Toward National Estimates of Effectiveness of Treatment for Substance Use

Carlos Blanco, MD, PhD[a]; Aimee N. Campbell, PhD[b]; Melanie M. Wall, PhD[b,c]; Mark Olfson, MD, MPH[b,*]; Shuai Wang, PhD[b]; and Edward V. Nunes, MD[b]

ABSTRACT

Objective: To estimate how results would have varied if a substance abuse clinical trial had been conducted with nationally representative adults with substance use and with representative adults receiving substance use treatment.

Methods: Results were analyzed from a multisite clinical trial comparing the effectiveness of the Therapeutic Education System to treatment as usual for outpatient addiction treatment (n = 507). Patients were recruited between June 2010 and August 2011. Abstinence was the primary outcome. The general population sample and general population–treated samples were derived from Wave 1 of the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) (n = 43,093). Propensity scores provided a standardized measure of the difference between clinical trial participants and the 2 NESARC samples. The clinical trial was reanalyzed by reweighting the sample with propensity scores derived from the 2 samples to obtain generalizable estimates of treatment effects.

Results: Before the clinical trial sample was reweighted, the odds ratio (OR) of response to Therapeutic Education System versus treatment as usual in the trial was 1.62 (95% CI, 1.12–2.35). After the sample was reweighted to be representative of the 2 NESARC groups, ORs were 1.33 (95% CI, 0.34–5.26) for the representative sample with any substance use and 1.64 (95% CI, 0.82–3.27) for the representative treated sample.

Conclusions: Applying propensity score weighting to clinical trial results provides a method for estimating the population generalizability of clinical trial findings that relies on effect moderators observed in the study sample and population. Broader confidence intervals in the reweighted samples do not necessarily indicate lack of efficacy of the Therapeutic Education System but rather greater uncertainty concerning effectiveness in general population samples.

J Clin Psychiatry 2017;78(1):e64–e70
https://doi.org/10.4088/JCP.15m10333

[a]National Institute on Drug Abuse, Rockville, Maryland
[b]Department of Psychiatry, New York State Psychiatric Institute/Columbia University, New York, New York
[c]Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, New York
**Corresponding author:* Mark Olfson, MD, MPH, Department of Psychiatry, New York State Psychiatric Institute/Columbia University, 1051 Riverside Dr, Unit 24, New York, NY 10032 (mo49@cumc.columbia.edu).

The separation of clinical research from practice raises concerns over whether research results can truly inform practice.[1] Because most efficacy studies use stringent selection criteria, study participants are relatively homogenous.[2] When study participants substantially differ from target populations with the disorder, however, trial-based estimates of treatment effectiveness may not directly translate into clinical practice.[3] Beyond documenting the extent to which trial participants represent target populations, clinical policy makers and health planners want to know the likely effectiveness of experimental treatments in target populations. We illustrate a new method to estimate from clinical trial results the effectiveness of interventions in target populations.[4–6]

Several studies[4,7,8] suggest that clinical trials in psychiatry, which by design exclude 60%–85% of individuals with the target disorder, have limited generalizability. Given concerns over the representativeness of trial participants in clinical psychiatric research, we selected a recent randomized controlled trial[9,10] of a behavioral intervention for substance use disorders to illustrate the estimation of treatment effectiveness in target populations. This 10-site clinical trial evaluated the effectiveness of the Therapeutic Education System, a 12-week Web-based behavioral intervention that includes motivational incentives, for adults with substance use disorders. The study compared the outcomes of subjects assigned to either the Therapeutic Education System and treatment as usual (n = 255) or treatment as usual alone (n = 252).

To assess the generalizability of the results, we drew on a nationally representative epidemiologic study[11] of psychiatric and substance use disorders and applied propensity score methods to participants from the Therapeutic Education System trial.[9,10] We first evaluated differences between the clinical trial participants and the nationally representative target population with substance users and a subgroup who sought treatment. We then applied propensity scores to estimate how the clinical trial results would have varied had the study been conducted in this nationally representative sample. The effect size estimates using the nationally representative sample were then compared with the effect size estimate in the clinical trial sample. The generalizability of this approach relies on the extent to which it is possible to measure and adjust for the intervention effect moderators in the study sample and population sample.

