Are Treatment Groups Led by Peers as Effective as Groups Led by Counselors for Treating Posttraumatic Stress Disorder and Substance Use Disorder?

Annette Crisanti, PhD, Cristina Murray-Krezan, Jessica Reno, MPH

1University of New Mexico, Albuquerque, NM

Original Project Title: Patient-Centered Trauma Treatment for PTSD and Substance Abuse: Is It an Effective Treatment Option?
PCORI ID: CE-12-11-4484
HSRProj ID: 20143239
ClinicalTrials.gov ID: NCT02081417

Abstract

**Background:** Seeking Safety (SS) is an evidence-based treatment for co-occurring trauma, posttraumatic stress disorder (PTSD), and substance use disorders (SUDs). Evidence of its effectiveness comes from studies that have used trained clinicians to implement the treatment. While the benefits of peer-delivered services have been well documented, the value of peer support workers in the provision of SS is unknown.

**Objectives:** The primary objective was to determine the effectiveness of peer-led (PL) SS groups compared with clinician-led (CL) SS groups in decreasing PTSD symptoms, substance use, and craving and improving coping skills, mental health, and physical health.

**Methods:** A randomized controlled comparative effectiveness trial was conducted from January 2014 to December 2016. Participants (N = 420) were randomized to weekly gender-specific PL-SS groups or CL-SS groups. Primary outcomes included PTSD symptoms, as measured by the PTSD Checklist—Civilian Version (PCL-C), and coping as measured by a scale designed specifically to measure the acquisition of the skills taught in SS. Data were collected at baseline, 3 months, and 6 months post baseline. Participants were recruited from a peer-run wellness center and a residential substance abuse treatment program. In all, 291 adults aged 18 years or older completed a baseline interview; 189 and 174 completed a 3-month and 6-month follow-up interview, respectively. Of the baseline population, 84% were Hispanic, 56% were male, and 78% were unemployed. Nine percent met criteria for PTSD only, 26% met criteria for SUD only, and 65% met criteria for both.

**Results:** Overall, no differences between the CL treatment arm and the PL treatment arm were found in the primary and secondary outcome measures. For the primary outcomes, changes over time between arms for PCL-C scores were measured from baseline to 3 months (mean difference of $\Delta_{PL, 3 \text{ mos} - BL} - \Delta_{CL, 3 \text{ mos} - BL} = 0.18$ [95% CI: –3.4, 3.8]) and from baseline to 6 months ($\Delta_{PL, 6 \text{ mos} - BL} - \Delta_{CL, 6 \text{ mos} - BL} = -0.24$ [95% CI: –4.0, 3.5]). Coping scores were measured similarly (mean difference of $\Delta_{PL, 3 \text{ mos} - BL} - \Delta_{CL, 3 \text{ mos} - BL} = 1.1$ [95% CI: –3.5, 4.3]; $\Delta_{PL, 6 \text{ mos} - BL} - \Delta_{CL, 6 \text{ mos} - BL} = -0.07$ [95% CI: –4.8, 4.7]). An exploratory examination of craving found that participants in PL-SS reported
significantly lower mean craving scores, $\bar{x} = 11.5$ (95% CI: [9.6, 13.4]), than participants in CL-SS, $\bar{x} = 13.5$ (95% CI: [11.6, 15.3]). Participants in both treatment arms experienced similar results, including a significant decrease in drug addiction severity from baseline to 6 months, significantly lower craving and PTSD symptoms over time, and significantly higher coping scores and mental health functioning over time.

**Conclusions:** Overall, PL-SS and CL-SS resulted in similar changes in outcomes and were not statistically different from one another. The study was designed to detect noninferiority between the treatment arms with respect to PTSD. While the estimate for the change score in the PL-SS arm fell fully within our prespecified margins of noninferiority as compared with the CL-SS arm, due to higher variability than expected, the associated confidence bounds fell beyond the hypothesized margins.

**Limitations and Strengths:** Limitations included lack of blinding, attrition, and baseline data not being true baseline in that these data were collected after exposure to 1 SS group. Study strengths included randomization, a large sample size, and patient and stakeholder involvement throughout the process.

---

**Background**

Clinical and general population studies on the prevalence and consequences of trauma have generated enough evidence to confirm 2 things. First, exposure to traumatic events is common among the general population and even higher among persons with mental illness (MI) and substance use disorders (SUDs).\(^1\)\(^,\)\(^2\) Second, the effects of trauma are substantial, affecting physical, mental, emotional, spiritual, social, and economic well-being.\(^3\)\(^-\)\(^7\) An estimated 70% of adults in the United States have experienced a traumatic event (eg, physical abuse and/or sexual abuse) at least once in their lives.\(^8\) The prevalence of trauma is even higher among persons with MI and/or SUDs. Approximately 90% of persons receiving public mental health services have a history of trauma.\(^9\)\(^,\)\(^10\) In the U.S., trauma is associated with a significant burden
in prevalence, mortality, morbidity, individual suffering, loss of productivity, and health and mental health care costs. While diagnoses of depression, anxiety and panic disorders, obsessive-compulsive disorder, and psychotic disorders are common among individuals who have experienced trauma, the most widely known consequences associated with trauma are post-traumatic stress disorder (PTSD) and SUDs. The relationships between trauma, PTSD, and SUDs are complex, but the presence of any 1 increases the risk for developing or increasing the severity of either of the other 2 syndromes.

Trauma-specific interventions are designed to address the psychological and behavioral consequences of trauma in the individual and to facilitate healing.\(^{11}\) Seeking Safety (SS), a present-focused cognitive-behavioral therapy designed to target trauma, PTSD, and SUDs, is the most evidence-based treatment intervention for co-occurring PTSD and SUDs\(^{12}\) and the only effective treatment intervention for co-occurring PTSD and SUD identified by the International Society for Traumatic Stress Studies Practice Guidelines.\(^{13}\) While no specific degree or experience level is required to conduct SS, all of the evidence comes from studies that have used trained clinicians to implement the treatment, including substance abuse or mental health counselors, social workers, psychologists, psychiatrists, case managers, and nurses.\(^{13-20}\) These research findings do not apply to underserved communities that lack mental health professionals. The lack of accessible health services is particularly acute for racial and ethnic minorities and low-income people who live in these shortages areas.\(^{21}\) Rio Arriba County, the location of this study, has a long history of being a Mental Health Professional Shortage Area.\(^{22}\)

Peer-delivered behavioral health services are provided by individuals who identify themselves as having an MI and/or SUDs, are receiving or have received services for their disorders, and have had training that exceeds that of typical layperson patient.\(^{23,24}\) In the behavioral health workforce, these individuals are often referred to as peer support workers (PSWs). In the past decade, peer-delivered services have proliferated. They are an important and accepted component to behavioral health services and programs in many states, including New Mexico, and an innovative way to address workforce shortages. Evidence suggests that peer-delivered services, particularly those provided in a group context, can be effective in
engaging individuals and in improving outcomes. Benefits from these groups have been reported in psychiatric symptoms, self-esteem, quality of life, perceived social support, medication adherence, criminal justice involvement, and substance use problems. Unfortunately, published randomized controlled trials (RCTs) are rare and, except for 1 published study, the benefit of peers in providing trauma treatment has not been examined. Our study addressed this research gap. It also addressed the gap in the SS literature about effectiveness according to the agent of delivery.

High levels of therapeutic alliance (TA)—briefly defined as a strong bond between therapist and patient—would be expected in peer-delivered services, given the trust and the connection between those who provide behavioral health services (ie, PSWs) and the recipients of those services. TA, a combination of patient attachment to the therapist and patient investment in the therapeutic process, is one of the most important and most frequently overlooked variables for predicting response to an intervention. With respect to trauma treatment, according to Ruglass, “the therapeutic relationship is often considered the primary vehicle through which successful trauma treatment occurs.” TA has been shown to be significantly related to decreases in PTSD symptoms and higher attendance in trauma-specific interventions. Little is known about the impact of peer-delivered services on levels of TA, and our study addressed this research gap.

We had 3 specific aims:

1. Determine the difference between peer-led SS (PL-SS) groups and clinician-led SS (CL-SS) groups in improving the lives of people with trauma, PTSD, and SUDs in 5 domains: (1) PTSD, (2) coping skills, (3) substance use, (4) mental health, and (5) physical health. Our hypothesis was that PL-SS compared with CL-SS groups would be as effective in improving outcomes. After funding, our patient partners and stakeholders recommended that we include a measure of craving, given the relevance of this outcome to the target population.

2. Determine whether levels of TA differed among PL-SS and CL-SS groups and the impact of TA on (1) PTSD, (2) coping skills, (3) substance use, (4) mental health, and (5) physical health. Our hypotheses were that levels of TA would be higher and would play more of a
role in affecting outcomes in the PL-SS compared with CL-SS groups. We also examined the impact of TA on craving.

3. Determine whether the standard SS Instructor Training was adequate for peers.

The following section provides an overview of the methodology and results for the first 2 specific aims. In alignment with PCORI Methodology Standard RQ-2, a formal study protocol is available from the PI. We then provide an overview of the methodology and results for the third specific aim.

II. PATIENT AND STAKEHOLDER INFLUENCE ON THE RESEARCH

In alignment with PCORI Methodology Standard PC-1, patients and stakeholders were involved in the identification of the problem, the intervention, the research design, the identification of outcomes, and the instruments for measurement. In addition, patients were involved in the recruitment and retention of subjects, including the identification of strategies to increase both. Patients were also involved in the implementation of the intervention and data collection, including intake, baseline data, and outcome data. Finally, both patients and stakeholders were involved, and will continue to be so, in the interpretation of results and dissemination of findings at local meetings and national conferences. Examples of how patients and stakeholders influenced these areas are presented in the following paragraphs.

Patients and stakeholders participated in a bimonthly steering committee. Stakeholders included local providers who had many years of experience working with this target population and family members. Family members were compensated for their time. The steering committee meetings provided an opportunity to review milestones and deliverables for PCORI. They also provided an opportunity to review staffing, project implementation, recruitment and retention rates, and preliminary data. Challenges with project implementation and recruitment and retention were discussed and strategies were identified. The meetings provided an opportunity for joint decision making in project implementation.
During the project, patients were staffed in various positions, including project director, group support coordinator, transportation coordinator/driver, and group facilitators.

- **Examples of Patient and Stakeholder Involvement in Identifying the Problem and the Intervention:** In 2012, the director of Inside Out contacted the PI, who presented on patient involvement in research at a local conference the director attended and asked to meet with her to talk about the high rates of trauma and addiction among residents in her community and clients at her agency. The PI drove 3 hours to Inside Out and spent the afternoon with the staff. During the meeting, staff shared stories about their clients’ struggle with trauma and addiction and identified the intervention that they thought would work best for their target population (ie, Seeking Safety). They questioned if there was anything they could do to show that peers could, in fact, deliver SS effectively. The PI reported that while there was a strong evidence base for SS, there was no evidence on the effectiveness of peer-delivered SS. The PI further explained that while the evidence on peer-delivered services in general was growing, the studies lacked methodological rigor. After a review of research designs, we decided that the RCT proposed in this study would provide strong evidence on whether peers could deliver SS effectively.

