SERVES Statistical Analysis Plan

ClinicalTrials.gov ID: NCT05900479

Date: July 2, 2026

Introduction

Background and rationale

Partnership with psychiatric service dogs has been associated with improvements in posttraumatic stress disorder (PTSD) severity, depression, and anxiety among U.S. military veterans. However, the effect of service dog partnership compared to a waitlisted control condition has not been evaluated in a randomized controlled trial.

Objectives

- Primary objective: Evaluate the effect of psychiatric service dog partnership compared to a waitlisted control condition on self-reported PTSD symptom severity, measured by the *PTSD Checklist for DSM-5* (PCL-5), 8 months post-randomization, among U.S. military veterans and service members (henceforth, veterans) with PTSD. The primary comparison is the mean difference in PCL-5 total score at 8 months between treatment groups, adjusted for baseline PCL-5 score and service dog provider organization.
- Secondary objectives: Evaluate the effect of psychiatric service dog partnership compared to a waitlisted control condition on clinician-rated PTSD symptom severity (*Clinician-Administered PTSD Scale for DSM-5 Revised* [CAPS-5-R]), depression symptoms (*Patient Health Questionnaire-9* [PHQ-9]), and anxiety symptoms (*Patient Reported Outcomes Measurement Information System Anxiety 8a* [PROMIS-A-SF]), 8 months post-randomization. The secondary comparisons are based on the mean difference in the total score approximately 8 months post-randomization, adjusted for the respective baseline score and service dog provider organization.

Estimand

We defined our primary research question using the estimand framework[1], in accordance with the 2025 SPIRIT and CONSORT guidelines.[2,3] The primary analysis targets a hybrid estimand designed to reflect real-world individual behavior by isolating the effect of *receiving* a service dog from external logistical challenges involved in trial administration. Specifically, our hybrid estimand uses the hypothetical strategy to address logistical obstacles to expedited partnership and early waitlist crossover, while retaining the impact of variable participant adherence via the treatment policy strategy. Estimand attributes, including intercurrent events and the operational details of the hybrid strategy are detailed in the Analysis methods section.

Study Methods

The **Ser**vice Dog and **V**eteran **E**xperiences **S**tudy (SERVES) is a multisite, parallel-group, two-arm, randomized controlled trial, with a superiority framework. Participants are randomly allocated in a 1:1 ratio to either psychiatric service dog partnership or waitlist control. All participants have unrestricted access to usual care throughout the trial.

Randomization is stratified by service dog provider organization to maintain balanced allocation across providers. Participants are randomly allocated to service dog partnership or waitlist control in batches with ad hoc block sizes based on the number of participants available for the next scheduled service dog training class, using a computer-generated randomization sequence. While study personnel know the ad hoc block size (i.e., the number of available participants for the next scheduled training class), they are unable to predict individual allocations because the participants are randomized simultaneously within a batch (i.e., class).

The sample size calculation considered power for the primary outcome, which is the PCL-5 symptom severity score. A test for the effect of intervention will be based on the difference in mean responses at 8 months follow-up in the intervention group compared to the control group, averaged over sites (service dog providers). The sample size determination was based on a fixed effects regression model with main effects for treatment group and site, a treatment-by-site interaction and baseline PCL-5 score. Assuming a sample size of 125 participants provide responses at 8 months, we have 80% power to detect an effect size of at least d=0.51, where the effect size is the mean difference between groups, averaged over sites (weighted by site sample sizes) divided by the within treatment-by-site variation. The detectable effect size is smaller than d=0.82 based on the data reported in Leighton et al.[4] for PCL-5 at 3 months. The calculation assumes that the split between intervention and control within each service dog provider is 1:1, that there is a 10:4:1 ratio of sample sizes for the dog providers, and that the baseline scores have identical sample means for the 6 treatment-by-provider combinations. The power is not impacted by variation in the ratios of site sample sizes but is affected to a modest amount by variation in the baseline sample means which are only guaranteed to agree when averaged over randomizations and there is no dropout. It is important to note that the analysis model will be a mixed model for repeated measures with unstructured covariance matrix, fixed effects for time, treatment, dog provider (each treated as a factor), baseline PCL-5 score, interactions for time-by-treatment, provider-by-treatment, and time-by-baseline PCL-5 score. We expect that the power based on the analysis model will exceed that given above due to the additional information on variation provided by the data at intermediate time points.

No interim analyses are planned.

Data management, including scoring of standardized instruments, analyses of primary and secondary outcomes, and manuscript preparation and submission are anticipated to occur within approximately 1 year following the last randomized participants reaching the primary endpoint assessment (8 months follow-up).