METHODS

Clinical Trial Sample

Treatment as usual included a minimum of 2 hours of face-to-face therapeutic group or individual sessions per week. The Therapeutic Education System consists of 62 computer-interactive, multimedia modules delivered at the clinic sites, covering skills

Clinical Points

- Clinical trials may not be representative of individuals with the target disorder.
- Reweighting clinical trials to make them more representative provides a better estimate of treatment effects that are expected in clinical practice

for achieving and maintaining abstinence, and prize-based motivational incentives contingent on abstinence and treatment adherence. Patients, which were recruited between June 2010 and August 2011, were eligible if they (1) were 18 years or older, (2) had been using illicit substances in the 30 days prior to baseline (or 60 days if the patient was exiting a controlled environment) to exclude participants with alcohol use disorders only, (3) were within 30 days of entering the treatment program, and (4) were planning to remain in the area and treatment program for ≥3 months. Patients were excluded if they were (1) prescribed opioid replacement therapy or (2) unable to provide informed consent.[9,10] The study was approved by the institutional review boards of all the participating sites. The primary outcome was abstinence from drugs and drinking as measured by weekly urine drug screens and self-reports.[9,10] Generalized estimating equations were utilized to adjust for the correlation of half weeks within patients.

General Population Sample

We used as the general population sample Wave 1 of the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). The NESARC is a nationally representative sample of the adult population of the United States conducted by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) that has been described in detail elsewhere.[12–15] The target population was the civilian noninstitutionalized population, 18 years and older, residing in households and group quarters in the United States. Face-to-face interviews were conducted with 43,093 respondents. The survey response rate was 81%. Blacks, Hispanics, and young adults (ages 18–24 years) were oversampled, with data adjusted for oversampling and nonresponse. The weighted data were then adjusted to represent the US civilian population based on the 2000 census. *DSM-IV* diagnoses were assessed with the Alcohol Use Disorder and Associated Disabilities Interview Schedule–*DSM-IV* Version (AUDADIS-IV),[16] a fully structured diagnostic interview for nonclinician interviewers. The high reliability and validity of the AUDADIS substance use disorder diagnoses (κ = 0.70–0.94) have been demonstrated in numerous clinical and general population studies[17–22] in the United States and abroad. The NESARC has been previously used by our group and others to estimate the a priori generalizability of clinical trials of several psychiatric disorders.[4–6,23–29] The NESARC research protocol received approval from the institutional review boards of the US Office of Management and Budget and the US Census Bureau.

Statistical Analyses

In the analysis of clinical trials, each participant is generally considered equally important in estimating the efficacy of the intervention and given a weight of 1. However, because participants in clinical trials may not represent the target population with the disorder, it is informative to reweight the clinical trial sample to better approximate the distribution of demographic and clinical characteristics of the target population.[30–32] Thus, we proceeded in 2 steps. First, we reweighted the sample and then we repeated the original analyses using this reweighted sample.

One way to achieve this first step, which involves reweighting the clinical sample, is through the use of propensity scores.[33] Specifically, the propensity score is the probability of membership in a particular target population for each individual in the clinical sample as a function of his or her baseline demographic and clinical characteristics. In this case, the propensity score provides an estimate of probability that participants in the clinical trial would have been randomly selected from a representative sample of the target population. We focused our first analyses on individuals with substance use and our second analysis on those with substance use who had sought treatment in the previous year.[34,35]

Separate logistic regressions (one for each of these 2 target populations) were used to obtain the propensity scores that combined all individuals from each of the NESARC target populations with all individuals from the clinical trial. The outcome was set to 0 if the individual was from the clinical trial and 1 if the individual was from the NESARC target population. Predictors included all demographic and clinical characteristics available in the NESARC and clinical trial data sets (Table 1). All statistically significant ($P < .05$) 2-way interactions were also included. Sampling weights from the NESARC were used, and weights for the individuals from the clinical trial members were fixed at 1 in the logistic regression. The analyses were conducted with SUDAAN 11 (RTI International, Research Triangle Park, North Carolina) to take into account the complex design of the NESARC.

