PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	A Randomized Clinical Trial of an Emergency Department-Based Peer Recovery Support Intervention to Increase Treatment Uptake and Reduce Recurrent Overdose among Individuals at High Risk for Opioid Overdose: Study Protocol for the Navigator Trial
AUTHORS	Goedel, William C; Marshall, Brandon D.L; Samuels, Elizabeth A; Brinkman, Mark G; Dettor, Debra; Langdon, Kirsten J; Mahoney, Linda A; Merchant, Roland C; Nizami, Tarek; O'Toole, George A; Ramsey, Susan E; Yedinak, Jesse L; Beaudoin, Francesca L

VERSION 1 – REVIEW

REVIEWER	Helena Chmura Kraemer Department of Psychiatary and Behavioral Sciences Stanford University
REVIEW RETURNED	24-Jul-2019

GENERAL COMMENTS	What is proposed here is a randomized clinical trial (RCT) to compare two interventions for dealing with an ED presentation for opioid overdose at two sites. Currently both interventions are used apparently haphazardly at both sites (page 7, line 38ff), and this RCT seeks to show that one treatment (by a certified peer recovery support specialist (RSS)) is overall better than the other (by a licensed clinical social worker, LCSW). #1. Since the training for the RSS is provided by the Anchor Recovery Community Center, who are involved in the design, execution and analysis of this RCT, conflict of interest issues must be considered. I would suspect that LCSWs would consider this study as loading the dice against their services (I am not a LCSW), and they are probably right. Generally the more contact an intervention has with a patient, the better the outcome. The LCSW has only one contact in the ED while the RSS has that one contact plus a “call every day for ten days…. and regular contact and follow-up”. If it is here found that the RSS gets better outcomes, it might be true that the LCSW required/allowed to have more intense contact with the patient over time, would have had even better outcomes, or that a non-trained but empathetic person required to do such follow up would get as good results as the trained RSS. #2. One of the advantages of this design is that the outcome measures will be taken from administrative data. This should minimize missing data due to dropout or loss to follow-up, and ensure “blinding”. However, it is also said that other information will be obtained via telephone by a research assistant at 30 days, 3 months, and 6 months post-discharge. Since this adds follow-up to both

	intervention groups, this may affect response and bias the study. Now we're not considering the usual LCSW, but that augmented by 3 post-discharge contacts. In studies of this kind, there is usually a great deal of dropout and missing data, and these are usually closely related to choice of treatment and to response to treatment, thus not "missing at random". It is not clear that having these follow-up data will help understand what mediates treatment outcome (presumably the purpose), because of the biases so introduced. #3. The two primary outcomes are presumably highly correlated. If it is found that one treatment improved engagement with a formal SUD treatment program, but the other reduced opioid overdose over 18 months, what could possibly be concluded, except possibly that the formal SUD treatment programs increased the risk of overdose? It would be far better to choose one outcome measure that reflected both outcomes. For example, the outcomes might be ranked from 1 (best): engagement + no overdose to 4 (worst): no engagement + overdose. Then you could clarify where engagement+ overdose and no engagement+no overdose ranked in between best and worst. This would reduce to one test, which would determine whether patient outcome is preferable in one versus another treatment. Not my field, but when I tried to do this rank ordering, I ended up dropping the engagement variable altogether, considering no overdose the best outcome (with or without engagement), overdose the worst (with or without engagement). Also note that you refer to "recurrent ED visits", but your analysis seems to be for "at least one ED visit" for opioid overdose in 18 months. That should be clarified. You could consider the number of ED visits as an ordinal outcome rather than the yes/no answer you did. That would yield more power and precision. You might instead consider time to first ED visit, and use survival methods to compare groups, with even greater power and precision. #4. You propose to sample 650 participants (page 10, line 22) at 2 sites (equal numbers per site?). Then you propose to stratify on age and gender (2 genders X how many age groups?). You propose block randomization with block sizes randomized between 4 and 8. OK. But do you propose to block randomize within site? Or within site X age X gender strata? If the last, will all those strata be large enough to sustain that block randomization? Since this is a two-site study, randomization separately at the two sites is preferable, but why stratify by age and gender? If you propose to employ procedures to match the two groups per site by age and gender, that should be specified, for then the analyses you propose are likely incorrect. I would strongly recommend randomization within the two sites, ignoring age and gender. With this sample size, there are unlikely to be major differences between the groups with randomization. #5. "success of randomization"??? I think what you are doing is checking how well matched the two groups are, and randomization does not produce two matched group, but two random samples from the same population. Two randomized groups should significantly differ ($p<.05$) on 5% of independent baseline variables, and perhaps more than 5% if those baseline variables are correlated. Either you randomized or you didn't, and surely you known which. No use testing a null hypothesis you know to be

