or SA situations. In 2020, the intervention name was changed to *Initial-Wingman Intervention* (WIT) Training Program. The main target population for the *WIT* was first time enlisted Airmen/Guardians at technical school or first duty installation. Additionally, the extra violence prevention modules were made optional depending on the Installation Commanders' request to provide prevention programming. These modules include topic areas such as domestic violence prevention, suicide prevention, and addressing drinking behaviors among Airmen/Guardians on installations. NORC is working with DAF staff to determine how frequently these additional modules are being used in the Air Force and if they are of a voluntary nature or are requirements for some Airmen/Guardians to receive.

The *WIT* program is taught by trained volunteer Airmen/Guardians who undergo a 4-hour training program prior to teaching *WIT* sessions to new Airmen/Guardians. The interactive training is designed to equip implementers with the necessary connection, knowledge, and skill to increase their WIT skills and proactive behaviors that set positive norms. It is designed to strengthen the deliberate use of peer influence and to harness natural leadership within subgroups. The applied training equips implementers to teach participants how to take immediate action and about the impact they can have on reducing SA in DAF, if they take action. The *Green Dot* bystander program, originally developed and implemented in high school and college settings, focuses on bystander training to engage students in actions to reduce sexual violence.[28, 29] As with the *Green Dot* curriculum, WIT seeks to empower potential bystanders to actively engage their peers in both reactive responses (e.g., helping victims of dating or sexual violence), and proactive responses (e.g., safely but effectively interacting with potentially violent peers and potential victims to reduce violence risk).[24] The four essential steps for *WIT* can follow the *Green Dot* metaphor or be presented independent of the metaphor. Participants are trained to recognize situations and behaviors that can contribute to violence and determine actions they could safely take to reduce the likelihood or effect of violence.[30] These active bystander behaviors are called "Green Dots" to distinguish them from "red dots" or behaviors that may contribute to violence.[30] *Green Dot* presentations oriented participants to their potential role as

engaged bystanders and explained how to recognize “red dots” and “green dots.”[30] The Air Force has altered the following terminology from the Green Dot program (Table 1):

Table 1. Adaptation of *WIT* Program Terminology

WIT Program Terminology	***Green Dot Program Terminology***
Warning signs	Red dots
Reactive *WIT* behaviors	Reactive green dots
Barriers	Barriers
Proactive behaviors	Proactive green dots

The *WIT Program* can take place in either a large lecture format or a small classroom format in either Technical School or First-Term Airmen/Guardians Centers. Therefore, class sizes can vary by the number of enrolled Airmen/Guardians in a given year. The maximum group size is 50, though ideally, DAF tries to limit group size at 25-35 Airmen/Guardians to keep the training relationship-based and personal. Smaller groups ensure everyone is engaged and that participating Airmen/Guardians get adequate practice applying newly learned skills.

The 60-minute *WIT* curriculum uses a range of different types of activities. As depicted in Table 2 (see below), the 60-minute curriculum has four main components: (1) Background and “Basic Bones” of the program, (2) introducing the 3 Ds of the program, (3) learning about “Proactives” and (4) building a sense of commitment to change for the Airmen/Guardians and a summary of the lessons learned. The curriculum is designed with the bystander lens as the central frame of reference and guided by the question, “What does the participant need to know in order to increase the likelihood of effective proactive and reactive bystander response to SA and other forms of interpersonal violence?”[31] The focus of *WIT* is on SA prevention but the DAF believes *WIT* can be useful for preventing related behavior of SH, intimate partner violence and possibly, self-directed violence. The curriculum includes a facilitated discussion approach to delivering the material and consists of lecture, scenario-based demonstrations (re: Three Ds) and practice, and videos. Upon completing the training, students are provided with a QR code and asked to complete a brief non-mandatory feedback survey about their experience in the training (see below under Process Data).

Program facilitators

WIT is facilitated by trained volunteers within DAF. *WIT* facilitators should be attending a 4-hour training provided by Violence Prevention Integrators to be able to implement the program. NORC is still exploring with DAF whether there is variation in the number of hours of training that the implementers receive. The training is designed to strengthen the deliberate use of peer influence and to harness the natural leadership within subgroups. Violence Prevention Integrators play an important role in observing and evaluating *WIT* facilitators to promote implementation fidelity. The Violence Prevention Integrators use the Direct Observation Tool to assess five key dimensions of implementation fidelity: Adherence, Exposure, Quality of Delivery, Participant Responsiveness and Program Differentiation.[32]

Program participants

The target population for the *WIT* evaluation will be newly enlisted Airmen/Guardians/Guardians (ages 17-24 years old); however, there are exceptions for older individuals who are new enlistees in DAF. Program participants receive the intervention in their First-Term Airmen/Guardians center (FTAC). The plan is to have program participants complete baseline surveys prior to exposure to the WIT, and then follow-up surveys shortly after the intervention at 6 months post intervention.

Table 2. Brief Descriptions of the *WIT* Activities:

	Activity	**Components**
Introduction (20 min)	Background and "Basic Bones" of program	Segment 1: PowerPoint presentation • Personal introduction. • Setting goals to reduce the number of people who experience interpersonal violence. • Setting the problem that too many Airmen/Guardians are impacted by sexual assault and domestic violence. • Changing the culture to reduce SH and SA. Segment 2: PowerPoint presentation • Introducing the U.S. map of SH and SA. • Transitioning to installation level SH and SA map. • Introducing the Red Dots. • Activity 1: Recognizing red dots. • Introducing Green Dots. Segment 3: PowerPoint presentation • Introducing barriers: Barriers are things that often stop us from getting involved. o Personal barriers, for instance: being introverted, or being afraid of a physical escalation or retaliation, or being unsure, or not wanting to make a scene or embarrass yourself. o Relationship or social barriers, for instance: not wanting to break an unspoken rule in your group, being perceived as a squeaky wheel, or feeling uncomfortable confronting a buddy. o Organizational barriers, for instance: rank, lack of support in your unit, or concerns about career.

Lecture (20 mins)	Introducing the 3 Ds	Introducing the 3 Ds • Discuss the 3 Ds and scenarios of what Airmen/Guardians can do it different situations: o *Direct*: Do something yourself, like ask someone to stop what they are doing or check on someone you might be worried about. o *Delegate*: If you can't do something directly because of your barriers, ask their friends to help. Talk to a trusted First Sergeant, Commander or fellow Airmen/Guardian in your unit. Tell the bartender or ask a family friend to check in. Leave an anonymous note for a trusted colleague or chaplain you think may be able to help. o *Distract*: If you don't want to address the situation directly or even acknowledge you see it, try to think of a distraction that will defuse the situation or calm things down in the moment. A distraction might be "accidentally" spilling a drink or asking to borrow the phone of someone who is in a risky situation, or asking for a ride, or starting an unrelated conversation. Activity: 3 Ds • Pick from the following scenarios and present to Airmen/Guardians on how to react: o At a party, someone looks really drunk as they are being led to a bedroom away from the group. o You notice a couple in a parking lot arguing loudly. He grabs her by the arm aggressively and you are getting concerned about the rapid escalation. o At an after-hours social gathering, you notice someone more senior leaning into and touching someone who looks uncomfortable and keeps trying to back away. Activity: Red Dot prompts • There are general scenarios below you can use as prompts for the 3 Ds activity. It is important that you adapt these scenarios to match your participants. Consider the context they are in, well-known reference points, language, types of people they interact with, etc.
Activities	Learning about "Proactives"	Segment 4: PowerPoint presentation There will be an Optional Clicker Question implemented by the trainer: • Emotional response reflects the anonymous disclosures of violence Activity: Proactives • PowerPoint slides with images and prompts to discuss what Proactive Green Dots are. They are things we can do to stop red dots before they even start. They begin to reset base norms and make it clear: (1) Interpersonal violence is not okay, and (2) Everyone is expected to do their part.
Review and Closing	Commitment to change & Summary	Trainers will end with discussion the importance of committing to change to reduce SH and SA within the DAF. • The important thing is not necessarily what you do, but **that you do something**! Be realistic about what you will and won't do and spend some time thinking about options that work for you. No one has to do everything, but everyone has to do something. • Prompt: Ask Airmen/Guardians to think about a proactive *Green Dot* they would do by the end of day and ask them to share their commitment with a person near them to reduce SH and SA.

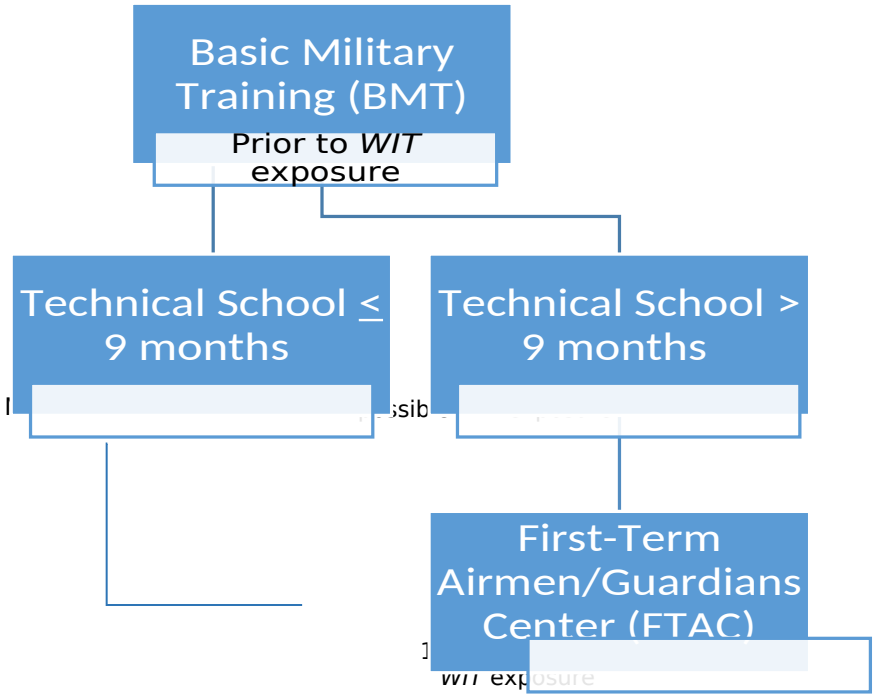
Exposure to WIT Program for Enlisted Personnel

WIT is a 60-minute training program currently implemented at technical school and/or within an Airmen/Guardians first duty installation or First-Term Airmen center (FTAC). The foundational phase of the SA prevention program required the Total Force to participate in the prevention training; however, the current participants of *WIT* Program are enlisted Airmen/Guardians in their first year. Given the COVID-19 pandemic, DAF has taken precautions to make sure regulations are in place for *WIT* to continue in-person during technical school and FTAC.

Figure 1 (see below) outlines when enlisted Airmen/Guardians (the program's and this evaluation study's target population) get exposed to *WIT* across most DAF bases within and outside the US. There are multiple touch points for enlisted Airmen/Guardians to receive *WIT*. Once individuals are enlisted into DAF, they go through Basic Military Training (BMT) where they spend seven weeks (although, prior to the COVID-19 pandemic, BMT required eight and a half weeks) learning the basics of DAF life to condition one's body both mentally and physically. Upon completion of BMT, enlisted Airmen/Guardians will attend technical school to train for their specific career path. DAF Technical Training varies in length depending on the chosen career field. Since training for various career paths differs, the length of technical school can span from 6-72 weeks.[26] Once technical school is completed, Airmen/Guardians then go to their first duty installation at FTAC to continue training for their career path. Exposure to the *WIT* program can occur more than once for first year Airmen/Guardians, depending on timing of individual Airmen/Guardians' enlistment day into DAF. Enlisted Airmen/Guardians enrolled in a technical school that lasts less than nine months will not receive *WIT* training until they reach their first installation, FTAC. Technical schools lasting longer than nine months *may* provide *WIT* training at the technical school. All enlisted Airmen/Guardians will be exposed to *WIT* training at FTAC; therefore, it is possible for the Airmen/Guardians to be exposed twice to the *WIT* program during a calendar year.

NORC has been conducting one-on-one meetings with Violence Prevention Integrators (VPIs) to understand the current status of *WIT* program implementation. As of January 21, 2022, NORC has met with fifteen bases that will potentially be part of the treatment group (and eight bases that will be comparison bases). We have learned that the vast majority of the service members receiving *WIT* are First-Term Airmen/Guardians. There may be some Guardians from the U.S. Space Force also participating in the evaluation. Since the WIT training cohorts that include both civilians and officers are usually small, the evaluation will not include these groups. Additionally, NORC has been meeting with VPIs at the potential comparison-group bases (those bases that are not currently implementing *WIT*). To date, we have met with eight bases that are not implementing *WIT* but are conducting other forms of SH or SA trainings, targeting response strategies for victims but not conducting SH or SA prevention programming. These trainings, for the most part, are the annual required trainings all Air Force personnel must complete as part of Sexual Assault Prevention and Response (SAPR) requirements.

