Design and methods of the Longitudinal Eating Disorders Assessment Project research consortium for veterans

Kelsie T. Forbush1 | Trevor J. Swanson1 | Melinda Gaddy2 | Mary Oehlert2 | Alesha Doan1 | Robert W. Morgan1 | Colin O’Brien1 | Yiyang Chen1 | Kylie Christian1 | Q. Chelsea Song3 | David Watson4 | Joanna Wiese5

Abstract

Introduction: Military service members must maintain a certain body mass index and body fat percentage. Due to weight-loss pressures, some service members may resort to unhealthy behaviors that place them at risk for the development of an eating disorder (ED).

Objectives: To understand the scope and impact of EDs in military service members and veterans, we formed the Longitudinal Eating Disorders Assessment Project (LEAP) Consortium. LEAP aims to develop novel screening, assessment, classification, and treatment tools for veterans and military members with a focus on EDs and internalizing psychopathology.

Methods: We recruited two independent nationally representative samples of post-9/11 veterans who were separated from service within the past year. Study 1 was a four-wave longitudinal survey and Study 2 was a mixed-methods study that included surveys, structured-clinical interviews, and qualitative interviews.

Results: Recruitment samples were representative of the full population of recently separated veterans. Sample weights were created to adjust for sources of non-response bias to the baseline survey. Attrition was low relative to past studies of this population, with only (younger) age predicting attrition at 1-week follow-up.

Conclusions: We expect that the LEAP Consortium data will contribute to improved information about EDs in veterans, a serious and understudied problem.

Keywords
 assessment, department of defense, disordered eating, eating disorders, longitudinal, mental health, nationally representative, psychiatric disorders, VADIR, veterans
(Allen et al., 2013), which is similar to the prevalence of depression or anxiety. EDs affect people of all ethnicities and socioeconomic levels, and occur across the lifespan (Mitchison et al., 2014; van Hoek & Hoek, 2020). EDs have one of the highest mortality rates among psychiatric disorders (Fichter & Quadflieg, 2016; Iwajomo et al., 2021), with standardized mortality ratios that are two times higher than mortality due to heavy smoking (Chesney et al., 2014). Moreover, EDs are associated with suicide-related mortality in the USA, with suicide rates similar to (or higher than) depression, opioid addiction, or schizophrenia (Chesney et al., 2014). Due to the effects of EDs on general physical health, civilians with EDs incur additional healthcare costs associated with medical events (e.g., myocardial infarction) or due to comorbid health conditions, such as obesity, which is more common among people with EDs than the general population (Striegel Weissman & Rosselli, 2017). Similar to estimates in civilian populations, veteran patients with an ED incurred 1-year adjusted total healthcare costs that were $18,152 higher than matched veterans without an ED (Striegel-Moore et al., 1999), suggesting that there is a critical need to identify and treat EDs in veterans.

Veteran men and women appear to be at risk for the development of an ED (Cuthbert et al., 2020; Masheb et al., 2021). For example, veterans have elevated rates of disordered eating, overweight, and obesity compared to non-veteran civilians (Cuthbert et al., 2020). Although there is a wide prevalence range, recent studies suggest that 14.1%–32.8% of veteran women and 4.1%–18.8% of veteran men have a current ED (Arditte Hall et al., 2017; Forman-Hoffman et al., 2012; Masheb et al., 2021). Military personnel must adhere to body mass index (BMI) and body fat percentage requirements, as well as meet physical performance standards that are assessed every six months. Failure to maintain these standards can result in referral to weight-loss programs and, eventually, discharge from service. The emphasis on body weight within the military may promote unhealthy attempts at weight loss and the development of eating disorders (EDs). For example, being on a weight loss diet while in the military is associated with an increased risk for ED symptoms (Mitchell et al., 2021). If left untreated, EDs result in significantly greater Veterans Affairs (VA) healthcare utilization and physical and psychiatric morbidity and mortality.

Despite the public-health importance of addressing EDs in active-duty and veteran populations, few nationally representative studies of veterans have included ED and assessments of related comorbid psychopathology. To address the need to better understand the scope and impact of EDs in military service members and veterans, the Department of Defense’s Congressionally Directed Medical Research Program (CDMRP) included EDs as an Area of Encouragement for the first time in 2017. Since 2017, the CDMRP has continued to fund studies on EDs, which led to the development of the Longitudinal Eating Disorders Assessment Project (LEAP) research consortium. LEAP brings together researchers and clinicians across state lines to develop novel screening, assessment, classification, and treatment tools for veterans and military members with a focus on EDs and other forms of internalizing psychopathology. Our consortium also includes researchers with expertise in qualitative methods and public policy, with a focus on identifying formal and informal policies and practices that can potentially inform ED treatment implementation efforts in military-relevant populations.

In this paper, we report the design, protocol, and methods of the first two datasets from the LEAP research consortium. We also report initial attrition analyses from the first wave of our longitudinal dataset (Study 1). As we describe in the paper, other data waves from Study 1 and Study 2 were ongoing at the time of this writing. Our primary objective was to develop and validate a military-specific transdiagnostic screening tool for eating and related disorders that can eventually be implemented in healthcare settings. The scientific premise that guided our research was that an integrated transdiagnostic framework for assessing and classifying EDs could lead to improvements in the ability to identify veterans who are most at-risk for poor post-discharge psychosocial adjustment and who need a referral to treatment services. Our primary aim was to validate/develop a transdiagnostic ED screening tool. Secondary aims were to (1) evaluate longitudinal relationships among ED symptoms, internalizing, and externalizing psychopathology, and (2) identify barriers to obtaining ED treatment during military service.

