

## Supporting Statement Outline – Sample

### SUPPORTING STATEMENT – PART B

#### B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS

##### 1. Description of the Activity

This section covers civilian participants in the study. These include 1) veterans of the U.S. military and 2) individuals who are close friends or family members of U.S. military service members and/or veterans. Participants will be recruited from the GfK KnowledgePanel®. We note also that U.S. military service members will also be included in the study. This subpopulation will not be described in detail, but the inclusion of these participants influenced our statistical calculations described throughout this statement.

GfK’s KnowledgePanel® is a probability-weighted panel of individuals aged 18 or older who participate in regular surveys over the Internet. Panel members are provided with technology to respond to surveys to improve representativeness of the whole U.S. population and not just those with Internet access.<sup>1</sup> Individuals recruited to participate in this study will be at least 18 years of age and English-language survey takers. KnowledgePanel® collects demographic and other data from its panel members annually; we will leverage these (deidentified) data for this study to minimize respondent burden.

KnowledgePanel® members are randomly selected through random-digit dialing (RDD) or address-based sampling (ABS).<sup>2</sup> The sample frame of residential addresses covers approximately 97% of U.S. Households. Samples are drawn from among active members using a probability proportional to size (PPS) weighted sampling approach. Individuals may join the panel only after being randomly selected; no one is allowed to “opt in” to the panel. Panel members may complete a maximum of one survey per week; most complete about two surveys per month. We will use a stratified random sampling plan to ensure enough respondents from our target groups ((i.e., veterans or family/friends of a veteran or service member) are included. Invitations for participation, as well as compensation for participation in the study, will come directly from GfK.

All data will be collected through an online survey administered via the GfK Knowledge Panel. Participants will complete a survey, view materials from a health public awareness campaign (random assignment will determine the campaign that each participant views), then complete more survey questions. The surveys administered before and after viewing health public awareness campaign materials contain items that will be used to determine our four main outcome variables: 1) increased knowledge of mental health symptoms and treatment, 2) more positive attitudes and beliefs about mental health and treatment-seeking, 3) feeling comfortable/supported in identifying and seeking help for mental health symptoms, 4) proximal indicators of treatment-seeking (e.g., intentions).

GfK will provide RAND NDRI study team members with a de-identified dataset. Thus, the study team will at no time possess personally identifiable information (PII) about participants.

---

<sup>1</sup> Those participants who do not have access to the Internet are provided a web-enabled computer and free Internet service so that they can also participate as panel members.

<sup>2</sup> GfK. (2013) *Knowledge Panel® Design Summary*.

[http://www.knowledgenetworks.com/knpanel/docs/knowledgepanel\(R\)-design-summary-description.pdf](http://www.knowledgenetworks.com/knpanel/docs/knowledgepanel(R)-design-summary-description.pdf)

Some identifiable information about the respondents is on file with the survey vendor (GfK) because this project will use members of their existing survey panel. However, the researchers will not obtain these identifiers and ask for any additional identifiers during the survey. After the survey has been fielded, GfK will securely send the RAND study team the data that has been stripped of any information that would identify the respondent, per their standard protocols. RAND will not have access to any identifying information.

Further, these policies conform to participant treatment protocols from the Belmont Report. Survey responses are secure and individual responses are not publicly released. When surveys are assigned to KnowledgePanel® panel members, they are notified in their password-protected email account that a survey is available for completion. Surveys are self-administered and accessible any time of day for a designated period. Participants can complete a password-protected survey only once. Members may withdraw from the panel at any time.

Our power calculations and planned analyses focus on the effects of the Real Warriors campaign for the primary (i.e., service members) and secondary target audiences (i.e., veterans, friends or family members of services members or veterans) for the campaign. We have also selected a public health campaign to serve as a control campaign in the study. This control campaign is the Centers for Disease Control (CDC) Let’s Stop HIV Together campaign, designed for the U.S. general population.

Based on a power analysis (described in more detail in section 2c below), we plan to survey 2,772 members of the public. Table 1 below describes the expected sample.

**Table 1. Anticipated sample: Members of the Public**

	T
TOTAL	2772
Veterans of the U.S. military	1378
Family/friends of U.S. military service members or veterans	1378
Pre-test survey respondents	16

Our sample will be drawn from GfK’s KnowledgePanel®. GfK reports that among their enrolled panel members typically “higher than 55%”<sup>3</sup> respond to a request to participate in ongoing surveys; thus, we estimate that approximately 60% of GfK panel members contacted for this study will respond. Since we are unable to calculate a true response rate, we will also conduct survey and item non-response analysis.