Figure 1. Schematic for possible locations for WIT program exposure



*Logic Model and Theory of Change

The *WIT program* is theory-based and supported by research drawing from bystander psychology and social-psychological research addressing active bystander behaviors and how and why some bystanders intervene. In addition, diffusion of innovations theory plays a role in how active bystander behaviors might be diffused from person to person to engage their peers to intervene as a bystander with them, thus increasing the likelihood that new norms will be spread across the community.[33] *WIT* also draws from research on how sexual violence perpetrators target victims, and the motivations and antecedents to sexual violence, helping bystanders better assess situations, identify potential risks for violence, view their options for action, and select safe active bystander behaviors that they are willing to carry out.[33] Bystander models call for all members of the community to have a role in shifting social norms to prevent violence. The bystander literature provides guidance on which factors or predictors increase the likelihood that a bystander will intervene to prevent violence.[33]

Coker outlines five key concepts in the theory of change related to the closely aligned *Green Dot program*, on which *WIT* is based, and notes how using a bystander framework attends to these factors to help prospective bystanders overcome barriers to become active bystanders:[33]

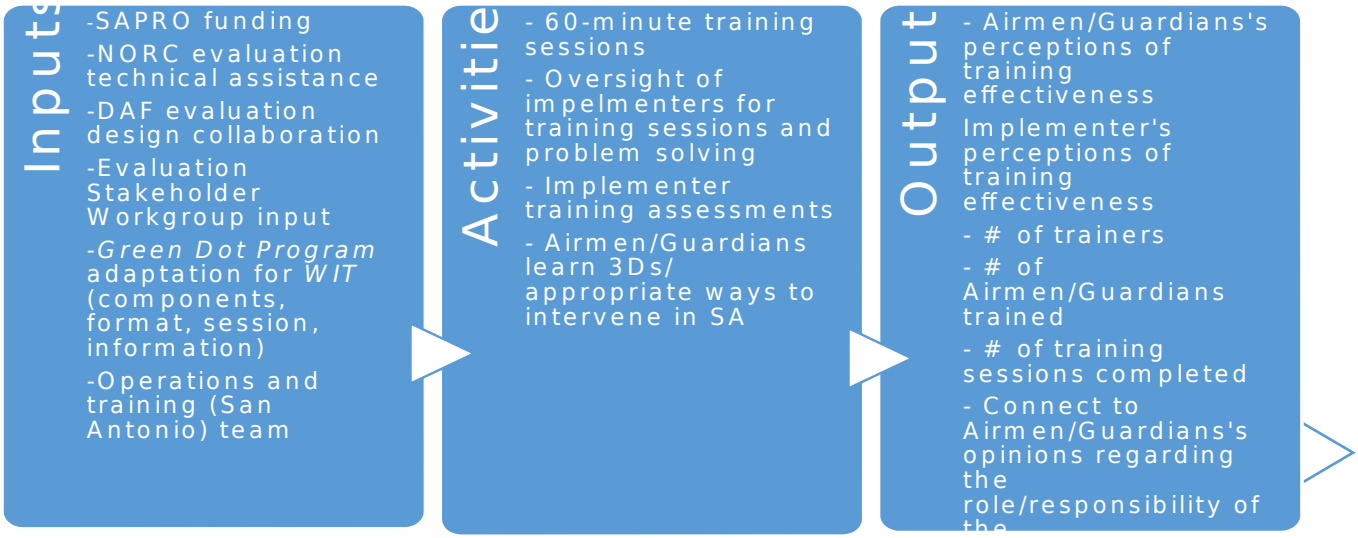
(1) *Diffusion of responsibility* or the concept that individuals are less likely to respond in a crisis when more people are present because each assumes that someone else will handle it.
(2) *Evaluation apprehension* in which individuals are reluctant to respond in a high-risk situation because they are afraid they will look foolish.
(3) *Pluralistic ignorance* or the likelihood that when faced with an ambiguous, but potentially high-risk situation, individuals will defer to the cues given by those around them when deciding whether to respond.

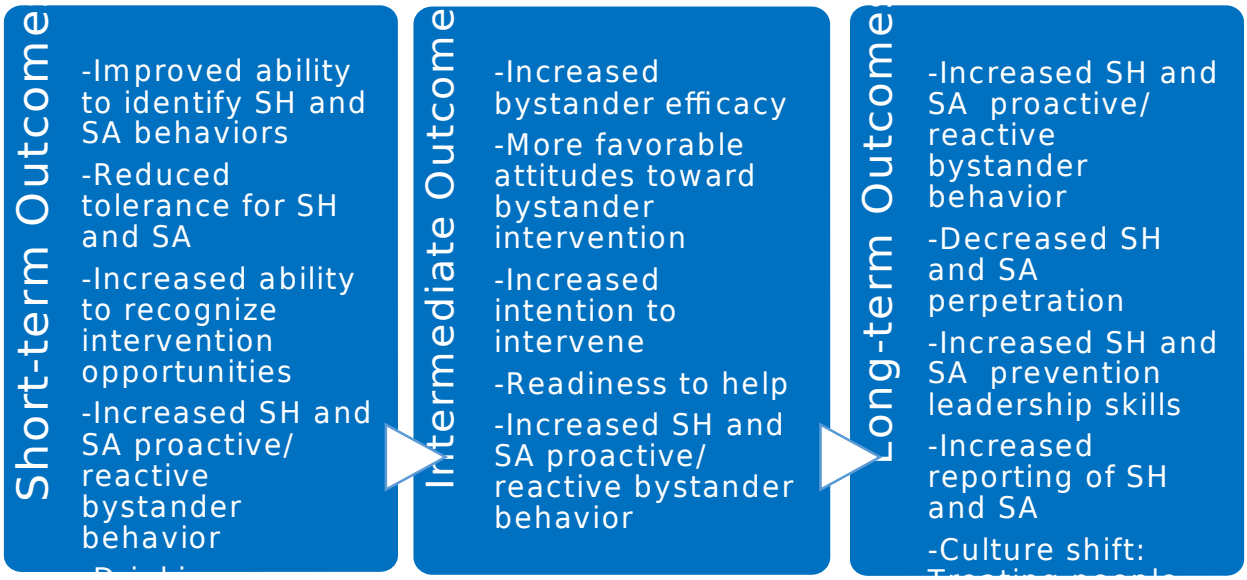
(4) *Confidence in skills* in which individuals are more likely to intervene in a high-risk situation when they feel confident in their ability to do so effectively.
(5) *Modeling* in which individuals are more likely to intervene in a high-risk situation when they have seen someone else model active bystander behaviors first.

The *WIT* logic model (see Figure 2) describes how the inputs and activities combine to create the program and details the short, intermediate, and long-term program goals (outcomes). The logic model is useful for understanding programmatic aims and vision and serves as a guide to identifying and determining specific outcome measures. The *WIT* curriculum is comprised of several in-person activities that focus on building understanding of when bystanders should intervene during potential acts of SH or violence. Learning aids such as the "3Ds" and scenarios are used to help build bystander knowledge, efficacy, and readiness to intervene. The inputs for the training include the training program developers who are in the DAF San Antonio office and Violence Prevention Integrators (VPIs) implementing training on installations, staff and other stakeholder feedback, and adaptations from the *Green Dot* Program. Outputs of interest include understanding Airmen/Guardians' perceptions of the effectiveness of the training curriculum, as well as the trainers' perceptions of information absorption and dissemination among Airmen/Guardians. Additionally, it is important to track how many volunteer Airmen/Guardians were trained, how many Airmen/Guardians attended the session and how many sessions were completed to assure fidelity of the program.

Short-term goals include improving Airmen/Guardians' ability to identify SH and SA behaviors when seen in the community, reducing Airmen/Guardians' overall tolerance for SH and SA behaviors, and increasing ability to recognize intervention opportunities. Intermediate goals include increasing bystander efficacy and overall confidence to intervene, having more favorable attitudes toward bystander intervention while increasing intention to intervene and one's readiness to help in a situation. Long-term goals of the training are increased bystander behavior, decreased SH and SA perpetration, increased SH and SA prevention leadership skills, and increased reporting of SH and SA.

Figure 2. Working Version of the WIT Program Logic Model





*Evaluation Stakeholder Workgroup

With DAF and SAPRO collaboration and input with the evaluation team, a stakeholder workgroup has been convened and operational since November 2020. The Evaluation Stakeholder Workgroup (ESW) continues to meet weekly with the core members to determine the evaluation timeline, establish project protocols, discuss the evaluation design, and assess issues with Airmen/Guardians recruitment to the study. The ESW also offers expert advice on when additional ESW members need to be engaged into the conversation, either directly or indirectly and sets up additional communication meetings as necessary. The ESW assists in the development of the evaluation plan and later evaluation. The ESW for the *WIT* evaluation is structured to ensure that we use the right metrics, guide the selection of protocols for implementation, interpret the results appropriately, and have an audience prepared to use the results. The purpose of including stakeholders into the evaluation planning and implementation is to have individuals and groups who are involved and invested in SH and SA prevention lend assistance in understanding the value of the *WIT* and discuss the short-term and long-term effects of such a program in DAF. The stakeholders include individuals within various parts of DAF and are not exclusively evaluation experts. The stakeholders are people or organizations invested in the program, interested in the results of the evaluation, and/or with a stake in what will be done with the results of the evaluation. The stakeholders have been providing feedback on program context, content, and implementation protocols, giving input on evaluation design decisions, engaging in evaluation implementation cooperative partnerships, and will later help interpret results and disseminate lessons learned. In partnership with DAF, NORC developed a working list of ESW members (see Figure 3) that has evolved over time. Major Pound (A1ZR), Chair for the ESW, joined the team in February 2021. Additional members of the ESW include:

- Dr. Melissa Lynes (A1Z),
- Ms. Amy Jensen (DPFZ),
- Ms. Wendy Link (DPFZ- San Antonio office),
- Ms. Edith Davis (DPFZ- San Antonio office),
- Ms. Shelby Jones (A1ZR), and

- Ms. Cynthia Jean-Baptiste (A1Z).

In addition to this leadership group, we are working with DAF to identify the MAJCOMs (VPPMs) and Installations (VPIs and commander leadership at the installation) to include in the outcome evaluation. Once those sites are confirmed, specific VPPMs, VPIs and base commanders will be invited to participate in the ESW.

As of January 21, 2022 the VPIs at the following 26 bases are part of the ESW (only 23 have agreed to formally participate in the evaluation): 1) Seymour Johnson AFB, 2) Goodfellow AFB, 3) Dover AFB, 4) JB McGuire Dix-Lakehurst, 5) Spangdahlem AFB, 6) Yokota AFB, 7) Misawa AFB, 8) Vance AFB, 9) Holloman AFB, 10) Eielson AFB, 11) JB Lackland, 12) Sheppard AFB, 13) Whiteman AFB, 14) Offutt AFB, 15) Beale AFB, 16) Fairchild AFB, 17) Moody AFB, 18) Kunsan AFB, 19)Laughlin AFB, 20) Kirtland AFB, 21) JB Randolph, 22) Minot AFB, 23) Hill AFB, 24) Dyess AFB, 25) Columbus AFB, 26) Peterson SFB. We are inviting base leader participation to make sure that these additional ESW members are specifically connected to the evaluation as their MAJCOM/base is enrolled in the evaluation. More generally, members were selected for their investment in SH and SA prevention strategies (especially familiarity with *WIT*) and for their necessary input in evaluation construction and implementation. Stakeholders will be engaged throughout the evaluation plan development process and their involvement will continue through evaluation implementation. Based on our conversations with DAF personnel the following groups are members of the ESW:

Figure 3. Expected Evaluation Stakeholder Workgroup membership for WIT evaluation

Violence Prevention Integrators (VPIs)	Installation level leader facilitating WI
Violence Prevention Program Managers (VPPMs)	MAJCOM level leader, working with multiple installations to facilitate WI
Commanders	High ranking officials at specific installations interested in WI
Operations and Training Personnel	San Antonio office personnel developing WI
Older Airmen/Guardians from Technical School	Upper class Airmen/Guardians who have experienced the WI
Chair & Co-Chairs	Leaders invested in the evaluation of the WI

Evaluation Plan

Research Questions (RQs)

The overarching research question of this outcome evaluation is whether the *WIT* program is effective in achieving its intended goals. The working hypothesis to be tested is that the *WIT* program will result in improvements on the short, intermediate, and long-term outcomes of U.S.

Airmen/Guardians. We have several process and outcome research questions included to assess the intermediate steps informing accomplishment of the overarching program goals, as follows:

RQ1. What is the impact of *WIT* on short-term outcomes?
RQ1a. Will the *WIT* impact Airmen/Guardians ability to identify SH and SA behaviors, to recognize intervention opportunities and to reduce tolerance for SH and SA?

Process questions:
RQ1b. Do trainers complete feedback forms?
RQ1c. How many Airmen/Guardians attend the sessions?
RQ2. What is the impact of *WIT* on intermediate-term outcomes?
RQ2a. Do Airmen/Guardians increase bystander efficacy through the *WIT*?
RQ2b. Are Airmen/Guardians attitudes towards intervening as a bystander impacted by the *WIT*?
RQ2c. Do Airmen/Guardians feel prepared to intervene as bystanders?
RQ3. What is the impact of *WIT* on long-term outcomes?
RQ3a. Do Airmen/Guardians indicate positive proactive bystander behavior to address SH and SA they observe?
RQ3b. Do Airmen/Guardians demonstrate better SH and SA prevention leadership skills?
RQ3c. Do participants report decreased incidence of SH (victimization/perpetration)?
RQ3d. Do participants report decreased incidence of SA (victimization/perpetration)?

Measurement

In this section we cover the instrumentation for measurement of outcomes that will be used in the *WIT* program evaluation plan. Each phase of the evaluation plan will include various data sources (i.e., background administrative data, survey data, qualitative data, etc.) that will be used to address the research questions of interest and conduct analysis for overall intervention outcomes. Instrumentation for the *WIT* program have been fully developed for the intervention in line for implementation in the fall of 2021. Instrumentation for the *WIT* evaluation has been refined and finalized with the ESW; we have conducted cognitive interviews prior to final Internal Review Board (IRB) review and implementation of the Airmen/Guardians baseline survey. NORC continues to work with leadership to ensure that all required administrative and IRB protocols are complete prior to any survey administration that is part of this evaluation.

Process Assessments

Several process measures will be tracked and documented, along with Airmen/Guardians survey response rates informing attrition analyses. Fidelity of *WIT* implementation will be documented in several ways. While parts of *WIT* is scripted and the volunteer trainers are trained on delivery, we will assess whether key content is conveyed in all sessions. The fidelity implementer feedback form, already implemented by *WIT*, will continue to provide insights to the fidelity of *WIT* implementation (e.g., tracking whether trainers are teaching all the components to the 60-minute *WIT* session). However, NORC, working with DAF, has revised the feedback forms to better support the evaluation of *WIT*. Updated versions of the feedback forms are included in the Appendix A.