	true! In any case, check the current CONSORT guidelines on testing baseline variables—not to be done. #6. Page 11, line 22ff. If you examine the data and use what you see to change the hypothesis, that is 'post hoc' hypothesis testing, and the usual tests and p-values are no longer valid. Moreover, if you chose to use those variables that differentiate the treatment groups at baseline as covariates in the analysis, you are deliberately introducing collinearities with treatment choice, which biases the treatment comparisons. This should not be done. Including any covariate changes the hypothesis to be tested. If, after you've completed testing the 'a priori' hypothesis of this study, you choose to explore for possible moderators of treatment outcome, that would be welcome, but that should only be done after you test your 'a priori' hypothesis, and you would check every baseline variable, not merely those that did not match between the groups. Because this is a 2 site study, you are obliged to consider site, site X treatment interaction as well as treatment in your analysis. With a binary outcome (as the two here), you might use Logistic Regression, coding the two sites +1/2 and the two treatments +1/2. Then the interaction effect tests whether the same treatment effect generalized over the two sites, and the treatment effect tests whether the mean effect over the two sites (the common effect if there is no interaction) is null. No covariates! I would strongly recommend including effect sizes and their confidence intervals. With a binary outcome, SRD=p1-p2, where p1 and p2 are the probabilities of the "good" outcome in the two groups, would be preferable. Since there are policy issues here involved, Number Needed to Treat NNT=1/SRD would also be useful. In your power considerations (page 10, line19ff), you set your critical effect size at .07 (.14-.07), which means that you think that the critical NNT=14. That means that if one had to treat 14 patients with one treatment to get better results for only 1 of those 14 (i.e, treating 13 unnecessarily), that would be acceptable (but no more than 14). Considering the differential costs to the two programs here, is that reasonable? If the differential effect were less than .07, even if statistically significant ($p<.05$), you would not recommend the better treatment over the other, instead considering them clinically equivalent? I don't know the answers to these questions, but they should be considered.

REVIEWER	Lauren Whiteside University of Washington, USA
REVIEW RETURNED	12-Aug-2019

GENERAL COMMENTS	Overall, the authors present a protocol for a randomized clinical trial or ED-based peer recovery support compared to a social work intervention (usual care. The manuscript provides details on a RCT that aims to address an important and timely issue and comparing peer recovery support compared to standard SW is innovative and timely given the growing use of peer recovery coaches. Opioid overdose continues to rise and thus this topic is extremely important. The Introduction section does a nice job of describing the scope of the problem. Since the main intervention being tested is a peer recovery support specialist, this section should provide more

	background on the benefit of this approach for OUD (or SUD) in other settings such as primary care or outpatient behavioral health. The authors describe previous ED-based interventions to reduce overdose but should include 'Bohnert AS, et al A pilot randomized clinical trial of an intervention to reduce overdose risk behaviors among emergency department patients at risk for prescription opioid overdose' published in Drug and Alcohol Dependence (2016) as well to complete this literature review. In the last paragraph of the Introduction please include the condition or specific population being tested (e.g. OUD, patients at risk for overdose). In the Methods section please include the dates of the study including the date of collection of all primary endpoints. In the paragraph that states the inclusion criteria, please note how '3) self-report of an opioid overdose in the previous 12-months' is captured. Is this systematic from RA universal approach or reported to the clinical team during the encounter and subsequently referred to the program? Please note if all recruitment occurs in the ED or if patients who become eligible for screening once they are not critically ill can be recruited/screened in other areas of the hospital (e.g. inpatient). This may be important depending on the rate of admission for overdose at these institutions. How does the study team manage a patient who is randomized to the peer support arm but requests a social work visit? How is fidelity to the intervention model being documented? Figure 1 suggests interventions will be done using principles of motivational interviewing. Please note if there is MI fidelity testing being done. The peer recovery specialist intervention has three specific components and is provided by a lay person. How is fidelity to these three elements being done across the various peer recovery coaches? Consider including the number of peer recovery coaches involved in this study. Please note if SW and peer recovery coaches are available 7 days a week and 24 hours a day. The Outcomes section states two primary outcomes. With regards to engagement in treatment outcome, it is a strength of the trial to use administrative data. Please note if BHOLD includes methadone maintenance administered at an OTP. Also, how will investigators handle patients who initiate buprenorphine from the ED, but do not engage with an outpatient provider in 30 days? Or likewise, have an interruption of buprenorphine coverage during the 30 days? While patients offered naltrexone will likely be small, these patients will not be captured in this data set. Consider noting this in the limitations/discussion section. The second primary outcome (recurrent visits) is also assessed using administrative data. Consider adding something to the limitations/Discussion section that ED visits outside of RI will not be captured and thus may be an underestimate. The clinicaltrials.gov link lists several secondary outcomes. Consider including these as well.

VERSION 1 – AUTHOR RESPONSE

Reviewer #1

Comment #1: Since the training for the RSS is provided by the Anchor Recovery Community Center, who are involved in the design, execution, and analysis of this RCT, conflict of interest issues must be considered. I would suspect that LCSWs would consider this study as loading the dice against their services (I am not an LCSW), and they are probably right. Generally, the more contact an intervention has with a patient, the better the outcome. The LCSW has only one contact in the ED while the RSS has the one contact plus a "call every day for ten days…and regular contact and follow-up." If it is here found that the RSS gets better outcomes, it might be true that the LCSW required/allowed to have more intense contact with the patient over time, would have had even better outcomes, or that a non-trained but empathetic person required to do such follow-up would get as good results as the trained RSS.

Response #1: We thank the reviewer for this comment. We respectively note that peer recover support services are increasingly being employed in a range of clinical settings to assist individuals with substance use disorder (White & Evans, 2014). Because of their own experiences with substance use disorder and recovery, peer recovery support specialists are experientially qualified to support their peers who are currently experiencing substance use disorder and related problems through mentoring, education, and support (Valentine, 2010). A recent systematic review concluded that there remains a more rigorous investigation to establish the efficacy, effectiveness, and cost-effectiveness of peer recovery support services (Eddie et al., 2019). Existing randomized controlled trials (e.g., Bernstein et al., 2005; Rowe et al., 2007) have been subject to several limitations, particularly poorly defined and non-manualized roles for peers in the interventions (Eddie et al., 2019). As such, we believe this trial fills an important gap in the literature on the utility of peer recovery support specialists by employing individuals who have been certified to provide these services after completing a standardized curriculum (Valentine, 2010). This strength has been added to the "Background and Rationale" section of the revised manuscript (Page 5).