2. Procedures for the Collection of Information

a. Statistical methodology for stratification and sample selection

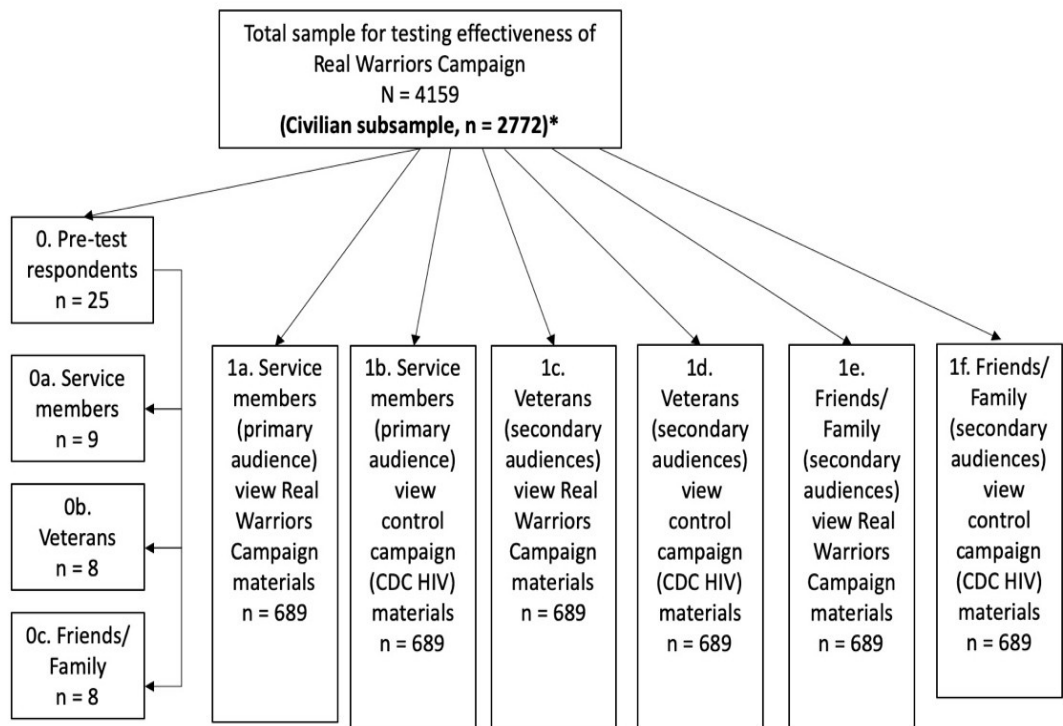
As noted briefly in section 1, KnowledgePanel® draws a sample of respondents using a probability proportional to size (PPS) weighted approach. A customized stratified random sample,

<sup>3</sup> [https://www.gfk.com/fileadmin/user\\_upload/dyna\\_content/US/documents/GfK\\_KnowledgePanel\\_Overview.pdf](https://www.gfk.com/fileadmin/user_upload/dyna_content/US/documents/GfK_KnowledgePanel_Overview.pdf)

with strata based on our specific audiences of interest, will be drawn to ensure adequate representation of individuals in each strata are equally represented.

To determine the necessary sample size for the study, we conducted a power analysis based on the planned analyses. Figure 1 shows the number of participants in each subsample of the study. Though this OMB submission focuses on civilian respondents, we have included information related to U.S. military service members for the sake of completeness.

**Figure 1. Sample size breakdown**



Non-civilian subsample calculation:  
Service member subsample (0a + 1a + 1b), n = 1387

**\*Civilian subsample = 2772** (sum of veteran and friends/family subsamples)  
Veteran subsample (0b + 1c. + 1d.), n = 1386  
Friends/Family subsample (0c + 1e. + 1f.), n = 1386

First, we will estimate the potential impact of the Real Warriors campaign – applied to its primary target group (i.e., military service members) – relative to a group of controls from the same target group (i.e., military service members). During this stage, we will account for multiple comparisons to account for the multiple hypothesis tests associated with the four primary outcome variables (increased knowledge of mental health symptoms and treatment, more positive attitudes and beliefs about mental health and treatment-seeking, feeling comfortable/supported in identifying and seeking help for mental health symptoms, and proximal indicators of treatment-seeking, such as intentions). 4134 individuals will be used to test the efficacy of the Real Warriors campaign, with half (2067 individuals) randomized to the campaign and half randomized to receive the control condition. These sample sizes accommodate multiple comparisons corrections that will control the probability of falsely rejecting the null hypothesis if no such effect exists. Because of

the randomized design, we do not have to be concerned about confounding due to observed or unobserved covariates.

Second, if the Real Warriors campaign is found to have a significant treatment effect for at least one of the primary outcomes, we will assess whether the campaign has a stronger impact when applied to their primary target population(s) than when applied to secondary population(s). We will expose 689 members of each of the secondary populations (veterans, friends and family members) to the campaign. For each outcome for which it was found to be effective, we will estimate whether the campaign has a stronger impact when applied to the primary versus secondary population using a difference in differences analysis, which assesses whether the difference relative to the controls in outcomes for the targeted population was larger than the equivalent difference was for the secondary population. This model will only be run for the campaign/outcome combinations for which the campaign had a significant effect for the primary target population (in step 1). Hence, we do not include a multiple comparison correction for this set of analyses.

In all, our study design for both sets of hypotheses requires 4134 participants (2772 civilian participants; 1378 of which are veterans and 1378 of which are family/friends of veterans or service members), which powers the study to detect “small” effect sizes (Cohen’s  $d = 0.2$ ; Cohen 1988)<sup>4</sup> with 80% power at the standard 95% level of significance throughout.