Evaluability Assessments

As part of our evaluability assessment, NORC developed and conducted a Qualtrics survey distributed to 92 VPIs to assess the implementation of *WIT*. Results from the VPI survey has informed the evaluation design. NORC has been working with the DAF to review the project logic model (Figure 2, above), which has been finalized by the DAF team. Based on data collected to date from the VPIs overseeing *WIT*, there are no other changes needed to the model and associated measures. If new information surfaces on the implementation of the *WIT* program during this planning period, NORC will update the logic model as necessary to align with current practices. Based on collaborative conversations with DAF, NORC has recommend several concrete steps to improve the evaluation (e.g., content review process, implementer readiness/consistency, stakeholder input/feedback, incorporating quality improvement cycle, developing data collection system, improving understanding of context through focus groups and interviews, etc.).

Process Measures

Feedback forms will be collected from Airmen/Guardians for tracking of engagement and satisfaction in training as well as fidelity of training sessions. In addition, fidelity feedback forms will be collected from implementers (VPIs) at the end of the session to ensure consistency of *WIT* program dissemination.

Key Participant Outcomes

The logic model details short-term, intermediate, and long-term outcomes targeted by the *WIT* program. Below we provide more details on these outcomes. Each of these outcomes has been reviewed by the ESW. The ESW expressed support for these measures but still would like to prioritize them to include a manageable number of items to keep the burden of completing the survey workable for the Airmen/Guardians. Below we have grouped the measures into short, intermediate, and long-term outcomes. Lastly, the ESW provided specific input while reviewing the survey instrument, and we have included a small number of items based on their suggestions.

Short-term outcomes

An overarching short-term outcome for Airmen/Guardians is to have better *understanding of SH and SA issues and connection to their peers*. In order to assess *ability to identify SH* and *SA behaviors*, SH victimization and perpetration as well as SA victimization and perpetration will be measured by separate scales.[34-36] The recently released IRC report notes that legal concerns regarding privacy in research participation can be addressed by following standard industry practices for collecting sensitive data (e.g., anonymous data collection and protection of confidentiality) that the DoD currently leverages when gathering data on other illicit forms of behavior.[8] NORC will use an anonymous survey approach to administer the project survey, inclusive of perpetration items. Measurement of *reduced tolerance for SH* and *SA attitudes* is a key precursor for active bystander intervention in and increased reporting of SH and SA behaviors. As recommended by the ESW, rape myth acceptance will not be measured, but self-efficacy and sexual assault will be measured to capture the impact of contextual factors, alcohol use and the relationship between the victim-offender on sexual self-efficacy.[37] Additionally, the ESW wanted to understand the impact leaders within DAF have on addressing SH and SA,

therefore a few items from the DEOCS Questionnaire (https://www.defenseculture.mil/Assessment-to-Solutions/Factor-Products/Protective-Factors/#transformational-leadership) on transformational leadership were included.[38, 39] Next, in order to fully capture how an Airmen/Guardians would react to reporting SA, empathy and support for victims will be measured.[40] Lastly, the ESW wanted to get a better understanding of unit culture and support, which will be captured by measuring scenarios Airmen/Guardians may have experienced regarding engaging in bystander intervention. These last set of bystander items were designed by NORC and the ESW.

Intermediate outcomes

Bystander Constructs: We will measure several outcomes that may contribute to the improved attitude and self-efficacy to enact a bystander intervention. *Increased ability to recognize intervention opportunities* when situations arise can be measured by a bystander intentions scale, focusing on Airmen/Guardians and friends.[41, 42] A bystander efficacy scale will measure individual level confidence among Airmen/Guardians regarding SH and SA scenarios.[42, 43] In addition, we will measure one's self-reported readiness to help to capture engagement in certain bystander behaviors.[44] Two scales for reactive bystander behaviors will also be used to assess reactions to low risk situations and scenarios.[45] Additionally, at the request of the Air Force, we are including the Sexual Self-Efficacy Rating (SER) scale to measure individual self-efficacy covering the domain of sexual behavior.[37, 46]

Descriptive Norms: Research suggests that our beliefs influence our behavior.[47] Indeed, human beliefs and perceptions about prevention behaviors, others' expectations, others' behaviors, control, and self-efficacy influence whether or not they will engage in a particular behavior.[48] Especially relevant for prosocial bystander behavior (i.e., supportive wingman behavior) are descriptive normative beliefs about perceptions of what the people around us actually do (as opposed to injunctive norms which are subjective beliefs about pressure to comply with norms and doing what others think one should do). To measure the social norms within the peer climate a descriptive social norms scale (modified from Banyard and colleagues work)[49] will be used. Additionally, as requested by the Air Force, we will be asking four items regarding unit culture and support. These items will assess the response Airmen/Guardians receive, from other Airmen/Guardians, when intervening during an unwanted sexual experience, as well as the chances of not responding due to stigma or shame when presented with an opportunity to intervene.

Long-term outcomes.

The *WIT* evaluation is not focused on measuring long-term outcomes, but we have two measures that will be assessed over time. First, *reductions in SH and SA* can be measured through DAF aggregate administrative data documenting SH and SA reports. While informal and formal reports are sensitive to Airmen/Guardians confidence in the value and safety of reporting, as much as in the prevalence of SH and SA over time, and there is the counterforce of increased safety resulting in increased reporting, this metric is likely to be an ongoing indicator of the climate at DAF. Second, we will have self-report survey measures on victimization and perpetration of SH and SA. While the ESW expressed some concerns about asking perpetration items in the Airmen/Guardians survey, through discussion, NORC got feedback that using a respondent-generated identification number (see human subjects' section for more details on this number), to make the data collection anonymous, would offer additional protections that will help in making participating Airmen/Guardians more comfortable to answer such items. A respondent-generated identification number may also make collection of perpetration measures more palatable for DAF leadership as a mechanism to protect Airmen/Guardians.

Contextual Measures

We will measure multiple constructs known to correlate with the dependent variables. These measures will be collected in the Airmen/Guardians surveys.[50] Some of the items will cover demographic characteristics of the Airmen/Guardians such as sex, sexual orientation, and race/ethnicity. Responses to the sexual orientation and race/ethnicity questions will be re-coded by algorithm, collapsing responses into binary straight/not straight and White/non-White categories, before data is viewed by the research team. This additional step will serve to (a) increase personal validation by allowing all respondents to select the sexual orientation and race/ethnicity that best describes them, and (b) ensure respondent anonymity. Under this protocol, it will not be possible to regain the detail on these two characteristics for later analyses (or in the unlikely event of a data breach).

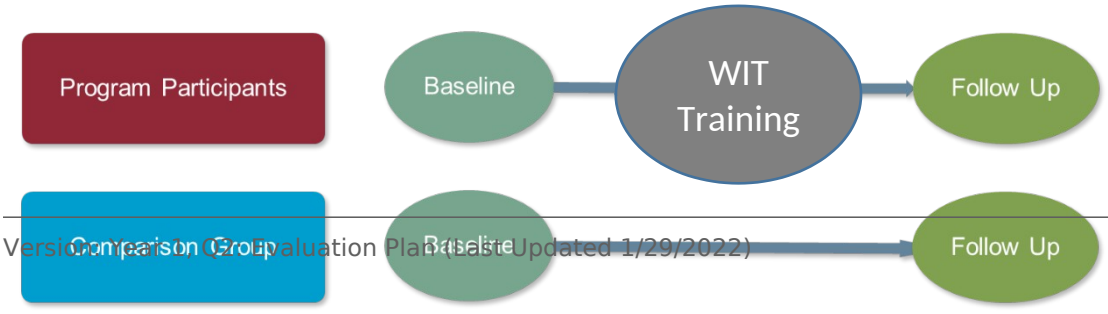
*Evaluation Research Design Plan

A key goal of this study is to determine the effectiveness of the *WIT* program for increasing bystander intervention and decreasing SH and SA using a rigorous evaluation design. We will produce methodologically rigorous evidence on the effect of *WIT* on key outcomes that will provide DAF the necessary information to make informed decisions and policies moving forward. Given SAPRO's call for rigorous evaluations, we explicitly looked at balancing the rigor of implementing a randomized controlled trial (RCT) or quasi-experimental design (QED) against the unique requirements of DAF and the need for feasibility within the identified settings and their level of "research readiness." The most critical element of this evaluation design is the selection of a valid counterfactual comparison group. Selecting a comparison group that is like the treatment participants on observed characteristics, allows us to determine that any systematic differences in their subsequent outcomes can be attributed to the treatment group having received program services, rather than to any systematic differences in the characteristics of the two groups (See Figure 4). The goal is to produce methodologically rigorous evidence on the effect of the *WIT* on the key outcomes that will provide DoD with the necessary information to make informed decisions and policies moving forward.

RCTs are the 'gold standard' in evaluation research. Subjects are randomly assigned into a treatment group and a control group. This leaves the likelihood of an individual's assignment to *WIT* or the control group to chance. As the study is conducted, the only expected difference between the control and treatment groups is the relative effectiveness of the *WIT* program. While a RCT is the gold standard for measuring program effectiveness, it is not feasible to withhold the *WIT* intervention as it is part of an ongoing DAF program.

With the RCT design deemed not feasible, as *WIT* is part of an ongoing DAF program and cannot be withheld, we considered other designs such as regression discontinuity, QEDs with property-score matching, difference-in-differences, comparative interrupted time series analyses, instrumental variable methods, and regression point displacement. After extensive consultation with the ESW, we decided to plan on using a QED. QEDs involve statistically matching program participants to similar individuals who do not receive the program. A QED will provide rigorous, unbiased estimates of the impact of *WIT* on the various outcomes of interest.

Figure 4: Comparison group design



To capture the breadth of implementation of the *WIT* program, as discussed above, NORC developed a survey for VPIs to complete to assess whether *WIT* is currently being conducted. Appendix B shows the results of the survey, along with proposed treatment and comparison bases. As seen in Appendix B, the response rate for the VPI survey was 77% and details regarding implementation of *WIT* by MAJCOMs and bases can be found in Tables A1 and A2. We found 35 bases reporting full implementation of *WIT* in our VPI survey. Table A3 indicates the average number of Airmen/Guardians receiving *WIT*, monthly, at Tech School, During FTAC, and After FTAC. NORC has narrowed down the number of DAF bases that would be good candidates to serve as *WIT* 'treatment sites' for the evaluation (Table A4). In addition to the bases reporting full *WIT* implementation, eight bases have many Airmen/Guardians who are receiving *WIT* at the base each month. Our survey also has helped to identify 15 bases that are not implementing any elements of the *WIT* program. After we formally select bases to serve as treatment sites, we will select a matched group of comparison bases not implementing *WIT* of a similar size and type. The hope is to select a relatively small number of large sites (n~5) with the strongest implementation of *WIT* to serve as treatment sites to keep the burden for DAF to a minimum (likewise for five comparison sites not implementing *WIT*). To facilitate comparability, we will attempt to find matched pairs of treatment and comparison sites within the same MAJCOM and likely only select bases in the US.

Of bases that indicated full implementation of *WIT*, almost half (45.7%) stated that they are currently conducting the training only once a month, while some (22.2%) stated that conduct the training once a week (Table A5). To pick the treatment sites for the outcome evaluation, the VPIs that responded to the survey were asked specifically about training for implementers and average number of Airmen/Guardians receiving the program on their bases. Twenty-five percent of the VPIs stated that approximately 11-20 implementers are trained annually on their own bases, and 20% of the VPIs project the same number of implementers trained in 2021 (Table A7). Approximately 48 (68%) installations reported that implementers lead 1-2 sessions per a month (Table A8). Approximately 50% of implementers phase out of the position usually within one-two years (Table A9). VPIs indicated this turnover is due to Permanent Change of Station (PCS), retirement, or deployment. VPIs were then asked about fidelity checks for conducting *WIT* sessions and feedback from Airmen/Guardians on the training sessions. VPIs reported that 38 installations (54%) completed implementer fidelity checks to verify that the sessions are being conducted consistently (Table A10). These fidelity checks are being completed by VPIs themselves, or Lead Implementers, and are done predominantly by paper at 23 installations (32%), while a few installations complete the fidelity checks in-person by visual observations (Table A10). Lastly, a majority of VPIs stated that 52 of the 71 installations (73%) do not have Airmen/Guardians complete fidelity feedback forms (Table A11). Airmen/Guardians will be asked to complete a separate feedback form during the main study to collect process measures to indicate areas of the *WIT* curriculum that may need to change.

To confirm the current implementation of the *WIT* training program on bases, NORC has been conducting one-on-one sessions with VPIs at bases that fit the criteria to be assigned either to the treatment group or the comparison group (or neither condition) to determine their possible interest in being involved in the upcoming evaluation. These meetings have proven to be very useful in gauging interest and suitability for the evaluation. Except for two bases that are not implementing *WIT* (i.e., JBSA Lackland, Columbus AFB), all the bases that were determined to be

likely good candidates for the treatment group based on the VPI survey informed NORC that they are still fully implementing *WIT*. Also, these VPIs plan to continue *WIT* implementation going forward except for Ramstein AFB due to current orders regarding refugee assistance, and JB Charleston due to VPI leaving the position. We also learned about slowdowns in *WIT* implementation during the pandemic, but that in-person *WIT* implementation has resumed with a renewed sense of vigor. Further, the VPIs with whom we have spoken view the evaluation in very positive terms and expressed strong interest in participating in the evaluation. The VPIs also indicated that our plans for data collection (surveys of Airmen/Guardians, fidelity forms completed by observers of the implementation and Airmen/Guardians feedback forms on *WIT*) are feasible, and that they are prepared to facilitate data collection. The list of potential treatment bases is shown below:

MAJCOM	Treatment Group Implementing WIT
Air Combat Command (ACC)	1. Seymour Johnson AFB
Air Combat Command (ACC)	2. Moody AFB
Air Education and Training Command (AETC)	3. Goodfellow AFB
Air Education and Training Command (AETC)	4. Vance AFB
Air Education and Training Command (AETC)	5. Laughlin AFB
Air Education and Training Command (AETC)	6. Sheppard AFB
Air Mobility Command (AMC)	7. Dover AFB
Air Mobility Command (AMC)	8. JB McGuire Dix-Lakehurst
Air Force Materiel Command (AFMC)	9. Hill AFB
U.S. Air Forces in Europe (USAFE)	10. Spangdalem AFB
Pacific Air Force Bases (PACAF)	11. Yokota AFB
Pacific Air Force Bases (PACAF)	12. Eielson AFB
Pacific Air Force Bases (PACAF)	13. Kunsan AFB
Air force Global Strike Command (AFGSC)	14. Dyess AFB
Air force Global Strike Command (AFGSC)	15. Kirtland AFB
Air Education and Training Command (AETC)	Holloman AFB (back-up if needed)

As of January 21, 2022, we have confirmed that 15 Air Force bases will be a part of the treatment group for the evaluation study.