We recruited two independent nationally representative samples of veterans from the VA/DoD Identity Repository (VADIR), a VA office with access to DoD records and the only comprehensive registry of military personnel and veterans in the USA. Study 1 was a four-wave longitudinal survey-based study that assessed veterans at baseline, 1-week, 3-months, and 6-months. Study 2 was a cross-sectional study that included surveys, diagnostic interviews, and qualitative interviews. Although the first two studies from the LEAP consortium were primarily focused on EDs and internalizing disorders, the datasets include a wide range of demographic, military-service related, and other psychopathology constructs that may be of interest to general psychiatric researchers. Given the DoD’s focus on publicly available data and our own commitment to open-science practices, de-identified data (with personal information removed) will be released to the public when it becomes available in compliance with the General Data Protection Regulation. Thus, the goal of this paper is to provide information on the design, protocol, methods for LEAP studies and initial attrition analyses for the first wave of the Study 1 LEAP dataset.

1 Methods

1.1 Ethics approval

Ethical approval for the project was provided by the University of Kansas, VA Eastern Kansas Health Care System (VAEKHCS), and the Department of Defense Human Research Protections Office. All participants provided informed consent prior to engaging in any study-related procedures. The University of Kansas received a HIPAA waiver to recruit veterans who separated from service within the past year from VADIR.
1.2 | Study design overview

Study 1 is a four-wave longitudinal study that included online and paper-based surveys. Participants were assessed at baseline, 1-week follow-up, 3-month follow-up, and 6-month follow-up. Study 2 is a cross-sectional study that includes self-report surveys, structured clinical interviews, and qualitative interviews. In Study 2, veterans will be recruited to participate in a 1-h structured clinical interview that assesses current and lifetime EDs, mood disorders, generalized anxiety disorder, and post-traumatic stress disorder. Veterans who were diagnosed with an ED during the structured clinical interview will be invited to participate in an additional qualitative interview. Veterans with EDs will be recruited until we achieve our target sample size for qualitative interviews (N = 50). Qualitative methods will include semi-structured, open-ended, in-depth interviews that range between 45 and 60 min. These exploratory interviews are designed to identify institutional barriers or informal practices in the military that created obstacles for participants seeking ED and trauma-related care during active duty. Qualitative interviews will be piloted with 10% of the total sample (N = 5) to assess the validity of the questions and modify them according to participants’ feedback. Interviews will be recorded and transcribed verbatim. Transcripts will be coded in vivo (i.e., using the words of participants) whenever possible, using an inductive, open coding format to catalog dominant and frequent domains (Charmaz, 2006). Taxonomic analysis, which entails developing a classification system to synthesize data, will be used to inventory each domain (Bradley et al., 2007).

1.3 | Enrollment status

Baseline and 1-week follow-up recruitment for Study 1 is complete, whereas the 3- and 6-month follow-ups for Study 1 are ongoing. Study 2 launched in March 2022 and recruitment is ongoing (current participant numbers are reported in Table 1).

1.3.1 | VADIR data extraction

Based on power calculations, anticipated attrition, and expected response rates, our Study 1 target sample size was N = 1000 veterans at baseline. For Study 2, our target sample size was N = 400. To achieve our targeted Ns, we recruited two separate, non-overlapping nationally representative samples of veterans from the VADIR that separated from service within the past year. Our rationale for recruiting veterans who discharged within the past year was to collect information as temporally close to their military service as possible to help minimize retrospective recall biases about military-relevant experiences. VADIR has been used in past studies of mental health to obtain nationally representative samples of veterans.

### Table 1 | Study 1 baseline sample characteristics

<table>
<thead>
<tr>
<th>VADIR</th>
<th>Sample</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Army</td>
<td>11936</td>
<td>4065</td>
</tr>
<tr>
<td>Air Force</td>
<td>4203</td>
<td>16905</td>
</tr>
<tr>
<td>Army</td>
<td>1790</td>
<td>7236</td>
</tr>
<tr>
<td>Navy</td>
<td>6031</td>
<td>24180</td>
</tr>
<tr>
<td>Air Force</td>
<td>4308</td>
<td>17235</td>
</tr>
<tr>
<td>National Guard</td>
<td>3208</td>
<td>1240</td>
</tr>
<tr>
<td>Reserve</td>
<td>12602</td>
<td>40068</td>
</tr>
<tr>
<td>Non-Commissioned Officer</td>
<td>6627</td>
<td>26704</td>
</tr>
<tr>
<td>Commissioned Officer</td>
<td>1580</td>
<td>6153</td>
</tr>
<tr>
<td>Warrior Officer</td>
<td>18</td>
<td>0.007</td>
</tr>
<tr>
<td>White</td>
<td>15629</td>
<td>4859</td>
</tr>
<tr>
<td>Native Hawaiian/ Pacific Islander</td>
<td>150</td>
<td>0.06</td>
</tr>
<tr>
<td>Race</td>
<td>1127</td>
<td>4583</td>
</tr>
<tr>
<td>Black or African American</td>
<td>5827</td>
<td>23965</td>
</tr>
<tr>
<td>Asian</td>
<td>414</td>
<td>1675</td>
</tr>
<tr>
<td>Multinational or Unknown Race</td>
<td>1585</td>
<td>6260</td>
</tr>
<tr>
<td>Hispanic</td>
<td>3876</td>
<td>15366</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>20986</td>
<td>84116</td>
</tr>
<tr>
<td>Non-Deployed</td>
<td>17217</td>
<td>6026</td>
</tr>
<tr>
<td>Deployed</td>
<td>7645</td>
<td>30756</td>
</tr>
<tr>
<td>1st Quartile</td>
<td>6841</td>
<td>2746</td>
</tr>
<tr>
<td>2nd Quartile</td>
<td>7618</td>
<td>30404</td>
</tr>
<tr>
<td>3rd Quartile</td>
<td>8496</td>
<td>35196</td>
</tr>
<tr>
<td>4th Quartile</td>
<td>5575</td>
<td>22441</td>
</tr>
<tr>
<td>Age at Discharge</td>
<td>24862</td>
<td>1775</td>
</tr>
</tbody>
</table>