Bernstein J, Bernstein E, Tassiopoulos K, Heeren T, Levenseon S, Hingson R. Brief motivational intervention at a clinic visit reduces cocaine and heroin use. Drug Alcohol Depend. 2005; 77(1): 49–59.

Eddie D, Hoffman L, Vilsaint C, Abry A, Bergman B, Hoepper B, Weinstein C, Kelly JF. Lived experiences in new models of care for substance use disorder: A systematic review of peer recovery support services and recovery coaching. Front Psychol. 2019; 10: 1052.

Rowe M, Bellamy C, Baranoski M, Wieland M, O'Connell MJ, Benedict P, Davidson L, Buchanan J, Sells D. A peer-support, group intervention to reduce substance use and criminality among persons with severe mental illness. Psychiatr Serv. 2007; 58(7): 955–961.

Valentine P. Peer-based recovery support services within a recovery community organization: The CCAR experience. In: Kelly JF, White WL, eds. Addiction Recovery Management. 1st ed. New York, New York: Springer, 2010. 259–279.

White WL, Evans AC. The recovery agenda: The shared role of peers and professionals. Public Health Rev. 2013; 35(2): 1–14.

Comment #2: One of the advantages of this design is that the outcome measures will be taken from administrative data. This should minimize missing data due to dropout or loss to follow-up and ensure blinding. However, it is also said that other information will be obtained via telephone by a research assistant at 30 days, 3 months, and 6 months post-discharge. Since this adds follow-up to both intervention groups, this may affect response and bias the study. Now we're not considering the usual LCSW, but that augmented by 3 post-discharge contacts. In studies of this kind, there is usually a great deal of dropout and missing data, and these are usually closely related to choice of treatment and to respond to treatment, thus not missing at random. It is not clear that having these follow-up data will help understand what mediates treatment outcome (presumably the purpose) because of the biases so introduced.

Response #2: We thank the reviewer for this comment on the data collection methods. The reviewer is correct in assuming that these follow-up assessments are being understand what mediates the effect of treatment. Although the completion of these follow-up assessments may be dependent on a participant's intervention allocation, we believe that the completion of these assessments is independent of the treatment outcomes. For example, individuals randomized to receive a behavioral intervention from a certified peer recovery support specialist may be more likely to complete these follow-up assessments because they receive more contact from the certified peer recovery support specialist following their initial emergency department visit. However, we believe that their completion of a follow-up assessment should not impact the primary outcomes of the trial, particularly because the first follow-up assessment does not occur until after the period of time used to define the primary outcome of engagement with substance use disorder treatment.

Comment #3: The two primary outcomes are presumably highly correlated. If it is found that one treatment improved engagement with a formal SUD treatment program, but the other reduced opioid overdose over 18 months, what could possibly be concluded, except possibly that the formal SUD treatment programs increased the risk of overdose? It would be far better to choose

one outcome measure that reflected both outcomes. For example, the outcomes might be ranked from 1 (engagement + no overdose) to 4 (no engagement + overdose). Then you could clarify where engagement + overdose and no engagement + no overdose ranked in between best and worst. This would reduce to one test which would determine whether patient outcome is preferable in one versus another treatment. Not my field, but when I tried to do this rank ordering, I ended up dropping the engagement altogether, considering no overdose the best outcome (with or without engagement), overdose the worst (with or without engagement). Also note that you refer to "recurrent ED visits" but your analysis seems to be for "at least one ED visit" for opioid overdose in 18 months. That should be clarified. You could consider the number of ED visits as an ordinal outcome rather than the yes/no answer you did. That would yield more power and precision. You might instead consider instead time to first ED visit, and use survival methods to compare groups, with even greater power and precision.

Response #3: We agree with the reviewer that the outcomes may be correlated, particularly because entering effective treatment for a substance use disorder may decrease engagement in risk behaviors that lead to drug-related harms while increasing access to primary care services, thus reducing emergency department utilization (Rockett, Putnam, Jia, Chang, & Smith, 2005). However, we believe that both of these outcomes should be considered independently of one another, as increases in engagement in substance use disorder treatment would represents the presence of a positive outcome for individuals who use drugs while decreases in recurrent emergency department visits would represent the absence of a negative outcome for this population. We believe that an exclusive focus on emergency department visits would perpetuate the stereotypes of individuals with substance use disorders as a cost burden on the health system and would exclude any potential understanding of the resilience and empowerment inherent in achieving recovery.

Rockett JRH, Putman SL, Jia H, Chang CF, Smith CS. Unmet substance abuse treatment need, health services utilization, and cost: A population-based emergency department study. Ann Emerg Med. 2005; 45(2): 118-127.

Comment #4: You propose to sample 650 participants (Page 10, Line 22) at 2 sites (equal numbers per site?). Then you propose to stratify on age and gender (2 genders X how many age groups?). You propose block randomization with block sizes randomized between 4 and 8. Ok. But do you propose to block randomize within site? Or within site X age X gender strata? If the last, will all those strata be large enough to sustain that block randomization? Since this is a two-site study, randomization separately at the two sites is preferable, but why stratify by age and gender? If you propose to employ procedures to match the two groups per age and gender,

that should be specified, for then the analyses you propose are likely incorrect. I would strongly recommend randomization within the two sites, ignoring age and gender. With this sample size, there are unlikely to be major differences between the groups with randomization.

Response #4: We thank the reviewer for this comment. In the revised manuscript, we have clarified that randomization is stratified by site. Each site is maintaining its own randomization scheme and individuals are no longer stratified on age and gender (Page 9). Allocations are assigned at random using the REDCap randomization feature.