MAJCOM	Comparison Group Not Implementing WIT
Air Force Global Strike Command (AFGSC)	1. Whiteman AFB
Air Combat Command (ACC)	2. Offutt AFB
Air Combat Command (ACC)	3. Beale AFB
Air Combat Command (ACC)	4. Fairchild AFB
Air Education and Training Command (AETC)	5. Columbus AFB
Pacific Air Force Bases (PACAF)	6. Misawa AFB
Air Force Global Strike Command (AFGSC)	7. Minot
United States Space Force (USSF)	8. Peterson/Schriever AFB

Air Education and Training Command (AETC)	JB Randolph
Air Education and Training Command (AETC)	JB Lackland

As of January 21, 2022, we have confirmed 8 Air Force bases will be a part of the comparison group for the evaluation study.

As noted later in the power analysis section, we would like to compare 1,000 DAF enlisted Airmen/Guardians receiving *WIT* to 1,000 enlisted Airmen/Guardians not receiving *WIT* (total n= 2,000). We are hoping to recruit five or fewer DAF bases implementing *WIT* to compare to five or fewer bases not implementing *WIT*. Using a four-month data collection window and a 50% attrition rate for a follow-up survey, we would need DAF bases that provide *WIT* training to about 100 Airmen/Guardians each month, all of whom we would need to recruit to the study and completed a baseline survey (100 Airmen/Guardians per month * 4 months= 400 Airmen/Guardians per base times 10 bases= 4,000 Airmen/Guardians * 50% attrition rate leads to 2,000 Airmen/Guardians available for a completed follow-up surveys) (see Table A6). The comparison sample of bases not implementing *WIT* would need to support recruitment of a similar sample size to generate the comparison group of Airmen/Guardians. Comparison group and *WIT* trained Airmen/Guardians would be matched post-hoc, following the methods of a quasi-experimental design (QED).

On November 5, 2021, NORC received feedback from DAF Survey Office about the proposed survey instrument. NORC provided a point-by-point response to address all the concerns presented by DAF Survey Office. On November 19, 2021, NORC and DOD SAPRO worked together to assemble an emergency Office of Management and Budget (OMB) for study approval. On December 9, 2021, SAPRO completed its review of NORC's draft of emergency review. SAPRO identified a presidential appointee to sign the emergency request to move on to the next stage of approval. On January 21, 2022, SAPRO requested the documentation for 30-day FRN from NORC for emergency approval. On January 21, 2022, NORC submitted the 30-day OMB FRN materials to SAPRO for review and submission.

For the QED evaluation design (see Figure 4), assessments are completed at baseline by both the program participants and comparison group, before the *WIT* training has occurred. After the training is completed, a follow-up assessment will be given to both program participants and the comparison group participants.

Timeline

	Date	Task	Responsibility
Base Year (BY) Contract Period: Sept 2020-Sept. 2021	Dec-Feb 2021	Identify ESW	DAF; NORC team
	Dec-Mar 2021	Determine evaluation design	NORC team; DAF
	Dec-April 2021	Identify & confirm comparison groups	NORC team; DAF
	Mar 2021	Create survey instrument (draft created)	NORC team; DAF
	Apr 2021-June 2021	Submit IRB protocol (IRB review was completed for the VPI survey and the team drafted the protocol for a *WIT* outcome evaluation study for IRB review within the next two months)	NORC team and NORC IRB
	July-Aug 2021	Revisions and final IRB approval secured	NORC team
	July-Aug 2021	Waiting on confirmation on the need for possible OMB review	NORC team; DAF
	July 2021	DAF signed MOU with NORC for Evaluation	NORC team; DAF

	Apr 2021-Aug 2021	Confirm NORC-approved IRB protocol	DAF
	Aug 2021	Cognitive interviewing and Pre-test/pilot-test instrument and IRB approval for revised instruments	NORC team
	Sept 2021	NORC IRB received approval of Phase 3	NORC team
	Sept 2021	NORC received feedback from DAF legal office about survey instrumentation items, and submitted edits	NORC team, DAF
	Sept 2021	Form 4453 was signed by the Commander of A1Z	DAF
	Aug 2021 – Sept 2021	Program survey	NORC team
	Aug 2021 – Sept 2021	Survey scenario testing and instrument check	NORC team
Option Year 1: Oct 2021-Sept. 2022	Nov 2021	DAF background aggregated administrative data	DAF Internal Research
	March 2022- September 2022	Baseline survey and follow-up surveys, qualitative data collection	DAF SAPR, NORC team
	March 2022-September 2022	Disseminate updated feedback forms at each *WIT* session	DAF SAPR
	Jan 2022-Sept 2022	Analyze baseline and immediate post-test survey data, write report, slide deck	NORC team
Option Y2: Oct 2022-Sept. 2023	Sept 2022 –March 2023	Follow-Up field survey (six-month r post-test), qualitative data collection	DAF SAPR, NORC team
	Sept 2022-Sept 2023	Analyze follow-up data, write report, slides	NORC team
Option Y3: Oct 2023-Sept. 2024	Oct 2023-Sept. 2024	Analyze across all waves of data and write Final report	NORC team
	Sept. 2024	Return confidential files to DAF	NORC team

Human Subjects Protection

Process: NORC and DoD SAPRO have consulted with DoD eIRB and have determined that DoD eIRB will not be involved with the human subjects review for the *WIT* evaluation. As NORC will use NORC data collection systems and staff to implement the evaluation data collection and analyses, the NORC IRB will serve as the IRB of record. Similar to the approach of the Office of People Analytics (OPA) in their administration of the Workplace and Gender Relations Survey of Active Duty Members (WGRA), this process will allow for DAF staff and the *WIT* program staff to remain distinctly separate for the collection of evaluation data, enhancing the perceived confidentiality of responses and thus the integrity of the data. As of this time, the DAF IRB is expected to conduct an administrative review of the NORC IRB protocol, inclusive of the survey instrumentation and consent forms to ensure they conform to DAF standards. On June 2, 2021, NORC presented the package of data collection protocols to the DAF leadership for their approval for use of the survey instrument and data collection protocols with the Airmen/Guardians/Guardians. DAF signed off on the MOU with DoD SAPRO on July 29, 2021. NORC is currently waiting for final DAF administrative approval from the DAF Survey Office. DAF legal review has been conducted and NORC received approval on September 21, 2021. OMB approval is not needed at this time.

Components: The IRB protocol includes the following attachments: (a) data collection plan; (b) modes of administration; (c) cognitive interview and pilot test consent form to review study instrumentation with first year Airmen/Guardians/Guardians in the Summer of 2021; (d) cognitive interview and pilot-testing guide; (e) longitudinal study participation consent form; (f) survey

instrumentation for baseline and follow-up administration; (g) recruiting and follow-up communications to Airmen/Guardians/Guardians; and (h) adverse event protocol.

Respondent-generated identification number: After consenting to participate in the Airmen/Guardians/Guardians survey, participants will be asked to respond to survey items that result in a self-generated unique ID code (the instructions guide the participant through a set of four questions that will facilitate replication of the same unique ID at the follow-up survey) based on protocols currently in place in other DoD research. The questions will be unique to the respondent, resulting in a combined unique numeric code. NORC will not generate the study IDs, and it will not be possible to link survey responses to any individual respondent. Further, the self-generated ID code has been tested by NORC statisticians to ensure that it is robust enough to minimize duplicates (very small chance) across the potential 4,000 respondents. To further protect anonymity, the demographic questions asked on the survey are limited in number, and will be recoded to ensure that no cell will contain fewer than five (5) unique individuals.

Timeline. NORC submitted to the IRB for review a VPI survey to study the implementation of *WIT* across all DAF bases. The IRB approved the VPI survey protocol as Non-Human Subject Research on February 22, 2021. Also, NORC submitted an initial Human Subjects protocol for Phase 2, the recruitment of nine or fewer Airmen/Guardians/Guardians to review and respond to the proposed survey items via cognitive interviews. This protocol was submitted to NORC IRB on May 19, 2021. We received approval from the NORC IRB on June 1, 2021, for the Phase 1 and 2 research activities with Airmen/Guardians/Guardians. The approved protocol was shared with DAF on June 1, 2021. Cognitive testing was completed in mid-August, and revised and secured NORC IRB approval for the updated instrument. As part of Phase 2, NORC pre-tested/pilot-tested the revised instrument to assess how well survey programming and survey delivery to Airmen/Guardians/Guardians email accounts is working, all before implementation of Phase 3, the *WIT* program outcome evaluation. An additional amendment to the IRB protocol and instrumentation was necessary after the cognitive testing for Phase 2 pilot work. NORC submitted the full *WIT* Evaluation Human Subjects Protocol (Phase 3), via amendment to Phase 1 & 2, to NORC IRB on September 17, 2021. We received approval from the NORC IRB on September 22, 2021. NORC submitted an OMB 30-day FRN emergency approval materials to SAPRO on January 21, 2022.

Office of Management and Budget (OMB)

NORC was notified of the possible requirement to apply for an emergency OMB approval towards the end of June 2021. There is a new OMB regulation for the DoD that requires OMB review of DoD research on any sensitive items of interest to the general public. NORC worked with the DAF to confirm the need for an OMB review. OMB review was deemed necessary in December 2021. After further discussion, NORC submitted OMB 30-day FRN emergency approval materials to SAPRO on January 21, 2022.

Data Sources

This evaluation design will draw on both individual-level survey data and, to the extent possible, background aggregated administrative data sources. Where possible, process data will be collected from the VPIs, trainers, and Airmen/Guardians/Guardians program participants and outcome data will be collected from the Airmen/Guardians/Guardians program participants via baseline and follow-up surveys.

Data Collection Procedures

For each survey data collection, we will use survey respondent self-generated identification numbers, so the data will already be anonymous. This will ensure privacy and confidentiality. NORC will document in a written summary the methods used for the evaluation, inclusive of instrumentation, modalities, sample recruitment and participation rates, and analytic plans, as well as the analytic results. Because of early collaborative work with leadership and the ESW, presentation of the analytic results will be consistent with expectations and supportive of the subsequent longitudinal comparisons of program impact. *Quantitative surveys* of program participants will cover constructs directly related to the intervention with known reliability and validity (where possible). We will also assess available background aggregated *administrative data sources* (e.g., where possible, reports of SA and service response) to compliment the survey data, mitigating the sample recruitment/attrition, and recall challenges that arise from survey research.

Informed Consent: As part of the IRB-approved protocols, NORC has developed informed consent forms that communicate the purpose of the study, content of the instrumentation, burden of participation, the risks and benefits for participants, and contact information for the study. Data collection will be collected via online surveys, and NORC has approved consent forms inclusive of a waiver of hard copy signatures. With NORC IRB and DAF IRB approval, NORC programed the approved informed consent language into the introduction of the online baseline data collection instrumentation, allowing participants to click a box to sign their e-based consent.

Conduct surveys: Baseline data collection will be administered remotely via online, web-based surveys. As we have been developing the data collection plans over the past few months, we have determined with DAF that the survey will be administered via web survey exclusively. Airmen/Guardians have good access to the internet for completing the survey and all ESW stakeholders believe that most Airmen/Guardians would prefer doing the survey online anonymously. All NORC data collection staff have been formally trained on protocols, including the study purpose, target population, frequently asked questions, and protection of sensitive data.

We have also decided, based on our work over the past months with DAF, on the bases to serve as the comparison group of Airmen/Guardians participants (see Table 1 below). That is, working with DAF, we are advancing a web-based survey methodology. We have considered the unique challenges faced by participants who may have varying duty assignments with differential travel, time flexibility, and private computing time access. Moreover, the COVID-19 pandemic and attention to physical distancing limits in-person contacts. Our survey system renders optimized survey presentation for mobile phone users. Each field period will be planned to include flexibility for extending data collection should unexpected duty assignments interfere with a large group of participants completing the relevant survey. NORC developed a detailed recruitment and follow-up plan as part of the IRB protocols, which includes tailored follow-up recruitment messages.