Outcome assessments are planned at baseline (pre-randomization) and at approximately 2 months, 5 months, and 8 months follow-up (primary endpoint) for the PCL-5, PHQ-9, and PROMIS-A-SF. The CAPS-5-R is planned at baseline and 8 months follow-up.

Statistical Principles

Confidence intervals and P values

Statistical tests will be two-sided with alpha = 0.05. P values for the primary and secondary outcomes will be calculated using Wald tests.

There is one primary outcome (PCL-5 Total Score at 8 months randomization); therefore no adjustment for multiplicity is needed. Secondary outcomes (CAPS-5-R, PHQ-9, and PROMIS-A-SF) will be analyzed without adjustment for multiple comparisons, and results will be interpreted cautiously.

Effect estimates for the adjusted mean difference between groups at 8 months follow-up will be presented with Wald 95% confidence intervals for the primary and secondary outcomes.

Adherence

In the service dog arm, adherence is defined as formal partnership with a service dog (operationalized as passing the Public Access Test at the end of the training class) following randomization and retention of the service dog through the end of the clinical trial. In the waitlist control arm, adherence is defined as remaining on the waitlist (i.e., absence of service dog partnership) during the trial period. Nonadherence may occur due to participant decision (e.g., declining service dog partnering class; returning service dog following partnering) or provider action (e.g., offering service dog early due to short waitlist; repossessing service dog for welfare or safety concerns).

Adherence will be assessed through provider documentation of Public Access Test completion (indicating service dog partnering took place and informing adherence in both groups) and date of service dog return or repossession (indicating maintenance of service dog partnership, informing adherence in the intervention group).

Adherence will be summarized according to the above criteria, with the number and percentage of participants in each group at each assessment time point.

Protocol Deviations

Major protocol deviations are defined as the research team:

1. Enrolling participants who do not meet eligibility criteria
2. Failing to obtain proper informed consent
3. Providing incorrect randomization allocation information to service dog provider
4. Compromising blinded outcome assessment
5. Not initiating collection of primary trial endpoint data

Protocol deviations will be summarized by type, with the number and percentage of participants in each group.

Analysis population

The analysis population will include all randomized participants.

Trial Population

Screening data

For all participants who completed screening, the following information will be presented: The month and year that recruitment started and ended; the number of individuals contacted for recruitment; the number of individuals who initiated screening (e.g., completed any of the screening activities); the number of participants enrolled; and reasons for non-enrollment. This information will be presented overall, with number contacted and number enrolled also presented according to service dog provider organization.

Eligibility

Eligibility criteria include: applied for and approved to receive a psychiatric service dog from a service dog organization participating in the study (alphabetically, America's VetDogs, Canine Companions, K9s For Warriors); no current or prior service dog partnership; military service; honorable discharge or current honorable service; current diagnosis of PTSD assessed by blinded, independent clinician evaluator; no conviction of crimes against animals; willing to follow the approved study timeline for randomization; and aged 18+ years. Acceptance onto the waitlists for the service dog organizations included additional eligibility screening (e.g., no active substance use disorder; no more than 2 dogs in the home), detailed in Table 1 of the manuscript.

Recruitment and Withdrawal/Follow-up

The CONSORT participant flow diagram will report the number of veterans assessed for eligibility, excluded prior to randomization with reasons (e.g., ineligible, did not complete screening, lost to follow-up), randomized, allocated (intervention vs. control group), received service dog (intervention group) or remained on waitlist (control group), service dog partnership ended (e.g., returned or repossessed with reasons), withdrew, and lost to follow-up.

Baseline participant characteristics

Baseline characteristics will be summarized, including age, gender identity, race, ethnicity, sexual orientation, relationship status, educational level, employment status, military deployment [y/n], branch of service, household characteristics (children [y/n], pet dog ownership [y/n]), clinical characteristics (military sexual trauma [MST], traumatic brain injury [TBI]), concurrent evidence-based treatment (CPT, EMDR therapy, PE therapy), suicidal behavior (C-SSRS lifetime), and service dog provider (stratification variable). Categorical data will be presented as

frequency (percentage), and continuous data (age) will be presented as mean (standard deviation) or median (interquartile range).