Once the logistic regression models were fit for each target population, a propensity score was calculated for each individual in the clinical trial corresponding to the predicted probability of being in each target population. The inverse of the propensity score was used as a weight to rescale the clinical sample. This rescaling resulted in multivariate distributions of demographic and clinical characteristics that were similar to each of the target populations. We normalized the propensity score weights so that the sum of the weights would be identical to the sample size of the clinical trial.

In the second step of our approach, we replicated the original analyses of the clinical trial by reweighting the clinical sample. In this analysis, the propensity score weights applied to the clinical trial data and the analytic model together provide an estimate of the effectiveness of the treatment in the target population. Specifically, a longitudinal logistic regression model was used to obtain the odds ratio (OR) and 95% confidence interval (CI) of abstinence in the

Table 1. Background Characteristics of Original Clinical Trials Network (CTN) Sample, NESARC Treatment-Seeking Substance Users, and Reweighted CTN Sample

Characteristic	Original CTN Sample (N = 507)		NESARC Treatment-Seeking Substance Users (n = 183)		Original CTN Sample vs NESARC Treatment-Seeking Substance Users, *P* Value[a]	CTN Reweighted to Approximate NESARC Treatment-Seeking Substance Users (N = 507)[b]		Original CTN Sample vs CTN Sample Reweighted to Approximate NESARC Treatment-Seeking Substance Users, *P* Value[a]
	Mean	SD	Mean	SD		Weighted Mean	Weighted SD	
Age, y	34.90	10.90	35.52	11.75	.52	36.25	10.47	.46
	n	%	n	%		Weighted n	Weighted %	
Female	192	37.90	75	39.29	.75	172.00	34.14	.21
Race					<.0001			.80
White	284	56.00	133	73.65		372.59	74.07	
Black/African American	116	22.90	33	14.52		79.61	15.83	
American Indian/Alaska Native	3	0.60	5	4.54		11.49	2.28	
Asian	13	2.60	1	0.76		3.19	0.63	
Native Hawaiian/Pacific Islander	12	2.40	2	1.64		12.92	2.57	
Multiracial	54	10.70	9	4.89		23.20	4.61	
Other	23	4.50						
Hispanic/Latino	55	10.80	34	14.10	.24	43.08	8.56	.03
Education					<.0001			.77
<High school degree	118	23.30	42	24.17		109.48	21.76	
High school degree/GED	310	61.10	59	30.44		152.72	30.36	
>High school degree	79	15.60	82	45.39		240.81	47.87	
Marital status					<.0001			.77
Single/never married	308	60.70	75	37.79		176.21	35.03	
Married/remarried	72	14.20	53	36.12		184.49	36.68	
Separated/divorced/widowed	127	25.00	55	26.09		142.30	28.29	
Underemployed	190	37.50	80	43.22	.17	226.82	45.09	.66
Substance dependence								
Alcohol	224	44.20	85	46.10	.66	274.29	54.53	.05
Cocaine	177	34.90	19	10.10	<.0001	56.49	11.23	.67
Stimulants	100	19.70	12	7.04	<.0001	39.82	7.92	.70
Cannabis	146	28.80	23	12.50	<.0001	60.67	12.06	.88
Opiates	158	31.20	19	10.46	<.0001	54.81	10.90	.87
Other	41	8.10	23	12.28	.09	57.90	11.51	.78

[a]Based on χ^2 and *t* tests as appropriate. [b]Weighting was done using propensity score weighting (see text for details).
Abbreviations: GED = general education development, NESARC = National Epidemiologic Survey on Alcohol and Related Conditions.

treatment versus control group. We obtained 2 summary measures of the difference for the clinical trial sample and the target NESARC sample: (1) the standardized difference (ie, the mean difference divided by the overall standard deviation [SD]) between the propensity scores between the clinical trial sample and the target populations[36] and (2) the overlap of the distributions of propensity scores of the clinical trial sample and the target populations.[37] The first measure provides an estimate of the mean difference in the values of the propensity scores in the clinical sample versus the general population sample, whereas the second measure is more focused on the overall distributions and may be less sensitive to the effect of extreme values.