Table 1: Potential Treatment and Comparison Bases

Maj Com	Treatment Group Implementing WIT	Maj Com	Comparison Group Not Implementing WIT
Air Combat Command (ACC)	1. Seymour Johnson AFB	Air Force Global Strike Command (AFGSC)	1. Whiteman AFB
Air Combat Command (ACC)	2. Moody AFB	Air Combat Command (ACC)	2. Offutt AFB
Air Education and Training Command (AETC)	3. Goodfellow AFB	Air Combat Command (ACC)	3. Fairchild AFB
Air Education and Training Command (AETC)	4. Vance AFB	Pacific Air Force Bases (PACAF)	4. Misawa AFB
Air Education and Training Command (AETC)	5. Sheppard AFB	Air Combat Command (ACC)	5. Beale AFB
Air Education and Training Command (AETC)	6. Laughlin AFB	Air Force Global Strike Command (AFGSC)	6. Minot AFB
Air Mobility Command (AMC)	7. Dover AFB	Air Education and Training Command (AETC)	7. Columbus AFB
Air Mobility Command (AMC)	8. JB McGuire Dix-Lakehurst	United States Space Force (USSF)	8. Peterson/ Schriever AFB
Air Force Materiel Command (AFMC)	9. Hill AFB	Air Education and Training Command (AETC)	JB Randolph
Pacific Air Force Bases (PACAF)	10. Eielson AFB	Air Education and Training Command (AETC)	JB Lackland
U.S. Air Forces in Europe (USAFE)	11. Spangdahlem AFB		
Pacific Air Force Bases (PACAF)	12. Yokota AFB		
Pacific Air Force Bases (PACAF)	13. Kunsan AFB		
Air Force Global Strike (AFGSC)	14. Kirtland AFB		
Air Force Global Strike (AFGSC)	15. Dyess AFB		
Air Education and Training Command (AETC)	Holloman AFB		

Follow-up surveys (about 15-20 minutes in length) will be conducted approximately 6 months after the *WIT* intervention, based on ESW feedback. We anticipate that each survey data collection intake period will take on average four months to complete. Based on the generally low response rates achieved in similar surveys of SH and SA (e.g., the WGRA), we will be exploring a variety of methods to maintain a strong response follow-up rate such as multiple reminders to participants, use of gift card incentives, and appeals to improving military life and the science of preventing SH and SA. While the WGRA surveys are achieving a 14-18% response rate (RR), we are confident we can exceed these follow-up rates based on our strong track record of high longitudinal participation rates from a variety of difficult-to-track populations.

Fidelity of trainers/facilitators' preparation: At baseline, for each of the participating treatment DAF bases during the likely four-month intake period, NORC will assess trainers' readiness to deliver the content with integrity and document their role in program delivery, so that inconsistent program delivery does not impact outcomes. NORC will conduct a brief online trainer survey to assess knowledge and attitude towards the program material and readiness to implement to identify significant gaps in preparation

Process evaluation data: NORC will document and monitor all intervention activities, including a review of program documentation.

Power Calculations

Statistical power provides an estimate of the probability of identifying a relationship through a significant statistical test when, in fact, such an impact exists.[51] To calculate our power estimates we used formulas for computing the expected test statistic found in many power analysis texts[52, 53] in conjunction with Microsoft Excel's routines for evaluating the standard normal curve.

Since the primary analyses will compare the treatment group against the comparison group, power estimates were computed for N=2,000 completed six-month follow-up survey participants (e.g., 1,000 DAF enlisted Airmen/Guardians receiving *WIT* and 1,000 enlisted Airmen/Guardians not receiving *WIT*), for the various effect sizes. To assess the effect of the intervention for treated Airmen/Guardians in a non-clustered design, with a projected sample of 1,000 treatment cases and 1,000 comparison cases, the statistical power of our evaluation will be .81 (.80 and greater is a typical power level sought in prevention experiments in public health) to identify a standardized mean difference/effect size of .11 (considered a small effect size)[51], based upon an alpha level of .05, a two-tailed statistical test, and covariates that explain 25% of outcome variation (say, a pre-test). For this main scenario, our power level is over .90 for any effect size of .125 or above (still in the small effect size range). This scenario of 1,000 treatment cases and 1,000 comparison cases will also provide ample power to explore subgroup differences (e.g., differences by gender).

Our power analyses revealed that the standardized mean difference of .3 (approaching a small to medium effect size) is statistically significant with a power level of above .9 for nearly all the comparisons we calculated from n=500 to n=4,000, for all the sample sizes in individual-level RCTs, and for most comparisons in clustered RCTs (e.g., clustered by military base/installation). However, selecting a clustered design (where entire groups are assigned treatment rather than individual assignment) would impact our ability to find statistically significant differences above the traditional power level of .8.[53, 54] This impact is prominent when planning for effect sizes of less than .20. For example, for a clustered experiment with 20 bases/installations and 2,000 six-month follow-up survey participants (1,000 treatment and 1,000 comparison cases) an effect size of .20 would result in an adequate .83 power level,

assuming two-tailed standard normal tests with α = 0.05, modest intra-class correlation of 0.1, covariate impacts assume R-square values of 0.25 for individuals and 0.5 for clusters. However, using the same assumptions, for an effect size of .19, our power would dip to .79 for those same 20 bases and 2,000 participants. In non-clustered designs, as discussed above, 2,000 participants (1,000 treatment and 1,000 comparison cases) would lead to a power level of .81 to detect an even smaller effect size of .11.

*Analysis Plan

Data cleaning

We will start with standard data cleaning to remove errors and inconsistencies in all data files. Errors will be detected by checking skip patterns, using descriptive statistics, scatterplots, and histograms. Our team uses Mplus, Stata, SAS, R, and SPSS, according to analytic task and SAPRO preferences.

Nonresponse and Missing Data Bias

We propose to compare responders with non-responders with basic aggregated demographic information and other information (e.g., demographics, company affiliation, athletic team participation, disciplinary records, etc.), 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. We will compare the impact of employing various methods to handle missing data. Item-level missing data will be addressed using widely accepted methods, e.g., Full Information Maximum Likelihood (FIML) and Multiple Imputation (MI) procedures.[55] We will compare the impact of employing various imputation-based procedures to fill in missing values for the surveys that are only partially completed. NORC is very experienced in various imputation methods (e.g., nearest neighbor “hot deck”), including multiple imputation. We will use Rubin’s multiple imputation strategy[56, 57] (as appropriate) to replace each missing value with a set of plausible values that represent the uncertainty about the correct value.

Baseline Equivalence

Baseline equivalence is a test to determine whether the intervention and comparison groups were similar enough (“equivalent”) on key variables before the start of the intervention (at “baseline”). Differences between the two groups at baseline could bias the estimated impact of the program – that is the impact could be attributed to the baseline differences and not the program. We will therefore measure the equivalence of the treatment and comparison groups at baseline using analytic samples. We use the final analytic sample to calculate baseline equivalence because of the potential for differential attrition over the course of the program. We calculate baseline equivalence for continuous variables using Hedges’ g, a common effect size index. It is the difference between the average characteristic for the intervention group and the average characteristic for the comparison group, divided by the pooled standard deviation (SD) of the characteristic. For dichotomous variables, we use Cox’s Index, which is more complex than Hedges’ g, but is designed to produce a comparable effect size. We will assess baseline equivalence on each outcome measure to determine whether baseline differences are: 1) Small (ES is smaller than .05), i.e., the groups are equivalent, 2) Moderate (ES is between .05 and .25),

i.e., the analysis requires a statistical adjustment, or 3) Large, (ES larger than .25) i.e., the differences at baseline are too large to allow the analysis to continue.

Attrition

While attrition between the baseline and follow-up surveys is a concern, NORC is well versed in longitudinal diagnostics and in conducting attrition analyses to handle the impact of missing data and study drop-outs. For QEDs, matching program participants to their counterparts in the counterfactual condition creates groups with similar characteristics at the start of the study (baseline). RCTs achieve the same equivalence at baseline through random assignment. When the two groups have similar characteristics at baseline, differences in outcomes between the groups at follow-up can be attributed to the intervention. However, if attrition occurs, the program participants and comparison group members may no longer be equivalent at baseline, preventing us from being able to attribute any differences in outcomes solely to the intervention. The monitoring and documentation of attrition is therefore a critical component of the impact evaluation. We will document two kinds of attrition by cohort 1) attrition for all study participants (overall attrition) and 2) differences in attrition between the intervention and comparison groups (differential attrition). When the combination of overall and differential attrition is high, it is possible that estimates of program impact may be biased. While we do not anticipate high levels of attrition at follow-up due to tracking systems available in DAF to find study participants, we will monitor and document attrition carefully, using the empirically derived What Works Clearinghouse standards.[58]

Nonresponse Adjustment

We propose to compare responders with non-responders on basic demographic information and other information (e.g., age, gender, income, etc.), and adjust for non-response bias with appropriate methods (e.g., non-response weights) if needed. We will use non-response weighting so that overall population estimates are not negatively affected by any differential response. Without non-response weighting, population estimates would be overly influenced by the particular subgroups that responded at a higher rate, skewing the estimates and making them less accurate or useful. We will use a response propensity approach to calculate non-response weights, which is calculated as the conditional probability that a particular respondent completed the survey given observed covariates.[59]

Process/Qualitative Analysis Plan

NORC will review all *WIT* intervention activities through documentation provided by DAF SAPR staff. NORC has mostly completed this task by reviewing all available program material and conducting a survey on program implementation. NORC has also requested from DAF all available records of *WIT* program session implementation and attendance records (provided by DAF), and session feedback from Airmen/Guardians receiving *WIT*. To date, it does not appear that DAF has maintained such historical records of implementation and attendance. Therefore, we conducted a survey with the VPIs to assess their recent knowledge of *WIT* implementation (ranging from currently to some questions asking about the past three years).

Should COVID-guidance and DAF staff allow it, NORC staff will conduct direct observation of the program delivery (n<9). Observations of the treatment will be analyzed using a structured observation protocol that will allow comparisons across data sources. Special attention will be

given to assessing differences across the various data sources. Observational data will validate Airmen/Guardians's perspectives and enhance description of the processes and experiences of change. In sum, NORC will collect data on dosage, fidelity/stability of delivery, and contextual factors that may impact program outcomes. Measures of dosage, attendance and/or program context with sufficient variability may suggest potential impact on variability in the *WIT* effects. For open-ended responses on session feedback forms, NORC will conduct separate content thematic analyses of the qualitative data through coding of themes,[60, 61] using NVivo (v12) software.[62]

Descriptive/Exploratory Analyses of Survey Data

Starting with our descriptive analyses, we will confirm the validity/reliability of our key constructs (e.g., attitudes) using confirmatory factor analysis and reliability analyses. We will examine the distribution of the data and run frequencies, measures of central tendency, and measures of dispersion with all study variables. We will provide single point estimates (with confidence intervals). We will perform bivariate analyses on relevant background variables to determine whether key factors (socio-demographics, relationship between perpetrator and victim) are statistically significant correlates of the selected outcome measures. NORC will estimate correlations to examine multi-collinearity between key covariates specified in our Conceptual Model. Variables that are significant in bivariate models will be entered into multivariable models of the key short, intermediate, and long-term outcomes.

Multivariable Models

NORC will test our hypotheses regarding short, intermediate, and long-term outcomes with multivariate models. OLS regression models will be used for continuous outcomes, multinomial regression for categorical, and Poisson/negative binomial models for count outcomes. In addition, the longitudinal nature of the data will support trend analyses using latent growth models.[65] If it turns out that we have Airmen/Guardians nested within groups when receiving the intervention, we will account for this clustering in groups by estimating robust standard errors with a Sandwich estimator or Hierarchical linear modeling (HLM).[66] HLM[67] provides a conceptual framework and a flexible set of analytic tools to analyze the special requirements of our data.[68] Nesting occurs when a unit of measurement is a subset of a larger unit and the units clustered in the larger unit might be correlated.

There will be multiple dependent variables (e.g., attitudes, skills, and behavior outcomes), therefore, we might also explore using structural equation modeling (SEM)[69] to examine mechanisms of the intervention, i.e., how does the intervention work and for whom the intervention works the best. SEM will allow us to simultaneously examine a set of relationships between multiple independent and dependent variables.[69] Stata 15 will allow our team to estimate simultaneous equations and recursive and non-recursive paths. By using a latent model, we will be in position to disentangle the effects of measurement error from true-score variation.[70]

Exploratory Analysis of Subgroup Variation

Our planned sample size for the *WIT* is strong for the main analyses. However, uncertainty about the direction of effects is greater when comparing some of the subgroups than when studying main effects. Thus, we will limit exploratory analysis of variation in subgroup impacts to a few key subgroups and outcomes, and we will apply two-sided tests. In addition to testing

whether effects for specific subgroups are significantly different from zero, we will also test whether effects are significantly different across subgroups. Key subgroups to be explored include gender, those with prior history of receiving SH and SA prevention programming, and those with an alcohol/drug use disorder.

Exploratory Analysis of Mediation and Moderation

A natural question is the extent to which program fidelity, and active participation by Airmen/Guardians receiving the intervention mediate or moderate the subsequent impacts. Mediation analysis is a way to check if a third variable *mediates* the relationship between the independent and dependent variables – explaining the reason for such a relationship to exist. In a perfect mediation, an independent variable leads to some kind of change to the mediator variable, which then leads to a change in the dependent variable.[71] Moderation analysis is a way to check whether a third variable influences the strength or direction of the relationship between an independent and dependent variable.[71] We will measure multiple constructs known to correlate with the dependent variables. These measures will be collected in the Airmen/Guardians surveys. Items of interest for mediation analysis include race/ethnicity, and alcohol use. Items of interest for moderation analysis may include prior history of SH and SA victimization and exposure to other SH and SA prevention interventions.

Anticipated Challenges & Solutions

First, a key challenge to this outcome evaluation plan is the difficulty of holding the *WIT* timing constant over the evaluation period and having a consistent approach to implementation across the DAF. Second, participation in surveys in the military are generally low, as evidence by the response rates below 20% for the WGRA surveys. Low response rates — because those who respond to the survey may be systematically different on key variables from those who did not respond[72] — can lead to bias, erroneous conclusions, and limited generalizability of findings.[73] Addressing low response rates proactively will ensure that we obtain program evaluation data that is trustworthy and useful in showing whether the *WIT* is having the intended impact. NORC, with approval from the DAF, will offer incentives to Airmen/Guardians based on their completion of the survey. We will be offering a $10 Amazon e-gift code for baseline survey completion and $15 Amazon e-gift code for follow-up survey completion.

