Assessing sample representativeness in randomized controlled trials: application to the National Institute of Drug Abuse Clinical Trials Network

Ryoko Susukida1, Rosa M. Crum1,2,3, Elizabeth A. Stuart1, Cyrus Ebnesajjad1 & Ramin Mojtabai1,3

Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA,1 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA2 and Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA3

ABSTRACT

Aims  To compare the characteristics of individuals participating in randomized controlled trials (RCTs) of treatments of substance use disorder (SUD) with individuals receiving treatment in usual care settings, and to provide a summary quantitative measure of differences between characteristics of these two groups of individuals using propensity score methods. Design Analyses using data from RCT samples from the National Institute of Drug Abuse Clinical Trials Network (CTN) and target populations of patients drawn from the Treatment Episodes Data Set—Admissions (TEDS-A). Settings  Multiple clinical trial sites and nation-wide usual SUD treatment settings in the United States. Participants  A total of 3592 individuals from 10 CTN samples and 1 602 226 individuals selected from TEDS-A between 2001 and 2009. Measurements  The propensity scores for enrolling in the RCTs were computed based on the following nine observable characteristics: sex, race/ethnicity, age, education, employment status, marital status, admission to treatment through criminal justice, intravenous drug use and the number of prior treatments. Findings  The proportion of those with \( \geq 12 \) years of education and the proportion of those who had full-time jobs were significantly higher among RCT samples than among target populations (in seven and nine trials, respectively, at \( P < 0.001 \)). The pooled difference in the mean propensity scores between the RCTs and the target population was 1.54 standard deviations and was statistically significant at \( P < 0.001 \). Conclusions  In the United States, individuals recruited into randomized controlled trials of substance use disorder treatments appear to be very different from individuals receiving treatment in usual care settings. Notably, RCT participants tend to have more years of education and a greater likelihood of full-time work compared with people receiving care in usual care settings.

Keywords  Clinical trials, sample representativeness, substance use disorders.

INTRODUCTION

Randomized controlled trials (RCTs) are generally considered the gold standard for establishing the effectiveness of new interventions. Decision-making for health policies and for clinical practice relies heavily on findings of RCTs. Furthermore, RCTs provide confidence in causal attribution of the effects of new interventions through eliminating threats to internal validity.

However, the study design of RCTs does not eliminate threats to external validity, which indicates how well findings from one particular setting can apply to the target population, i.e. ‘the group of persons for whom an intervention is planned’ [1]. Lack of external validity is a concern when the RCT participants are different from the target population. Findings from recent studies have heightened these concerns by showing that RCT samples might not represent the types of patients encountered in usual clinical practice settings [2,3]. In the context of substance use disorders (SUD), recent studies have shown that tight exclusion criteria commonly employed in RCTs might have resulted in RCT samples that are different with regard to sex and race distribution from the treatment-seeking populations in usual care settings [4–7].

Past research has mainly examined what proportion of a putative target population would be excluded from RCTs
based on the formal eligibility criteria. One study found that commonly used exclusion criteria in the RCTs of alcohol use disorder treatments excluded approximately 20–33% of patients with these disorders [5]. This study also found that commonly used exclusion criteria tend to exclude more female and African American patients. Another study found that common eligibility criteria in cannabis treatment RCTs would exclude 80% of patients with cannabis dependence [7]. These findings indicate substantial selection bias in RCT samples.

However, past studies have rarely compared the characteristics of actual RCT participants and the intended target populations. In addition to exclusion criteria, refusal to participate in RCTs impacts representativeness. Refusal to participate in RCTs is a particular concern in SUD treatment studies, as a large proportion of clients are referred to treatment from the criminal justice system and are not seeking treatment voluntarily. Direct comparisons of RCT samples with the target populations for whom the treatments are intended could provide guidance as to the potential threats to the generalizability of findings of RCTs. To our knowledge, no previous studies have directly compared the RCT samples and target populations in the context of SUD treatments.

Recruiting representative samples of target populations into RCTs can be challenging [8]. Participation decisions may vary across different socio-demographic groups. Individuals with more education and higher socio-economic status might hold a more positive attitude towards scientific research and be more willing to participate in RCTs [9,10]. It may also be difficult to enroll individuals who have jobs, and thus have higher opportunity costs of participation in RCTs than unemployed people [11]. Importantly, these characteristics are likely to be associated with attrition and treatment outcomes [12–15].