**Note:** This table displays the frequencies and percentages associated with the levels of each demographic variable used in the stratification procedure. For each study, stratification was performed separately for males and females to sample a 1:1 ratio. The meaning of the quartiles for Age at Discharge differed depending on the group. In Study 1, the quartiles for females included ages: (1) 17-24, (2) 25-27, (3) 28-30, (4) 31-66, whereas the quartiles for males included ages: (1) 17-25, (2) 26-27, (3) 28-30, (4) 31-68. In Study 2, the quartiles for females included ages: (1) 17-20, (2) 21-24, (3) 25-28, (4) 29-66, whereas the quartiles for males included ages: (1) 17-21, (2) 22-26, (3) 27-28, (4) 29-79. Note that age-at-discharge quartiles are not exactly 25% due to rounding.
Before beginning the stratification process, we modified the stratification algorithm. Given that there were 93 possible values for age at discharge whose rank was "ROTC," this left a total of 139,718 individuals in the population of interest. Address, telephone number, race, ethnicity, sex, whether the veteran was deployed post-9/11, branch of service (Army, Navy, etc.), pay category (enlisted, non-commissioned officer, officer, or warrant officer), personnel category (reserve, active duty, or national guard) and military rank. We requested two random samples of veterans (i.e., one sample for each study), to ensure that each sample was independent from the other sample. Date of birth and service end dates were used to calculate age at discharge.

1.3.2 Data sharing and accessibility

De-identified and unclassified data collected from this study will be made publicly available through the Center for Open Science (OSF). Our procedures for sharing data will be similar to those used in other large-scale mental-health studies, such as the Minnesota Twin Family Study (Iacono et al., 2006). Investigators interested in using the data should contact the Principal Investigator (KTF) to provide a brief description of the intended use and proposed project timeline. These project descriptions will eventually be posted on OSF. Investigators will also be required to sign a data-use agreement. Such documentation will prevent duplication of research and provide a greater level of protection of intellectual property.

1.3.3 Stratification procedures

**Study 1**

For Study 1, we obtained data for all 179,111 post-9/11 veterans separated from military service between October 1, 2018 and September 30, 2019. From this population, we selected a random sample of 4500 veterans stratified by race, ethnicity, personnel category, rank (which we operationalized through pay grade), branch, age at discharge, and deployment status. We oversampled women to recruit a 1:1 ratio of men to women to achieve an adequate sample of women. We achieved an equal sex ratio by dividing the full population into a group of women and a group of men, and then performed the stratification process separately for each group. To perform the stratification, we developed an R package called "sampleVADIR" (Swanson & Forbush, 2021). The stratification procedure implemented by the sampleVADIR function is described below.

First, we removed individuals from the dataset for whom ethnicity, address, or zip code was missing, as well as individuals whose rank was "ROTC." This left a total of 139,718 individuals in the population from which to sample (24,862 women, and 114,856 men). Before beginning the stratification process, we modified the rank and age at discharge variables into alternative formats to better suit the stratification algorithm. Given that there were 93 possible values for the rank variable (which created too many possible groups to stratify), we converted the rank values into four categories corresponding to pay grades. Specifically, we created groups for enlisted individuals (E), non-commissioned officers (NCO), warrant officers (W), and commissioned officers (O). Given that age at discharge was defined as a continuous variable, we created four groups associated with quartiles observed in the population. The four quartiles were defined by the ages: (1) 17–25, (2) 25–27, (3) 28–30, and (4) 31–68. The first two quartiles differed slightly for men and women: for women, the upper limit of the first quartile was 24 and, for men, the lower limit of the second quartile was 26. A breakdown of the age at discharge quartiles for both studies is provided in Table 2.

After preprocessing, the "sampleVADIR" package computed the number of unique strata defined by every combination of the seven variables used in the stratification. For women, there were 1305 unique strata. The proportion of each stratum within the population of women was then computed and multiplied by the number of women to be sampled; this value was then rounded to the nearest integer. This procedure resulted in a sample of 4500 individuals (n = 2258 women and n = 2242 men).

Unexpectedly, approximately 48% of our initial 4500 recruitment mailings were returned due to invalid addresses. Due to the high proportion of invalid addresses, we used a proprietary stratification procedure to select an additional 14,750 cases from the original Study 1 VADIR data. To identify individuals with invalid addresses, we began mailing postcards to let veterans know they would receive an invitation letter. After implementing this procedure, 24% of 14,750 (3546) had invalid addresses. From the 11,204 valid addresses, we selected 10,000 by randomly removing members of overrepresented demographic groups, which mostly included randomly removing veterans who identified as White. Over-represented groups were identified as those strata that had larger representation in the reduced sample, proportionally, than in the original sample of 14,750. Combining across all Study 1 initial mailings, our incorrect address rate was 38.56%. Of the remaining 8909 individuals that had valid addresses, 2.55% declined to participate and 87.16% did not respond to any mailings. Thus, our rate of enrollment from participants who received mail was 10.29%.

Finally, 642 participants completed the 1-week survey, which was a 70.01% retention rate (or 29.99% attrition rate at 1-week follow-up). Demographic and military service characteristics of the individuals who were selected for recruitment are provided in Tables 1 and 2 (4500 from the initial sample and 10,000 from the additional sample).