Third, as of this point in time, adjustments to program and research implementation to accommodate restrictions arising from the COVID pandemic may interfere with participant learning, participant risk of exposure to SH and/or SA, and observation by the research team. However, unlike many DAF activities during the pandemic, the SAPR office and DAF leadership have negotiated continuation of in-person *WIT* activities. Barring reversal of this decision, the pandemic impact on program delivery has been minimized to the extent possible. Fourth, the evaluation as currently designed relies on the collaborative participation from either another group of comparable Airmen/Guardians not receiving *WIT* or another service. Should there be an interruption to the full participation of the comparison group, the analytic plans and options for interpretation of outcome analyses will need to be adjusted. In that case, NORC will rely on a pre-post data collection and analysis plan. Fifth, if there is a significant change in DoD or DAF policies affecting programing or discipline related to SH and SA over the course of the two-year evaluation period, such policy changes would interfere with the test of the direct effects of the *WIT* on Airmen/Guardians outcomes. This risk cannot be controlled but the impact of any policy

changes will be discussed as contextual background in the interpretation of the evaluation results.

*Dissemination Plan

Per contractual obligation, NORC will deliver interim and final reports to DoD SAPRO. NORC will prepare briefing decks at key points during the evaluation design and implementation process. First, NORC, in collaboration with the DAF has been developing a briefing deck that captures the decisions that have occurred regarding the logic model, research questions, and ESW participants. This briefing deck continues to be updated regularly and at the close of the project will serve as a record of final decisions and products. Second, at key decision points during the evaluation planning and implementation phases, NORC will prepare upon request briefing decks to share plans with stakeholders who have not been involved in the weekly decision-making process and may need targeted information to make additional decisions (e.g., for DAF leadership to understand the role and activities of a comparison population, or for other DoD leadership to review the evaluation project activities, or for DoD to brief Congress as needed). Third, should DAF or DoD SAPRO request it, NORC will prepare a briefing deck of the final evaluation results for the DAF and DoD internal use.

NORC is prepared to support DoD SAPRO and DAF in the preparation of other dissemination products, following mutual agreement on timeline and parameters. For example, NORC has produced a briefing deck for a presentation that occurred on May 25, 2021, summarizing the study results for the VPIs. DAF may wish to share the results of the evaluation with other stakeholders which could be accomplished with a slide deck, an infographic, or a research report. In addition, because of the confidential nature of NORC data collection efforts, NORC will provide DAF SAPR with the project survey data while withholding and destroying data about Airmen/Guardians victimization and perpetration (*pending final data management decisions*). This will enable DAF to continue to use the NORC-collected data in future analyses without contaminating DAF SAPR's role. Aggregate findings regarding SH and SA victimization and perpetration, and the impact of the WIT on these outcomes, will be provided to both DoD SAPRO and DAF SAPR through reports.

References

1. Brown AL, Testa M, Messman-Moore TL. Psychological consequences of sexual victimization resulting from force, incapacitation, and verbal coercion. *Violence Against Women*. 2009;15doi:10.1177/1077801209335491
2. Campbell R, Dworkin E, Cabral G. An ecological model of the impact of sexual assault on women's mental health. Review. *Trauma Violence Abus*. Jul 2009;10(3):225-246. doi:10.1177/1524838009334456
3. Najdowski CJ, Ullman SE. PTSD symptoms and self-rated recovery among adult sexual assault survivors: The effects of traumatic life events and psychosocial variables. *Psychology of Women Quarterly*. 2009;33(1):43-53.
4. The National Intimate Partner and Sexual Violence Survey (NISVS): 2010 Summary Report (National Center for Injury Prevention and Control, Centers for Disease Control and Prevention) (2011).
5. Lévesque S, Rodrigue C, Beaulieu-Prévost D, Blais M, Boislard M-A, Lévy JJ. Intimate partner violence, sexual assault, and reproductive health among university women. *The Canadian Journal of Human Sexuality*. 2016;25(1):9-20.

6. Pegram SE, Abbey A. Associations between sexual assault severity and psychological and physical health outcomes: Similarities and differences among African American and Caucasian survivors. *Journal of interpersonal violence*. 2016:0886260516673626.
7. Brown AL, Testa M, Messman-Moore TL. Psychological consequences of sexual victimization resulting from force, incapacitation, or verbal coercion. *Violence Against Women*. 2009;15(8):898-919.
8. Independent Review Commission (IRC) on Sexual Assault in the Military. Hard Truths and the Duty to Change: Recommendations from the Independent Review Commission on Sexual Assault in the Military. 07/02/2021 2021:299.
9. U.S. Department of Defense. *Fiscal year 2009 annual report on sexual assault in the military.* 2009.
10. Walsh K, Koenen KC, Cohen GH, et al. Sexual violence and mental health symptoms among National Guard and Reserve soldiers. *Journal of general internal medicine*. 2014;29(1):104-109.
11. Tiet QQ, Leyva YE, Blau K, Turchik JA, Rosen CS. Military sexual assault, gender, and PTSD treatment outcomes of US veterans. *Journal of Traumatic Stress*. 2015;28(2):92-101.
12. Allard CB, Nunnink S, Gregory AM, Klest B, Platt M. Military sexual trauma research: A proposed agenda. *Journal of Trauma & Dissociation*. 2011;12(3):324-345.
13. Bell ME, Reardon A. Experiences of sexual harassment and sexual assault in the military among OEF/OIF veterans: Implications for health care providers. *Social Work in Health Care*. 2011;50(1):34-50.
14. Castro CA, Kintzle S, Schuyler AC, Lucas CL, Warner CH. Sexual assault in the military. *Current psychiatry reports*. 2015;17(7):54-67.
15. Basile KC, DeGue S, Jones K, et al. *STOP SV: A technical package to prevent sexual violence*. 2016.
16. Orchowski LM, Berry-Cabán CS, Prisock K, Borsari B, Kazemi DM. Evaluations of Sexual Assault Prevention Programs in Military Settings: A Synthesis of the Research Literature. *Military Medicine*. 2018;183(suppl_1):421-428. doi:10.1093/milmed/usx212
17. Website ODAF. Military Demographics. United States Air Force. 2020. https://www.afpc.af.mil/About/Air-Force-Demographics/
18. Breslin R, Davis L, Hylton K, et al. *2018 Workplace and Gender Relations Survey of Active Duty Members: Overview Report*. 2019. https://www.sapr.mil/sites/default/files/Annex_1_2018_WGRA_Overview_Report_0.pdf
19. 2016 Workplace and Gender Relations Survey of Active Duty Members Overview Report. (2017).
20. Castro CA, Kintzle S, Hassan A. The state of the American veteran: The Los Angeles county veterans study. *University of Southern California Center for Innovation and Research on Veterans and Military Families*. 2014;
21. Dichter ME, True G. "This Is the Story of Why My Military Career Ended Before It Should Have" Premature Separation From Military Service Among US Women Veterans. *Affilia*. 2015;30(2):187-199.
22. U.S. Department of Defense. *Enclosure 3: Department of the Air Force*. 2019. *The Department of Defense Fiscal Year 2019 Annual Report on Sexual Assault in the Military*.
23. Holland KJ, Rabelo VC, Cortina LM. Sexual assault training in the military: Evaluating efforts to end the "invisible war". *American Journal of Community Psychology*. 2014;54(3-4):289-303.
24. Coker AL, Fisher BS, Bush HM, et al. Evaluation of the Green Dot Bystander Intervention to Reduce Interpersonal Violence Among College Students Across Three Campuses. *Violence Against Women*. August 14, 2014 2014;doi:10.1177/1077801214545284
25. Green Dot Military Applications. (2016).
26. Training Curriculum: Bystander Training (Air Force) (2018).
27. 4 Hour Training: Wingman Intervention (2019).
28. Banyard VL, Moynihan MM, Crossman MT. Reducing sexual violence on campus: The role of student leaders as empowered bystanders. *Journal of College Student Development*. 2009;50(4):446-457.
29. Banyard VL, Moynihan MM, Plante EG. Sexual Violence Prevention through Bystander Education: An Experimental Evaluation. *Journal of Community Psychology*. 05/01/ 2007;35(4):463-481.
30. Coker AL, Bush HM, Cook-Craig PG, et al. RCT testing bystander effectiveness to reduce violence. *American journal of preventive medicine*. 2017;52(5):566-578.

31. 60 Minute Training:Wingman Intervention Initial (2019).
32. Wingman and Leadership Intervention Evaluation Guide (2019).
33. Coker AL, Cook-Craig PG, Williams CM, et al. Evaluation of Green Dot: An active bystander intervention to reduce sexual violence on college campuses. *Violence Against Women*. 2011:1077801211410264.
34. Schell TL, Cefalu M, Morral AR. *Development of a Short Form Measure of Sexual Harassment Risk in the Military*. 2019.
35. Koss MP, Abbey A, Campbell R, et al. Revising the SES: A collaborative process to improve assessment of sexual aggression and victimization. *Psychology of Women Quarterly*. 2007;31(4):357-370.
36. Koss MP, Abbey A, Campbell R, et al. *The Sexual Experiences Short Form Victimization (SES-SFV)*. 2006.
37. McCauley JL. *Self-Efficacy and Sexual Assault: The Impact of Victim-Offender Relationship and Alcohol*. University of Georgia; 2006.
38. Barling J, Loughlin C, Kelloway EK. Development and test of a model linking safety-specific transformational leadership and occupational safety. *J Appl Psychol*. 2002;87(3):488.
39. Bass BM, Avolio BJ, Jung DI, Berson Y. Predicting unit performance by assessing transformational and transactional leadership. *J Appl Psychol*. 2003;88(2):207.
40. White House Task Force to Protect Students from Sexual Assault. Not alone: The first report of the White House Task Force to Protect Students from Sexual Assault. https://www.justice.gov/ovw/page/file/905942/download2014.
41. Banyard V, Moschella E, Grych J, Jouriles E. What happened next? New measures of consequences of bystander actions to prevent interpersonal violence. *Psychology of Violence*. 2019;9(6):664-674. doi:10.1037/vio0000229
42. Prevention Innovations Research Center. *Evidence-Based Measures of Bystander Action to Prevent Sexual Abuse and Intimate Partner Violence-Resources for Practitioners (Short Measures)*. 2015. https://www.unh.edu/research/sites/default/files/media/2019/09/bystander_program_evaluation_measures_-_short_version_compiled.pdf
43. Banyard VL. Measurement and correlates of prosocial bystander behavior: The case of interpersonal violence. *Violence and victims*. 2008;23(1):83-97.
44. Banyard VL, Moynihan MM, Walsh WA, Cohn ES, Ward S. Friends of Survivors The Community Impact of Unwanted Sexual Experiences. *Journal of Interpersonal Violence*. Feb 2010;25(2):242-256. doi:10.1177/0886260509334407
45. McMahon S, Banyard VL. When can I help? A conceptual framework for the prevention of sexual violence through bystander intervention. *Trauma Violence Abuse*. Jan 2012;13(1):3-14. doi:10.1177/1524838011426015
46. Calhoun KSGCA. Self-efficacy as a predictor of revictimization. Poster session presented at the annual meeting of the Association for the Advancement of Behavior Therapy, Reno, NV. 2002;
47. Berkowitz AD. Applications of social norms theory to other health and social justice issues. *In: Perkins HW, ed The social norms approach to preventing school and college age substance abuse: A handbook for educators, counselors, and clinicians: San Francisco, Jossey-Bass; 259-279*. 2003:259-279.
48. Berkowitz AD. Fostering healthy norms to prevent violence and abuse: The social norms approach. *The prevention of sexual violence: A practitioner's sourcebook*. 2010:147-171.
49. Banyard V, Edwards K, Rizzo A. "What would the neighbors do?" Measuring sexual and domestic violence prevention social norms among youth and adults. *Journal of Community Psychology*. 2019;47(8):1817-1833. doi:https://doi.org/10.1002/jcop.22201
50. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. *Archives of internal medicine*. 1998;158(16):1789-1795.
51. Cohen J. *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates; 1988.
52. Moerbeek M, Teerenstra S. *Power analysis of trials with multilevel data*. CRC Press; 2015.
53. Hedberg EC. *Introduction to Power Analysis: Two-group Studies*. vol 176. Sage Publications; 2017.
54. Hedges LV, Hedberg EC. Intraclass correlation values for planning group-randomized trials in education. *Educ Eval Policy Anal*. 2007;29(1):60-87.

55. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple Imputation by Chained Equations: What is it and how does it work? *International journal of methods in psychiatric research*. 2011;20(1):40-49. doi:10.1002/mpr.329
56. Rubin DB. Inference and Missing Data. *Biometrika*. 1976;63(3):581-590.
57. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons; 1987.
58. What Works Clearinghouse. *Reporting Guide for Study Authors: Group Design Studies*. 2020.
59. Jo B, Stuart EA, MacKinnon DP, Vinokur AD. The use of propensity scores in mediation analysis. *Multivariate Behav Res*. 2011;46(3):425-452.
60. Creswell J, ed. *Research design: Qualitative and quantitative approaches*. Sage; 1994. Miles M, Huberman A, eds. *Qualitative data analysis: An expanded sourcebook (2nd edition)*.
61. Strauss A, Corbin J. *Basics of qualitative research: Techniques for developing grounded theory (2nd edition)*. Sage; 1998.
62. Bazeley P, Jackson K. *Qualitative data analysis with NVivo*. Sage Publications Limited; 2013.
63. Miles MB, Huberman AM. *Qualitative Data Analysis*. Sage; 1994.
64. Patton MQ. *Qualitative Evaluation and Research Methods*. Sage; 1990.
65. Muthen B. Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In: Kaplan D, ed. *Handbook of quantitative methodology for the social sciences*. Sage Publications; 2004:pp.345-368.
66. White H. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*. 1980;48(4):817-838. doi:10.2307/1912934
67. Raudenbush SW, Bryk AS, Cheong YF, Cogdon R. *HLM 6: Hierarchical Linear and Nonlinear Modeling*. Scientific Software International; 2004.
68. Tate RL, Pituch KA. Multivariate hierarchical linear modeling in randomized field experiments. *Journal of Experimental Education*. Sum 2007;75(4):317-337.
69. Kline RB. *Principles and Practice of Structural Equation Modeling (3rd Edition)*. The Guilford Press; 2010.
70. Hu LT, Bentler PM. Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*. 1998;3(4):424-453.
71. Kenny BR. 0 (1986). *The moderator-mediator variable distinction in social psychological research: conceptual, strategic and statistical considerations} Pm Soc Psychol51*.1173-1182.
72. Hendra R, Hill A. Rethinking Response Rates: New Evidence of Little Relationship Between Survey Response Rates and Nonresponse Bias. *Evaluation Review*. 2019;43(5):307-330. doi:10.1177/0193841x18807719
73. Guo Y, Kopec JA, Cibere J, Li LC, Goldsmith CH. Population Survey Features and Response Rates: A Randomized Experiment. *American Journal of Public Health*. 2016;106(8):1422-1426. doi:10.2105/ajph.2016.303198

Appendix A. Implementer & Airmen/Guardians Feedback Forms

60-Minute Training: Implementer WIT Fidelity Assessment

As a Violence Prevention Integrator (VPI) or third-party observer, please complete this form while the implementer is either delivering the WIT session or shortly after the completion of the WIT session.