- **Examples of Patient and Stakeholder Influence on the Research Design:** Patients and other stakeholders influenced the approach in 2 ways—both of which had major implications on recruitment and retention. First, we originally proposed a closed-group format that we pilot-tested during the first 3 months of the study. We noticed that recruitment was very difficult and that we were not going to achieve the targeted sample size. A closed-group format meant that some participants had to wait up to 12 weeks in some cases to start groups. This was an obvious problem to the patients and stakeholders, who informed the PI that it would be very unlikely for individuals with severe addiction, once determined eligible, to come back in 12 weeks to start groups/treatment. They recommended that we change the methodology to an open-group format, which immediately improved recruitment. The second way in which patients and other stakeholders influenced the
approach was the decision on when to collect baseline data. To minimize the potential for bias, baseline data are typically collected immediately after randomization, which is what we originally proposed. But after participants received their $20 merchandise card for completing the baseline interview, they never returned to participate in their first treatment group. This did not surprise the patient partners and stakeholders. They recommended that we wait to collect baseline data until after participants attended their first group—knowing that the likelihood of treatment engagement would increase substantially if participants completed even just 1 treatment session. Although this strategy increased the potential for bias, we decided that the methodological limitation was better than not being able to recruit or retain participants.

**Examples of Patient and Stakeholder Influence on the Selection of Outcomes and Data Collection:** Patients and stakeholders also influenced the types of outcomes that were collected and the data collection instruments. For example, the patient partners understood that researchers and clinicians were very interested in measuring medication compliance among this target population; however, they recommended that we focus on other, more relevant outcomes, such as PTSD symptoms and coping skills. After funding, they also recommended that we include a measure of craving, given the prevalence of this problem among the target population and the importance of reducing this outcome. Patient partners reviewed all the instruments for acceptability and cultural appropriateness, and we incorporated their suggestions. A major recommendation was for briefer instruments whenever possible, given the low attention span among the target population and the fear that longer assessments would increase the likelihood of frustration and result in low participation rates. Because 80% of the population was expected to be Hispanic, the patient providers and stakeholders also recommended that we selected instruments that were available in Spanish. Patient partners also ensured that the literacy level and the wording of questions on the instruments were appropriate for the target population. For example, 1 question on the SF-36 asks respondents to indicate how much during the past 30 days they
felt downhearted and blue. Knowing that the target population might not understand the word *blue*, we came up with additional terms in case participants asked for clarification.

- **Examples of Patient and Stakeholder Influence on Recruitment and Retention:** Patients were involved in identifying outreach methods and locations to recruit participants. They revised recruitment flyers originally developed by the PI so that the material would be more appropriate (eg, literacy level and culture) and attractive to the target population. When it seemed unlikely that we would achieve our proposed target sample size, patient partners and stakeholders recommended the site to which we eventually expanded to recruit more subjects. The director of Inside Out (ie, the patient partner) had a connection with a different agency and held the initial meeting with the director of the other agency and received his buy-in. Because of the patient partner’s established level of trust with the local provider, we were able to expand the study to another location and achieve our proposed sample size. One patient was hired to be the group support coordinator; this person was responsible for calling all participants to remind them about their group as well as following up with those who did not show up for groups. Participants said that this type of follow-up by a patient in recovery was much better received than phone calls by a non-peer. The patient partners were also responsible for writing the motivational sayings and identifying the gifts to be used in the PAMI protocol (described later under Strategies to Increase Retention) that would be appropriate for their target population. Finally, patient partners were involved in identifying the best time and days for groups, to increase the likelihood of retention. Rates of recruitment and retention were shared in bimonthly steering committee meetings and strategies were constantly discussed and identified to increase both rates.

- **Examples of Patient and Stakeholder Involvement in the Implementation of the Intervention:** Both patients and stakeholders were involved in identifying the 12 of 25 possible sessions that were used in this study. This process was done through consensus. Each group (patients and stakeholders) was given the full list of 25 sessions and asked to select the 12 that they thought would be most valuable for the target population. There
was consensus among 11 of the 12 sessions. After a lengthy, open, and safe discussion about the discrepancy, the 12th session was agreed on. The 12 sessions included (1) Safety, (2) PTSD: Taking Back Your Power, (3) Detaching From Emotional Pain (Grounding), (4) When Substances Control You, (5) Asking for Help, (6) Red and Green Flags, (7) Integrating the Split Self, (8) Setting Boundaries in Relationships, (9) Discovery, (10) Coping With Triggers, (11) Healthy Relationships, and (12) Healing From Anger. Finally, patients were involved in the implementation of the intervention by facilitating groups.

• **Examples of Patient and Stakeholder Involvement in Data Collection:** Data were collected by trained research assistants who were patients in recovery from a mental illness and/or SUD. This approach presented several advantages. First, peers can put participants at ease during the consent process and research interviews, which increases the likelihood that individuals will agree to participate in the study and complete the data collection process. Second, researchers with “lived experience” understand the barriers individuals may face in participating in research and therefore can help identify strategies to increase the likelihood of participation and decrease the likelihood of attrition for follow-up interviews. Finally, because of their personal experiences, patient interviewers are often able to build a good rapport with participants, who are then more comfortable with sharing their honest opinions.

• **Examples of Patient and Stakeholder Involvement in the Interpretation of Results and Dissemination of Findings:** Data were shared at quarterly steering committee meetings, which provided an opportunity for patients and stakeholders to help interpret preliminary results. In one instance, this exercise proved especially valuable. After a review of some preliminary ASI data at a steering committee meeting, a patient partner said that the results were not making sense to her—they were not aligning with her experiences with the target population. She asked if we could review the data for errors, and it was through this process that we identified the problem with the ASI discussed later (see the Limitations section in the Discussion). In another example, patients were able to help the PI understand findings contrary to expectations (see the discussion of SF-36 physical health scores in the
Discussion). With respect to the dissemination of findings, the patient partners and stakeholders organized a data dissemination meeting held October 21, 2016, in Santa Fe, New Mexico. The patient partners identified the list of stakeholders as well as the format for the meeting. They also helped present the preliminary findings at this meeting. More than 40 stakeholders attended the data dissemination meeting, including the director of the Behavioral Health Collaborative New Mexico Human Services Department and other directors of local treatment agencies and managed health care organizations.

III. METHODS: SPECIFIC AIMS #1 AND #2

A mixed methods research methodology was used with an RCT design. Participants were randomized to CL-SS groups or PL-SS treatment groups. SS groups were delivered once a week, with each group lasting 1.5 hours. An open-group format was used, which allowed patients to join an SS group immediately (ie, at any time). All groups were led by the one and only assigned facilitator. While SS has 25 sessions, 12 were selected by consensus by the patient and stakeholders most familiar with the target population. This resulted in 11 cohorts of 12 sessions over the 3-year study period. Decisions about the dose of SS (12 sessions), the frequency (once per week), the length of each session (90 minutes), and the administration (group format) were based on patient and stakeholder input and the evidence about the most effective way to implement SS, especially to maximize patient participation, motivation, and retention. Based on input from our patient partner, some of the original work by the author of SS on its effectiveness, and a more recent study that examined the efficacy of a condensed SS intervention among a similar population, we defined treatment completers as those who attended 6 or more sessions.29,30

Quantitative data were collected thoroughly and systematically through structured face-to-face interviews with patients using standardized measures (Methodology Standard PC-2). Based on patient input (described later in the Patient and Stakeholder Engagement section), baseline data were obtained immediately following the first treatment episode or within the
first week. Participants were followed to assess the effectiveness of the intervention (PL-SS) compared with the standard treatment (CL-SS) where outcome data were collected at 3 months and 6-months post baseline. Participants received a $20 merchandise gift card for every interview they completed. This study was approved by the University of New Mexico’s Human Research Protections Office Institutional Review Board.

1. Participant Selection

Subjects were recruited from 2 sites located in Rio Arriba County: Inside Out and Hoy Recovery Services. Inside Out is a nonprofit organization (501(c)(3) operated by peers (ie, people with lived experience) and serving individuals with SUDs. Primary services include individual peer counseling, treatment groups, and referral and transportation to detox and rehabilitation centers. Other services include a children’s corner for parents who need to access services but cannot afford daycare, a clothing and food bank, and hygiene kits for the homeless. Services are offered free of charge and no one is denied help. Services at Inside Out are provided by certified peer support workers (CPSWs) who are in recovery from SUDs. Subjects who were enrolled at the Inside Out site were recruited through ongoing outreach to local health, mental health, and criminal justice community-based agencies (eg, community corrections, family health services). In alignment with PCORI Methodology Standard PC-2, our widespread outreach and recruitment process ensured that study participants were representative of the spectrum of the population of interest. Due to concerns about reaching our targeted sample size, recruitment was extended to Hoy Recovery Program Residential Treatment. Hoy Residential Treatment provides multidisciplinary treatment, including individual counseling, group counseling, family counseling, individual treatment plans, and skills training for adults 18 years of age or older with a SUD. Participants were recruited from Inside Out between January 2014 and May 2016 and from Hoy between October and December 2015.
2. Inclusion and Exclusion Criteria

The target population included male and female trauma survivors aged 18 and older, who had symptoms related to PTSD and/or substance use, and who sought outpatient behavioral health services from a peer-run wellness center or inpatient services for SUDs in Rio Arriba County, a rural county in northern New Mexico. Exclusion criteria included a psychiatric hospitalization or suicide attempt in the past 2 months and the inability to provide informed consent to participate in the study. Exclusion criteria were limited so that the study population would be as representative of the target population as possible.

A brief intake was completed by a CPSW through a face-to-face interview of all individuals who sought peer-delivered services from Inside Out during the study period or were admitted to Hoy. The purpose of the intake was to collect basic demographic data, determine criminal justice involvement, identify problem areas (eg, drug use, trauma, mental health problems), determine needs, and identify potential eligibility and interest in the study. Individuals were told about the PCORI-funded study; those interested were then scheduled for an assessment by a licensed clinical mental health counselor (LPCC) with an MA in counseling to determine PTSD and/or SUD eligibility. Although individuals at Hoy were eligible because of their SUD, they still completed an assessment by an external psychologist to determine eligibility because of co-occurring PTSD. Clinical assessments were completed immediately after or within a week of the intake and involved a battery of standardized assessment tools, including:

- The Life Events Checklist (LEC) for DSM-5, to determine a trauma history. The LEC is a self-report checklist that includes 16 events known to potentially result in PTSD or distress. Respondents are asked to indicate their level of experience with each event among the following categories: happened to me, witnessed it, learned about it, part of my job, not sure, and doesn’t apply. One additional question asked about exposure to any other stressful event or experience not covered by the 16 items. Three subscores were generated: number of events experienced, number of events witnessed, and
number of events learned about. The events experienced, and events witnessed were combined for a total number of types of lifetime traumatic events experienced or witnessed, from 0 to 16.

- The Mini International Neuropsychiatric Interview (M.I.N.I.) Section I Alcohol Dependence/Abuse and Section J Substance Dependence/Abuse (Non-Alcohol), to determine substance abuse or dependence, and Section H Post Traumatic Stress Disorder, to determine PTSD.\textsuperscript{32}

Individuals determined to be eligible were randomized to either a gender-specific PL-SS or CL-SS group. Rolling accrual into the full study was implemented and treatment arm allocation was concealed from the study staff, except for the therapist who was conducting the assessments to determine eligibility. Within the 2-gender strata (males and females), cohorts of size $n = 16$ were treated as blocks and the randomization was performed as a factorial randomization wherein we selected random block sizes of 4 and 8 within each cohort of 16 with equal allocation to either treatment arm (PL-SS or CL-SS). In April 2016, an allocation plan for 1 additional cohort of $n = 16$ was created for each of males and females at Inside Out in blocks as described earlier. In September 2015, a second site, Hoy Recovery, was added and a separate randomization was created for that site. We stratified by gender (male and female, $n = 50$ each) and randomized in block sizes of 5 within each stratum, with equal allocation to both treatment arms. The randomization was written in the statistical software SAS 9.3 and 9.4 (Cary, North Carolina) and outputted to Word documents that were saved as PDFs. The allocations were also saved as SAS tables. Randomization seeds were used for reproducibility. Allocation tables were provided to the therapist as hard copies of the PDF documents. The therapist kept the list in a locked drawer in his office, which was also kept locked.