This study compared differences in the characteristics of individuals who participated in a number of RCTs of SUD treatments with those drawn from target populations for whom these treatments are intended. The aims of this study were twofold:

1 To perform a pairwise comparison of socio-demographic characteristics between each RCT sample and its corresponding target population; and
2 To estimate differences between the RCT samples and target populations by calculating propensity scores based on nine measured characteristics of participants in each RCT and its corresponding target population.

METHODOLOGY

Data source

The NIDA CTN studies are multi-site clinical trial projects to evaluate the efficacy of treatments for SUD. At the time of this study, data from 27 CTN RCTs were publically available [16].

The TEDS-A is part of a national census data system collecting data on admissions to SUD treatment facilities annually. All States that receive public funds for SUD treatment programs are required to provide the data to the TEDS-A. The TEDS-A compiles annual national data on more than 1.5 million patients aged 12 years and older, thus providing a relevant target population for samples recruited to specific RCTs. We identified a separate target population drawn from TEDS-A for each RCT based on the characteristics of the patients who were the target of the intervention for that RCT. We considered age and the target substance in defining the target populations. Age was considered because some interventions were intended specifically for young adults. We also attempted to draw samples from TEDS-A for the years corresponding to each RCT. For example, the target population for the CTN00010, an RCT of buprenorphine/naloxone facilitated rehabilitation for heroin-addicted adolescents and young adults aged 14–21 years recruited between July 2003 and December 2005, was drawn from the population of patients in TEDS-A 2003–2006 who were aged between 14 and 21 years and received SUD treatment for opioid use disorders. If an RCT was clearly intended for a more specific target population (e.g., pregnant women), we identified the corresponding target population according to these additional criteria. Supp1, provides descriptions of corresponding target populations for each CTN RCT.

We could not define target populations for all 27 CTN data sets available in the CTN database at the time of this study because of limited information available in TEDS-A. For instance, TEDS-A does not contain information regarding HIV status. Therefore, we could not define target populations for CTN studies involving HIV-positive patients. This study utilizes 10 CTN RCTs for which TEDS-A target populations could be matched based on the RCT inclusion criteria.

Five of these 10 studies assessed the effectiveness of buprenorphine/naloxone detoxification (Bup/Nx-Detox) for opioid dependence either in in-patient (CTN0001 [17]) or out-patient settings (CTN0002 [17], CTN0003 [18], CTN0010 [19], CTN0030 [20]). Most focused on Bup/Nx-Detox in adults aged ≥18 years (CTN0001, CTN0002, CTN0030). One study included those aged ≥15 years (CTN0003); another included youth aged 14–21 only (CTN0010). Three studies aimed to assess the effectiveness of motivational enhancement/interviewing (MEI) for SUD in out-patient settings (CTN0004 [21], CTN0005 [22], CTN0013 [23]). While CTN0004 and CTN0005 targeted men and women aged ≥18 years, CTN0013 targeted only pregnant women. The other two studies aimed to assess the effectiveness of motivational incentives (Incentives) for current cocaine, methamphetamine or amphetamine use.
<table>
<thead>
<tr>
<th></th>
<th>Buprenorphine/naloxone (Bup/Nx) detoxification</th>
<th>Motivational enhancement/interviewing</th>
<th>Motivational incentives</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CTN0001</td>
<td>CTN0002</td>
<td>CTN0003</td>
</tr>
<tr>
<td></td>
<td>RCT</td>
<td>TEDS</td>
<td>RCT</td>
</tr>
<tr>
<td>Total number</td>
<td>113</td>
<td>3111</td>
<td>230</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>39.8</td>
<td>25.9</td>
<td>28.3</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>55.8</td>
<td>52.5</td>
<td>40.0</td>
</tr>
<tr>
<td>Black</td>
<td>19.5</td>
<td>15.9</td>
<td>36.5</td>
</tr>
<tr>
<td>Hispanic</td>
<td>15.9</td>
<td>30.1</td>
<td>20.0</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12–17</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>18–29</td>
<td>28.3</td>
<td>30.9</td>
<td>22.2</td>
</tr>
<tr>
<td>30–39</td>
<td>26.6</td>
<td>38.4</td>
<td>22.6</td>
</tr>
<tr>
<td>40–49</td>
<td>35.4</td>
<td>24.8</td>
<td>37.8</td>
</tr>
<tr>
<td>50 and over</td>
<td>9.7</td>
<td>5.9</td>
<td>17.4</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 years</td>
<td>87.6</td>
<td>63.0</td>
<td>73.5</td>
</tr>
<tr>
<td>Employment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-time</td>
<td>58.0</td>
<td>5.6</td>
<td>54.3</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>33.6</td>
<td>12.8</td>
<td>24.3</td>
</tr>
<tr>
<td>Criminal justice admission</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>8.9</td>
<td>3.9</td>
<td>3.1</td>
</tr>
<tr>
<td>No. of prior treatments</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 5 times</td>
<td>31.1</td>
<td>18.5</td>
<td>26.7</td>
</tr>
<tr>
<td>No. prior treatments</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 5 times</td>
<td>15.4</td>
<td>6.1</td>
<td>20.4</td>
</tr>
</tbody>
</table>