#### b. Estimation procedures

For each of the four planned outcome variables, we will calculate an individual mean across items and subscales to be included in the outcome measures. In preliminary analyses, we will calculate Cronbach alphas for each domain to assess the internal consistency of the subscales. If any of the alphas are below 0.8, and prior to estimating treatment effects, we will assess whether any of the subscales do not in fact measure the underlying construct of interest and should therefore be dropped. In situations where the response scales differ within a domain (e.g., a dichotomous variable versus a Likert scale), we will re-scale each subscale so that the observed standard deviation is one prior to taking a mean.

Because the study randomizes the treatment assignments, we do not have to be concerned with confounding variables (observed or unobserved) for the first set of hypothesis tests (i.e., the first set of analyses detailed in section 2a). Hence, to test the impact of the Real Warriors campaign applied to its primary target group, we will perform a series of two-sample t-tests to assess whether, on average, the primary outcomes would be significantly different if a population had received the treatment versus control condition. We will perform 12 such statistical tests, for each primary outcome. Because we are testing multiple hypotheses simultaneously, there would be greater than a 5% chance of rejecting at least one of the null hypotheses of at least one unadjusted t-test if, in fact, the campaign did not have an impact on any of the outcomes. Hence, we will correct the t-test p-values using the Holm (1979)<sup>5</sup> correction to account for the multiple comparisons.

---

<sup>4</sup> Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences, 2nd Edition*. Hillsdale, N.J.: Lawrence Erlbaum

<sup>5</sup> Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**, 65–70.

To test whether the campaign applied to their primary target population has a stronger significant treatment effect relative for the campaign applied to a secondary population, we will estimate difference-in-differences analyses to assess whether the treatment effect is different when applied to primary versus secondary target audiences. The form of this model will be

$y_i = \beta_0 + x_p\beta_1 + x_t\beta_2 + x_px_t\beta_3 + x'_c\beta_4 + \varepsilon_i$ . In this regression model,  $x_p$  is an indicator of the population,  $x_t$  is an indicator of exposure to the campaign versus the control, and  $x_c$  is a vector of covariates (such as whether one has personally had a mental health condition), and  $\varepsilon$  is the error term. In this model, we will focus on the statistical significance of  $\beta_3$ , which reflects potential differences in the treatment effect between the primary and secondary populations.

c. Degree of accuracy needed

As noted in the sample size calculation described in detail in section 2a, we powered the study to detect small effects (i.e., Cohen's  $d = 0.20$ ). This effect size was selected based on prior literature. Specifically, one study of mass media campaigns designed to change perceptions of people with mental health disorders found that such campaigns had a "small-to-medium effect."<sup>6</sup> Another study of mental health stigma-reduction interventions found that interventions involving education and some form of contact with a person with a mental health disorder (both characteristics associated with the campaigns being evaluate) found overall effect sizes on outcomes ranging from 0.15 to 0.36.<sup>7</sup>

d. Unusual problems requiring specialized sampling procedures

We do not anticipate unusual problems requiring specialized sampling procedures.

e. Use of periodic or cyclical data collections to reduce respondent burden This is for a one-time data collection.

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

GfK has standard methods to minimize non-response, including limiting the number of survey invitations received in a given month by a single panel member, email-based reminders, and a modest point-based incentive program. We will also plan for survey and item non-response analysis.

4. Tests of Procedures

Participants will first complete a screening questionnaire to determine if they are eligible to participate. The screening questionnaire ensures that individuals are in one of the target populations for the study (i.e., service members, veterans, and their friends and family).

---

<sup>6</sup> Clement et al 2013 Cochrane review

<sup>7</sup> Corrigan et al 2012 meta-analysis

Eligible participants will then complete a pre-survey containing questions about knowledge, attitudes, and beliefs about mental health and HIV. Participants will be randomly assigned to one of the experimental conditions: the Real Warriors Campaign or the control campaign (CDC's Let's Stop HIV Together). All participants will be exposed to media campaign materials (i.e. videos, images, digital versions of print materials like brochures/ articles) corresponding to the condition to which they are assigned. Participants will then complete a post-survey containing questions about their knowledge, attitudes, and beliefs about mental health and HIV; their perceptions of campaign materials; and questions about their own experiences with mental health services.

#### 5. Statistical Consultation and Information Analysis

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

The development of the statistical plan was led by Drs. Joie Acosta, consulting with Dr. Lane Burgette.

Joie Acosta

RAND Corporation

1200 South Hayes Street Arlington, VA 22202-5050

(703) 413-1100

Lane Burgette RAND Corporation

4570 Fifth Ave., Suite 600

Pittsburgh, PA 15213-2665

(412) 683-2300

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

The GfK staff member who will oversee GfK's collection of the data will be named at a later date (when a formal contract is signed between RAND and GfK). GfK will securely transmit a de-identified data set securely to RAND, where Drs. Acosta and Burgette will oversee analysis of the collected information.