**Study 2**

The VADIR variables requested for Study 2 were identical to those requested for Study 1 (see the previous section). For Study 2, we attained a representative sample of 7316 cases from the 119,380 individuals separated from military service between October 1, 2019 and September 30, 2020. Table 2 contains descriptive information regarding the demographic and military service characteristics associated with the total sample of 7316 individuals selected for post-9/11 veterans.
TABLE 2  Study 1 baseline sample characteristics

<table>
<thead>
<tr>
<th></th>
<th>VADIR Sample</th>
<th>Baseline</th>
<th>Weighted</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>7175 (49.48%)</td>
<td>507 (62.13%)</td>
<td>404 (49.51%)</td>
</tr>
<tr>
<td>Male</td>
<td>7325 (50.52%)</td>
<td>309 (37.87%)</td>
<td>412 (50.49%)</td>
</tr>
<tr>
<td><strong>Branch</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Army</td>
<td>7302 (50.36%)</td>
<td>437 (53.55%)</td>
<td>411 (50.37%)</td>
</tr>
<tr>
<td>Air Force</td>
<td>2158 (14.88%)</td>
<td>122 (14.05%)</td>
<td>121 (14.83%)</td>
</tr>
<tr>
<td>Marines</td>
<td>1980 (13.72%)</td>
<td>91 (11.15%)</td>
<td>112 (13.73%)</td>
</tr>
<tr>
<td>Navy</td>
<td>3051 (21.04%)</td>
<td>166 (20.34%)</td>
<td>172 (21.08%)</td>
</tr>
<tr>
<td><strong>Personnel Category</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active Duty</td>
<td>4286 (29.56%)</td>
<td>217 (26.59%)</td>
<td>241 (29.53%)</td>
</tr>
<tr>
<td>National Guard</td>
<td>1908 (13.16%)</td>
<td>110 (13.48%)</td>
<td>107 (13.11%)</td>
</tr>
<tr>
<td>Reserve</td>
<td>8300 (57.28%)</td>
<td>489 (59.93%)</td>
<td>467 (57.25%)</td>
</tr>
<tr>
<td><strong>Rank/Pay Category</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enlisted</td>
<td>9551 (65.87%)</td>
<td>427 (52.33%)</td>
<td>537 (65.81%)</td>
</tr>
<tr>
<td>Non-Commissioned Officer</td>
<td>4210 (29.03%)</td>
<td>309 (37.87%)</td>
<td>237 (29.04%)</td>
</tr>
<tr>
<td>Commissioned Officer</td>
<td>727 (5.01%)</td>
<td>77 (9.44%)</td>
<td>41 (5.02%)</td>
</tr>
<tr>
<td>Warrant Officer</td>
<td>12 (0.80%)</td>
<td>3 (0.37%)</td>
<td>1 (0.12%)</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Native Hawaiian/Pacific Islander</td>
<td>52 (0.36%)</td>
<td>1 (0.12%)</td>
<td>3 (0.37%)</td>
</tr>
<tr>
<td>Asian</td>
<td>605 (4.17%)</td>
<td>36 (4.41%)</td>
<td>34 (4.17%)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>2671 (18.42%)</td>
<td>98 (12.01%)</td>
<td>150 (18.38%)</td>
</tr>
<tr>
<td>American Indian or Alaskan Native</td>
<td>165 (1.14%)</td>
<td>10 (1.25%)</td>
<td>9 (1.13%)</td>
</tr>
<tr>
<td>Multinational or Unknown Race</td>
<td>701 (4.83%)</td>
<td>42 (5.15%)</td>
<td>39 (4.78%)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>2017 (13.91%)</td>
<td>97 (11.89%)</td>
<td>114 (13.97%)</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>12483 (86.09%)</td>
<td>719 (88.11%)</td>
<td>702 (86.03%)</td>
</tr>
<tr>
<td><strong>Deployment Status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Deployed</td>
<td>8839 (60.96%)</td>
<td>452 (55.39%)</td>
<td>497 (60.91%)</td>
</tr>
<tr>
<td>Deployed</td>
<td>5661 (39.04%)</td>
<td>364 (44.61%)</td>
<td>319 (39.09%)</td>
</tr>
<tr>
<td><strong>Age at Discharge</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Quartile</td>
<td>3822 (26.36%)</td>
<td>127 (15.56%)</td>
<td>215 (26.35%)</td>
</tr>
<tr>
<td>2nd Quartile</td>
<td>4339 (29.86%)</td>
<td>237 (29.04%)</td>
<td>244 (29.95%)</td>
</tr>
<tr>
<td>3rd Quartile</td>
<td>3006 (20.73%)</td>
<td>182 (22.33%)</td>
<td>169 (20.71%)</td>
</tr>
<tr>
<td>4th Quartile</td>
<td>3342 (23.05%)</td>
<td>270 (33.09%)</td>
<td>188 (23.04%)</td>
</tr>
<tr>
<td><strong>Total Size</strong></td>
<td>14500</td>
<td>816</td>
<td>816</td>
</tr>
</tbody>
</table>

Note: Shows the frequencies and percentages associated with the levels of each demographic variable used in the stratification procedure. For each study, stratification was performed separately for males and females to sample the sexes with a 1:1 ratio. An implication of this was that the meaning of the quartiles for Age at Discharge differed depending on the group. In Study 1, the quartiles for females included ages: (1) 17-24, (2) 25-27, (3) 28-30, (4) 31-66, while the quartiles for males included ages: (1) 17-25, (2) 26-27, (3) 28-30, (4) 31-68. In Study 2, the quartiles for females included ages: (1) 17-20, (2) 21-24, (3) 25-28, (4) 29-66, while the quartiles for males included ages: (1) 17-21, (2) 22-26, (3) 27-28, (4) 29-79.

recruitment for Study 2. The same stratification strategy used to obtain the Study 1 recruitment sample was used to obtain the Study 2 recruitment sample.