*Pearson’s χ² test was conducted. Numbers shown in bold type indicate statistically significant differences between RCT and TEDS-A samples at P < 0.05. bNot included in the analyses, as these variables were not available for CTN0010. cNot included in the analyses because of a large number of missing values for this variable in TEDS-A. i.v. = intravenous.
among adults aged ≥ 18 in out-patient settings (CTN0006 [24], CTN0007 [25]). This study included a total of 3592 individuals from 10 CTN samples and 1 602 226 individuals from TEDS-A between 2001 and 2009.

Assessments
Baseline characteristics of the CTN RCT patients were assessed at the time of enrollment. In TEDS-A, patients’ information was collected at treatment onset. Nine comparable variables were recorded both in the RCT samples and the TEDS-A: sex, race-ethnicity (white, black, Hispanic and other), age (recoded into 12–14, 15–17, 18–20, 21–24, 25–30, 31–34, 35–40, 41–44, 45–50, 51–54, ≥ 55 years), education (< 8 years, 9–11, 12, 13–15, ≥ 16 years), employment (full-time, part-time, out of labor force (students, homemakers and those without jobs not looking for work) and unemployed (those without jobs who are actively looking for work), marital status (never married, married, separated, divorced/widowed), admission through criminal justice, intravenous drug use and the number of prior treatments for SUD.

Statistical analysis
We first compared each RCT sample with its corresponding target population with regard to the baseline characteristics noted. There was a non-negligible amount of missing data, particularly in the TEDS-A. Supporting information, Table S2 presents the percentage of missing observations in each RCT and target population. Missing values ranged from 0.1% for sex in the TEDS-A target populations for CTN0004 RCT to 70.5% for the number of prior treatments in the TEDS-A target population for CTN0001, we did not use this variable in the imputation of CTN0001 data and its corresponding TEDS-A data.

Using the imputed data sets, we calculated propensity scores modelling being enrolled in each RCT based on the characteristics of the RCT and target population: the propensity score is the conditional probability of an individual being in the RCT. We computed the propensity score-based index $\Delta p$, introduced by Stuart et al. [27], to aid researchers in assessing the representativeness of RCT samples compared to target populations. $\Delta p$ is defined as the difference between the average propensity scores of the RCT and the target population. Divided by the pooled standard deviation of the propensity scores, the standardized $\Delta p$ provides a summary index of differences between samples with regard to all variables used for computing the propensity scores. In the context of observational studies [28,29], propensity score mean values that differ by more than 0.25 standard deviations (standardized $\Delta p$) indicate significant differences between the samples, requiring a large amount of extrapolation [30,31]. Other investigators have adopted a more stringent value of 0.1 and larger as indicating significant differences [32]. In this study we computed both $\Delta p$ and standardized $\Delta p$. In addition, we conducted two-sample $t$-tests for comparison of the propensity scores.

We used the non-parametric random forests approach to calculate the propensity scores [33,34]. The R package ‘randomForest’ [35] was used for these analyses. Random forests have several advantages over a parametric approach, including higher predictive accuracy [36] and the ability to reduce misclassification error through bootstrap resampling methods, which is especially useful when comparing class-imbalanced data, such as the present case, where target populations are much larger than the RCT samples [37]. Through this bootstrap resampling method, the same number of observations from the larger group is drawn to match the number of the smaller group to balance the sizes of the groups. This down-sampling of the majority class has been shown to work well for class-imbalanced data [38].

Because 50 imputations were generated, we obtained 50 different sets of propensity scores for each comparison. The variances of the mean propensity scores were estimated using the formula introduced by Rubin [39] in order to take into account both within- and between-imputation variance.