Comment #5: "Success of randomization"? I think what you are doing is checking how well matched the two groups are, and randomization does not produce two matched groups, but two random samples from the same population. Two randomized groups should significant differ (*p* < .05) on 5% of independent baseline variables, and perhaps more than 5% if those baseline variables are correlated. Either you randomized or you didn't, and surely you know which. No use testing a null hypothesis you know to be true. In any case, check the current CONSORT guidelines on testing baseline variables – not to be done.

Response #5: We thank the reviewer for this comment on our statistical methods. In line with the CONSORT guidelines (Schulz, Altman, & Moher, 2010), the analytical plan has been revised to remove significance testing of variables measured at the baseline assessment following treatment randomization (Page 10–11).

Schultz KF, Altman DG, Moher D. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomized trials. BMJ. 2010; 340: c332.

Comment #6: Page 11, Line 22. If you examine the data and use what you see to change the hypothesis, that is 'post hoc' hypothesis testing, and the usual tests and *p* values are no longer valid. Moreover, if you chose to use these variables that differentiate the treatment groups at baseline as covariates in the analysis, you are deliberately introducing collinearities with treatment choice, which biases the treatment comparisons. This should not be done. Including any covariate changes the hypothesis to be tested. If, after you've completed testing the 'a priori' hypothesis of this study, you choose to explore possible moderators of treatment outcome, that would be welcome, but that should only be done after you test your 'a priori' hypothesis, and you would check every baseline variable, not merely those that did not match between the

groups. Based this is a two-site study, you are obligated to consider site, site X treatment interaction as well as treatment in your analysis. With a binary outcome (as the two here), you might use logistic regression, coding the two states +1/2 and the two treatments +1/2, then the interaction effect tests whether the same treatment effect generalized over the two sites, and the treatment effect tests whether the mean effect over the two sites (the common effect if there is no interaction) is null. No covariates! I would strongly recommend include effect sizes and their confidence intervals. With a binary outcome, SRD = p1-p2, where p1 and p2 are the probabilities of the "good" outcome in the two groups, would be preferable. Since there are policy issues here involved, number needed to treat, NNT = 1/SRD, would also be useful. In your power considerations (Page 10, Line 19), you set your critical effect size at 0.07 (0.14-0.7), which means that you think the critical NNT = 14. That means that if one had to treat 14 patients with one treatment to get better results for only 1 of those 14 (i.e., treating 13 unnecessarily), that would be acceptable (but no more than 14). Considering the differential cost to the two programs here, is that reasonable? If the differential effect were less than 0.07, even if statistically significant (*p* < .05), you would not recommend the better treatment over the other, instead considering them clinically equivalent? I don't know the answers to these questions, but they should be considered.

Response #6: We thank the reviewer for these recommendations on our statistical methods. In line with this suggestions, we have updated the analytical plan to describe an assessment of the treatment effect that uses logistic regression for each of the binary outcomes and includes terms for an indicator of treatment allocation (licensed clinical social worker versus certified peer recovery support special), an indicator for study site (Site 1 versus Site 2), and a term representing the interaction of treatment allocation and study site (Pages 10). Further, we have specified two subgroup analyses *a priori* that intended to investigate heterogeneity of treatment effects by age and gender (Pages 10). In addition, we have added 95% confidence intervals for the anticipated effect sizes in the "Sample Size" section (Pages 9).

Reviewer #2

Comment #1: The introduction section does a nice job of describing the scope of the problem. Since the main intervention being tested is a peer recovery support specialist, this section should provide more background on the benefit of this approach for OUD (or SUD) in other setting such as primary care or outpatient behavioral health. The authors describe previous ED-based interventions to reduce overdose but should include 'Bohnert AS et al. A pilot randomized clinical trial of an intervention to reduce overdose risk behaviors among emergency department patients at risk for prescription opioid overdose' published in Drug and Alcohol Dependence (2016) as well to complete this literature review. In the last paragraph of the introduction, please

include the condition or specific population being tested (e.g., OUD, patients at risk for overdose).

Response #1: We thank the reviewer for these comments on the introduction section of the manuscript.

In the revised submission, we have added the following text to describe the benefits of peer-supported interventions (Page 5), "A recent systematic review by Ramchand and colleagues (2017) found that group-based interventions that use peers as educators commonly improve knowledge, attitudes, beliefs, and perceptions and improve connectedness and engagement with health promotion activities (Ramchand, Ahluwalia, Xenakis, Apaydin, Raaen, & Grimm, 2017)."

Further, we have added the following text to provide a complete picture of the literature on this topic (Page 4), "One of the only randomized studies of overdose education combined with a brief motivational interviewing-based intervention delivered in the ED by licensed mental health counselors resulted in reductions in overdose risk behaviors among individuals presenting to the ED and reporting non-medical prescription opioid use (Bohnert et al., 2016). However, a trial deploying a similar intervention did not reduce overdose rates or prevent subsequent ED visits or hospitalizations compared to usual care (Banta-Green et al., 2019)."

In the last paragraph of the introduction ("Objectives"), we have revised the statement of our aim to specify the condition or specific population being tested (Page 5). This sentence now reads, "We aim to test the effectiveness of behavioral interventions delivered in the ED by certified peer recovery support specialists in improving outcomes for patients at high risk of opioid overdose relative to those delivered by licensed clinical social workers (LCSWs)."

Banta-Green CJ, Coffin PO, Merrill JO, Sears JM, Dunn C, Floyd AS, Whiteside LK, Yanez ND, Donovan DM. Impacts of an opioid overdose prevention intervention delivered subsequent to acute care. Inj Prev. 2019; 25(3): 191–198.

Bohnert ASB, Bonar EE, Cunningham R, Greenwald MK, Thomas L, Chermack S, Blow FC, Walton M. A pilot randomized clinical trial of an intervention to reduce overdose risk behaviors among emergency department patients at risk for prescription opioid overdose. Drug Alcohol Depend. 2016; 163: 40–47.