RESULTS

Before the propensity score weights were applied, individuals in the clinical trial tended to be older and were less likely to be white and to have lower educational attainment than the NESARC sample of all individuals with substance use regardless of treatment-seeking behavior. They were more likely to be single, underemployed, and dependent on all substances (data not shown). When the sample was narrowed to those seeking treatment in the past year, individuals in the clinical trial were less likely than treatment-seeking individuals in the NESARC to be white, to have achieved greater than a high school education, and to be married. They were more likely to be single and to be dependent on cocaine, stimulants, cannabis, and opiates (Table 1).

The standardized difference in propensity scores between the clinical trial sample and the nationally representative treatment-seeking sample was 1.4, whereas the difference with the nationally representative sample of substance users that included respondents without respect to past-year treatment was 2.1. The overlaps in the propensity score distributions between the clinical trial sample and the nationally representative sample of treatment seekers and substance users were 0.86 and 0.73, respectively (see Supplementary eFigures 1 and 2). After the propensity score weights were applied, each reweighted sample had a distribution that more closely resembled the target NESARC subsample (Table 2).

Prior to reweighting the sample, the OR of response to Therapeutic Education System versus treatment as usual was 1.62 (95% CI, 1.12–2.35), as previously reported.[9] Interactions between background characteristics and study group assignment on the abstinence outcome are presented in Supplementary eTable 1. After the sample was reweighted to be representative of those who had sought treatment in

Table 2. Background Characteristics of Original Clinical Trials Network (CTN) Sample, NESARC Substance Users, and Reweighted CTN Sample

Characteristic	Original CTN Sample (n = 507)		NESARC Substance Users (n = 2,461)		Original CTN Sample vs NESARC Substance Users, *P* Value[b]	CTN Reweighted to Approximate NESARC Substance Users (n = 507)[c]		Original CTN Sample vs CTN Sample Reweighted to Approximate NESARC Substance Users, *P* Value[b]
	Mean	SD	Mean	SD		Weighted Mean	Weighted SD	
Age, y	34.90	10.90	33.23	13.39	.0027	33.38	10.79	.78
	n	%	n	%[a]		Weighted n	Weighted %	
Female	192	37.90	1,095	39.91	.41	265.00	52.59	< .0001
Race					< .0001			< .0001
White	284	56.00	1,845	80.27		366.92	72.95	
Black/African American	116	22.90	409	10.89		115.06	22.87	
American Indian/Alaska Native	3	0.60	34	1.37		5.54	1.10	
Asian	13	2.60	45	2.36		2.41	0.48	
Native Hawaiian/Pacific Islander	12	2.40	18	0.71		1.35	0.27	
Multiracial	54	10.70	110	4.40		11.73	2.33	
Other	23	4.50						
Hispanic/Latino	55	10.80	420	9.71	.43	13.55	2.69	< .0001
Education					< .0001			< .0001
< High school degree	118	23.30	413	15.58		162.35	32.28	
High school degree	310	61.10	701	29.34		151.30	30.08	
> High school degree/GED	79	15.60	1,347	55.08		189.35	37.64	
Marital status					< .0001			.0013
Single/never married	308	60.70	1,158	45.77		194.84	38.74	
Married/remarried	72	14.20	783	38.36		236.59	47.04	
Separated/divorced/widowed	127	25.00	520	15.88		71.57	14.23	
Underemployed	190	37.50	793	31.36	.01	129.66	25.78	.01
Substance dependence								
Alcohol	224	44.20	538	24.46	< .0001	101.81	20.24	.04
Cocaine	177	34.90	49	2.17	< .0001	13.80	2.74	.43
Stimulants	100	19.70	27	1.13	< .0001	5.10	1.01	.82
Cannabis	146	28.80	133	5.23	< .0001	19.13	3.80	.18
Opiates	158	31.20	41	1.83	< .0001	7.42	1.48	.58
Other	41	8.10	49	1.76	< .0001	9.06	1.80	.95

[a]Percentages of the NESARC sample reflect design weights. [b]Based on χ^2 and *t* tests as appropriate. [c]Weighting was done using propensity score weighting (see text for details).
Abbreviations: GED = general education development, NESARC = National Epidemiologic Survey on Alcohol and Related Conditions.