1. Please identify the MAJCOM or component you are based out of.
   - ☐ Air Combat Command (ACC)
   - ☐ Air Education and Training Command (AETC)
   - ☐ Air Force District of Washington (AFDW)

- ☐ Air Force Global Strike Command (AFGSC)
- ☐ Air Force Materiel Command (AFMC)
- ☐ Air Force Reserve Command (AFRC)
- ☐ Air Force Special Operations Command (AFSOC)
- ☐ Air Mobility Command (AMC)
- ☐ Pacific Air Force Command (PACAF)
- ☐ United States Air Forces in Europe-Air Forces Africa (DAFE-AFAFRICA)
- ☐ United States Space Force (USSF)

2. Please indicate the installation you are based out of.
   - ☐ [Open-ended question] → Needs hard check

Please indicate your level of agreement on how the implementer delivered the following components.

Introduction:	Strongly Disagree	(2)	(3)	(4)	Strongly Agree
3. The implementer clearly stated that the goals of the training are to describe how sexual assault may be a common problem within the U.S. Department of the Air Force (DAF), and to reduce the number of Airmen/Guardians who experience violence.	❍	❍	❍	❍	❍
4. The implementer emphasized the importance of shifting the culture to bystanders intervening and provided examples of how negative culture and violence (red dots) can spread by sharing, liking, or commenting on social media posts.	❍	❍	❍	❍	❍

Introducing the Red Dots & Green Dots	Strongly Disagree	(2)	(3)	(4)	Strongly Agree
5. The implementer provided a clear definition of red and green dots.	❍	❍	❍	❍	❍
6. The implementer provided useful examples of red dots, and both reactive and proactive green dots.	❍	❍	❍	❍	❍
7. The implementer asked Airmen/Guardians to vividly imagine the impact a single green dot can have on preventing sexual assault within the U.S. Department of the Air Force.	❍	❍	❍	❍	❍

Barriers to Reactive Green Dots & the Three Ds	Strongly Disagree	(2)	(3)	(4)	Strongly Agree
8. The implementer gave useful examples of the types of barriers (personal, relationship, organizational) that can hinder or promote intervention.	❍	❍	❍	❍	❍
9. The implementer provided a clear definition and examples of the 3Ds (Direct, Delegate, and Distract).	❍	❍	❍	❍	❍

Proactive Green Dots	Strongly Disagree	(2)	(3)	(4)	Strongly Agree
10. The implementer clearly defined two social norms: • Violence will not be tolerated. • Everyone needs to do their part to help.	❍	❍	❍	❍	❍

"The Commitment" and Closing	Strongly Disagree	(2)	(3)	(4)	Strongly Agree

11. The implementer clearly reminded Airmen/Guardians/Guardians they have many options for reducing sexual assault and domestic violence.	❍	❍	❍	❍	❍
12. The implementer definitively asked Airmen/Guardians/Guardians to commit to intervene when they notice situations that cause them concern or discomfort.	❍	❍	❍	❍	❍
13. The implementer adjusted their delivery style and tone to ensure Airmen/Guardians/Guardians stayed engaged.	❍	❍	❍	❍	❍

WIT Program Session Feedback Form for Airmen/Guardians/Guardians

Please provide your feedback on the WIT session you have just completed. Your feedback is greatly appreciated and will allow the U.S. Department of the Air Force (DAF) to provide the best training experience to service members.

1. Please identify the MAJCOM or component you are based out of.
   - ☐ Air Combat Command (ACC)
   - ☐ Air Education and Training Command (AETC)
   - ☐ Air Force District of Washington (AFDW)
   - ☐ Air Force Global Strike Command (AFGSC)
   - ☐ Air Force Materiel Command (AFMC)
   - ☐ Air Force Reserve Command (AFRC)

- ☐ Air Force Special Operations Command (AFSOC)
- ☐ Air Mobility Command (AMC)
- ☐ Pacific Air Force Command (PACAF)
- ☐ United States Air Forces in Europe-Air Forces Africa (DAFE-AFAFRICA)
- ☐ United States Space Force (USSF)

2. Please indicate the installation you are based out of.
   - ☐ [Open-ended question] → Needs hard check

3. Please indicate your position in the U.S. Department of the Air Force.
   - ☐ Officer
   - ☐ Enlisted
   - ☐ Civilian
   - ☐ Other _____

Please indicate how useful each of the topics of the WIT training was, or indicate "not covered" if you did not receive any of the activities specified below.

	(1) Not at all Useful	(2)	(3)	(4)	(5) Very Useful	(6) Not Covered
4. Introduction of sexual assault risk factors and prevalence in the U.S. and U.S. Department of the Air Force.						
5. Introduction of red dots (e.g., single cases of harmful actions that can add up to hurt someone else), and the spread of red dots on your installation map.						
6. Introduction of proactive/reactive green dots and identification of barriers (e.g., personal, relationship, social, organizational) to acting as green dots.						
7. Introduction of the 3Ds (Direct, Delegate, and Distract) as a method to address the barriers that						

arise when violence occurs.						
8. "The ask" – that Airmen/Guardians will make the commitment to do their part in reducing sexual assault within the U.S. Department of the Air Force.						

Please indicate to what extent you disagree or agree with the following statements below.

	(1) Strongly Disagree	(2)	(3)	(4)	(5) Strongly Agree
9. The implementer encouraged you to participate in discussions and activities during the training.					
10. Enough time was provided for questions and discussion during the training.					
11. The implementer inspired you to contribute to creating a positive climate in the U.S. Air Force.					
12. The training accurately portrayed realistic situations an Airmen/Guardian may face in the U.S. Air Force.					
13. The training helped you understand how to apply the steps of the 3Ds (Direct, Delegate, and Distract) to intervene as a bystander.					
14. The training will help you identify sexual harassment or sexual assault behaviors in the U.S. Department of the Air Force in the future.					
15. The training will help reduce and prevent sexual harassment and sexual assault within the U.S. Air Force.					

16. Please provide any other comments or suggestions to improve future WIT trainings.

Appendix B. VPI Survey Results

As part of our evaluability assessment, NORC conducted a survey with Violence Prevention Integrators (VPIs) at each of the 92 Installations within the US Air Force. Of the 92 Installations invited to participate in the survey, 71 Installations (77%) completed the questionnaire about WIT implementation as well as details pertaining to implementer fidelity of WIT and Airmen/Guardians retention.

Table A1: Current Implementation of WIT Program by MAJCOM

	Yes		**Yes, but not all components**		**No**		**Total Completes**
	N	**%**	**N**	**%**	**N**	**%**	
ACC	8	61%	3	23%	1	15%	**12**
AETC	2	16%	3	25%	7	58%	**12**
AFDW	0	0%	1	50%	1	50%	**2**

AFMC	6	100%	0	0%	0	0%	**6**
AFSGC	2	28%	3	42%	2	28%	**7**
AFRC	2	50%	0	0%	2	50%	**4**
AFSOC	1	50%	1	50%	0	0%	**2**
AMC	4	67 %	2	28%	0	0%	**6**
DAFE	6	87%	1	12%	0	0%	**7**
DAFA	1	100%	0	0%	0	0%	**1**
PACAF	2	25%	1	12%	5	62%	**8**
USSF	1	25%	0	0%	3	75%	**4**
	0	**50%**	0	**20%**	0	**28%**	**71**

Table A1 Summary: This table highlights the implementation status of all of the installations that participated in the VPI Questionnaire. A total of 35 installations reported “Yes” to the implementation of WIT, 14 installations reported “Yes, implemented WIT, but not all components”, and 21 installations reported “No” to the implementation of the WIT program.

Table A2: Current Implementation of WIT Program by MAJCOM and Installation level

	Yes, definitely implementing	**Yes, but not all components→** Verification Status	**No, definitely not implementing** **WIT**	Stated No, not implementing (Needs verification on implementation)	**Total Surveys Complete**
ACC	• Creech AFB • Grand Forks AFB • Moody AFB • Mountain Home AFB • Seymour Johnson AFB • Tyndall • Beale AFB • JBLE	• Davis-Monthan AFB →Needs to be verified • Nellis AFB→ Not implementing • Shaw AFB→ Not implementing		• Offuit AFB	12
AETC	• Good Fellow AFB • Maxwell AFB	• Atlus AFB → Not implementing • JBSA Lackland → Needs to be Verified • Joint Base San Antonio → Needs to be Verified	• Columbus AFB • Holloman AFB • Laughlin AFB • Luke AFB • Vance AFB	• Sheppard AFB • JBSA Randolph	12
AFDW		• Joint Base Andrews→ Implementing	• Bolling Air Force Base		2
AFMC	• Edwards AFB • Eglin AFB • Hill AFB • Robins AFB • Tinker AFB • Wright-Patterson AFB				6
AFSGC	• Francis E Warren AFB • Malmstorm AFB	• Barksdale AFB → Needs to be Verified • Dyess AFB→ Needs to be Verified • Minot AFB → Needs to be Verified	• Whiteman AFB	• Kirtland AFB	7
AFRC	• Niagra Falls Reserve Station • Youngstown Air Reserve Base		• Minneapolis Saint Paul Joint Air Reserve Base	• Fort Worth AFB	4
AFSOC	• Hurlburt Field AFB	• Cannon AFB → Needs to be Verified			2
AMC	• Joint Base McGuire Dix-Lakehurst • Little Rock AFB • Macdill AFB • Travis AFB	• Fairchild AFB → Implementing • Joint Base Lewis-McChord →Implementing			6
DAFE	• Alconbury • Aviano AFB • Lakenheath • Midenhall • RAF Welford • Spangdahlem AFB	• Ramstein AFB→ Not implementing			7
DAFA	• DAFA				1

PACAF	• Joint Base Elemndorf-Rich AFB • Kadena AFB	• Anderson AFB→ Needs to be Verified	• Eielson AFB • Joint Base PRL-HBR-Hickman • Kunsan AFB • Misawa AFB • Yokota AFB		8
USSF	• Buckley AFB		• Peterson AFB • Schriever AFB	• Vandenberg AFB	4

Table A2 Summary: Table 2 provides the names of the installations (within each major command) on how they provided answers to questions on their WIT implementation status.

- Yes, definitely implementing WIT → 35 installations
- No, definitely not implementing WIT → 15 installations
- Reported not implementing WIT and needs verification → 6 installations
- Reported implementing partial WIT and needs verification → 7 installations
- Reported implementing partial WIT, but actually implementing full WIT → 4 installations
- Reported implementing partial WIT, but not implementing WIT→ 4 installations

Installations Implementing WIT (N=35)

Table A3: Average number of Airmen/Guardians receiving WIT among those reported "Yes, currently implementing the WIT program"

MAJCOM	Installation	How many Airmen/Guardians on average participate in the WIT program monthly?			Total Airmen/Guardians participating in WIT
		Tech School	During FTAC	After FTAC	
Air Combat Command (ACC)	Moody Air Force Base	28	31	30	89
	Mountain Home Air Force Base	50	0	0	50
	Other, (Beale)	-	60	10	70
	Seymour Johnson Air Force Base	-	50	-	50
	Creech Air Force Base	-	-	6	6
	Grand Forks Air Force Base	-	-	5	5
	Tyndall Air Force Base	-	-	30	30
Air Education and Training Command (AETC)	Good fellow Air Force Base	70	17	23	110
	Maxwell Air Force Base	6	-	-	6
Air Force Global Strike Command (AFGSC)	Francis E. Warren Air Force Base	-	25	-	25
	Malmstorm Air Force Base	-	20	-	20
Air Force Materiel Command (AFMC)	Edwards Air Force Base	9	10	10	29
	Tinker Air Force Base	0	15	-	15
	Wright-Patterson Air Force Base	19	20	21	60
	Eglin Air Force Base	-	6	6	12
	Robins Air Force Base	-	19	18	37
	Hill Air Force Base	-	-	62	62
Air Force Reserve Command (AFRC)	Niagra Fall Air Reserve Station	-	-	-	100
Air Force Special Operations Command (AFSOC)	Hurlburt Field Air Force Base	38	100	100	238
Air Mobility Command	Joint Base McGuire Dix-Lakehurst	0	20	-	20
	Little Rock Air Force Base	29	41	41	111

(AMC)					
	Macdill Air Force Base	-	50	-	50
	Travis Air Force Base	-	-	30	30
United States Air Forces in Europe-Air Forces Africa (DAFE-Air Forces Africa)	Alconbury	0	9	-	9
	LakenHeath	0	49	10	59
	RAF Welford	0	8	-	8
	Spangdahlem Air Base	0	0	20	20
	Aviano Air Base	-	60	-	60
	Milden Hall	-	60	-	60
US Air Force Academy (DAFA)	DAFA	-	-	20	20

Table A3 Summary: Of the 35 installations that stated full implementation of WIT, the installations with the largest average number of Airmen/Guardians receiving the program were: AETC: Goodfellow AFB, AFSOC: Hurlburt Field AFB, AM: Little Rock AFB, and ACC: Moody AFB.