Ramchand R, Ahluwalia SC, Xenakis L, Apaydin E, Raaen L, Grimm G. A systematic review of peer-supported interventions for health promotion and disease prevention. Prev Med. 2017; 101: 156–170.

Comment #2: In the methods section, please include the dates of the study including the date of collection of all primary endpoints. In the paragraph that states the inclusion criteria, please note how '3) self-report of an opioid overdose in the previous 12-months' is captured. Is this systematic from RA universal approach or reported to the clinical team during the encounter and subsequently referred to the program? Please note if all recruitment occurs in the ED or if patients who become eligible for screening once they are not critically ill can be recruited/screened in other areas of the hospital (e.g., inpatient). This may be important depending on the rate of admission for overdose at these institutions.

Response #2: We thank the reviewer for these comments on the methods section of the manuscript. In the revised manuscript (Page 9), we have clarified that participants are being identified as eligible by screening the electronic medical record (EMR) of consecutive patients presenting to the emergency department for the relevant conditions (e.g., being treated for an opioid-involved overdose, receiving treatment related to opioid use disorder) and referral from treating providing (e.g., those identifying patients who self-report an opioid-involved overdose within the previous 12 months). Given the specificity of the pool of licensed clinical social workers (LCSWs) and certified peer recovery support services to the emergency department, recruitment will be limited to patients in the emergency department only.

Comment #3: How does the study team manage a patient who is randomized to the peer support arm but requests a social work visit? How is fidelity to the intervention model being documented? Figure 1 suggests interventions will be done using principles of motivational interview. Please note if there is motivational interviewing fidelity testing being done. The peer recovery specialist intervention has three specific components and is provided by a lay person. How is fidelity to these three elements being done across the various peer recovery coaches? Consider including the number of peer recovery coaches involved in this study. Please note if social workers and peer recovery coaches are available 7 days a week and 24 hours a day.

Response #3: We thank the reviewer for these comments. At enrollment, participants agree to be randomized to receive one of these services (either a peer recovery support specialist or a licensed clinical social worker [LCSWs]). Patients are able to request to see the other service as they would be allowed to as part of usual clinical care. These individuals would be noted as a protocol violation

and would be excluded from the per protocol analysis. In addition, in the revised manuscript, we have clarified that this trial is designed as a pragmatic trial (Thorpe et al., 2009) where we aim to determine the real-world effectiveness of the behavioral interventions delivered by the certified peer recovery specialists. As this is a pragmatic trial, there is no measurement of practitioner adherence to intervention protocols and no special strategies will be used to maintain or improve adherence to these protocols. Nonetheless, both the certified peer recovery support specialists and LCSWs receive ongoing training from their respective leadership (Page 6). At the outset of the study, both groups of interventionists will receive refresher training (Page 6).

Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, Tunis S, Bergel E, Harvey I, Magid DJ, Chalkidou K. A pragmatic-explanatory continuum indicator summary (PRECIS): A tool to help trial designers. J Clin Epidemiol. 2009; 62(5): 464–475.

Comment #4: The outcomes section states two primary outcomes. With regards to engagement in treatment outcome, it is a strength of the trial to use administrative data. Please note if BHOLD includes methadone maintenance administered at an OTP. Also, how will investigators handle patients who initiate buprenorphine from the ED, but do not engage with an outpatient provider in 30 days? Or likewise, have an interruption of buprenorphine coverage during the 30 days? While patients offered naltrexone will likely be small, these patients will not be captured in this data set. Consider noting this in the limitations/discussion section. The secondary primary outcome (recurrent visits) is also assessed using administrative data. Consider adding something to the limitations/discussion section that ED visits outside of RI will not be captured and thus may be an underestimate. The ClinicalTrials.gov link lists several secondary outcomes. Consider including these as well.

Response #4: We thank the reviewer for these comments on the discussion section of the manuscript. In the revised manuscript, we have clarified that this outcome focuses on treatment engagement *after* the initial emergency department visit (Page 8). As such, individuals who initiate buprenorphine from the emergency department will not be considered engaged in treatment without receiving a prescription in the community. In addition, we have clarified that BHOLD includes individuals who receive methadone, buprenorphine, or naltrexone at a certified opioid treatment program and that data from the PDMP will be used to identify individuals receive either buprenorphine or naltrexone in an office-based setting (Page 8). The limitation associated with treatment engagement or emergency department utilization in states outside of Rhode Island has been added to the discussion section (Page 13).

VERSION 2 – REVIEW

REVIEWER	Helena Chmura Kraemer Stanford University, Stanford, CA USA
REVIEW RETURNED	09-Sep-2019