Analysis

Outcome definitions

- Primary outcome: PTSD symptom severity via self-report, assessed with *PTSD Checklist for DSM-5* (PCL-5)[5] at 8 months [range 0 to 80]
- Secondary outcomes
    - PTSD symptom severity via blinded clinician rater, assessed with the revised *Clinician-Administered PTSD Assessment for the DSM-5* (CAPS-5-R)[6] at 8 months [range 0 to 200]
    - Depression, assessed with the *Patient Health Questionnaire-9* (PHQ-9)[7] at 8 months [range 0 to 27]
    - Anxiety, assessed with *Patient Reported Outcomes Measurement Information System Anxiety 8a* (PROMIS-A-SF)[8] at 8 months [T-score with a mean of 50 and a standard deviation of 10; range 37.1 to 83.1]
- Exploratory outcomes
    - CAPS-5-R PTSD diagnosis per DSM-5 diagnostic criteria[9]
    - PCL-5 subscales
        - Intrusion [range 0 to 20]
        - Avoidance [range 0 to 8]
        - Negative Alterations in Cognition and Mood [range 0 to 28]
        - Arousal and Reactivity [range 0 to 24]
    - CAPS-5-R subscales
        - Intrusion [range 0 to 50]
        - Avoidance [range 0 to 20]
        - Negative Alterations in Cognition and Mood [range 0 to 70]
        - Arousal and Reactivity [range 0 to 60]

Analysis methods

We designated the effect of *receiving* a service dog as the **target of estimation (i.e., primary estimand)**, and aligned our strategies for handling intercurrent events accordingly.[1] Specifically, we anticipated unavoidable non-compliance with the protocol-specified timeline, due to external factors that determine the fluctuating length of organization-specific waitlists, as well as logistical challenges involved in expedited partnership, and define these as intercurrent events. The hypothetical strategy will be applied to intercurrent events that affect when participants receive a service dog (i.e., too early in the waitlist group, too late in the service dog group), while treatment discontinuation due to side effects or lack of efficacy will be handled with the treatment policy strategy to avoid adherence bias resulting from intrinsic properties of the treatment (e.g., participant returns service dog due to allergies).

Estimand attributes	
Target population	US military members and veterans with PTSD
Treatment and comparator	Psychiatric service dog partnership vs waitlist control condition
Variable (outcome/endpoint)	PCL-5 Total Score, 8 months post-randomization
Population-level summary	The adjusted mean difference in PCL-5 Total Score between treatment groups at 8 months post-randomization
Intercurrent events	**Strategy**
1. Participant in the waitlist control group receives a service dog before the primary endpoint assessment because the provider waitlist was shorter than planned in the protocol (early waitlist crossover)	Hypothetical (counterfactual outcome: waitlist condition is maintained through study follow-up)
2. Participant in the service dog group does not receive a service dog before the primary endpoint assessment due to logistical challenges of expedited partnership (e.g., service dog shortage, training class conflict with employment, childcare, family illness)	Hypothetical (counterfactual outcome: participants receive a service dog during the study as assigned)
3. Participant in the service dog group returns the service dog (e.g., allergies, dog behavior, etc.) or the dog is repossessed by the provider (e.g., concern for human or dog safety) before the primary endpoint assessment	Treatment Policy

The hypothetical strategy involves censoring observations collected after specified intercurrent events and considering the counterfactual outcomes had the intercurrent event not occurred (e.g., had the participant assigned to the waitlist control condition not received a service dog early due to provider-specific waitlist conditions).

The analysis will include all randomized participants, in the groups to which they were randomly assigned. The primary outcome, the PCL-5 total score, will be analyzed with an approach that involves multiple imputation (MI) followed by a mixed model for repeated measures (MMRM), with fixed effects for treatment group (service dog vs. waitlist condition), timepoint (categorical), service dog organization, baseline score of the outcome measure (PCL-5), treatment-by-timepoint interaction, treatment-by-organization interaction, and baseline score-by-timepoint interaction, with an unstructured covariance matrix and Kenward-Roger degrees of freedom adjustment. Since the model includes the baseline PCL-5 score as a covariate rather an outcome, a MMRM in absence of MI would drop any participants who were lost to follow-up or withdrew

from the study prior to their first post-baseline assessment. MI will be used to simultaneously impute missing data due to dropout and to implement the hypothetical strategy by imputing post-censoring outcomes, assuming missing and hypothetical outcomes are missing at random (MAR), conditional on observed data. Multiple imputation by chained equations (MICE) will be conducted with predictive mean matching (PMM). The imputation will accommodate the longitudinal nature of the data. The imputation model will include all variables from the analytical model, including treatment group, service dog provider, timepoint, and baseline PCL-5 total score, as well as pre-specified secondary outcomes as auxiliary variables (i.e., CAPS-5-R total score, PHQ-9 depression score, PROMIS-A-SF score), as they are expected to be correlated. Following MI, a MMRM will be fitted to each of 50 imputed datasets, and the pooled adjusted mean difference at 8 months, with Wald 95% confidence interval and p-value, will be calculated based on Rubin's rules. Model assumptions, including linearity and normality of residuals, will be assessed based on diagnostic plots.