1.4  Description of recruitment procedures and materials

After receiving approval from our local and national VA privacy officers, VAEKHCS shared the stratified sample data with the University of Kansas through encrypted email. The stratified sample data were saved on a dual-authenticated research server at the University of Kansas, and veterans listed in the sample were sent recruitment mailings.

We used a modified version of Dillman’s Tailored Design Method (Dillman, 2011; Hoddinott & Bass, 1986) to recruit participants. This three-stage procedure has been shown to maximize responses to survey-based research (Dillman et al., 2009). Across this three-stage procedure, a total of five mailings were sent to participants. Past studies that used the Tailored Design Method to recruit veterans from VADIR achieved response rates of 25%-44% (Coughlin et al., 2011). An initial letter (or postcard), brochure, and decline-to-participate card invited veterans to participate in an online Qualtrics® study (Stage 1). The initial letter also included a $1.00 incentive payment—regardless of whether the invited veteran chose to participate in the study—because this technique has been shown to increase survey response rates in previous research (Coughlin et al., 2011). We attempted to send non-responders a reminder letter two weeks after the initial mailing (Stage 2); however, due to printing delays and COVID-19 pandemic-related labor shortages at the U.S. postal service that impacted mail delivery speed, the time between each reminder varied (see Table S1). Next, non-responders were sent up to three mailings (Stage 3) that contained a reminder letter, two informed consent documents, decline postcard, study questionnaires, and a pre-paid return envelope. On August 12, 2021 the contents of reminder mailing packets were adjusted due to printing and postage costs. After this date, the first two reminder mailings contained a letter and decline postcard only. However, the final (third) reminder included a letter, two informed consent documents, decline postcard, study questionnaires, and a pre-paid return envelope.

1.5  Measures

A list of measures is included in Table 3. Internal consistency for each measure is provided in Table 4. Unless otherwise noted, measures were administered in both Studies 1 and 2. Study measures are described below:
1.5.1 | Alcohol Use Identification Test (AUDIT)

The AUDIT is a 10-item self-report screen for problematic alcohol use (Babor et al., 1992). The AUDIT has shown high internal consistency across a variety of samples and has a median sensitivity of 0.86 and median specificity of 0.89 (Reinert & Allen, 2002). The AUDIT was administered in Study 2.

1.5.2 | Clinical Impairment Assessment (CIA)

The CIA is a 16-item measure evaluating the severity of psychosocial impairment secondary to an ED (Bohn et al., 2008). This scale has shown evidence for high internal consistency (α = 0.97) and test-retest reliability (r = 0.86; Bohn et al., 2008).

1.5.3 | Demographics, body mass index, and weight history

We adapted the National Survey of Veterans (Department of Veterans Affairs, 2010) to collect basic and military-specific demographic information from veterans. Additional ad hoc demographic measures included: lifetime history of ED behaviors, body mass index (BMI) kg/m², and participants’ weight history. We also administered questions adapted from Garber et al. (2008) and the Eating Pathology...
Symptoms Inventory (Forbush et al., 2013) to assess past ED behaviors to meet military requirements.

1.5.4 | Deployment Risk and Resilience Inventory-2 (DRRI-2)

We used three scales from the original DRRI-2 inventory (Vogt et al., 2013): Pre-deployment Life Events, Combat Experiences, Relationships During Deployment. DRRI-2 scales have shown evidence for good internal consistency ($\alpha > 0.80$), criterion validity, and test-retest validity ($\alpha = 0.61-0.94$; Maoz et al., 2016; Vogt et al., 2013). The DRRI-2 scales showed high discriminant validity across scales (Vogt et al., 2013).

1.5.5 | Eating Disorder Examination Questionnaire (EDE-Q)

The EDE-Q (Fairburn & Beglin, 1994) is a 36-item questionnaire that assesses ED psychopathology. The EDE-Q has four subscales: Eating Concern, Weight Concern, Shape Concern, and Restraint. Subscales...
can be averaged to form a Global Score. The EDE-Q has demonstrated evidence for high internal consistency, test-retest reliability, and criterion-related validity (Berg et al., 2012), although concerns have been raised about factor structure replicability and discriminant validity (Forbush et al., 2013; Thomas et al., 2014).

### 1.5.6 | Eating Pathology Symptoms Inventory (EPSI)

The EPSI (Forbush et al., 2013) is a 45-item questionnaire designed to measure ED behaviors and cognitions. Initial development and validation of EPSI scales demonstrated evidence for strong internal consistency ($\alpha = 0.84–0.89$) and test-retest reliability ($r = 0.73$). The EPSI has shown excellent convergent and discriminant validity (Forbush et al., 2013, 2014, 2020).

### 1.5.7 | EPSI-Clinician-Rated version (EPSI-CRV)

The EPSI-CRV (Forbush et al., 2020) is a semi-structured clinical interview that generates both dimensional scores and current DSM-5 ED diagnoses. The EPSI-CRV has shown high inter-rater reliability, discriminant and convergent validity, and criterion-related validity (Forbush et al., 2020). The EPSI-CRV was administered in Study 2.

### 1.5.8 | Externalizing Spectrum Inventory—Brief (ESI-Brief)

The ESI-Brief is a 145-item assessment tool that evaluates three higher-order dimensions of externalizing psychopathology, including: Callous Aggression, Substance Use, and General Disinhibition, which demonstrated high internal consistency (Patrick et al., 2013). The ESI-Brief was administered in Study 1.

### 1.5.9 | Generalized anxiety disorder-Q7

The GAD-7 (Spitzer et al., 2006) is a seven-item measure of generalized anxiety with strong support for its construct validity and reliability. Recent research found the GAD-7 had poor sensitivity and a high false-positive rate for identifying generalized anxiety disorder and other anxiety disorders (Beard & Björgvinsson, 2014). However, the GAD-7 performed well as a measure of anxiety symptom severity in a large psychiatric sample (Beard & Björgvinsson, 2014). The GAD-7 was administered in Study 2.