The clinician-led groups were facilitated by a male licensed clinical mental health counselor (LPCC) with an MA in counselling (male group) and a female licensed Alcohol and Drug Abuse Counselor (LADAC) with an MS in developmental psychology (female group). The peer-led groups were facilitated by a male and a female CPSW. In New Mexico, CPSWs must
complete 40 hours of classroom training and a certification exam through the New Mexico Credentialing Board for Behavioral Health Professionals. All facilitators received the same level of training and supervision in SS.

3. Outcome Measures

In alignment with PCORI Methodology Standard RQ-6, we selected outcomes that people in the population of interest notice and care about. The measures were selected through patient input (Methodology Standard PC-1) in terms of cultural appropriateness in the target population. When selecting the measures, we also considered practicality, brevity, and psychometric validations. After funding, the patient partners and stakeholders recommended that we include a measure of craving, given the relevance of this outcome to the target population.

**PTSD Symptoms:** DSM-IV PTSD symptoms were assessed by the self-report 17-item Post-traumatic Stress Disorder Checklist—Civilian (PCL-C). The PCL-C is a sensitive tool to monitor symptom change during and after treatment. The PCL-C is useful because it can be used with any population and has strong psychometric properties. The symptoms endorsed may not be specific to just 1 event, which can be helpful when assessing survivors who have symptoms due to multiple events. Responses are summed to yield a total severity score, with scores ranging from 17 to 85 (higher scores mean higher severity). The National Center for PTSD suggests a change in an overall score of 5 as “responding to treatment” while a change in score of 10 is “clinically meaningful.”

**Coping:** The Coping Scale developed for SS was used to assess coping skills. The Coping Scale directly assesses the degree to which participants report using 18 specific coping skills from SS, scaled from 0 (not at all) to 5 (extremely). Higher scores indicate greater frequency of use of coping skills, with the highest possible score being 90. This scale was selected because it is the most widely used measure of coping in the SS literature and has high internal consistency (Cronbach’s alpha of 0.90).
**Substance Use:** Drug and alcohol problem severity was assessed using the drug and alcohol subscales of the Addiction Severity Index (ASI) Lite. Items assess frequency of drug and alcohol use and abuse within the past 30 days, how bothered the individual is by his/her drug or alcohol problems, and the importance of treatment. Higher composite scores indicate more severe problems. The ASI questions focus on 2 distinct time periods: the past 30 days and lifetime. A number of studies have confirmed the reliability and validity of the ASI. Two composite scores are generated from the ASI-Lite—the Alcohol Composite Score and the Drug Composite Score.

**Mental Health:** Overall mental health was assessed by the Mental Health Summary Scale of the SF-36. The SF-36 questions measure functional health and well-being from the patient’s point of view. The SF-36 has proven useful in differentiating the health benefits produced by different treatments. The Mental Component Summary (MCS) includes an overall measurement of mental health comprised of an assessment of psychological distress and well-being, social and role functioning, and overall vitality. MCS scores range from 0% to 100%, with higher scores indicative of higher functioning. Previous research on the SF-36 defines a clinically significant change as greater than or equal to 1 standard error of measurement.

**Physical Health:** Overall physical health was assessed by the Physical Health Summary Scale of the SF-36. The SF-36 Physical Component Summary (PCS) provides an overall measurement of physical health, which includes both functioning and evaluations of one’s ability to perform physical activity. The Component Summary scores range from 0% to 100%, with higher scores indicative of higher functioning. Previous research on the SF-36 defines a clinically significant change as greater than or equal to 1 standard error of measurement.

**Craving:** The Cocaine Craving Questionnaire Brief (CCQ-B) was used to measure craving but was revised to include cravings specific to the individual’s drug of choice, rather than just cocaine.

**Therapeutic Alliance:** The Revised Helping Alliance questionnaire II (HAq-II) is a well-validated measure of the counselor/patient alliance construct rated from the patient point of view.
HAq-II consists of 19 items rated on a 6-point Likert scale and was used to measure therapeutic alliance. Total scores range from 19 to 114, with higher scores indicative of higher TA.

Session Attendance: Sign-in sheets were used to track session attendance for each participant. Treatment Adherence and Competence: The SS Adherence Scale was used to assess fidelity to SS. Fidelity assessments occurred 9 times over the study for both peer-led and clinician-led SS groups. Fidelity assessments showed that all facilitators were implementing SS with high fidelity to the SS model.

4. Strategies to Increase Retention
   Participants were encouraged to complete all 12 SS sessions. Several strategies were used to encourage retention in SS groups, including providing the following: light refreshments at all groups, a $10 dollar gift card at the 6-session milestone and a second gift card at treatment completion (ie, 12-month session), transportation to and from SS groups, weekly reminder and follow-up phone calls, and child care services. In addition, motivational incentives using the fishbowl method based on an affordable contingency management approach (ie, a low-cost prize incentive system) were used during every session in alignment with the Promoting Awareness of Motivational Incentives (PAMI) guidelines developed by NIDA and SAMHSA/CSAT. The fishbowl method involved patients picking a chit out of a bowl at the end of every group. The chit had varying levels of worth ($0, $1, $5) and patients could go to a file cabinet and select a prize that corresponded to the value on the chit. The $0 chits had motivational sayings on them. Participants received prizes on the spot, resulting in immediate reinforcement. The fishbowl also included 1 chit for a large prize, which was a DVD player.

5. Sample Size and Power Estimates
   In our original proposal, Aim 1 had 5 primary outcomes: (1) PTSD, (2) coping, (3) substance use, (4) mental health, and (5) physical health. However, due to high levels of attrition, PCORI asked us to recalculate our sample size and power to measure our proposed hypotheses. In our Statistical Action Plan (March 2016), we proposed to modify our approach to
include 2 primary outcomes—(1) PTSD and (2) coping—and 3 secondary outcomes—(1) substance abuse, (2) mental health, and (3) physical health. We reestimated our sample size requirements based on the 2 primary outcomes, with the same design and same expected differences.

Therefore, the first aim of this study evaluated 2 primary outcomes—PTSD and coping—and 3 secondary outcomes—substance use, mental health, and physical health. Sample size calculations were calculated in the power analysis software PASS 14.50 We hypothesized that we would not detect significant differences in change outcome measure scores between PL-SS and CL-SS. Estimates were based on the PCL-C for the primary outcome because PTSD symptoms were the most resistant to change among the outcomes. By using the PCL-C to determine power, we ensured that our sample was adequate to detect significant differences in the outcome expected to change the least. Clinically significant improvement in PCL-C scores is a 10- to 20-point decrease from baseline to postintervention.37 After consultation with Dr. Bonham, a psychiatrist in the UNM Psychiatry Department and expert in PTSD, we determined that the change in PCL-C scores in both the PL-SS groups and CL-SS groups from baseline to postintervention (ie, 3 months) would be considered clinically nonsignificant if they fell within 2.5 points of each other. A noninferiority power analysis based on the PCL-C change scores with a standard deviation for change = 5 points for both groups was performed and the results are displayed in Table 1. We assumed a hypothesized mean difference in PCL-C change to be 0 (in PASS, D = 0, the true difference). In the final analysis, if the upper bound of the 2-sided (1– 2α) × 100% confidence interval for the mean difference in change score is less than or equal to the hypothesized noninferiority margin, we can declare PL-SS to be noninferior to CL-SS.49 A positive difference between change scores indicates CL-SS has a larger decrease in PCL-C over time, while a negative difference indicates PL-SS has a larger decrease over time.

Table 1. Sample Size Estimates by Power for Specific Aim #1

<table>
<thead>
<tr>
<th>Power</th>
<th>Sample Size</th>
<th>Noninferiority Margin for PCL Δ Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peer-led SS</td>
<td>Clinician-led SS</td>
<td></td>
</tr>
</tbody>
</table>


Because 2 primary outcome measures were assessed, to address potential multiplicity issues a Bonferroni correction was applied to the type I error rate in the power analysis so that \( \alpha = 0.05/2 = 0.025 \).

While we included many strategies to reduce attrition, treatment dropout and lost to follow-up was high among our extremely difficult-to-treat population. Even with our observed approximate 40% attrition rate, we hypothesized that we would be able to recruit enough participants that have a pretreatment interview (ie, baseline) and posttreatment interview (3 months post baseline) to have 80% power to detect the differences posed for the primary outcome measures (based on lost-to-follow-up trends observed in the first 2.5 years). By the end of the study, we anticipated 144 total participants, 72 per group.

The second aim of this study was to evaluate the effect of intervention groups on TA. We expected the PL-SS arm to report scores at least comparable to the CL-SS arm, if not higher. Based on work published by Ruglass and colleagues, who compared alliance in SS to a Women’s Health Education program, we estimated that HAq-II scores in the PL group at 3 months were considered noninferior to CL if their scores were within 4 points of each other at 3 months (Group 1 SD = 13.0, Group 2 SD = 15.0). Ruglass calculated the mean of the 19 questions, whereas we calculated a total score, so we estimated total HAq-II scores by multiplying their achieved mean score for each group by 19. Using this difference as the hypothesized margin of noninferiority of the PL-SS arm to the CL-SS arm with the given standard deviations, we calculated the power necessary to declare noninferiority for the sample sizes and noninferiority margins (see Table 2). We assumed a hypothesized mean difference in HAq-II change to be 0 (in PASS, \( D = 0 \), the true difference).
Table 2. Power Estimates by Sample Size for Aim #2

<table>
<thead>
<tr>
<th>Power</th>
<th>Noninferiority Margin for HAq-II Score</th>
<th>Sample Size</th>
<th>Peer-led SS</th>
<th>Clinician-led SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.64</td>
<td>–5.0</td>
<td>64</td>
<td>64</td>
<td>64</td>
</tr>
<tr>
<td>0.48</td>
<td>–4.0</td>
<td>64</td>
<td>64</td>
<td>64</td>
</tr>
<tr>
<td>0.75</td>
<td>–5.0</td>
<td>86</td>
<td>86</td>
<td>86</td>
</tr>
<tr>
<td>0.59</td>
<td>–4.0</td>
<td>86</td>
<td>86</td>
<td>86</td>
</tr>
<tr>
<td>0.81</td>
<td>–5.0</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>0.64</td>
<td>–4.0</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>0.90</td>
<td>–5.0</td>
<td>133</td>
<td>133</td>
<td>133</td>
</tr>
<tr>
<td>0.75</td>
<td>–4.0</td>
<td>133</td>
<td>133</td>
<td>133</td>
</tr>
</tbody>
</table>

Subgroup analyses similar to those described in Aim 1 were performed but were exploratory in nature, with the intent to guide future research. The third aim was qualitative and did not require a power analysis.

6. Analysis

Data were entered into Access Databases developed specifically for this project. Random checks of data entry accuracy were conducted to identify data entry errors. As outlined in our proposal, data were analyzed as intention-to-treat, where all participants were analyzed with their assigned intervention arm. Descriptive statistics, such as means,
standard deviations, medians, quartiles, frequencies, and percentages, were used to summarize participant characteristics, overall and by intervention groups. Arms were first compared with t tests, Wilcoxon rank sum tests, chi-square tests, and/or Fisher exact tests, as appropriate. All analyses were performed in SAS 9.4. All analyses for Aims 1 and 2 were prespecified in the proposal and were adhered to per Methodology Standard IR-3.