Table 3. Treatment Effect for Abstinence of Web-Based Therapeutic Education System Versus Treatment as Usual in the Clinical Trials Network (CTN) and in NESARC Target Generalizability Samples

Sample	OR	95% CI	*P*
Original CTN sample	1.62	1.12–2.35	.01
CTN reweighted to approximate NESARC treatment-seeking substance users	1.64	0.82–3.27	.16
CTN reweighted to approximate NESARC substance users	1.33	0.34–5.26	.68

Abbreviations: NESARC = National Epidemiologic Survey on Alcohol and Related Conditions, OR = odds ratio.

the previous year, the OR was 1.64 (95% CI, 0.82–3.27). The corresponding ORs obtained after reweighting the sample to be representative of individuals with substance use without respect to past year treatment was 1.33 (95% CI, 0.34–5.26) (Table 3).

DISCUSSION

Concerns over the representativeness of clinical trials raise questions about the generalizability of their results to populations of clinical and policy interest. In this study, we used propensity scores to obtain a standardized measure of the difference between participants in a recent clinical trial and a nationally representative sample of individuals who had used illicit substances and sought treatment in the prior year. The clinical trial sample was reweighted with propensity score weights to make the distribution of baseline characteristics resemble the nationally representative sample and then reanalyzed using those weights to obtain generalizable estimates of the treatment effects. The point estimate of this reweighted treatment-seeking sample was very similar to the estimate of effect derived from the original (ie, unweighted) clinical trial data, although its confidence interval was broader. The estimates of efficacy for the intervention were substantially lower when the clinical trial sample was reweighted by a national representative sample of individuals with substance use regardless of treatment-seeking behavior.

To our knowledge, this study is the first to use propensity scores to reweigh the results of a clinical trial for the treatment of a mental disorder. In accord with previous work,[4,7,8] the composition of the clinical trial participants prior to reweighting differed from the community target populations. As a result, the standardized differences between the clinical trial sample and both nationally representative samples were greater than the recommended upper limit (0.25–0.50 standard deviations) for observational studies,[38]

although they were closer to the 0.73 standard deviations found in the only other study[36] that, to our knowledge, has used similar methods. Furthermore, the overlap between the propensity score distribution was 0.73 for the substance use sample, ie, close to the recommended 0.80, and 0.86 for the treatment-seeking sample. Because of the larger standardized differences between the clinical trial and community substance use samples than between the clinical trial and community treatment-seeking samples, balancing covariates was difficult, even after reweighting. Our findings suggest that randomized trials, which have relatively small sample sizes and rely on the participation of volunteers, face challenges recruiting representative samples. Reweighting may partially but not fully compensate for incomplete representativeness of clinical trial samples. As more investigators apply propensity scores to examine the generalizability of clinical trials, it may be possible to calibrate the range of standardized differences between clinical samples and target populations and to examine the impact of these differences on the generalizability of study results.

Differences in the composition of the clinical trial sample and the target populations support the need to estimate the effectiveness of interventions in various target populations. The use of propensity scores to reweight clinical trial samples offers a new approach to obtain these estimates. In our study, the effectiveness of the intervention varied with the target population. The point estimate of the nationally representative treatment-seeking sample was 1.64, very close to the estimate in the unweighted clinical trial sample, suggesting that, in this case, the clinical trial sample provided a reasonable estimate of effectiveness of the intervention in the target populations of interest at the national level (ie, treatment-seeking individuals who used illicit drugs). By contrast, the point estimate in the nationally representative sample that included individuals without respect to recent treatment-seeking behavior was 1.33, an almost 20% difference compared to the treatment-seeking sample. The variation in estimates is in accord with prior work demonstrating that results of clinical trials are highly sensitive to their inclusion and exclusion criteria and that the effectiveness of an intervention depends on the target population.[39–41] Our findings illustrate the importance of carefully selecting the eligibility criteria when planning clinical trials and of defining the target population of interest.