Table A4: Potential treatment groups (large installations) with approximate number of Airmen/Guardians receiving WIT

MAJCOM	Installation	Number of Airmen/Guardians Receiving WIT monthly at Tech School, During and After FTAC
Air Combat Command (ACC)	Seymour Johnson Air Force Base	50
	Moody Air Force Base	89
Air Education and Training Command (AETC)	Good fellow Air Force Base	110
Air Mobility Command (AMC)	Macdill Air Force Base	50
	Little Rock Air Force Base	111
Air Force Special Operations Command (AFSOC)	Hurlburt Field Air Force Base	238
Air Force Materiel Command (AFMC)	Wright-Patterson Air Force Base	60
	Hill Air Force Base	62

Table A4 Summary: There are eight potential installations that have large cohorts of enlisted Airmen/Guardians. These eight installations may serve as the treatment group, once confirmed by DAF.

Table A5: Frequency of VPI or someone else implementing the WIT program among installations currently fully implementing WIT

	N	%
A few times per week	2	5.6
A few times per year	7	19.4
Once a month	16	45.7
Once per week	8	22.2
Once per year or less	2	5.6
Total	35	100.0

Table A5 Summary: Sixteen of the 35 installations (46%) that reported implementing WIT stated that they implemented the training session once a month.

Table A6: Average number of Airmen/Guardians receiving WIT program monthly and a 4 month projection of number of Airmen/Guardians who can potentially receive WIT.

MAJCOM	Installation	Total Airmen/Guardians participating in WIT monthly, from tech school, during FTAC, after FTAC	Total Airmen/Guardians participating in WIT, projection for 4 month period
Air Combat Command (ACC)	Moody Air Force Base	89	356
	Mountain Home Air Force Base	50	200
	Other, (Beale)	70	280
	Seymour Johnson Air Force Base	50	200
	Creech Air Force Base	6	24
	Grand Forks Air Force Base	5	20
	Tyndall Air Force Base	30	120
Air Education and Training Command (AETC)	Good fellow Air Force Base	110	440
	Maxwell Air Force Base	6	24
Air Force Global Strike Command (AFGSC)	Francis E. Warren Air Force Base	25	100
	Malmstorm Air Force Base	20	80
Air Force Materiel Command (AFMC)	Edwards Air Force Base	29	116
	Tinker Air Force Base	15	60
	Wright-Patterson Air Force Base	60	240
	Eglin Air Force Base	12	48
	Robins Air Force Base	37	148
	Hill Air Force Base	62	248

Air Force Reserve Command (AFRC)	Niagra Fall Air Reserve Station	0	0
Air Force Special Operations Command (AFSOC)	Hurlburt Field Air Force Base	238	952
Air Mobility Command (AMC)	Joint Base McGuire Dix-Lakehurst	20	80
	Little Rock Air Force Base	111	444
	Macdill Air Force Base	50	200
	Travis Air Force Base	30	120
United States Air Forces in Europe-Air Forces Africa (DAFE-AFAFRICA)	Alconbury	9	36
	LakenHeath	59	236
	RAF Welford	8	32
	Spangdahlem Air Base	20	80
	Aviano Air Base	60	240
	Milden Hall	60	240
US Air Force Academy (DAFA)	DAFA	20	80
Total Sum of Airmen/Guardians Receiving/ Potentially Receiving WIT		0	0

Table A7: Average number of implementers trained over 3 year period, and projection of number of trained implementers in 2021

	Q2. How many implementers are trained on average annually over the last three years?		Q2a. How many implementers do you plan on training in 2021?	
	N	%	N	%
0-1	9	12.7	12	16.9
2-5	7	9.9	**14**	**19.7**
6-10	10	14.1	6	8.5
11-20	**18**	**25.4**	**14**	**19.7**
21-30	9	12.7	7	9.9
31-40	2	2.8	0	0.0
41-50	3	4.2	6	8.5
> 51 or more	11	15.5	9	12.7
Missing	2	2.8	3	4.2
Total	71	100.0	71	100.0

Table A7 Summary: Among the 71 installations that completed the survey, 18 installations (25%) reported that approximately 11-20 implementers were trained on average annually over the last three years. When asked how many implementers are projected to be trained in 2021, 14 installations stated 2-5 implementers and 14 installations stated 11-20 implementers.

Table A8: Number of WIT sessions implementers lead per month

	N	%
1-2 sessions	48	67.6
3-5 sessions	9	12.7
6-10 sessions	1	1.4
Missing	13	18.3
Total	71	100.0

Table A8 Summary: Approximately 48 (68%) installations reported that implementers lead 1-2 sessions per a month.

Table A9: Number of implementers phased out of role and length of time implementer remains in role

	Q4a. As of January 2020, how many implementers have phased out or dropped out of their role?			Q5. What is the average length of time an implementer remains a facilitator at your installation? Meaning, once they are trained, how long do they typically lead the WIT intervention?	
	N	%		N	%
0	6	8.5	1-3 months	2	2.8
1-3	5	7.0	10-12 months	3	4.2
4-5	10	14.1	2 years	21	29.6
6-10	9	12.7	3 or more years	3	4.2
> 10	37	52.1	7-9 months	2	2.8
Missing	4	5.6	One year	19	26.8
Total	71	100.0		71	100.0

Table A9 summary: Of the 71 installations, approximately 51% of implementers have phased out or dropped out of their role as implementer after serving for 1-2 years within the position. The reason for phasing out of the position as implementer is mostly because of the following reasons: permanent change of station (PCS), retirement, deployment, or an implementer asks to be removed from position.

Table A10: Implementer fidelity check of WIT program delivery

	Do you (the VPI or lead implementer) complete fidelity check forms regarding the WIT program delivery at your installation?			How are the fidelity checks being completed?	
	N	%		N	%
No	26	36.6	Online survey	1	1.4
Yes	38	53.5	Paper	23	32.4
Missing	7	9.9	Missing	33	46.5
Total	71	100.0			

Table A10 Summary: VPIs were then asked about fidelity checks for conducting WIT sessions and feedback from Airmen/Guardians on the training sessions. VPIs reported that 38 installations (54%) complete implementer fidelity checks to verify that the sessions are being conducted consistently. These fidelity checks are being completed predominantly by paper at 23 installations (32%), and also in-person by visual observations.

Table A11: Airmen/Guardians feedback of WIT Program (Are Airmen/Guardians completing locally developed fidelity check forms?)

	N	%
No	52	73.2
Yes	12	16.9
Missing	7	9.9
Total	71	100.0

Table A11 Summary: The majority of VPIs (52 of the 71 installations) do not have Airmen/Guardians completing fidelity feedback forms.

Appendix C. Data Security

Approaches to Data Collection – Systems, security, and confidentiality

A key consideration as we are building the plans for each ISAPPP evaluation is whether these sensitive data can be collected confidentially in a way that protects the survey participants' rights and does not create a legal liability for the DoD. We recognize that exposure of sensitive data could potentially lead to investigations, disciplinary actions, and criminal liability. However, we believe a number of safeguards can be implemented to address these concerns.

An important consideration will be the data collection approach, specifically as it relates to information technology (IT) systems and survey platforms. This also feeds into the general perceptions that respondents will have when considering their participation in an evaluation and providing potentially sensitive information.

NORC has a robust IT infrastructure that uses a number of data collection (survey) tools and platforms built specifically to accommodate rigorous data collection efforts, with a primary focus on precision, security, and confidentiality. Additionally, our systems offer the ability to seamlessly integrate multiple modes of administration, such as computer assisted personal interviews (CAPI), computer assisted telephone interviews (CATI), and web (online) interviews, all supported by a custom-built project-specific Case Management system (CMS) used to manage production. This CMS allows for the application and monitoring of non-response methodologies, survey returns across modes, electronic prompting (such as SMS/text or email outreach), and regular custom or ad hoc reporting. We have not ruled out the use of telephone prompting and a telephone version of the Airmen/Guardians survey. As we continue developing the evaluation plan, we will learn more about the feasibility of reaching Airmen/Guardians by phone for the follow-up survey (all Airmen/Guardians will be more easily reached by email for the baseline survey since they will be at a base when WIT would be offered or not offered in the case of the comparison condition). If a phone modality is adopted for the follow-up survey NORC's web and telephone (CATI) survey systems are

integrated. That is, the IT systems NORC proposes are already integrated into our telephone centers, which allow for hundreds of staff to dial per day, while accommodating a web mode.*

NORC's infrastructure framework is compliant with the Federal Information Security Management Act (FISMA) to ensure that all data, operations, and assets are protected from security threats. As such, we follow the standards and guidelines set by the National Institute of Standards and Technology (NIST) Special Publication 800-53 rev 4 (Recommended Security Controls for Federal Information Systems and Organizations) at the Moderate level, and the Federal Information Processing Standards (FIPS). Upon request, NORC can deliver, among others, a security plan outlining the storage, backup and recovery procedures and a list of personnel responsible for the security of the systems. All personnel maintaining the systems are trained according to the policies set by each project and client to comply with the data security requirements and manage the usage of data, including personally identifiable information (PII).

The most critical consideration for utilizing NORC's IT infrastructure is that NORC offers an unbiased "outsider" approach to the effort, potentially increasing the sample member's comfort level in responding to a survey. We have received feedback in our work on other projects that people are more forthcoming when they know that their information is collected by an independent organization, using its own software and systems, and who is forthcoming in its confidentiality, anonymity, and security statements. We have seen higher comfort level in surveys we have done at NORC with law enforcement officers on multiple U.S. Department of Justice studies on a range of sensitive topics (e.g., police officer observations of community violence and officer use-of-force against suspects). Participants in the ISAPPP evaluation studies may have concerns if the information they provide resides on DoD servers and within DoD systems, or can be easily accessed by commanders or others to whom they report or with whom they routinely engage. Providing clear assurances that participant information (which is limited in this study to duty email addresses) will be stored separately. Response data will reported only in the aggregate to superiors and DoD leadership, and these assurances will be critical to achieving the response rates needed for reliable outcome evaluations. Additional assurances will be provided through the recruitment process. Individuals recruited for ISAPPP study participation may still have concerns about confidentiality and security of their response data. NORC will use an IRB-approved informed consent form to assure that only military personnel who have voluntarily agreed can respond to the evaluation survey. The informed consent will also notify the participants in plain English about the security measures NORC takes (per above) to maintain all data on approved secure servers used only for research purposes.

* Our online surveys are also able to use unique login PINs and passwords for each sample member such that survey data is securely accessed and stored. The surveys for this study will be anonymous, with no individual sample member identification, therefore not requiring a unique login/PIN.

NORC will also apply for a Certificate of Confidentiality (CoC) from the DoD to protect the evaluation data. The NORC team has applied for and received CoCs (as well as Privacy Certificates for the USDOJ) for numerous DHHS research studies. A CoC will protect the privacy of research participants by prohibiting disclosure of identifiable, sensitive research information to anyone outside of the approved NORC research team except when the subject explicitly releases NORC to do so. The CoC will protect military personnel who have reported sexual assault perpetration through a NORC survey response from disciplinary and legal action. We would include the following type of language in our informed consent statement for the participants:

> "All answers that you give will be kept private. This study has been given a Certificate of Confidentiality. This means that NORC cannot share any of your survey responses with anyone, unless you direct us to release the information or release is mandated by court order. But under the law, we must report to the proper authorities if you tell us you are planning to cause serious harm to yourself or others."

A process for respondents to self-generate their own identification numbers (SGID) has been developed for this survey to further protect the anonymity of respondents. The process for generating these IDs has been reviewed by NORC statisticians to ensure a low likelihood of duplication or identification of a respondent, and a high likelihood that SGIDs remain stable between the first and second survey implementation. At both baseline and follow-up survey administrations, Airmen/Guardians who consent to participate will be advanced to a secure survey webpage to answer four questions on durable personal information (i.e., not personally identifying, but memorable to facilitate exact replication at the follow-up survey [e.g., first two letters of the city where you graduated high school], creating components of a "self-generated unique ID" [SGID]). This process incorporates a series of four questions at the beginning of the survey, the responses to which will be unique to the individual respondent, but combined, will not individually identify them. The 10 component letters/digits of the response data will create a unique 10-digit SGID. Thus, respondents will not know their own SGID; they will not need to remember an SGID; and they will not be able to risk a breach of respondent identity by having it written down somewhere for someone to find. However, by answering the same way to the same four questions at the follow-up survey, the same SGIDs will be generated. NORC staff who are covered under the DoD non-disclosure agreement (but are not directly part of the research team analyzing the data) will develop and maintain a confidential algorithm which will re-sort respondent's SGID into an unidentifiable number and letter string, further ensuring anonymity of respondents. The anonymous survey data will be programmed to save SGID component responses, along with the scrambled unique SGID into one confidential data file (not accessible to the analysis team) accessible to the few programming team members who are privy to the algorithm. This step is required to ensure that the SGID components are being 'scrambled" and masked correctly prior to the

analysis team seeing data. The substantive survey responses are saved into another confidential file for analyses by the research team.

Identifying respondents to deliver incentives: While participants self-generate their ID (“SGID”) in one survey, upon completion of the survey, and submission of their responses, participants are directed to a separate, new survey where they may log a duty email address for NORC to deliver the small incentive that had been offered for completing a survey. The dataset of email addresses is distinct from the dataset of substantive survey responses, which has no personally identifying information (PII). Neither NORC, nor anyone else, will ever know which survey data belongs to which participant (i.e., there are no possible crosswalks linking email addresses to substantive survey responses).

In sum, using NORC systems, and communicating the unbiased, independent collection of data, and providing legal assurances of confidentiality reduces concerns about retaliation or ramifications to reporting sensitive information. NORC is following industry best practices and believe this is the strongest SGID protocol to date to protect anonymity in longitudinal research, both in terms of length of the ID and the additional scrambling algorithm that will be secured separately from the survey response data set.