Missing data were carefully evaluated. Due to the nature of this population, we expected considerable attrition and missing data. We assessed the missingness for systematic problems and for missing not at random (MNAR) to determine whether multiple imputation for missing values could be considered. If one scale was missing from a given visit, typically all of the scales were missing (ie, the interview did not occur). In such a case, some of the missing data may be considered missing at random. The main assumption for using multiple imputation is that missingness does not depend on any observed or unobserved values of variables (MNAR). To assess this assumption, we calculated the number of total missed interviews and compared them by a chi-square test, ANOVA, or t test, as appropriate, with each of the dependent and independent variables included in our models as well as with all demographic data. We also calculated a dichotomous variable that indicated whether the participant was missing any or no interviews and compared it with the dependent and independent variables as above. We confirmed that, for each outcome, missingness was likely related to both observed and unobserved data, thus invalidating the use of multiple imputation to adjust for missingness. For example, participants who missed the 3-month follow-up visit had higher craving scores at baseline than those who did not miss that visit. It is likely that individuals who missed the 3-month craving scores had higher values than those who were not missing the scores. There may be other unmeasured variables that contributed to missingness as well. Because assumptions for multiple imputation were not met, we instead opted to include other collected data in our models that could account for some of the variability due to the missingness: the dichotomous variable missing any interviews versus missing no interviews, site, and living situation. Also considered was the goal to show that PL-SS is as effective as CL-SS and missingness was not related to intervention arm.
Specific Aim 1: Determine the effectiveness of PL-SS groups compared with CL-SS groups in decreasing PTSD and substance use, and improving coping skills, mental health, and physical health. Change scores were calculated for each of the primary and secondary outcome measures by subtracting each participant’s 3-month and 6-month scores from baseline scores. Summary statistics were calculated for the outcome measures at each time point and for the change scores. Linear mixed models were fitted to each of the outcome measures to assess the effect of the intervention group (PL versus CL). Logistic regression mixed models were fitted to ASI Alcohol and Drug Composite scores that were dichotomized at the median scores for each. For all models, a first-order autoregressive (AR[1]) covariance structure was assumed to account for observations closer in time to each other as being more likely to have a higher correlation than observations further apart in time. Full models included covariates for age, gender, ethnicity (Hispanic versus not Hispanic), completion of SS (did not complete, completed by 3 months, completed by 6 months), any missed interviews (any versus none), site (Hoy versus Inside Out), living status (house/apartment/group home/halfway house; homeless/shelter/friend or family’s home; institution: prison/school/hospital or detox center), and time point (baseline, 3 months, and 6 months). Two-way interactions were included between intervention group and each of time points, gender, and ethnicity; between gender and SS program completion; and between site and any missed interview. The interactions between intervention arm and gender and ethnicity were included in the models as our prespecified subgroup analyses to understand any possible heterogeneity of treatment effects. Pseudo-residuals were assessed for normality and quality of model fitting. For the primary outcome of PTSD, if no intervention arm effect was found, meaning that for the outcome measure, the least squares mean difference between the 2 intervention arms fell within the margin of noninferiority, then superiority was assessed with significance set to Bonferroni-corrected type I error rates: \( \alpha = 0.05/2 = 0.025 \) for the primary outcome measures. We did not identify previous literature that reported an expected change over time in coping. Thus, we expected the 2 arms to not differ significantly from each other, but we did not specify a noninferiority endpoint. The Bonferroni correction was used to control for multiple
comparisons due to 2 primary endpoints. For the secondary outcome measures and all subgroup analyses, we did not explicitly hypothesize noninferiority margins between arms but expected the arms to be similar. We report 95% confidence intervals for the secondary outcome measures. We assessed an additional exploratory outcome, craving, in a similar way. The ASI (measuring alcohol and drug use) had highly skewed distributions of scores ranging from 0 to 1. Adequate models for these continuous outcomes could not be fitted; instead, the scores were dichotomized into < median values and ≥ median values followed by analysis using logistic regression mixed models, as described above. Modeling details for the primary, secondary, and subgroup analyses are specified in the Appendix.

Specific Aim 2: Compare levels of therapeutic alliance among PL-SS and CL-SS groups and examine the impact of TA on outcomes. TA was evaluated at 3 months and 6 months. Of primary interest was TA at 3 months followed by whether TA persisted through 6 months. Additionally, change scores were calculated for TA by subtracting each participant’s score at 6 months from his/her 3-month score to assess a time effect. Summary statistics were calculated for the TA scores at both time points, the TA change score, and for the mean TA score. Models similar to those described for Specific Aim 1 were fitted to the Aim 1 outcome measures but included the mean-centered TA score (centered by subtracting overall mean from each participant’s mean) as a fixed covariate along with previously specified interactions and the interactions between mean TA score and intervention group. The mean of the 3- and 6-month TA scores for each participant was used in the linear modeling because TA was collected only at these interviews. We felt that including this variable in the modeling was important to account for variability related to participants’ receptivity to the SS program and leader. The impact of TA was assessed at the α = 0.05 level of significance.

IV. RESULTS: SPECIFIC AIMS #1 AND #2

A total of 431 assessments for eligibility were conducted: 375 at Inside Out and 56 at Hoy. Ninety-eight percent (n = 420) were determined eligible and randomized to a gender-specific CL (n = 208) or PL group (n = 212). Of those determined eligible, 69% (n = 291) attended
their first treatment session and were enrolled in the research: 145 in the CL group and 146 in the PL group. Figure 1 is the CONSORT diagram of participant flow. Of those who completed a baseline interview, 222 participants completed at least 1 follow-up interview. Sixty-nine participants (24%) were lost to follow-up.

Of the 420 individuals eligible for study inclusion and randomized to a treatment arm, 291 participated in the study by attending a session and responding to a baseline interview. Table 3 summarizes the demographic information for the 291 enrolled participants and the 129 eligible, randomized individuals who were not enrolled in the study. No significant differences between the 2 groups were observed.
Figure 1: Consort Diagram of Participant Flow

Intake to Determine Study Interest and Collect Basic Demographic Information and Service Need
N = 533

Brief Screen to Determine Study Eligibility
N = 431

Eligible and Randomized
N = 420

Not Eligible
N = 8

Not Randomized
N = 3

Allocated to Clinician-led Group
N = 208

Allocated to Peer-led Group
N = 212

Attended First Group
N = 166

Attended First Group
N = 163

Baseline Interview
N = 145

Baseline Interview
N = 146

Reasons for Loss to Follow-up
-Could Not Locate: N = 23
-Moved Away: 2
-Incarcerated: 2
-Deceased: 3

Reasons for Loss to Follow-up
-Could Not Locate: N = 26
-Moved Away: 1
-Incarcerated: 2
-Voluntary Withdrawal: 4
-Deceased: 4
-In Residential Treatment: 2

Follow-up Interviews:
3-Month Only: 23
6-Month Only: 20
3- and 6-Month: 72
No Interview: 30

Follow-up Interviews:
3-Month Only: 25
6-Month Only: 13
3- and 6-Month: 69
No Interview: 39
Table 3. Demographics of Randomized Individuals by Enrollment in Study

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Study Enrollment</th>
<th>Overall (N = 420)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes (n = 291)</td>
<td>No (n = 129)</td>
<td></td>
</tr>
<tr>
<td>Age, Mean [Range]</td>
<td>35 [18-63]</td>
<td>34 [18-64]</td>
<td>34 [18-64]</td>
</tr>
<tr>
<td>Traumatic Events, Mean [Range]</td>
<td>7 [0-14]</td>
<td>7 [0-14]</td>
<td>7 [0-14]</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>129 (44)</td>
<td>60 (47)</td>
<td>189 (45)</td>
</tr>
<tr>
<td>Male</td>
<td>162 (56)</td>
<td>69 (53)</td>
<td>231 (55)</td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>231 (82)</td>
<td>110 (88)</td>
<td>341 (84)</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>51 (18)</td>
<td>15 (12)</td>
<td>66 (16)</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>78 (58)</td>
<td>37 (65)</td>
<td>115 (60)</td>
</tr>
<tr>
<td>Native American</td>
<td>27 (20)</td>
<td>9 (16)</td>
<td>36 (19)</td>
</tr>
<tr>
<td>Asian</td>
<td>1 (1)</td>
<td>0 (0)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>African American</td>
<td>3 (2)</td>
<td>0 (0)</td>
<td>3 (2)</td>
</tr>
<tr>
<td>Native Hawaiian/Pacific Islander</td>
<td>7 (5)</td>
<td>1 (2)</td>
<td>8 (4)</td>
</tr>
<tr>
<td>Multiracial or Other</td>
<td>18 (13)</td>
<td>10 (18)</td>
<td>28 (15)</td>
</tr>
<tr>
<td>Diagnosis, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTSD only</td>
<td>26 (9)</td>
<td>14 (11)</td>
<td>40 (10)</td>
</tr>
<tr>
<td>SUD only</td>
<td>77 (26)</td>
<td>30 (23)</td>
<td>107 (25)</td>
</tr>
<tr>
<td>PTSD and SUD</td>
<td>188 (65)</td>
<td>85 (66)</td>
<td>273 (65)</td>
</tr>
</tbody>
</table>

(1) P-value reported from a Wilcoxon-Mann-Whitney test.
(2) P-value reported from a 2-sample t test.
(3) P-value reported from the chi-square test.
(4) P-value reported from the Fisher’s exact test.

Among the 291 participants enrolled into the study, 85% were Hispanic, 50% were male, and 77% were unemployed. The mean age was 35 years (range: 18-64). Nine percent met
criteria for PTSD only, 26% met criteria for SUD only, and 65% met criteria for both. See Table 4 for a summary of demographic characteristics.

| Table 4. Demographics of Participants by Treatment Arm |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Characteristic                  | Treatment Arm   |                  | Overall (N = 291) | p-value         |
|                                 | Clinician-led (n = 145) | Peer-led (n = 146) |                  |                 |
| Age, Mean [Range]              | 35 [18-64]      | 35 [18-60]      | 35 [18-64]      | 0.87(1)         |
| Traumatic Events, Mean [Range] | 7 [0-13]        | 7 [1-14]        | 7 [0-14]        | 0.14(1)         |
| Gender, n (%)                  |                 |                 |                 | 0.59(2)         |
| Female                         | 62 (43)         | 67 (46)         | 129 (44)        |                 |
| Male                           | 83 (57)         | 79 (54)         | 162 (56)        |                 |
| Ethnicity, n (%)               |                 |                 |                 | 0.73(2)         |
| Hispanic                       | 121 (83)        | 124 (85)        | 245 (84)        |                 |
| Non-Hispanic                   | 24 (17)         | 22 (15)         | 46 (16)         |                 |
| Race, n (%)                    |                 |                 |                 | 0.78(3)         |
| Caucasian                      | 90 (62)         | 86 (59)         | 176 (60)        |                 |
| Native American                | 16 (11)         | 17 (12)         | 33 (11)         |                 |
| Asian                          | 0 (0)           | 2 (1)           | 2 (1)           |                 |
| African American               | 1 (1)           | 2 (1)           | 3 (1)           |                 |
| Multiracial or Other           | 38 (26)         | 39 (27)         | 77 (26)         |                 |
| Employment, n (%)              |                 |                 |                 | 0.50(2)         |
| Unemployed                     | 116 (80)        | 112 (77)        | 228 (78)        |                 |
| Employed                       | 29 (20)         | 34 (23)         | 63 (22)         |                 |
| Education, n (%)               |                 |                 |                 | 0.14(2)         |
| 4-year College or Higher       | 6 (4)           | 6 (4)           | 12 (4)          |                 |
| Some College                   | 32 (22)         | 42 (29)         | 74 (25)         |                 |
| High School Graduate/GED       | 59 (41)         | 41 (28)         | 100 (34)        |                 |
| Some High School               | 43 (30)         | 46 (32)         | 89 (31)         |                 |
| Eighth Grade or Less           | 5 (3)           | 11 (8)          | 16 (5)          |                 |
| Living Situation, n (%)        |                 |                 |                 |                 |
The mean and median number of sessions for those in the CL-SS group was 5.7 and 4, respectively. The mean and median number of sessions for those in the PL-SS groups was 6.4 and 5, respectively. No significant differences in attendance were observed. Among those enrolled in the research \( (N = 291) \), 37% of participants in the CL group and 38% of participants in the PL groups completed treatment \( \text{ie, 6 SS sessions} \) by 3 months. The percentage of those who completed treatment by 6 months was 58% and 54% among CL and PL group participants, respectively.