The effect size of the original Therapeutic Education System study, 1.62 (95% CI, 1.12–2.35), was statistically significant; the estimated effect size for the nationally representative treatment-seeking sample, 1.64 (95% CI, 0.82–3.27), was not. The wider confidence interval of the nationally representative treatment-seeking sample is the result of the variance inflation generated by the differences between the clinical trial sample and the nationally representative sample and the need to apply propensity score weights, particularly the larger weights, to recalibrate the clinical trial sample. The wider confidence intervals in the nationally representative sample do not necessarily mean that Therapeutic Education System is not efficacious for the treatment of substance use disorders. Instead, the width of the confidence intervals reflects increased uncertainty associated with extrapolating the results of the clinical trial sample to the broader populations. This increase in uncertainty may help to explain variations in effectiveness experienced by clinicians who apply clinical interventions to patients from populations that are more heterogeneous (ie, have greater variability in their treatment response) than those in whom the intervention was originally tested. An increase in this variability and uncertainty associated with differences in the composition of the study populations may contribute to challenges in reproducing clinical and basic research findings.[42] By minimizing the variance inflation associated with reweighting the study sample, recruitment of more representative samples may help narrow confidence intervals of the estimates of effectiveness of the intervention in the target population. More detailed descriptions of clinical trial participants might also help facilitate the reweighting procedure and narrow confidence intervals of reweighted samples.[43] A complementary approach to narrowing confidence intervals would involve combining and reweighting to the same nationally representative sample several studies that test the same intervention. The resulting estimates could then be jointly examined by using meta-analytic techniques to assess more precisely the effectiveness of the intervention. It would even be possible to adapt approaches to interim analysis of clinical trials to determine when interventions have accumulated sufficient evidence to be considered effective at the population level.[44–46]

Our study should be understood in the context of several limitations. First, we recalibrated the clinical trial sample using propensity score weights. However, reweighting will yield only unbiased estimates of treatment effects if all of the treatment effect modifiers are adjusted for in the analysis. If unmeasured variables, such as motivation to participate in treatment, moderate treatment effects and are related to selection into the trial, the reweighted estimates may still be biased. Second, the assessments of substance use in the NESARC are based on self-report and were not confirmed with biological testing or collateral information. Third, our estimates assume that the intervention would be conducted under identical conditions as those in the clinical trial. Variation in clinical settings, treatment intensity, or other system-level variation could influence applied treatment effectiveness. Fourth, some of the study eligibility criteria were not available in the NESARC (ie, being on opioid replacement therapy and a willingness to provide consent to participate in the clinical trial) or had to be estimated using a different timeframe (use of illicit substances and treatment seeking was assessed in the NESARC in the last 12 months rather than in the last 30 days as in the clinical trial).

Despite these limitations, our study exemplifies a novel approach for estimating the population generalizability of clinical trial results. This method is flexible and may find applications in a range of disorders and target populations. We hope that it helps to refine clinical trial methods, improve estimation of population-level effect of interventions, and advance personalized and precision medicine.

Submitted: August 18, 2015; accepted March 9, 2016.

Author contributions: Dr Wang had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Drs Blanco, Campbell, Olfson, and Wall designed the study. Drs Wall and Wang and conducted the data analyses. Dr Blanco wrote the initial draft of the manuscript. All authors contributed to manuscript revision and approved the final version.

Potential conflicts of interest: **Dr Nunes** has received medication for research studies from Alkermes/Cephalon, Duramed, and Reckitt-Benckiser. **Drs Blanco**, **Campbell**, **Wall**, **Olfson**, and **Wang** declare no conflicts of interest.

Funding/support: Work on this article was supported by National Institute of Health grants DA023200, MH0760551, and MH082773 (Dr Blanco), and the New York State Psychiatric Institute (Drs Blanco, Nunes, Olfson, and Wall).

Role of the sponsor: The sponsors had no additional role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Disclaimer: The views and opinions expressed in this report are those of the authors and should not be construed to represent the views of any of the sponsoring organizations, agencies, or the US government.