Two of the secondary outcomes (PHQ-9 and PROMIS-A-SF) will be analyzed with the same approach specified for the primary outcome above. Since the CAPS-5-R is only assessed at baseline and 8 months, a linear regression will be fitted for the CAPS-5-R total score, with the following fixed effects: treatment group, CAPS-5-R total score at baseline, service dog organization, and treatment-by-service dog organization interaction.

As a **sensitivity analysis**, we will include additional auxiliary variables in the imputation model that could plausibly influence missingness, including baseline variables that may reflect socioeconomic status, social support, and wellbeing, such as age, gender identity, race, ethnicity, sexual orientation, relationship status, educational level, employment status, pet dog, household income, military sexual trauma, traumatic brain injury, concurrent evidence-based treatment, and suicidal behavior (C-SSRS lifetime).

We also planned a **supplemental estimand** that targets the effect of *assignment* to service dog with a treatment policy strategy for handling all intercurrent events (i.e., all follow-up observations included in analysis). However, external operational barriers, such as fluctuations in the length of organization-specific waitlists that preclude compliance with protocol timelines, may dilute the estimated effect of service dog partnership with noise that is specific to the administrative complexities of this trial, and irrelevant to the larger issue of informing policy and funding the provision of benefits to veterans with PTSD.

Missing data

Missing data will be reported descriptively, including the number and percentage of participants with missing data at each assessment time point, by treatment group, with reasons when available. The primary analysis assumes data are missing at random (MAR) conditional on observed covariates and outcomes in the model, as well as variables in the imputation model, while the sensitivity analysis additionally includes baseline variables that have the potential to inform missingness.

Additional analyses

Analyses of CAPS-5-R PTSD diagnosis, PCL-5 subscales, and CAPS-5-R subscales will be considered exploratory and interpreted accordingly.

The CAPS-5-R PTSD diagnosis outcome will be modeled using logistic regression, with treatment group (service dog vs waitlist control), CAPS-5-R total score at baseline, service dog provider, and treatment-by-service dog provider interaction. The results will be reported as an estimated odds ratio with 95% confidence interval, as well as the predicted probability of CAPS-5-R PTSD diagnosis given service dog group, the predicted probability of CAPS-5-R PTSD diagnosis given waitlist control group, and the absolute difference in predicted probabilities with a 95% confidence interval.

Exploratory analyses of the PCL-5 subscales will use the same MI and MMRM approach described for the primary outcome (PCL-5 total score). Exploratory analyses of the CAPS-5-R subscales will use the linear regression approach described for the CAPS-5-R total score.

Harms

Adverse events (AE) and serious adverse events (SAE) are collected via passive reporting. The number (and percentage) of AEs and SAEs will be reported by treatment group, overall, and with breakdowns based on whether they occurred at the service dog provider site or elsewhere, and whether the AE/SAE was designated as related to study participation (‘unrelated’, ‘possibly related’, ‘definitely related’). No formal statistical analysis will be conducted.

Statistical software

Analyses will be conducted using R. The mixed model for repeated measures will be fit using the mmrm package, and multiple imputation will be conducted with the mice package.

References

1. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. Guideline E9 (R1). Addendum on estimands and sensitivity analysis in clinical trials. 2019.
2. Hrobjartsson A, Boutron I, Hopewell S, et al. SPIRIT 2025 explanation and elaboration: updated guideline for protocols of randomised trials. BMJ. Apr 28 2025;389:e081660. doi:10.1136/bmj-2024-081660
3. Hopewell S, Chan AW, Collins GS, et al. CONSORT 2025 explanation and elaboration: updated guideline for reporting randomised trials. BMJ. Apr 14 2025;389:e081124. doi:10.1136/bmj-2024-081124
4. Leighton SC, Rodriguez KE, Jensen CL, et al. Service Dogs for Veterans and Military Members with Posttraumatic Stress Disorder: A Nonrandomized Controlled Trial. JAMA Network Open. 2024/6/4 2024;7(6):e2414686-e2414686. doi:10.1001/jamanetworkopen.2024.14686
5. Weathers FW, Litz BT, Keane TM, Palmieri PA, Marx PB, Schnurr PP. PTSD Checklist for DSM-5 (PCL-5). General Information. 2013 2013;
6. Weathers FW, Byrne CA, Marx BP, Schnurr PP, Keane TM. The Clinician-Administered PTSD Scale for DSM-5, Revised (CAPS-5-R)–Past Month–Initial Interview Version. 2022;
7. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatric Annals. 2002/09/01/ 2002;32(9):509-515. doi:10.3928/0048-5713-20020901-06
8. Ader DN. Developing the patient-reported outcomes measurement information system (PROMIS). Medical care. 2007;45(5):S1-S2. doi:10.1097/01.mlr.0000260537.45076.74
9. Association AP. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). American Psychiatric Pub; 2013:1679.