GENERAL COMMENTS	This paper describes the rationale and design for a RCT already well underway. From the clinicaltrials.gov registration, recruitment began Nov. 17, 2019 and is to go on until 2020. Nothing in the sampling and design can now be changed. Thus this paper would serve as background when the results are finally to be published. Additionally, the comments reviewers make on this paper may prepare the investigators for questions that might arise in publishing those results, perhaps even to encourage the investigators to take steps to avoid the criticisms. #1. As I noted in my earlier review, the duration of interventions is here crucial. It will not be possible to tell whether any differences between treatment outcomes (or lack thereof) are due to the peer vs. professional contact, or to the duration of contact (clearly seen in Figure 1). Moreover, the addition of several assessment time points (30 days, 3 months, 6 months post-discharge) (page11, line 25) may affect the results, and should be considered part of the treatment protocol for both treatments. Since the primary outcomes are both drawn from records (a distinct advantage here), I wonder why these assessment points were added. There are likely to be a great deal of missing data there, and non-randomly missing. It would be very helpful if you included the assessment times (30 days for engagement, 18 months for recurrent ED visit, as well as 30 days, 3 months, 6 months for other measures) in Figure 1. That, I think, would make the timing issue clearer. #2. I still have problems with your primary outcome measures. I appreciate that you now do the primary analysis with Treatment, Site, Treatment X Site interaction, with both Treatment and Site coded +1/2 and -1/2, so that the Main effect of Treatment tests the average effect over the two sites. You set your critical value for the power computations at an increase from 7% to 14% for engagement and a decrease from 15% to about 7.5% for a repeat ED visit. I've quickly checked your power, and provided you use a 5% test for each primary outcome, 650 would probably yield about 80% power. In both cases, the NNT is about 14, meaning that only 1 of 14 patients would do better with the new treatment than with the old. It would be useful if you were to eventually consider the cost per patient and evaluate the cost-effectiveness. I am puzzled by the two Effect Sizes and their confidence intervals you provided. Where did these come from? What effect size is that? They seem irrelevant here, but perhaps I miss their importance? Again, the two primary outcomes are correlated outcomes. Should you not be using the 2.5% significance level rather than the 5%? Finally I wonder why you are using a binary outcome (yes/no) at all rather than time to outcome, i.e., time to engagement, and time to recurrent ED visit. Comparing survival curves (e.g., using the Cox Model) would give you greater power (even with a 2.5% significance level), and the survival curves would provide much more information (the outcomes at 30 days for engagement, and that at 18 months for recurrent ED visit would be seen, as well as

	the outcomes at various other time points—which might be very informative.) #3. On page 11, why do the invalid comparison that you describe as a "sensitivity analysis"? You will have no missing data on your primary outcomes. Thus ITT analysis will give the answer to the primary question. No "sensitivity analysis" needed. Please do NOT do subgroup analysis (Check the literature for problems from subgroup analyses). Do moderator analyses in an exploratory mode. Using survival methods (e.g. Cox models) rather than logistic regression would be helpful. You would then add to the Treatment, Site, Treatment X Site, Age, Age X Treatment, Age X Site, Age X Treatment X Site. If the 3 way interaction came out, that would mean different moderation at the two sites. If the Age X Treatment came out, without that 3 way interaction, that means that the effect of treatment changes with age. You might then draw the survival curves for the different age ranges to see exactly how and why. This type of exploratory work requires fairly large sample size, and is less likely to succeed with binary outcomes. #4. Your final caveats are exactly why two-tailed tests are needed. The issue of removal to other states should similarly impact both treatment groups, would it not?

REVIEWER	Lauren Whiteside University of Washington Seattle, WA
REVIEW RETURNED	14-Sep-2019

GENERAL COMMENTS	Thank you for the opportunity to review this revised submission. The authors provide a manuscript that describes the protocol for a randomized clinical trial of a peer-based recovery support intervention compared to usual care (social work) form the ED for patients with overdose or risk for overdose with a goal of increasing retention in subtance use treatment at 30 days and decreasing ED visits over 18 months. I believe the introduction is strengthened with more robust description of peer-naviagors and the enhanced literature review on overdose intervention from the ED. A minor grammatical change to consier would include: Delete the word 'that' from line 11 to now read….'in the US (5) the crisis is expected to worsen…' I agree with the authors that it is more pragmatic to not track fidelity to the intervention, however I also believe it is impossible to know what was actually delivered without some research infrastructure in place such as fidelity tracking. Process fidelity (e.g. a checklist of activities) could be done with limited resources and intrusion into clinical work and still be pragmatic. Therefore, I would recommend the authors consider noting this in the limitations paragraph in the Discussion section as lack of fidelity monitoring or process checking limits understanding of what was actually done if this intervention proves effective. On page 8 under the section 'Behavioral Intervetnion Delivered via Licensed Clinical Social Workers' the authors note there are 35 full and part-time LCSW. It would be helpful to have a similar count of

	people or FTE employed as peer recovery support specialists to do this intervention. The first reviewer of this manuscript noted 'that you refer to recurrent ED visits but your analysis seems to be for 'at last one ED visit'. I would recommend being more explicit about this in the statistical methods section (Page 11) as it seems subsequent overdoses is really one or more overdoses vs none. Overall, I believe the edits and changes made by the authors strengthen this manuscript, which describes the protocol for an important public health investigation.

VERSION 2 – AUTHOR RESPONSE

Reviewer #1

Comment #1: As I noted in my earlier review, the duration of interventions here is crucial. It will not be possible to tell whether any differences between treatment outcomes (or lack thereof) are due to the peer vs. professional contact, or to the duration of contact (clearly seen in Figure 1). Moreover, the addition of several assessment time points (30 days, 3 months, 6 months post-discharge [Pg. 11, Line 12]) may affect the results, and should be considered part of the treatment protocol for both treatments. Since the primary outcomes are both drawn from records (a distinct advantage here), I wonder why these assessment points were added. There are likely to be a great deal of missing data there, and non-randomly missing. It would be very helpful if you included the assessment times (30 days for engagement, 18 months for recurrent ED visit, as well as 30 days, 3 months, 6 months for other measures) in Figure 1. That, I think, would make the timing issue clearer.

Response #1: We agree with the reviewer's comment that the duration of the interventions provided may be a crucial factor in determining their success. However, we respectfully note that the differences in intervention length are inherently part of the interventions themselves – licensed clinical social workers (LCSWs) operate only in the emergency department (ED) setting, while peer recovery support specialists provide services in the ED and community settings, allowing them to remain in touch following discharge. This is one hypothesized benefit of using peer recovery support specialists. We are also in agreement regarding the inclusion of the follow-up assessments. The follow-up assessments at 30 days, 3 months, and 6 months post-discharge were added to the trial to gather additional, exploratory self-reported outcomes from participants with the awarding of additional funding. However, due to limited success (only 25 participants out of 154 eligible participants have been completed any assessment at the time of this response letter) and the ending of this funding resource, the follow-up assessments have been discontinued. Discussion of these assessments has been removed from the revised manuscript (see Page 10), as suggested by the reviewer.