Additional information: The original data for the NESARC are available from the NIAAA (http://www.niaaa.nih.gov). Data from NESARC are available for downloading on the NESARC website at http://niaaa.census.gov.

Supplementary material: See accompanying pages.

REFERENCES

1. Califf RM, Platt R. Embedding cardiovascular research into practice. *JAMA*. 2013;310(19):2037–2038.
2. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA*. 2003;290(12):1624–1632.
3. Glasgow RE, Magid DJ, Beck A, et al. Practical clinical trials for translating research to practice: design and measurement recommendations. *Med Care*. 2005;43(6):551–557.
4. Blanco C, Olfson M, Goodwin RD, et al. Generalizability of clinical trial results for major depression to community samples: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *J Clin Psychiatry*. 2008;69(8):1276–1280.
5. Blanco C, Olfson M, Okuda M, et al. Generalizability of clinical trials for alcohol dependence to community samples. *Drug Alcohol Depend*. 2008;98(1–2):123–128.
6. Okuda M, Hasin DS, Olfson M, et al. Generalizability of clinical trials for cannabis dependence to community samples. *Drug Alcohol Depend*. 2010;111(1–2):177–181.
7. Zimmerman M, Chelminski I, Posternak MA. Generalizability of antidepressant efficacy trials: differences between depressed psychiatric outpatients who would or would not qualify for an efficacy trial. *Am J Psychiatry*. 2005;162(7):1370–1372.
8. Westen D, Morrison K. A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: an empirical examination of the status of empirically supported therapies. *J Consult Clin Psychol*. 2001;69(6):875–899.
9. Campbell AN, Nunes EV, Matthews AG, et al. Internet-delivered treatment for substance abuse: a multisite randomized controlled trial. *Am J Psychiatry*. 2014;171(6):683–690.
10. Campbell AN, Nunes EV, Miele GM, et al. Design and methodological considerations of an effectiveness trial of a computer-assisted intervention: an example from the NIDA Clinical Trials Network. *Contemp Clin Trials*. 2012;33(2):386–395.
11. Grant BF, Stinson FS, Dawson DA, et al. Prevalence and co-occurrence of substance use disorders and independent mood and anxiety disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Arch Gen Psychiatry*. 2004;61(8):807–816.
12. Grant BF, Dawson DA, Stinson FS, et al. The Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-IV): reliability of alcohol consumption, tobacco use, family history of depression and psychiatric diagnostic modules in a general population sample. *Drug Alcohol Depend*. 2003;71(1):7–16.
13. Grant BF, Moore T, Shepard J, et al. *Source and Accuracy Statement: Wave 1 National Epidemiologic Survey on Alcohol and Related Conditions (NESARC)*. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism; 2003.
14. Grant BF, Dawson DA, Hasin DS. *The Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions Alcohol Use Disorder and Associated Disabilities Interview Schedule-DMS-IV*. Bethesda, MD: National institute on Alcohol Abuse and Alcoholism; 2004.
15. Grant BF, Stinson FS, Dawson DA, et al. Co-occurrence of 12-month alcohol and drug use disorders and personality disorders in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Arch Gen Psychiatry*. 2004;61(4):361–368.
16. Grant BF, Dawson DA, Hasin DS. *The Alcohol Use Disorder and Associated Disabilities Interview Schedule–DSM-IV Version, National Institute on Alcohol Abuse and Alcoholism*. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism; 2001.
17. Canino G, Bravo M, Ramírez R, et al. The Spanish Alcohol Use Disorder and Associated Disabilities Interview Schedule (AUDADIS): reliability and concordance with clinical diagnoses in a Hispanic population. *J Stud Alcohol*. 1999;60(6):790–799.
18. Chatterji S, Saunders JB, Vrasti R, et al. Reliability of the alcohol and drug modules of the Alcohol Use Disorder and Associated Disabilities Interview Schedule—Alcohol/Drug-Revised (AUDADIS-ADR): an international comparison. *Drug Alcohol Depend*. 1997;47(3):171–185.