We used meta-analytical techniques to compute pooled $\Delta p$s for the three groups of different SUD treatment studies (Bup/Nx-Detox, MEI and Incentives) included in the sample and to compare these values across the study groups. Additionally, heterogeneity in $\Delta p$s was assessed.
among individual studies using $I^2$, which is defined as the percentage of total variation across studies that is due to heterogeneity rather than chance [40]. An $I^2$ of 75% is considered to indicate a high level of heterogeneity among studies. The metaan [41] command of Stata version 13 was used for these calculations.

**Role of the funding source**

The funding organizations had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; preparation, review or approval of the manuscript; or decision to submit the manuscript for publication.

**RESULTS**

**Comparison of characteristics of RCT samples and target populations**

For all 10 RCTs examined, RCT participants were more likely to have ≥ 12 years of education than target populations (Table 1). The proportion with ≥ 12 years of education was significantly higher among RCT-enrolled patients than among patients in target populations in seven of the 10 trials.

Patients in RCTs were also more likely to be employed full time than patients in target populations. The proportion of those who had full-time jobs was significantly higher among patients enrolled in the RCTs than among patients in target populations in all nine trials in which employment status was measured.

Except for CTN0005 (MEI for SUD) and CTN0030 (Bup/Nx-Detox for prescription opioid dependence), patients who were enrolled into the RCTs were more likely to have had a larger number of prior treatments before entering the trials than patients in the target populations. The proportion of those with ≥ 5 prior treatments was significantly higher among RCT patients than those in target populations in six of the eight trials in which the number of prior treatments was available or used in the analyses.

Individual RCTs and their target populations also differed with regard to other characteristics. There were statistically significant differences in proportions of females, specific race-ethnicity groups, age groups, married individuals, admissions through criminal justice system and i.v. drug use between individual RCTs and target populations (Table 1).

**Estimation of propensity scores**

Table 2 presents the propensity scores associated with being enrolled in each RCT, based on observable characteristics. These propensity scores represent summary measures of differences in these characteristics between RCTs and target populations.

Across all CTN studies examined, the estimated propensity scores for RCTs were significantly higher than for the target populations. The Δp indices, computed as the difference between the two propensity scores, ranged from 0.25 to 0.60 and standardized Δp indices ranged from 1.06 to 2.08 standard deviations (pooled average $\Delta p = 1.54$). The pooled Δps were larger for Bup/Nx-Detox studies [1.90, 95% confidence interval (CI) = 1.81–1.98] than for MEI studies (1.28, 95% CI = 1.18–1.37) and Incentives studies (1.31, 95% CI = 1.21–1.42) (test for comparison of groups of studies, $\chi^2 = 121.40$, d.f. = 2, $P < 0.001$). However, these differences should be interpreted with caution because of significant heterogeneity among studies within each of the three groups of studies, as indicated by the high $I^2$ values (Table 2, lower panel).

Figure 1 presents the density plots of propensity scores for each RCT and its target population. A larger overlapping area between the density plot for RCT and its target population indicates that the RCT sample had similar characteristics to the target population, whereas limited overlap indicates fewer similarities between the two. Studies that have relatively smaller standardized Δps (CTN0004, CTN0005, CTN0006, CTN0007) have larger overlapping areas between density plots of RCTs and the target populations. In contrast, studies that have relatively larger standardized Δps (CTN0001, CTN0010, CTN0030) have smaller overlapping areas.

**DISCUSSION**

We found significant differences between patients participating in RCTs and the target populations of patients receiving SUD treatment in usual care settings. RCT patients had higher levels of education and were more likely to have full-time employment than those in the target populations. This is consistent with past research, which suggests that those with higher socio-demographic status have greater trust in the benefits of scientific research and more willingness to participate in trials [9,10]. Moreover, except for one RCT, those who were included in the RCTs had larger numbers of prior treatments than the target populations, which could be due to greater reliance on the formal treatment service system or less successful experiences in previous treatments. The larger Δp values for medication trials compared to behavioral trials may reflect more stringent eligibility criteria in medication trials. Even though pregnant or lactating women were excluded from the clinical trials, it was not always the case that women were under-represented in these trials. There was also no systematic pattern of under-representation of racial/ethnic minorities, which is consistent with the intention of the CTN to recruit more racial/ethnic minorities into clinical trials. This was particularly the case in Incentives studies, where the percentage of the non-white individuals was higher in the trial...
suggest that positive attitude towards SUD treatment might be associated with better outcome or more timely recovery [44]. Furthermore, prior studies indicate that these characteristics may impact relapse and response to SUD treatment [12–15,45,46]. Future research needs to assess empirically whether these differences between RCT samples and target populations indeed contribute to a more favorable response to experimental interventions delivered through RCTs. It may be feasible to adjust the analyses of trials using propensity score weights or flexible regression models in order to estimate effects in the target populations of interest, if the differences in the propensity scores between the trial samples and the target populations are relatively small [27,47]. This weighting-based approach would presumably correct biases in estimated effects of treatments and improve generalizability of RCT results.