No significant differences were observed.

Specific Aim #1

**PTSD Symptoms:** To review, PTSD symptoms were measured by the PCL-C, with higher total scores indicating higher PTSD symptom severity. Improvement in PTSD symptoms did not change differentially over time between PL-SS and CL-SS groups \( \text{(interaction between intervention group and time point: } F = 0.02, p = 0.98, \text{ ndf = 2, ddf = 357).} \) The CL arm had a slightly larger decrease than the PL arm in PCL score from baseline to 3 months \( \text{(mean} \)
difference of change scores, $\Delta_{PL, 3\ mos - BL} - \Delta_{CL, 3\ mos - BL} = 0.18$ [95% CI: $-3.4, 3.8$]). However, the PL arm had a modestly larger decrease than the CL arm from baseline to 6 months (mean difference of change scores, $\Delta_{PL, 6\ mos - BL} - \Delta_{CL, 6\ mos - BL} = -0.24$ [95% CI: $-4.0, 3.5$]). The upper confidence limit of the 3-month baseline change score fell beyond the hypothesized noninferiority margin (NIM; UCL = 3.8 > NIM = 2.5). The main effect of intervention group was not significant ($F = 0.26, p = 0.61, ndf = 1, ddf = 280$). See Tables A.1-A.3 in the Appendix for modeling summaries and least squares mean and mean change estimates for PCL total score adjusted by covariates.

Subjects in both arms reported overall improved PTSD symptoms. In both arms, between baseline to 3 months the mean change in PCL-C scores was $\Delta_{3\ mos - BL} = -3.4$, 95% CI: (5.6, $-1.2$) and from baseline to 6 months, the mean change was $\Delta_{6\ mos - BL} = -5.0$, 95% CI: ($-7.2, -2.7$). In general, women had higher PCL scores than men (mean difference = 3.8, 95% CI: [0.3, 7.4]). On average, for every 10-year increase in age, the PCL score increased by 2.7 points (95% CI: [1.1, 4.3]). Overall, subjects attending the SS program at the residential treatment center Hoy had lower PCL scores than those who attended the program at Inside Out (mean difference = $-13.4$, 95% CI: $[-19.1, -7.6]$). There were no significant interactions between intervention arm and gender or ethnicity.

**Coping Skills**: Coping skills were measured by the Coping Scale developed specifically for SS skills. The 2 treatment arms did not have statistically different coping scores over time. The PL arm had a modestly larger increase than the CL arm in coping score from baseline to 3 months (mean difference of change scores, $\Delta_{PL, 3\ mos - BL} - \Delta_{CL, 3\ mos - BL} = 1.1$ [95% CI: $-3.5, 4.3$]). However, the CL arm experienced the same increase as the PL arm from baseline to 6 months (mean difference of change scores, $\Delta_{PL, 6\ mos - BL} - \Delta_{CL, 6\ mos - BL} = -0.07$ [95% CI: $-4.8, 4.7$]). Overall, coping scores in both arms increased significantly between baseline to 6 months (mean $\Delta_{6\ mos - BL} = 5.5$, 95% CI: [2.6, 8.5]). The increase from baseline to 3 months was not significant with Tukey adjustment (mean $\Delta_{3\ mos - BL} = 2.4$, 95% CI: $[-0.4, 5.2]$). Lastly, coping scores decreased significantly between 3 and 6 months (mean $\Delta_{6\ mos - 3\ mos} = -3.2$, 95% CI: $[-6.3, -0.1]$). The main intervention arm effect was not statistically significant. ($F = 0.07, p = 0.79, ndf = 1, ddf = 280$).
See Tables B.1-B.3 in the Appendix for modeling summaries and least squares mean and mean change estimates for coping total score adjusted by covariates.

Subjects who completed the SS program by the 3-month follow-up interview had higher coping scores than those who did not complete (mean difference = 5.1, 95% CI: [−0.1, 10.2], \( p = 0.05 \)), but this effect disappeared by the 6-month follow-up. For the gender and ethnicity subgroup analyses, males and females had similar coping scores throughout the study, as did Hispanics and non-Hispanics.

**Substance Use:** Substance use was measured by the ASI Lite. Problems with the administration of the ASI were observed during the data cleaning phase prior to analysis; as a result, Composite ASI Scores are available only for a subset of the sample (a further description of this problem is included in the Discussion section). Because of that issue and because a large proportion of subjects did not report usage or key variables required for scoring, only 197, 125, and 107 ASI Alcohol Use and Drug Use Composite Scores were available for analysis at baseline, 3 months, and 6 months, respectively. Overall, the median ASI alcohol use composite score was 0.0111, which was used to dichotomize the scores into above and below that value. No change occurred in the proportion of subjects who scored above or below the median alcohol composite score in either arm over the course of the study (time \( \times \) intervention arm \( F = 0.81, \ p = 0.45 \), ndf = 2, ddf = 318). The odds ratio (OR) comparing the 2 arms from baseline to 3 months was 1.5 (95% CI: [0.6, 3.6]) in favor of the CL arm and from baseline to 6 months the OR was 1.8 (95% CI: [0.7, 4.4]), also in favor of the CL arm—neither statistically significant. The average effect of alcohol use did not change over time, nor was it different between the arms when assessed together. When looking at the average effect of gender, ethnicity, and program completion, no differences occurred in alcohol use, whereas older subjects and those who attended the SS program at Inside Out were more likely to have higher average alcohol use composite scores. See Appendix C for details.

The median ASI drug use composite score was 0.1077. An analysis similar to that described above for alcohol use was performed. No change occurred in the proportion of subjects who scored above or below the median drug use composite score in either arm over
the course of the study (time × intervention arm $F = 0.67$, $p = 0.51$, ndf = 2, ddf = 191). The OR comparing the 2 arms from baseline to 3 months was 1.3 (95% CI: [0.4, 4.0]) in favor of the CL arm and from baseline to 6 months the OR was 2.0 (95% CI: [0.6, 6.7]), also in favor of the CL arm—neither statistically significant. Subjects in both arms reported 73% lower ASI drug scores by the end of the study (6 months versus baseline OR = 0.27, 95% CI: [0.11, 0.65]). Subjects who were older, female, Hispanic, or attended SS at Inside Out were more likely to have higher average drug use composite scores. See Appendix D for details.

Summaries of the ASI Composite Scores and the changes over time are presented in Tables C.1-3 and D.1-3 in the Appendix. Adjusted odds ratios and their 95% confidence intervals comparing treatment arms were calculated for alcohol and drug use changes between baseline and 3 months and baseline to 6 months and are reported in Tables C.2 and D.2, respectively. To assess the effect of gender, ethnicity, and treatment arm, interactions between these variables were assessed but no significant effects were present.

**Mental Health Functioning:** The changes in mental health function over time, as measured by the SF-36 Mental Component Summary (MCS) subscale, were not significantly different between the 2 arms. From baseline to 3 months, the mean difference between change scores was, $\Delta_{PL, 3 \text{ mos} - BL} - \Delta_{CL, 3 \text{ mos} - BL} = 0.4$ (95% CI: [–3.4, 4.1]). From baseline to 6 months, the mean difference between changes scores was, $\Delta_{PL, 6 \text{ mos} - BL} - \Delta_{CL, 6 \text{ mos} - BL} = 0.5$ (95% CI: [–3.4, 4.4]). No average intervention arm effect occurred ($F = 0.41$, $p = 0.52$, ndf = 1, ddf = 279). In both intervention arms, the SF-36 MCS scores increased from baseline to 3 months (mean $\Delta_{3 \text{ mos} - BL} = 4.7$, 95% CI: [2.5, 7.0]) and from baseline to 6 months (mean $\Delta_{6 \text{ mos} - BL} = 6.3$, 95% CI: [4.0, 8.7]). The change between the 3- and 6-month follow-up interviews was not significantly different.

Subjects who completed the SS program by the 3-month follow-up visit had higher SF-36 MCS scores than those who did not complete (mean difference = 2.7, 95% CI: [–1.7, 6.7], $p = 0.059$). SF-36 MCS scores did not differ significantly between those who did not complete and those who completed by the 6-month follow-up. For every 10-year increase in age, subjects’ SF-36 MCS scores lowered by 1.5 points (95% CI: [–2.7, –0.3]). Women had lower SF-36 MCS scores
than men (mean difference = –5.3, 95% CI: [–7.9, –2.6]). Additionally, subjects who attended the SS program at Hoy had higher SF-36 MCS scores than those who attended at Inside Out (mean difference = 10.9, 95% CI: [6.7, 15.2]). For the gender and ethnicity subgroup analyses, males and females had similar SF-36 MCS scores throughout the study, as did Hispanics and non-Hispanics. See Tables E.1-3 in the Appendix for modeling summary and least squares mean estimates of MCS scores adjusted for covariates. Table E.2 reports the least squares mean estimates and their 95% confidence intervals for the total score at each time point by treatment arm.

Physical Health Functioning: The SF-36 Physical Component Summary subscale measured reported physical health functioning. Comparing the changes in SF-36 PCS scores over time between the 2 arms, from baseline to 3 months, the mean difference between change scores was, $\Delta_{PL, 3 \text{ mos} - BL} - \Delta_{CL, 3 \text{ mos} - BL} = 0.6$ (95% CI: [–1.7, 2.9]). From baseline to 6 months, the mean difference between changes scores was, $\Delta_{PL, 6 \text{ mos} - BL} - \Delta_{CL, 6 \text{ mos} - BL} = 0.8$ (95% CI: [–1.6, 3.2]). No average intervention arm effect occurred ($F = 2.25, p = 0.13, ndf = 1, ddf = 279$).

Overall, a marginal decrease occurred in SF-36 PCS scores in both arms between baseline to 6 months (mean $\Delta_{6 \text{ mos} - BL} = –1.4$, 95% CI: [–2.8, 0.06], $p = 0.06$). The decrease from baseline to 3 months was not significant, nor was the change between the 3- and 6-month follow-up visits.

