Cover Page: Statistical Analysis Plan
We The Village Family Support Study
NCT04250077
Document Date: 8.30.20

Statistical Analysis Plan

Our statistical analysis focuses on a primary endpoint as well as a number of secondary endpoints of interest. The data derive from questionnaires collected from participants in the study. The final enrollment involved three groups: CRAFT-A, CRAFT-C, and PEER, with 13, 11, and 15 subjects enrolled, respectively.

1. **CRAFT-A:** Participants assigned to the CRAFT-A will have access to a 12-module on-line CRAFT intervention and asked to complete one module weekly for 12 weeks. Modules introduce CRAFT concepts and provide workbooks to assist participants in learning and applying the concepts. The modules include: 1)Introduction to CRAFT; 2) Communication Training; 3) Functional Analysis of Drug Using; 4) Positive Reinforcement; 5) Withdrawing Reinforcement; 6) Allowing Natural Consequences; 7) Problem-solving; 8) Life Enrichment; 9) Suggesting Treatment; 10) Recovery and Relapse; 11) Relationship; and 12) Recap of Skills.
   CRAFT-A participants also attend a weekly 60-minute online group sessions facilitated by a CRAFT-certified coach. During weekly group sessions concepts are briefly reviewed, questions are answered, and skills are practiced through role-plays of common situations.
2. **CRAFT-C:** Similarly CRAFT-C has access to the same online materials, but CRAFT-C participants attend a weekly 60-minute individualized on-on-one coaching session with a CRAFT certified coach. During weekly individual sessions concepts are briefly reviewed, questions are answered, and skills are practiced through role-plays of common situations. One-on-one sessions involve roleplays that are tailored to the participants' specific circumstances
3. **PEER:** Participants assigned to the PEER group will participate in an online peer support forum with other CSOs.
   Members of the forum post questions or comments to weekly peer-led discussions and receive responses and feedback from other CSO forum members. Members typically express concerns regarding their IP's wellbeing and ask other members to share any strategies they have employed when dealing with their IPs. Interactions typically, are based either in 12-Step strategies members have learned (usually through Al-Anon or Nar-Anon Family Groups or Family Training Workshops provided by treatment programs) or in CRAFT skills learned (usually from treatment programs or other We The Village members). A staff member from We The Village monitors forum interactions to ensure members are interacting respectfully. This individual also will report any adverse or severe adverse events that members mention online.

The CRAFT-C group is of interest primarily for exploratory and program design points of view, and is not a part of our analysis plan. Our comparisons are all treatment/control comparisons of CRAFT-A and PEER.

Our primary endpoint is the proportion of IPs of participants who enter a new treatment by 3 months after program initiation. We hypothesize that, compared to IPs of participants who receive the PEER control intervention, IPs of participants who receive CRAFT-A experimental intervention will have higher rates of new treatment entry by 3 months after the program. Alternatively, our null hypothesis is that IPs of participants who receive CRAFT-A experimental intervention will have lower rates of new treatment entry.

We also have a number of secondary efficacy endpoint(s):

A. We hypothesize that, compared to IPs of participants who receive the PEER control intervention, IPs of participants who receive CRAFT-A experimental intervention will have higher rates of new medication by 3 months after the program. Alternatively, our null hypothesis is that IPs of participants who receive CRAFT-A experimental intervention will have lower rates of new medication.

B. We hypothesize that, compared to participants who receive the PEER control intervention, participants who receive CRAFT-A experimental intervention will have lower mood disturbance as measured by Profile of Mood States (POMS) global scores by 3 months after the program. Alternatively, our null hypothesis is that participants who receive CRAFT-A experimental intervention will have higher mood disturbance as measured by POMS global scores

C. We hypothesize that, compared to participants who receive the PEER control intervention, participants who receive CRAFT-A experimental intervention will have better physical health as measured by SF-12 physical health subscale (SF-12PH) scores by 3 months after the program. Alternatively, our null hypothesis is that participants who receive CRAFT-A experimental intervention will have worse physical health as measured by SF-12PH scores.

D. We hypothesize that, compared to participants who receive the PEER control intervention, participants who receive CRAFT-A experimental intervention will have higher relationship happiness scores by 3 months after the program as measured by the Relationship Happiness Scale (RHS). Alternatively, our null hypothesis is that participants who receive CRAFT-A experimental intervention will have lower relationship happiness scores.

E. We hypothesize that, compared to participants who receive the PEER control intervention, participants who receive CRAFT-A experimental intervention will have better physical health as measured by SF-12 physical health subscale (SF-12PH) scores by 3 months after the program. Alternatively, our null hypothesis is that participants who receive CRAFT-A experimental intervention will have worse physical health as measured by SF-12PH scores.

F. We hypothesize that, compared to participants who receive the PEER control intervention, participants who receive CRAFT-A experimental intervention will have improved scores on work productivity and decreased work activity impairment by 3 months after the program as measured by the ***SAR (Jane help add something here?)*** Alternatively, our null hypothesis is that participants who receive CRAFT-A experimental intervention will have worse SAR scores.

Descriptive statistics of primary endpoint will be presented as percent of treatment entry events for each study arm, along with 95% confidence intervals of the percentages. For secondary endpoints of RHS, SF-12PH and POMS, we will present means, standard deviations, and 95%

confidence intervals on the means.

Confidence intervals (95%) for proportions, proportion differences, means, and mean differences will be presented, as will one-sided p-values testing the null that the treatment is worse than the control.

Our test statistic for the primary endpoint is proportion difference between treatment and control groups. From a hypothesis testing point-of-view, our null hypothesis is that proportion difference is non-positive and the alternative is that the proportion difference is positive. We will use the Agresti-Caffo two-proportion comparison to perform this analysis. We will also use this approach to analyze the proportions of subjects that receive new medications.

The other secondary endpoints (POMS, RHS, SF-12PH, SAR-SR) test statistics are all mean difference between treatment and control groups. From a hypothesis testing point-of-view, our null hypothesis is that proportion difference is non-positive and the alternative is that the proportion difference is positive. We will use the independent-samples T-statistic comparison to perform this analysis.

Our reporting includes actual p-values, rather than just notes of significance at a given level. Multiple testing inferences can thus be made directly by considering the number of secondary hypotheses.

Based on prior research we expect a treatment entry rate of 60-70% in the CRAFT groups and 15-30% in the control group, as CRAFT studies typically show large effects on Treatment entry (Sisson & Azrin 1986, Miller et al. 1999, Kirby et al. 1999, Meyers et al. 2002, Manuel et al. 2012, Kirby et al. 2017). Our sample size of 13 subjects per group is sufficient for 80% power for the larger of these effect sizes, but it has only 35% power to detect the smallest effect size in this range. We recognize this potential shortcoming within the context of conducting a proof-of-concept pilot study.