### 1.5.10 | Inventory of Depression and Anxiety Symptoms-II (IDAS-II)

The IDAS-II is a 99-item self-report questionnaire that contains 11 non-overlapping scales assessing depression, anxiety, mania, obsessive-compulsive, and trauma-related psychopathology. The IDAS-II has shown good-to-excellent internal consistency with median $\alpha$’s ranging from 0.84 to 0.87 across a variety of samples (Watson et al., 2012). IDAS-II scales have shown strong discriminant, convergent, and criterion validity (Watson et al., 2012).

### 1.5.11 | The Post-Traumatic Stress Disorder Checklist (PCL-5)

The PCL-5 (Blevins et al., 2015) is a 20-item screening measure. Studies show the PCL-5 has good stability, internal consistency, test-retest reliability, and convergent and discriminant validity (Bovin et al., 2016; Wortmann et al., 2016). The PCL-5 was administered in Study 2.

### 1.5.12 | Patient Health Questionnaire Depression Scale (PHQ-9)

The PHQ-9 (Kroenke et al., 2001) is a nine-item self-report measure of depression that is well-validated and reliable. The PHQ-9 has sensitivity and specificity of 0.88 for identifying major depression in primary care and obstetric-gynecology clinics (Kroenke et al., 2001). The PHQ-9 was administered in Study 2.

### 1.5.13 | SCOFF

The SCOFF (Morgan et al., 1999) is a five-item self-report screening measure for EDs. The SCOFF was administered in Study 2. Each letter in SCOFF corresponds to one of the items: making oneself sick; losing control over eating; losing at least one stone in weight; feeling fat; and believing food has control over one’s life. The SCOFF is a widely utilized instrument, with evidence for good specificity and sensitivity (see Kutz et al., 2020, for a review).

### 1.5.14 | The Structured Clinical Interview for DSM-5 Disorders (SCID-5)

The SCID-5 (First et al., 2015) is a semi-structured clinical interview that generates current and lifetime psychiatric diagnoses. Studies have demonstrated the specificity/sensitivity, interrater reliability, and predictive validity of the interview (Osório et al., 2019). We administered modules to assess the following lifetime and current psychiatric diagnoses: mood disorders, generalized anxiety disorder, trauma- and stress-related disorders, and eating disorders. The SCID-5 was administered in Study 2.

### 1.5.15 | Qualitative interview

A qualitative interview will be administered to participants in Study 2. Participants will answer non-numerical questions concerning
experiences in the military. Specifically, they will be asked about institutional and organizational barriers related to seeking or receiving treatment for eating and trauma-related psychopathology.

1.5.16 | World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0)

The WHODAS 2.0 (Üstün et al., 2010) is a 36-item measure of psychiatric impairment. Previous studies have found evidence for strong internal consistency ($\alpha = 0.86$) and test-retest reliability ($r = 0.98$) as well as concurrent and predictive validity (Gold, 2014; Üstün et al., 2010).

2 | RESULTS

2.1 | Sample representativeness

To assess sample representativeness, we computed correlations by creating a single variable to represent all unique strata within the full population of veteran men and women. Next, we correlated the frequencies of strata observed in the population with the strata we observed in our sample. In both Studies 1 and 2, the correlation of demographic characteristics between the stratified sample and the full population of veterans was 0.95 or higher for each sex; thus, our stratified samples were representative of the full populations from which they were drawn. Further details are provided in Table 4. As a supplementary analysis, we also tested representativeness of our sample across each of the seven categorical demographic variables by comparing the extracted samples to the full VADIR populations for each demographic variable using chi-square tests (see Supporting Information S1).

2.2 | Response bias and attrition analyses

For the Study 1 (longitudinal) sample, we conducted two analyses to investigate the extent to which demographic characteristics that were used to draw stratified samples from VADIR were associated with (1) non-response to our initial recruitment [response bias analysis], (2) retention from our initial recruitment at 1-week follow-up [attrition analysis], (3) missing data in the baseline survey for Study 1, and (4) invalid address. The objectives of these analyses were to assess whether certain demographic groups were significantly less likely to (1) participate (respond), (2) continue participation after completing the baseline survey for Study 1 (attrition), (3) have more than 10% of missing data, and (4) have invalid address. For all analyses, we used a logistic regression model with all seven demographic variables used for stratification, along with sex, as predictors. The first analysis was performed on the full recruitment group for Study 1 ($N = 14,500$), wherein we used the logistic regression model to predict participation in the baseline survey. For the second analysis, we examined the baseline sample to identify whether any demographic characteristics were associated with participation in the 1-week follow-up survey. For the third analysis, we examined the baseline sample to identify whether any demographic characteristics were associated with invalid addresses. Analyses included the following procedures: First, we fitted the saturated model (which included all predictors) and a reduced model, which included all terms except one predictor. Next, we conducted a likelihood-ratio test (LRT) comparing the saturated model to each reduced model to determine the significance of each variable in predicting participation at (1) baseline and (2) 1-week follow-up. We conducted post-hoc tests for variables with significant LRT results to assess specific levels of each variable that exhibited significant pairwise differences.