Subjects who completed the SS program by the 3-month follow-up visit had higher SF-36 PCS scores than those who did not complete (mean difference = 5.4, 95% CI: [2.2, 8.6]). SF-36 PCS scores did not differ significantly between those who did not complete and those who completed by the 6-month follow-up. For every 10-year increase in age, subjects’ SF-36 PCS scores lowered by 4.6 points (95% CI: [–5.5, –3.6]). Non-Hispanics had higher SF-36 PCS scores than Hispanics (mean difference = 4.0, 95% CI: [1.1, 6.8]). Subjects who missed at least 1 follow-up interview had significantly lower SF-36 PCS scores than those who missed none (mean difference = –3.3, 95% CI: [–5.8, –0.9]). For the gender and ethnicity subgroup analyses, males and females had similar SF-36 PCS scores throughout the study, as did Hispanics and non-Hispanics. See Tables F.1-3 in the Appendix for modeling summary and least squares mean estimates of SF-36 PCS scores adjusted for covariates. Table F.2 reports the least squares mean
estimates and their 95% confidence intervals for the total score at each time point by treatment arm.

**Craving:** Both intervention groups experienced a decrease in craving, as measured by a revised version of the Craving Questionnaire-Brief, between baseline to 3 months (mean $\Delta_{3 \text{ mos} - \text{BL}} = 2.2$, 95% CI: $[-3.3, -1.0]$) and baseline to 6 months (mean $\Delta_{6 \text{ mos} - \text{BL}} = -3.1$, 95% CI: $[-4.3, -2.0]$). Although the PL group had larger decreases in craving scores over time, the differences of the change scores between arms were not statistically different from each other. From baseline to 3 months, the mean difference between change scores was, $\Delta_{\text{PL, 3 mos} - \text{BL}} - \Delta_{\text{CL, 3 mos} - \text{BL}} = -1.7$ (95% CI: $[-4.0, 0.5]$). From baseline to 6 months, the mean difference between change scores was, $\Delta_{\text{PL, 6 mos} - \text{BL}} - \Delta_{\text{CL, 6 mos} - \text{BL}} = -1.0$ (95% CI: $[-3.3, 1.3]$). An overall significant treatment effect occurred, with the PL group having lower craving scores than the CL group (mean difference = $-2.0$, 95% CI: $[-3.8, -0.2]$).

When looking at other covariates, in general, for every 10-year increase in age, the craving score decreased by 1.2 points (95% CI: $[-2.0, -0.5]$). Subjects who completed the SS program by the 3-month follow-up visit had lower craving scores than those who did not complete (mean difference = $-2.3$, 95% CI: $[-4.5, -0.2]$). Craving scores did not differ significantly between those who did not complete and those who completed by the 6-month follow-up. Overall, subjects attending the SS program at the residential treatment center Hoy had lower craving scores than those who attended the program at Inside Out (mean difference = $-6.2$, 95% CI: $[-9.0, -3.3]$). For the gender and ethnicity subgroup analyses, males and females had similar craving scores throughout the study, as did Hispanics and non-Hispanics. See Tables G.1-3 for modeling details and summaries. Table G.2 reports the least squares mean estimates and their 95% confidence intervals for the total score at each time point by treatment arm.

**Specific Aim #2**

**Therapeutic Alliance:** The HAq-II (hereafter referred to as HA) was administered at 3 months and 6 months to measure therapeutic alliance. Overall a significant average treatment effect
occurred, with the PL group having higher HA scores than the CL group (mean difference = 4.0, 95% CI: [0.3, 7.7]); however, inspecting the 3-month and 6-month time points individually, the differences between arms was not significant ($\Delta_{PL, 3 \text{ mos} - \text{BL}} - \Delta_{CL, 3 \text{ mos} - \text{BL}} = 4.0$ (95% CI: [−9.4, 1.4]; $\Delta_{PL, 6 \text{ mos} - \text{BL}} - \Delta_{CL, 6 \text{ mos} - \text{BL}} = 4.0$ (95% CI: [−9.7, 1.6]). The primary objective was to assess the noninferiority of HA scores in the PL-SS group compared with the CL-SS group at 3 months. While the point estimate for the overall treatment effect was higher in the PL-SS groups, the lower confidence limit for the differences of the changes scores at 3 months fell beyond the hypothesized noninferiority margin (NIM; LCL = −9.4 < NIM= −5.0).

Subjects who completed the SS program by the 3-month follow-up visit had higher HA scores than those who did not complete (mean difference = 9.2, 95% CI: [4.8, 13.6]). Subjects who completed the SS program by the 6-month follow-up visit had even higher HA scores than those who did not complete (mean difference = 11.5, 95% CI: [5.1, 18.0]). For the gender and ethnicity subgroup analyses, males and females had similar HA scores throughout the study, as did Hispanics and non-Hispanics. Tables H.1-3 in the Appendix present summaries for the independent variables included in the modeling.

To assess the effect of TA on each of the primary aim’s outcome measures, subjects’ mean HA score was added as a covariate to the mixed models described earlier. Models for the outcomes coping and SF-36 PCS and MCS subscales showed a significant effect of HA on these outcomes. Subjects with higher mean HA scores had higher coping total scores, but the rate of increase over the study was lower in the CL group than in the PL group (interaction between intervention arm and mean HA: $F = 7.05$, beta = −0.37, 95% CI: [0.01, 0.67], $p = 0.0008$). Mean HA was positively associated with SF-36 PCS (interaction between intervention arm and mean HA: $F = 3.74$, beta = 0.18, 95% CI: [−0.003, 0.4], $p = 0.05$) such that increased HA scores were associated with increased SF-36 PCS scores, but with a higher rate of improvement in the PL group. The main effect of mean HA was positively associated with SF-36 MCS (centered mean HA: $F = 12.42$, beta = 0.19, 95% CI: [0.08, 0.3], $p = 0.0005$). HA was not associated with PTSD symptoms (PCL-C scores), craving, nor ASI drug and alcohol composite scores. Table H.2 in the
Appendix reports the least squares mean estimates and their 95% confidence intervals for the total score at each time point by treatment arm.

**Dissemination of Findings:**

Data on participant flow (eg, number screened, number enrolled, number of interviews) and preliminary findings were shared monthly with the steering committee (consisting of patients, stakeholders, family members, and researchers) throughout the study period. In addition, study overview, methodology, and preliminary findings were shared at several local and national conferences listed in the final progress report. Preliminary findings were also shared at a data dissemination meeting October 21, 2016, in Santa Fe, New Mexico. More than 40 stakeholders attended the data dissemination meeting, including senators, policymakers, providers, family members, and patients who were involved in the implementation of the study and who were involved in the study as participants from which we collected data. Now that we have completed the research, study findings will be disseminated widely through local meetings, presentations at national conferences, and publications.

---

**V. METHODS: SPECIFIC AIM #3**

The third specific aim was to determine whether the standard Seeking Safety Instructor Training (SS-IT) was adequate for peers. A post-training evaluation design was used with data collected immediately following completion of a 1-day, 6-hour training on SS. The training focused on teaching participants of various training and educational backgrounds how to implement the intervention. Dr. Najavits, author of the SS book, conducted the standard 1-day SS training.

1. **Participants**

Thirty-seven individuals completed the SS training: 16 peer support workers (PSWs) and 21 behavioral health practitioners (BHPs).
2. Data Collection

A survey developed specifically for the training was used to collect feedback from trainees. The brief survey included questions on demographics (ie, gender and ethnicity/race), training content, training delivery, and experience with SS prior to the training. Changes in knowledge and skill attainment (ie, training content) were assessed through 5 questions shown in Table 5.

Table 5. Evaluation Survey: Questions on Training Content

<table>
<thead>
<tr>
<th>How would you rate your...</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ability to counsel clients about the topic(s) covered in this training</td>
<td>Before this training</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>After this training</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2. Ability to manage clients regarding topic(s) covered in this training</td>
<td>Before this training</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>After this training</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3. Ability to implement Seeking Safety</td>
<td>Before this training</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>After this training</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4. Comfort level in providing services to clients in relation to the topic(s) covered in this training</td>
<td>Before this training</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>After this training</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>5. Overall knowledge of the topic(s) covered in this training</td>
<td>Before this training</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>After this training</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

For each question, trainees were asked to provide 2 ratings: a retrospective estimate and a post-training assessment. The analysis examined changes between the 2 estimates. The
questions related to training delivery asked trainees to rate on a 4-point Likert-type scale, ranging from strongly agree to strongly disagree: (1) how comfortable they felt asking questions, (2) the extent to which they understood all of the material that was presented, and (3) how satisfied they were with the material they learned in the training. A final question related to training delivery asked about whether the amount of material covered during the training was too much, too little, or just right.

With respect to prior experiences with SS, participants were asked to indicate whether they (1) ever attended an SS training and, if so, for how many hours; (2) read the SS manual and, if so, to what extent; (3) watched the SS training DVDs and, if so, to what extent; or (4) implemented SS and, if so, approximately how many sessions. The survey concluded with an open-ended question that encouraged trainees to list any additional feedback, including aspects of the training they found to be most useful, challenges that they might face in implementing SS, and whether additional SS-related training was needed.

3. Analysis

Descriptive statistics were calculated to summarize the demographic, background, and responses to the survey items on training content and delivery. Medians and quartiles were calculated for Likert scales and change scores, and frequencies and percentages for categorical variables. We also calculated chi-square or Fisher exact tests to compare participant demographics between PSWs and BHPs, and Wilcoxon rank-sum tests to compare change scores on the items related to training content (eg, skill attainment and knowledge acquisition) and training delivery (eg, satisfaction and comfort level between the 2 groups of trainees).

We adjusted for multiple comparisons between the variables related to training content and delivery via a Bonferroni correction such that 2-sided \( p \)-values were considered significant if \( p \leq 0.01 (\alpha = 0.05/5) \). We performed ordinal logistic regression on 2 of the change scores of particular interest related to training content (ie, ability to counsel and manage clients about the topics covered in the training) to assess whether PSWs felt as comfortable implementing the SS program as BHPs after controlling for previous exposure to SS material (reading SS books,
watching SS videos, previously conducting SS sessions with clients, or previously attending SS training).

**VI. RESULTS: SPECIFIC AIM #3**

Thirty-five participants completed the questionnaire, resulting in a 95% response rate comprising 15 PSWs and 20 BHPs. The BHPs included 13 addiction counselors (65%), 4 social workers (20%), 2 psychologists (10%), and 1 psychiatrist (5%). Both genders were equally represented: 51% male and 49% female. Males comprised 53% of PSWs and females were 47% of PSWs (chi-square test \( p = 0.85 \)). The ethnic/racial composition of the set was diverse among the groups. PSWs were more likely than BHPs to be Hispanic (53% versus 16%) and BHPs were more likely than PSWs to be white (79% versus 20%; Fisher’s exact test \( p = 0.0021 \)).

Significant change scores for PSWs and BHPs were observed for all 5 items that asked about perceived effectiveness of training on the ability to implement SS and knowledge of the subject matter (ie, training content). Table 6 summarizes the median, first-quartile, and third-quartile ratings for 2-time points (retrospective assessment of before and after), and for the change between time points for PSWs and BHPs.