Several limitations should be considered when interpreting this study’s findings. First, the number of variables that were available to assess for both RCTs and the TEDS-A were relatively small. The RCTs and target populations might differ on other characteristics. For example, the presence of mental and physical disorders and severity of symptoms could influence the probability of RCT enrollment and impact treatment outcomes. More importantly,

samples than the target populations, suggesting that these CTN studies successfully recruited more racial/ethnic minorities than other studies [42].

The differences between RCT and target population propensity scores (Δp) ranged from 1.06 to 2.08 standard deviations. These numbers far exceed the 0.25 standardized Δp cut-off proposed by Stuart [31], indicating significant differences between samples. This point was confirmed by the density plots of the propensity scores for the RCTs and target samples, which showed large differences. The interpretation of the standardized Δp is similar to the interpretation of Cohen’s d effect size [43]. Cohen’s d of 1.06 and 2.08 indicate a 59% and 29% probability assuming normal distributions, respectively, that the two groups will overlap.

The findings have implications for generalizability of treatment effects. For example, employed individuals with higher levels of education typically have more socioeconomic resources. Even though we could not directly assess attitude towards SUD treatment in this study, a higher participation rate in the trials among individuals with higher levels of education and full-time jobs might be associated with more positive attitudes towards SUD treatment. Studies suggest that positive attitude towards SUD treatment might

Table 2 Comparison of propensity scores between samples in 10 National Institute of Drug Use Clinical Trial Network (CTN) randomized controlled trials (RCTs) and target samples from the Treatment Episodes Data-Admission (TEDS-A).

<table>
<thead>
<tr>
<th>Study type</th>
<th>Study number</th>
<th>RCT</th>
<th>TEDS-A</th>
<th>Pooled standard deviation</th>
<th>Standardized Δp</th>
<th>t-Test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buprenorphine/naloxone</td>
<td>CTN0001</td>
<td>0.69</td>
<td>0.19</td>
<td>0.50</td>
<td>0.25</td>
<td>2.07</td>
<td>22.94</td>
</tr>
<tr>
<td>(Bup/Nx) detoxification</td>
<td>CTN0002</td>
<td>0.64</td>
<td>0.25</td>
<td>0.39</td>
<td>0.23</td>
<td>1.67</td>
<td>24.99</td>
</tr>
<tr>
<td></td>
<td>CTN0003</td>
<td>0.70</td>
<td>0.25</td>
<td>0.45</td>
<td>0.26</td>
<td>1.72</td>
<td>38.99</td>
</tr>
<tr>
<td></td>
<td>CTN0010</td>
<td>0.69</td>
<td>0.21</td>
<td>0.48</td>
<td>0.24</td>
<td>2.03</td>
<td>25.32</td>
</tr>
<tr>
<td></td>
<td>CTN0030</td>
<td>0.80</td>
<td>0.20</td>
<td>0.60</td>
<td>0.28</td>
<td>2.04</td>
<td>53.12</td>
</tr>
<tr>
<td>Motivational enhancement/</td>
<td>CTN0004</td>
<td>0.61</td>
<td>0.30</td>
<td>0.31</td>
<td>0.23</td>
<td>1.35</td>
<td>28.65</td>
</tr>
<tr>
<td>interviewing</td>
<td>CTN0005</td>
<td>0.57</td>
<td>0.33</td>
<td>0.24</td>
<td>0.23</td>
<td>1.06</td>
<td>21.74</td>
</tr>
<tr>
<td></td>
<td>CTN0006</td>
<td>0.57</td>
<td>0.32</td>
<td>0.25</td>
<td>0.21</td>
<td>1.20</td>
<td>25.23</td>
</tr>
<tr>
<td>Motivational incentives</td>
<td>CTN0007</td>
<td>0.62</td>
<td>0.28</td>
<td>0.34</td>
<td>0.23</td>
<td>1.46</td>
<td>27.59</td>
</tr>
</tbody>
</table>