Results of the first analysis suggested that several demographic variables were significant predictors of response to the baseline survey, including gender, rank, race, ethnicity, age of discharge (see Figure 1). Men were significantly less likely to respond to our mailings than women (Odds Ratio [OR] = 0.54, 95% CI = [0.46, 0.62]). Rank (operationalized by pay grade) exhibited a statistically significant omnibus effect (Chi-squared = 26.10, df = 3, $p < 0.001$), such that enlisted individuals were less likely to participate than non-commissioned officers ($p < 0.001$), commissioned officers ($p < 0.001$), and warrant officers ($p = 0.02$). Race exhibited a statistically significant omnibus effect (Chi-squared = 36.26, df = 5, $p < 0.001$), such that people who identified as Black or African American were significantly less likely to participate than those who identified as White ($p < 0.001$), Asian ($p = 0.03$), and multiracial or unknown ($p = 0.004$). For ethnicity, people not of Hispanic origin were significantly more likely to participate in the baseline survey (OR = 1.30, 95% CI = [1.04, 1.63]). Finally, we found a statistically significant omnibus effect for age of discharge (Chi-squared = 22.92, df = 3, $p < 0.001$), such that the fourth quartile was significantly more likely to participate than the first ($p < 0.001$) and second ($p = 0.01$) quartiles. Moreover, the second and third quartiles were significantly more likely to participate than the first quartile ($p = 0.003$, $p < 0.001$, respectively).

For the second analysis, we found that only age at discharge was statistically significantly related to retention in the 1-week follow-up (Chi-squared = 12.81, df = 3, $p = 0.005$) (see Figure 2). Individuals whose age at discharge was in the fourth quartile were significantly more likely to participate than the first ($p = 0.001$), second ($p = 0.02$) and third ($p = 0.01$) quartiles.

For the third analysis, we found that demographic characteristics were not related to missing data. Thus, there was no evidence of demographic differences between participants who had >10% of missing data and those who did not.

For the fourth analysis, several variables were significant predictors of invalid addresses, including branch, personnel category, age-at-discharge, and rank (see Figure 3). Branch exhibited a statistically significant omnibus effect (Chi-squared = 77.67, df = 3, $p < 0.001$), such that Navy veterans were more likely to have invalid...
addresses than Army ($p < 0.001$), Air Force ($p = 0.017$), and Marine veterans ($p < 0.001$). Personnel category exhibited a statistically significant omnibus effect (Chi-squared $= 13.86$, $df = 2$, $p < 0.001$), with Active Duty individuals being more likely to have invalid addresses than the National Guard ($p < 0.001$) and Reservists ($p = 0.047$). Age at discharge had a statistically significant effect (Chi-squared $= 41.09$, $df = 3$, $p < 0.001$), in which the 1st Quartile was more likely to have invalid addresses than the 2nd Quartile ($p < 0.001$); the 2nd Quartile was not significantly more likely to have invalid addresses than the 3rd Quartile ($p = 0.086$); and the 3rd Quartile was more likely to have invalid addresses than the 4th Quartile ($p = 0.016$). Rank had a significant effect (Chi-squared $= 9.53$, $df = 3$, $p = 0.023$)—Enlisted individuals were more likely to have invalid addresses than Non-Commissioned Officers ($p = 0.028$).

2.2.1 Post-stratification weights

To control for non-response bias, we created post-stratification weights, which can be used in future analyses to adjust for non-response bias. We applied a survey raking procedure (also known as "sample balancing") using the R package "anesrake" (Pasek, 2018). Raking is a method used to adjust survey data to ensure the distributions of sample characteristics closely mirror the distributions of population characteristics. In our case, the full population of interest was the group of veterans sampled from the VADIR data who were considered as candidates for participation (e.g., $N = 14,500$ in Study 1). Although Studies 1 and 2 drew samples from VADIR that represented the total population of recently separated veterans, we observed certain sources of non-response bias in which some demographic groups were more or less likely to respond to the surveys than other groups. Therefore, we needed to weight the data collected in both studies to match the characteristics of the recruitment groups to ensure that the final collected samples were representative of the population. Given that data collection for Study 2 is ongoing, we only applied this procedure to Study 1.

The raking algorithm implemented in the "anesrake" package (Pasek & Pasek, 2018) used the seven variables used for stratification (and sex) in the total population and evaluated which distributions deviated by 5% or more in the recruited samples. The algorithm iteratively adjusted the frequencies of stratified variables to produce...
FIGURE 2 One-week follow-up attrition analysis. Plot of odds ratios with 95% confidence intervals for baseline demographic characteristics predicting participation in the 1-week follow-up survey. For each predictor, the associated reference group is listed in parentheses. For instance, the second coefficient from the bottom reflects the odds ratio for Black or African American (vs. White) individuals, and indicates that Black or African American participants were significantly less likely to participate in the 1-week follow-up than White participants. One case was removed from the analysis, as there was only one individual who identified their race as Native Hawaiian/Pacific Islander in the sample that completed the baseline survey. The total sample size at the 1-week follow-up was N = 560.

A weight for each case in the sample until the distributions were aligned with the population distributions for each demographic characteristic. Weighted frequencies of demographic variables were then calculated for each level of each variable to assess the effectiveness of the raking procedure. These values are provided for Study 1 in Table 1. This procedure was highly successful, as indicated by an average error rate of less than 0.01%, with the maximum error for a given variable being 0.06%.

3 | DISCUSSION

The LEAP Consortium is focused on developing novel methods for screening, classifying, and treating veterans and military members with EDs and other internalizing disorders. The purpose of this paper was to provide information on the design, protocol, methods, response bias, and initial attrition analyses for the first two datasets from the LEAP Consortium. Study 1 is nearly complete and Study 2 began in March 2022. Study 1 is a four-wave longitudinal survey design and Study 2 is a cross-sectional study that will include both surveys, structured clinical interviews, and qualitative interviews designed to identify barriers to seeking ED and trauma-related treatment while in the military. Results showed our recruitment samples were highly representative of the full population of recently separated veterans.