When change scores for PSWs were compared with change scores for BHPs, only 1 significant difference was observed (see the third section of Table 3). PSWs reported a significantly greater improvement in their ability to counsel clients with SS because of the training compared with BHPs (Wilcoxon rank sum test: \( S = 351.00, p < 0.01 \)).
Table 6. Perceived Effectiveness of Training for PSWs and BHPs

<table>
<thead>
<tr>
<th>Survey Item</th>
<th>Pretraining Median (Q1, Q3)</th>
<th>Post-training Median (Q1, Q3)</th>
<th>Change Median (Q1, Q3)</th>
<th>Change Scores p-Value</th>
<th>Comparison of Change Between PSWs and Professionals p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PSWs (n = 15)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to counsel clients about the topic(s) covered in the training</td>
<td>2 (1, 3)</td>
<td>4 (4, 5)</td>
<td>2 (1, 2)</td>
<td>0.0001*</td>
<td></td>
</tr>
<tr>
<td>Ability to manage clients regarding topic(s) covered in the training</td>
<td>3 (1, 3)</td>
<td>4 (3, 5)</td>
<td>1 (1, 2)</td>
<td>0.0002*</td>
<td></td>
</tr>
<tr>
<td>Ability to implement Seeking Safety</td>
<td>2 (1, 3)</td>
<td>4 (3, 5)</td>
<td>2 (1, 3)</td>
<td>0.0001*</td>
<td></td>
</tr>
<tr>
<td>Comfort level in providing services to clients in relation to topics covered in the training</td>
<td>2 (2, 3)</td>
<td>4 (4, 5)</td>
<td>2 (1, 2)</td>
<td>0.0005*</td>
<td></td>
</tr>
<tr>
<td>Overall knowledge of the topic(s) covered in the training</td>
<td>3 (2, 3)</td>
<td>4 (4, 5)</td>
<td>2 (1, 2)</td>
<td>0.0002*</td>
<td></td>
</tr>
<tr>
<td><strong>BHPs (n = 20)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to counsel clients about the topic(s) covered in the training</td>
<td>4 (2, 5)</td>
<td>4 (4, 5)</td>
<td>1 (0, 1)</td>
<td>0.0027*</td>
<td></td>
</tr>
<tr>
<td>Ability to manage clients regarding topic(s) covered in the training</td>
<td>3 (3, 5)</td>
<td>5 (4, 5)</td>
<td>1 (0, 2)</td>
<td>0.0002*</td>
<td></td>
</tr>
<tr>
<td>Ability to implement Seeking Safety</td>
<td>3 (2, 4)</td>
<td>4 (4, 5)</td>
<td>1 (1, 2)</td>
<td>0.0001*</td>
<td></td>
</tr>
<tr>
<td>Comfort level in providing services to clients in relation to topics covered in the training</td>
<td>4 (3, 5)</td>
<td>4 (4, 5)</td>
<td>1 (0, 2)</td>
<td>0.0020*</td>
<td></td>
</tr>
<tr>
<td>Overall knowledge of the topic(s) covered in the training</td>
<td>3 (3, 4)</td>
<td>4 (4, 5)</td>
<td>1 (0, 2)</td>
<td>0.0005*</td>
<td></td>
</tr>
</tbody>
</table>
Overall, PSWs and BHPs equally reported being satisfied with the delivery of the training, feeling comfortable asking questions about the training, and understanding all of the material that was presented during the training. No differences occurred between PSWs and BHPs in how they felt about the amount of material covered during the training. Sixty percent of the PSWs and 65% of the BHPs reported that the amount of material covered was just right.

We were especially interested in assessing PSWs’ and BHPs’ comfort with counseling clients and implementing the SS program. Therefore, 2 ordinal logistic regression models were fitted to the change scores for each of these outcomes. We controlled for background exposure to the SS, including previously reading the SS book, watching the SS DVDs, number of previously conducted SS sessions, attendance at previous SS training, and demographic characteristics. We found that after controlling for these covariates, the improvement in comfort with counseling and implementing the SS program did not differ significantly between PSWs and BHPs.
Dissemination of Findings:

In alignment with PCORI Methodology Standards patients were involved in the dissemination of the data resulting from this specific aim through a co-author role on the following publication: Crisanti AS, Murray-Krezan C, Karlin LS, Sutherland-Bruaw K, Najavits LM. (2016). Evaluation of an evidence-based practice training for peer support workers in behavioral health care. *Cogent Psychol.* 3 10.1080/23311908.2016.121245.

VII. DISCUSSION

For ease of discussion, we first provide a discussion of Specific Aims #1 and #2 followed by a discussion of Specific Aim #3.

Discussion for Specific Aims #1 and #2

*Context for Study Results:* This was the first RCT of the effectiveness of trauma-specific treatment delivered by peer-support workers compared with the way it has been delivered traditionally—that is, by clinicians with master’s degrees or higher. We found no differences in primary and secondary outcomes between PL-SS groups and CL-SS groups after controlling for treatment arm, treatment completion, ethnicity, age, gender, living situation, site, any missed follow-ups, and time. As hypothesized, we found no significant differences in effectiveness between PSWs and clinicians in decreasing PTSD and substance use and in increasing coping skills and mental health functioning over time. An analysis of craving, a measure added to the study by recommendation from the partner stakeholders (including patients and treatment providers), also showed no differences between PL and CL groups—craving decreased significantly in both arms between baseline and both follow-up visits. These findings are consistent with the only other study of the effectiveness of SS delivered by PSWs. In the 2014 pilot study conducted by Najavits and colleagues, 18 women in a residential substance-abuse treatment program participated in SS delivered by PSWs. Participants were assessed at baseline and end of treatment, and improvements were observed in trauma-related symptoms,
psychopathology, functioning, self-compassion, and coping skills. The advantages of this study over the pilot study are the much larger sample size and the inclusion of a comparison group through randomization.\textsuperscript{26}

Regardless of treatment arm, participants experienced a significant decrease in drug addiction severity from baseline to 6 months. Furthermore, all participants experienced significantly lower craving and PTSD symptoms over time. They also experienced significantly higher coping scores from baseline to 6 months and mental health functioning over time. Our findings add to the large evidence base on SS.\textsuperscript{13-20}

A reliable change in PCL scores is 5-10 points, while a clinically significant improvement in PCL scores is a 10- to 20-point decrease from baseline to postintervention.\textsuperscript{37} On average, we saw a 5.0-point decrease on the PCL (95% CI: $[-7.2, -2.7]$) among all subjects between baseline and 6 months, with no difference between PL-SS and CL-SS groups. Compared with other populations, the chronic PTSD symptoms among this population may be more treatment resistant. There are many reasons for this, including the social environment in which many of the participants lived. Intergenerational trauma is a major problem in this community and participants’ recovery was often challenged by ongoing relationships with family members that fostered dysfunction. In addition, based on conversations with the SS groups leaders, many participants were involved in unhealthy marital or domestic relationships that hindered and sometimes thwarted the healing process. A clinically significant change on the SF-36 is greater than or equal to 1 standard error of measurement (SEM) between time points.\textsuperscript{44} In this study, there were no differential intervention effects in change of the MCS and PCS scales. However, the MCS scale increased by an average of 6.3 points (95% CI: $[4.0, 8.7]$) from baseline to 6 months. The change in SEM from baseline to 6 months associated with the mean MCS scores at these time points was 1.3, so this increase was clinically significant. The mean change in PCS scores did not differ between groups and, overall, the modest decrease of 1.4 points from baseline to 6 months was not statistically significant. The change in SEM from baseline to 6 months was 0.8, which implies that the change in PCS scores was also not clinically meaningful. Clinically meaningful parameters are not available for the coping and craving scales.
used in this study. However, our patient and stakeholder partners affirmed that the changes observed in coping skills and craving from baseline to 3-month and 6-month follow-up in this difficult-to-treat population were meaningful.

Overall TA as measured by the HAq-II (referred to in this document as HA) stayed consistent between 3 months to 6 months. Completion of SS has a positive impact on TA. Participants who completed the SS program by their 3-month interview had higher average HA scores than those who did not finish the program, and participants who did not complete the SS program within 3 months but did so by 6 months had a significantly higher average HA score than those who did not complete the program. As predicted, on average, participants in the PL-SS groups had significantly higher HA scores than participants in the CL-SS groups. TA also had a significant impact on 3 outcomes, and this impact varied by treatment arm for 1 of these outcomes. Participants with higher mean HA scores had higher coping total scores, but the rate of increase over the study was lower in the CL group than in the PL group. The main effect of mean HA was positively associated with SF-36 MCS. Lastly, higher HA scores were associated with lower craving scores. HA was not associated with PTSD symptoms (PCL-C scores), SF-36 PCS scores, nor ASI drug and alcohol subscale scores. While TA as measured by the HA was assessed only at months 3 and 6, we included the average of the 2 measures as a fixed covariate in the analyses for the primary outcomes. We did not measure TA at baseline because the participants had not yet established a relationship with a provider; however, we felt TA would vary significantly between participants and would be highly associated with reception to the SS program, ultimately influencing the outcomes of some, if not all, of our primary outcomes.

Research has shown a strong relationship between culturally competent providers and satisfaction with treatment and outcomes. All of the facilitators in this study were from the target population. They understood the culture, the language, and the challenges of the community (eg, high rates of poverty, addiction, and intergenerational drug use and trauma). The cultural competency of all group facilitators likely contributed to high positive findings,
especially therapeutic alliance, observed overall among this predominantly Hispanic population as well as to the lack of differences in outcomes between peer-led and clinician-led groups.

Generalizability of the findings

These findings are generalizable to similar populations—including Hispanic, substance use being primarily heroin, and underserved rural communities—with the understanding that several strategies (discussed earlier) were used for engagement and retention, and retention in services is related to outcomes. While many of the monetary-based strategies may be a challenge to implement in outpatient settings with limited and decreasing funding, strategies that require no direct funding, such as reminder and follow-up phone calls, may be worth consideration. Quantitative data on the effectiveness of each strategy implemented in this study are not available; however, in-depth interviews were conducted with 14 participants at the end of the study and many of them reported that the weekly phone calls were helpful. Finally, fidelity and mastery in the implementation of an evidence-based practice are related to outcome. In this study, fidelity was assessed quarterly and all group facilitators received intensive and ongoing supervision in the implementation of SS. Therefore, the generalizability of our findings may be limited to agencies that not only implement SS but monitor for adherence to the model and provide clinical supervision for implementation. Generalizability may also be limited by the small number of facilitators who led SS groups. Finally, it is important to note that the target population was recruited from a peer-run wellness center and a residential treatment program, and findings may not extend to populations that receive services from outpatient community-based behavioral health settings.

Implementation of study findings: Our finding that PL-SS facilitators achieved the same positive outcomes as CL-SS facilitators has major implications for the widespread dissemination and implementation of this evidence-based practice (EBP), especially in rural communities that lack access to mental health professionals. Even in urban areas, clients with trauma/PTSD and SUD
are some of the most disadvantaged and unable to access professional help. PL-SS offers a safe, low-cost or free option. Notably, no routinely disseminated or implemented PL trauma/PTSD model currently exists, which represents a major public health gap. There is a long and successful history of peer-led addiction models, such as AA and SMART Recovery, but those do not address trauma or PTSD explicitly.

The treatment of trauma and SUD requires complex decision making that relates to the patient’s level of severity, the capacity of providers to attend to the patient’s needs, and the level of care the setting provides. Treatment often includes a mix of both behavioral and pharmacologic options and is typically constrained by costs. Our findings help providers and program administrators (ie, primary decision makers) to make a more informed choice that allows them to provide a stronger level of care (SS) at lower cost (peer delivery) than traditional treatment options for this comorbidity. Notably, we showed strong results for PL-SS, even for severe patients in a rural setting with limited additional supports.