Comment #2: I still have problems with your primary outcome measures. I appreciate that you now do the primary analysis with Treatment, Site, Treatment x Site Interaction, with both Treatment and Site coded as +1/2 and -1/2, so that the main effect of treatment tests the average effect over the two sites. You set your critical value for the power computations at an increase from 7% to 14% for engagement and a decrease from 15% to 7.5% for a repeat ED visit. I've quickly checked your power, and provided you use a 5% test for each primary outcome, 650 would probably yield about 80% power. In both cases, the NNT is about 14, meaning that only 1 of 14 patients would do better with the new treatment than with the old. It would be useful if you were to eventually consider the cost per patient and evaluate the cost-effectiveness. I am puzzled by the two effect sizes and their confidence intervals you provided. Where did these come from? What effect size is that? They seem irrelevant here, but perhaps I miss their importance? Again, the two primary outcomes are correlated outcomes. Should you not be using the 2.5% significance level rather than the 5%? Finally, I would why you are using a binary outcome (yes/no) at all rather than time to outcome, i.e., time to engagement and time to recurrent ED visit. Comparing survival curves (e.g., using the Cox model) would give you greater power (even with a 2.5% significance level), and the survival curves would provide much more information (the outcomes at 30 days for engagement, and that at 18 months for recurrent ED visit would be seen, as well as the outcomes at various other time points – which might be very informative).

Response #2: We thank the reviewer for this comment. The use of binary outcomes rather than time to event outcomes was informed by our community stakeholders, who felt that the results with a binary outcome (e.g., any recurrent ED visits for opioid overdose versus no recurrent ED visits for opioid overdose) were more relevant and compelling for policy and advocacy efforts. As such, we have retained our initial analytical plan in this regard. However, as recommended by the reviewer, we have added exploratory survival analyses for both outcomes to the analytical plan in the revised manuscript (see Page 10), with text reading, "Analyses with binary outcomes were selected based on input from community stakeholders. In exploratory analyses, we will use survival methods (e.g., Cox proportional hazards models) to assess the impact of treatment on the time to events for both outcomes (i.e., days from discharge to enrollment in formal SUD treatment and days from discharge to first recurrent ED visit for opioid overdose)."

Further, we have retained independent analyses for both of the primary outcomes. Our community stakeholders have expressed a desire to focus on both outcomes, as engagement in substance use treatment represents the presence of a positive outcome, while a lack of recurrent ED visits for opioid overdose represents the absence of a negative outcome, thus providing a fuller picture of

the experiences of individuals beginning their recovery journey through these behavioral interventions. Further, we acknowledge that divergent results are possible and would be informative to future research, policy, and practice. In the case that the intervention delivered by a peer recovery support specialist improves treatment engagement but has no impact on subsequent ED visits for opioid overdose relative to the intervention delivered by a licensed clinical social worker, these findings would highlight a need for additional harm reduction interventions for individuals at high risk for opioid overdose in addition to promoting engagement in substance use treatment.

The effect sizes reported in the previous version of the manuscript were added in response to Reviewer #1, Comment #6 from the first round of review ("I would strongly recommend including effect sizes of their confidence intervals"), where the reported effect sizes are success rate differences (SRD) and their associated confidence intervals [1].

We thank the reviewer for the additional recommendation in considering the cost-effectiveness of the two intervention arms and will plan to do so in future studies.

1. Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. Biol Psychiatry. 2006; 59(11): 990–996.

Comment #3: On Page 11, why do the invalid comparison that you describe as a "sensitivity analysis"? You will have no missing data on your primary outcomes. Thus, ITT analysis will give the answer to the primary question. No "sensitivity analysis" needed. Please do NOT do subgroup analysis (check the literature for problems from subgroup analyses). Do moderator analyses in an exploratory mode. Using survival methods (e.g., Cox models) rather than logistic regression would be helpful. You would then add the Treatment, Site, Treatment x Site, Age, Age x Treatment, Age x Site, Age x Treatment x Site. If the three-way interaction came out, that would mean different moderation at the two sites. If the Age x Treatment interaction came out, without the three-way interaction, that means the effect of treatment changes with age. You might then draw the survival curves for the different age ranges to see exactly how and why. This type of exploratory work requires fairly large sample size and is less likely to succeed with binary outcomes.

Response #3: We thank the reviewer for this comment. The per-protocol (sensitivity) analysis has been removed from the analytical plan. Further, formal subgroup analyses have been removed from the analytical plan, but moderation analyses will be conducted on an exploratory basis as suggested by the reviewer. The text in the revised manuscript (see Page 10) now reads, "In

addition, on an exploratory basis, we will conduct moderation analyses to understand potential heterogeneity of treatment effects by age and gender."

Further, we have clarified that, in all analyses, we are not concerned with estimating site-specific treatment effects. In the revised manuscript (see Page 10), we have added text that reads, "Given the multicenter design of the trial, the effect of treatment site will be quantified by the intraclass correlation coefficient (ICC), representing the variance due to the between-center variability [1]. Should the ICC suggest that a large portion of the variance is explained by between-center variability, we will control for treatment site using a generalized estimating equations (GEE) approach to estimate population-average treatment effects across the two sites."