Despite the high correspondence between our recruitment sample of veterans invited to participate and the total population of recently separated veterans, not all invited veterans responded to our surveys. We identified several sources of response bias in our initial survey responses. For example, individuals who identified as men and enlisted personnel were less likely to participate in Study 1 compared to women and non-commissioned, commissioned, and warrant officers. Individuals in minoritized populations were also less likely to participate in Study 1 compared to veterans who identified as White. Finally, veterans who were 31–68 years of age at discharge were more likely to participate than veterans who were discharged at 17–25 or 25–27. Attrition analyses also revealed that veterans 31–68 years of age were the least likely to drop out of the study. No other demographic characteristics predicted attrition, suggesting that, overall, attrition largely was random rather than related to specific demographic features. To account for response bias, we created sample weights using a validated raking procedure to ensure
that researchers using LEAP data in the future will be able to
generalize data back to the population from which they were drawn
to ensure representative sampling.

Certain limitations of the study should be noted. First, we
experienced a higher proportion of returned mail than initially expe-
ected and encountered slow response time for mailed surveys due
to U.S. postal service delays. These issues resulted in fewer baseline
participants in Study 1 than we had initially expected. To address this
concern, we recruited additional participants from VADIR and
adjusted our protocol to send an initial postcard to check for inac-
curate addresses prior to sending our first recruitment invitation
letter and enrollment materials. These procedures were successful in
ensuring we were able to recruit sufficient numbers of veterans to
test our longitudinal research aims. Indeed, we observed 33% invalid
addresses—versus 48%—after implementing initial postcards to
check for invalid addresses. However, our procedural change result-
ed in some veterans participating later than the initial sample, often
past the 1-year mark from when they separated from service. For
example, the initial sample was sent their first recruitment letter on
9/15/2020 and the additional sample was sent their first recruitment
letter on 4/23/2021. We also found demographic differences be-
 tween individuals who had an invalid address and individuals who did
not; however, the potential impact of this bias was minimized by
post-stratification weights that correct for non-response bias. Sec-
ond, although we sampled men and women veterans in a 1:1 ratio,
women were more likely to enroll in Study 1. We attempted to

address this problem by creating post-stratification survey weights to
adjust our Study 1 response data to ensure our final sample matched
the full population of veterans recruited from VADIR on demographic
characteristics. Third, due to pandemic-related study issues that
prevented staff from in-person work and unexpected mailing delays,
we were delayed in initiating Study 2. To manage our time con-
straints with budgetary limitations that prevented us from hiring
additional staff to increase the rate of weekly interviews, we chose to
focus diagnostic interviews on current EDs, mood disorders, gen-
eralized anxiety disorder, and post-traumatic stress disorder because
these conditions were the focus of our newly developed self-
reported screening tool. Due to our study design, it was not
possible to conduct analyses on the prevalence or correlates of DSM-
5 defined externalizing disorders and certain internalizing disorders.
However, researchers who are interested in testing hypotheses
related to externalizing psychopathology or additional forms of
internalizing psychopathology could use Study 1 or Study 2 self-
reported data from the ESI-Brief (Patrick et al., 2013) or the IDAS-
II, which comprehensively assess a range of externalizing and
internalizing problems, respectively. Fourth, related to our previous
concerns, we were unable to assess lifetime diagnoses for most
mental-health disorders, given time and budget constraints. Fifth,
Study 2 was ongoing and we were unable to identify (or correct for)
non-response bias in this sample. Future research will focus on
creating post-stratification weights for Study 2, when data collection
is complete. Finally, it is possible that there were retrospective recall
biases that impacted the reporting of questions about barriers to accessing ED care while serving in the military. We attempted to mitigate this issue by recruiting veterans separated from service within the past year, to minimize recall biases. However, future studies are needed to assess access to ED services among active duty military service members (Table 5).

In conclusion, the current study offers a rich source of information about ED and internalizing symptoms and diagnoses, military-relevant variables, and self-reported externalizing psychopathology. Study 1 will represent one of the first longitudinal nationally representative samples of US veterans to include EDs and comorbid disorders, which will provide much-needed information about the temporal relationships among EDs, mood, anxiety, and trauma-related symptoms. Study 2 will provide prevalence estimates for current and lifetime EDs in veterans and identify barriers to seeking ED and trauma-informed care. We expect that the LEAP Consortium datasets will, therefore, contribute to improved information about EDs in veterans, an important and understudied problem.

ACKNOWLEDGMENTS

We thank Danielle A. N. Chapa, Sonakshi Negi, Sarah V. Nelson, and Marianna Thomeczek for their work on IRB submissions, survey distribution, and database management. We thank Sarah N. Johnson and Brianne Richson for participant interviewing. The U.S. Army Medical Research Acquisition Activity, 820 Chandler Street, Fort Detrick MD 21702-5014 is the awarding and administering acquisition office. This work was supported by the Department of Defense, in the amount of $1,721,698, through the Congressionally Directed Medical Research Program’s Peer-Reviewed Medical Research Program, Investigator-Initiated Award, under Award No. W81XWH-19-1-0207. PI: Forbush. Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the Department of Defense.

CONFLICT OF INTEREST

We have no conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT

De-identified and unclassified data collected from this study will be made publicly available through the Center for Open Science (OSF).

Our procedures for sharing data will be similar to those used in other large-scale mental-health studies, such as the Minnesota Twin Family Study (Lacono et al., 2006). Investigators interested in using the data should contact the Principal Investigator (KTF) to provide a brief description of the intended use and proposed project timeline. These project descriptions will eventually be posted on OSF. Investigators will also be required to sign a data-use agreement. Such documentation will prevent duplication of research and provide a greater level of protection of intellectual property.

ORCID

Kelsie T. Forbush https://orcid.org/0000-0002-5900-4204
Robert W. Morgan https://orcid.org/0000-0002-0188-2709

REFERENCES


Bradley, E. H., Curry, L. A., & Devers, K. J. (2007). Qualitative data analysis for health services research: Developing taxonomy, themes, and


**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.