The biggest challenge to this intervention that we encountered during the study was retention. While treatment engagement and completion is a challenge among those with substance use disorders in general, it is even more so among minorities, especially Hispanics, and among those who have less education and are addicted to heroin. These characteristics describe most of our study population. In a study of persons referred to publicly funded substance abuse treatment, primary heroin users were significantly less likely than others to enter treatment programs to which they were referred, to complete treatment, and to abstain from use. In our study, participants were encouraged to complete all 12 SS sessions. Several strategies were used to encourage treatment completion, including providing the following: light refreshments at all groups, a $10 dollar gift card at the 6-session milestone and a second gift card at treatment completion (ie, 12-month session), transportation to and from treatment, weekly reminder and follow-up phone calls, and child care services. In addition, motivational incentives using the fishbowl method based on an affordable contingency management approach (ie, a low-cost prize incentive system) were used during every session in alignment with the Promoting Awareness of Motivational Incentives guidelines developed by NIDA and
While the average number of SS sessions observed in this study is consistent with other studies that offered 12 sessions of SS to similar populations, it took a lot of effort and financial resources—both of which are unlikely in community-based mental health agencies.

Research on the reasons that people leave outpatient treatment for SUDs has shown that those who drop out of treatment are less satisfied with it. However, being lost to follow-up and dropping out of treatment may have nothing to do with less satisfaction but more to do with practical reasons (e.g., treatment may be difficult to get to, difficult to schedule) and current life events that may take precedence over treatment. When participants were asked through feedback surveys about what they would change about the treatment program, some of the qualitative responses indicated that the session times conflicted with other schedules. For example, 1 participant recommended changing the session times because they interfered with his/her schedule. This information is valuable for program directors who are planning SS groups with the goal of maximizing retention and engagement.

Research on barriers to D&I has repeatedly identified the lack of skilled personnel as one of the most important to address to improve EBP implementation. In our study, all facilitators received initial and ongoing training in the implementation of SS, and in an evaluation of this training, peers reported it to be more beneficial. Knowing that this high level of training may not be available in practice, a supplement to the SS manual that includes guidance specific for peers would increase the likelihood of successful PL-SS implementation and the replication of positive outcomes observed in this study.

Subpopulation considerations: For all outcome measure analyses, the inclusion of interaction terms between treatment arm and gender and ethnicity yielded no significant differences in these subgroups. These data suggest that PL-SS is a viable treatment option for males and females as well as Hispanics and non-Hispanics, but further research is warranted.

Study strengths and limitations: This study had several strengths. First was the team’s ability to successfully carry out an RCT, which is particularly challenging in a rural community. Second,
all of the facilitators received the same training in SS and clinical supervision during the study. Third, high patient and stakeholder involvement contributed to the study’s success, particularly for recruitment, retention, and data collection. The value of including patients and stakeholders in the research process has been previously identified. Other strengths of this study included a large sample size, an ethnically diverse sample, 2 follow-up time points, examination of multiple outcomes, and regular and ongoing confirmation of fidelity to the implementation of the evidence-based practice.

This study also had several limitations. The validity of the findings may be threatened by selection bias because the sample of patients is not a probability sample. In all, 420 individuals were determined eligible and randomized but only 69% participated in the research. While an examination of age, gender ethnicity/race, eligibility criteria, and history of trauma showed no significant differences between those who participated in the research (n = 291) and those who did not (n = 129), other unexamined variables may exist that may have biased the results.

Similarly, 3- and 6-month follow-up data were not available for 69 participants who completed the baseline. While it may be that those who were lost to follow-up were worse off than those who participated in follow-up interviews (eg, they may have been hospitalized, been incarcerated, or died), it is also possible that those lost to follow-up may have been doing well. In a study of attrition, Crisanti and her colleagues (2014) found that participants who were employed were less likely to complete a follow-up interview than those not employed.

Another problem is that the baseline data are not a true baseline. We originally proposed to collect data before the first group; however, during our pilot study we found that many participants who received $20 for completing the baseline interview before the first group never came back for treatment. Based on our patient and stakeholders’ familiarity with the target population, we changed the data collection methodology and waited until participants completed at least 1 group session to collect baseline data. Although this procedure biased the baseline data, we decided it was more important to have participants exposed to the treatment.
Another limitation of our data is that drug use and alcohol use composite scores from
the ASI were available for only a subset of subjects due to problems related to incorrect
administration of the survey by 1 research assistant. Most of the questions on the ASI ask about
substance use in the past 30 days. However, 4 questions intermixed within the questions that
ask about number of days of use in the past 30 days required a Likert-type scale response,
which ranged from 0 to 4. A random check of the data entry for accuracy identified several
instances in which respondents who completed an interview with a specific research assistant
were answering the Likert-type questions incorrectly. That is, they were responding in terms of
number of days instead of using the Likert-type scale. For example, responses for the question
“How troubled or bothered have you been in the past 30 days by alcohol problem(s)” had
responses greater than 4. Responses such as 20, 23, and 29 made it obvious that respondents
were responding in terms of number of days. Once this problem was identified, every ASI
completed during the study period was reviewed and incorrect responses were identified. It
was confirmed that all mistakes were confined to interviews administered by 1 research
assistant. All of the responses on the ASI are summed to create a total score of substance use.
The unfortunate result of 4 items being answered incorrectly is that a total ASI score could not
be generated for those records that had items answered incorrectly.

Another possible threat to the validity of the findings was the lack of blinding, a critical
methodological feature of RCTs. While randomization minimized the differences between
groups at the beginning of the trial, it did nothing to minimize the potential for biased
assessments that may have resulted from participants knowing the status of their group leader.
Given the small community of the target population and the familiarity with the group
facilitators, blinding to the intervention was not possible in this study. As pointed out by Kahan
et al. (2014), blinding is not always feasible in research and this was the case in this rural
community. As recommended by Karanicolas et al. (2009), many strategies were
incorporated to minimize the potential for bias, including the identical delivery of the
intervention, the identical treatment of all participants, and the objective collection of all data.
While we selected data collection tools that had high psychometric properties and that were
appropriate for the target population, there is lack of information on clinical significance metrics for the coping scale—which is one of our primary outcomes.

Another limitation is that we did not have information on whether the clinicians had previous personal experiences with PTSD and/or SUD. Previous studies have found the percent of counselors in recovery to range from 37% to 57%. The lack of differences observed between treatment arms could have been influenced by clinicians having lived experience with PTSD and/or addiction and being therefore similar, to a certain extent on a personal level, to the peers who facilitated the peer-led groups.

A technical limitation that affected the analyses was the inclusion of the following variables in the linear modeling: mean TA score (averaged over 3 and 6 months for each subject), any missed interviews, and SS program completion. We included each of these variables as fixed covariates (no change over time) in the models despite that none of the variables could have been observed at baseline and possibly not by the 3-month time point. However, the steering committee recommended that our results be adjusted for number of missed interviews and program completion because those participants who engage in treatment and research may have some characteristics and lifestyles different from those who do not (referred to as consent, participation, or volunteer bias).

A final limitation of this study is the reliance on self-reported data, which may be subject to recall bias. The accuracy of the self-reported data may also be questionable because respondents were asked to report about events that might be highly sensitive, such as substance use.

**Future research:** While there was no difference between PL-SS and CL-SS, participants in both groups experienced decreased physical health over time. While this was not originally hypothesized, this finding was not surprising to our patient partners. Based on their lived experiences, they explained that once they stopped using, they too started to feel the physical pain that was masked by alcohol and drugs. Another explanation was that substance use takes a toll on physical health. As people gain sobriety, they are more likely to be engaged in
health care and develop more awareness of physical health problems that may have developed as a consequence of their SUD. Future research on the impact of behavioral health treatment on physical health is warranted.

Overall, our results add further support to the increasing focus on PSWs in the delivery of behavioral health care, and in particular to the feasibility of SS as a model that peer facilitators have responded to positively both in this study and in a prior pilot outcome study on peer-led SS. Given the sheer numbers of traumatized individuals, as well as SUD issues that are prominent in this population, a model such as SS delivered by PSWs can add to the goal of improving care for this vulnerable segment of the population. Further research on the delivery of SS and other EBPs by peers is, however, warranted.

Discussion for Specific Aim #3

This study also sought to evaluate whether the standard Seeking Safety Instructor Training was adequate for peers. We were interested in learning how PSWs compared with “traditionally trained” behavioral health practitioners (including social workers, psychologists, and psychiatrists) responded to the standard SS training. Would they report comparable satisfaction with the training? Comparable learning? Comparable comfort with delivering SS? Overall, the training was perceived as beneficial and no differences were found between PSWs and BHPs. These findings are positive news. They can be interpreted as indicating that PSWs and BHPs can attend the same SS training, which can make for an efficient and less costly method than having to provide separate training for each. Moreover, being trained together can help build positive connections between PSWs and BHPs that can infuse their work—encouraging mutual respect and common understanding. Only 1 item evidenced a difference
between PSWs and BHPs: Compared with BHPs, PSWs reported a significantly greater improvement in their ability to counsel clients with SS because of the training. This is a highly positive finding in that PSWs by definition started out with less training and education in behavioral health than BHPs and reported a greater increase in their ability to use a behavioral health model such as SS. For many PSWs, this was likely their first exposure to training in how to deliver an EBP, and it was encouraging to see that their comfort level with doing so increased so rapidly. Perhaps reflecting their excitement about the opportunity for professional development training, every PSW responded to the open-ended question on the survey, in contrast to the BHPs. PSWs indicated an appreciation for the background information presented on the research behind the model. Role playing, grounding exercises, videos, and practice sessions were also noted by PSWs as being some of the most useful aspects of the training. With the growing recognition of PSWs as assets in the delivery of behavioral health care, better understanding of their training needs can help inform how to further advance their successful integration into practice.

Study Strengths and Limitations

This initial, exploratory study was the first study to evaluate whether the standard SS-IT was adequate for peers. The sample was relatively small; thus, the absence of differences could reflect simply a lack of statistical power. The small sample also limits generalizability of the findings. The retrospective estimate of pretest knowledge at the end of the 1-day training was also a limitation, and our results would have been strengthened had trainees completed the questions on training content prior to the onset of the training, to reduce the influence of recall bias. While pilot-tested before implementation, the post-training survey was developed specifically for this training with no established psychometric properties. We also did not conduct a knowledge test of SS principles and thus relied on their perceptions of learning the model. Future research would benefit from an objective knowledge test of SS principles.
Overall, we found no differences between the peer-led and clinician-led treatment arms for our primary and secondary outcome measures. However, we cannot conclude noninferiority of PL-SS to CL-SS based on the hypotheses we specified during the study design. We received a higher variability in responses than anticipated, so while our mean changes in outcome measures over time did not differ between arms, future work to better understand this variability in the arms is needed to make the strong statement that PL-SS is not inferior to CL-SS. An exploratory examination of craving found that participants in PL-SS reported significantly lower mean craving scores. With respect to our second aim, we found an overall significant treatment effect, with the PL-SS group having higher TA scores than the CL-SS group. Regardless of treatment arm, participants experienced a significant decrease in drug addiction severity from baseline to 6 months. Furthermore, all participants experienced significantly lower craving and PTSD symptoms over time. They also experienced significantly higher coping scores from baseline to 6 months and mental health over time.

Given their ability to produce positive outcomes and high therapeutic alliance, PSWs can increase access to patient-centered trauma-specific treatment in underserved rural communities and improve the lives of those suffering from PTSD and/or an addiction. Ongoing clinical supervision and regular fidelity assessments were a major component of this study and highly recommended for peer-support workers implementing evidence-based practices, especially for those who are newly employed with limited training and experience. Implementation research has shown the value of ongoing training in the successful delivery of EBPs, especially as it relates to fidelity, and this may be particularly true for PSWs who may be implementing an EBP for the first time.79
References