Pooled results

<table>
<thead>
<tr>
<th>Heterogeneity statistics</th>
<th>Mean propensity score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pooled standardized Δp</td>
<td>95% CI</td>
</tr>
<tr>
<td>Overall</td>
<td>1.54</td>
</tr>
<tr>
<td>Buprenorphine/naloxone</td>
<td>1.90</td>
</tr>
<tr>
<td>(Bup/Nx) detoxification</td>
<td>Motivational enhancement/</td>
</tr>
<tr>
<td>interviewing</td>
<td>Motivational incentives</td>
</tr>
</tbody>
</table>

*Δp is difference between propensity scores of CTN RCT sample and TEDS-A samples. Standardized Δp is computed as Δp divided by pooled standard deviation. I² is the percentage of observed total variation across studies that is due to heterogeneity rather than chance. It is calculated as I² = 100% × (Q – d.f.)/Q, where Q is Cochran’s heterogeneity statistic and d.f. the degrees of freedom (see text for further information). CI = confidence interval. © 2016 Society for the Study of Addiction
attitudes towards recovery and readiness for change may differ significantly between RCT participants and typical treatment populations. Furthermore, lack of information regarding HIV status and the presence of co-occurring mental disorders as well as substance use disorders could have biased the socio-economic distribution of the trial samples, because individuals with these disorders tend to have lower socio-economic status [48,49]. With a larger range of variables included in propensity score models, the differences between RCTs and TEDS-A would probably be even larger. Secondly, TEDS-A had a non-negligible number of missing observations that we addressed with multiple imputation, which could possibly bias the composition of the target populations if the missingness at random assumption is incorrect. Thirdly, because of limitations in reported characteristics for TEDS-A participants, we could not delineate target populations a number of other CTN RCTs. The RCTs included were limited to studies of Bup/Nx-Detox, MEI and Incentives. The results may not generalize to other CTN trials or other interventions, which have different inclusion criteria.

In the context of these limitations, findings from this study provide a first glimpse into differences between participants of SUD treatment RCTs and target patient populations in usual care settings based on direct comparisons of these groups. The results support past research that compared the exclusion criteria of RCTs with characteristics of target populations [5], indicating that these RCTs are highly selective and do not represent the target populations adequately.

In order to ensure generalizability of RCT findings to relevant populations.

Figure 1 Density plots of propensity scores in 10 National Institute of Drug Use Clinical Trial Network (CTN) randomized controlled trials (RCTs) and target samples from the Treatment Episodes Data-Admission (TEDS-A).
target populations, future studies should examine the implications of these differences. Some differences between RCT and target populations may arise from the strict eligibility criteria for RCTs, leading to exclusion of many potential participants. Some of these eligibility criteria may be necessary and justifiable from a patient safety perspective, such as pregnancy and medication allergies. However, stringent eligibility criteria may exclude individuals who are less responsive to treatments, leading potentially to an overestimation of the effectiveness of the interventions. For example, comorbid mental disorders may be associated with lower educational attainment which, in turn, is associated with a lower trial retention rate [50] and higher prevalence of substance use disorders [51].

Representativeness of the trial samples in future CTN trials should be considered carefully, particularly because the primary mission of the CTN was to improve the nation-wide quality of drug abuse treatment. The findings also have implications for other trial networks, such as the National Cancer Institute (NCI) Clinical Trial Network [52]. The movement towards ‘practical clinical trials’ has produced important insights regarding the real-world effectiveness of psychiatric medications [53]. These trials generally had less stringent exclusion criteria. A similar move towards less stringent exclusion criteria in SUD treatment trials might also improve generalizability of these RCTs. The growing demand for comparative effectiveness data on SUD treatments from policymakers and program developers may motivate future moves towards more representative samples in SUD treatment RCTs.

Declaration of interests

R.S. and C.E. have nothing to disclose. R.M.C., E.A.S. and R.M. report grants from National Institute on Drug Abuse and National Institute of Mental Health during the conduct of the study. R.M. has received research funding and consulting fees from Bristol-Myers Squibb and Lundbeck Pharmaceuticals.

Acknowledgements

This work was supported by grant R01 DA036520 from the National Institute on Drug Abuse (NIDA). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References


Supporting information

Additional supporting information may be found in the online version of this article at the publisher’s web-site.

**Table S1** Description of study aims and target population for 10 National Institute of Drug Use Clinical Trial Network (CTN) randomized controlled trials (RCTs).

**Table S2** Number and percentage of cases with missing values for each covariate in the 10 CTN RCT studies and specified target population.