1. Moerbeek M, van Breukelen GJ, Berger MP. A comparison between traditional methods and multilevel regression for the analysis of multicenter intervention studies. J Clin Epidemiol. 2003; 56(4): 341–350.

Reviewer #2

Comment #1: I believe the introduction is strengthened with more robust description of peer navigators and the enhanced literature review on overdose intervention from the ED. A minor grammatical change to consider would include delete the word 'that' from Line 11 to now read, "…in the US (5) the crisis is expected to worsen…"

Response #1: We thank the reviewer for this comment. This grammatical error has been corrected.

Comment #2: I agree with the authors that it is more pragmatic to not track fidelity to the intervention, however I also believe it is impossible to know what was actually delivered without some research infrastructure in place such as fidelity tracking. Process fidelity could be done with limited resources and intrusion into clinical work and still be pragmatic. Therefore, I would recommend the authors consider noting this in the limitations paragraph in the discussion section as a lack of fidelity monitoring or process checking limits understanding of what was actually done if this intervention proves effective.

Response #2: We thank the reviewer for this comment. In the revised manuscript, the lack of fidelity monitoring or process checking limits an understanding of what components of the intervention

were performed. Text reading, “Given the pragmatic nature of the trial, fidelity monitoring and process checking are not being conducted, thus limiting potential understanding of what components of the intervention were performed should they be deemed efficacious” has been added to the discussion section of the revised manuscript (see Page 13).

Comment #3: On Page 8, under the section ‘Behavioral Intervention Delivered via Licensed Clinical Social Workers’, the authors note there are 35 full and part-time LCSW. It would be helpful to have a similar count of people or FTE employed as peer recovery support specialists to do this intervention.

Response #3: On Page 7 of the revised manuscript, we have added additional text stating that there are 30 peer recovery support specialists available between the two study sites to deliver these interventions.

Comment #4: The first reviewer of this manuscript noted ‘that you refer to recurrent ED visits, but your analysis seems to be for ‘at least one ED visit’. I would recommend being more explicit about this in the statistical methods section (Pg. 11) as it seems subsequent overdoses is really one or more overdoses vs. none.

Response #4: On Page 8 of the revised manuscript, we have clarified our description of the outcome measure as a binary variable indicating any or no recurrent ED visits for opioid-involved overdose.

VERSION 3 – REVIEW

REVIEWER	Helena Chmura Kraemer Stanford University, USA
REVIEW RETURNED	11-Oct-2019

GENERAL COMMENTS	This now seems a valuable paper as the basis of a RCT yet to come. I have but one very small request: On page 10, when you say “Effect Size” could you specify what effect size that is. There are an infinite number of effect sizes for 2X2 association (here Treatment X binary outcome), ranging from odds ratio, to two risk ratios, to a range of weighted kappas to the phi coefficient. Moreover, many interpret “Effect size” to mean Cohen’s d, which is here assuredly not what you mean. I think your “Effect Sizes” are Risk Differences. This is important, because a critical effect size of .07 for Cohen’s d would be smaller

	than small, while a critical effect size of .07 for odds ratio would be very large. Risk Difference=.07 means Number Needed to Take=1/.07=14, i.e., you would have to provide the preferred treatment to 14 people to have one more "success" than if you had given them the non-preferred treatment. Just define what you mean by "Effect Size".

REVIEWER	Lauren Whiteside University of Washington, USA
REVIEW RETURNED	10-Oct-2019

GENERAL COMMENTS	Overall, this manuscript is improved from previous and the authors have done a nice job responding to all the reviewer responses. The description of the trial protocol is succinct and clear and the description of the two arms of the trial (Peer Support vs ED LCSW) is well done. The primary outcomes are relevant and well defined. The 'Data' Management section mentions REDCap and follow-up assessments to be conducted by phone (Page 33, line 33) which should be deleted since the self-report outcomes are no longer being reported. The Discussion section is well written and places the trial in context for local and state policy. The limitations paragraph is important and will allow for interpretation of trial results. I believe this paragraph is complete.

VERSION 3 – AUTHOR RESPONSE

Reviewer #1

Comment #1: This now seems to be a valuable paper as the basis of an RCT yet to come. I have but one very small request: On Page 10, when you say, "Effect Size", could you specify what effect size that is? There are an infinite number of effect sizes for 2x2 association, ranging from odds ratio, to two risk ratios, to a range of weighted kappas to the phi coefficient. Moreover, many interpret "effect size" to mean Cohen's *d*, which is here assuredly not what you mean. I think your "Effect Sizes" are risk differences. This is important, because a critical effect size for 0.07 for Cohen's *d* would be smaller than small, while a critical effect size of 0.07 for odds ratios would be very large. Risk difference = 0.07 means Number Needed to Treat = 1/0.07, i.e., you would have to provide the preferred treatment to 14 people to have one more "success" than if you had given them the non-preferred treatment. Just define what you mean by "Effect Size".

Response #1: In the revised manuscript, we have specified that these effect sizes are risk differences.

Reviewer #2

Comment #1: Overall, this manuscript is improved from previous and the authors have done a nice job responding to all the reviewer responses. The description of the trial protocol is succinct and clear and the description of the two arms of the trial (Peer Support vs. ED LCSW) is well done. The primary outcomes are relevant and well-defined.

Response #1: We thank the reviewer for these comments.

Comment #2: The 'Data Management' section mentions REDCap and follow-up assessments to be conducted by phone (Page 33, Line 33) which should be deleted since the self-report outcomes are no longer being reported.

Response #2: This text has been removed in the revised manuscript.

Comment #3: The discussion section is well written and places the trial in context for local and state policy. The limitations paragraph is important and will allow for interpretation of trial results. I believe this paragraph is complete.

Response #3: We thank the reviewer for these comments.