19. Cottler LB, Grant BF, Blaine J, et al. Concordance of *DSM-IV* alcohol and drug use disorder criteria and diagnoses as measured by AUDADIS-ADR, CIDI and SCAN. *Drug Alcohol Depend*. 1997;47(3):195–205.
20. Grant BF, Harford TC, Dawson DA, et al. The Alcohol Use Disorder and Associated Disabilities Interview schedule (AUDADIS): reliability of alcohol and drug modules in a general population sample. *Drug Alcohol Depend*. 1995;39(1):37–44.
21. Hasin D, Carpenter KM, McCloud S, et al. The alcohol use disorder and associated disabilities interview schedule (AUDADIS): reliability of alcohol and drug modules in a clinical sample. *Drug Alcohol Depend*. 1997;44(2–3):133–141.
22. Ustün B, Compton W, Mager D, et al. WHO Study on the reliability and validity of the alcohol and drug use disorder instruments: overview of methods and results. *Drug Alcohol Depend*. 1997;47(3):161–169.
23. Hoertel N, Le Strat Y, Blanco C, et al. Generalizability of clinical trial results for generalized anxiety disorder to community samples. *Depress Anxiety*. 2012;29(7):614–620.
24. Hoertel N, Le Strat Y, Lavaud P, et al. Generalizability of clinical trial results for bipolar disorder to community samples: findings from the National Epidemiologic Survey on Alcohol and Related Conditions. *J Clin Psychiatry*. 2013;74(3):265–270.
25. Hoertel N, Falissard B, Humphreys K, et al. Do clinical trials of treatment of alcohol dependence adequately enroll participants with co-occurring independent mood and anxiety disorders? an analysis of data from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). *J Clin Psychiatry*. 2014;75(3):231–237.
26. Hoertel N, Le Strat Y, De Maricourt P, et al. Are subjects in treatment trials of panic disorder representative of patients in routine clinical practice? results from a national sample. *J Affect Disord*. 2013;146(3):383–389.
27. Hoertel N, López S, Wang S, et al. Generalizability of pharmacological and psychotherapy clinical trial results for borderline personality disorder to community samples. *Pers Disord*. 2015;6(1):81–87.
28. Hoertel N, Le Strat Y, Limosin F, et al. Prevalence of subthreshold hypomania and impact on internal validity of RCTs for major depressive disorder: results from a national epidemiological sample. *PLoS One*. 2013;8(2):e55448.
29. Hoertel N, de Maricourt P, Katz J, et al. Are participants in pharmacological and psychotherapy treatment trials for social anxiety disorder representative of patients in real-life settings? *J Clin Psychopharmacol*. 2014;34(6):697–703.
30. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: the ACTG 320 trial. *Am J Epidemiol*. 2010;172(1):107–115.
31. Stuart EA, Bradshaw CP, Leaf PJ. Assessing the generalizability of randomized trial results to target populations. *Prev Sci*. 2015;16(3):475–485.
32. Hartman E, Grieve R, Ramsahai R, et al. From SATE to PATT: combining experimental with observational studies to estimate population treatment effects. *J R Stat Soc Ser A Stat Soc*. 2013;10:1111.
33. D'Agostino RB Jr. Propensity scores in cardiovascular research. *Circulation*. 2007;115(17):2340–2343.
34. Blanco C, Iza M, Rodríguez-Fernández JM, et al. Probability and predictors of treatment-seeking for substance use disorders in the US. *Drug Alcohol Depend*. 2015;149(1):136–144.
35. Lopez-Quintero C, Pérez de los Cobos J, Hasin DS, et al. Probability and predictors of transition from first use to dependence on nicotine, alcohol, cannabis, and cocaine: results of the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). *Drug Alcohol Depend*. 2011;115(1–2):120–130.
36. Stuart EA, Cole SR, Bradshaw CP, et al. The use of propensity scores to assess the generalizability of results from randomized trials. *J R Stat Soc Ser A Stat Soc*. 2001;174(2):369–386.
37. Tipton E. How generalizable is your

experiment? an index for comparing experimental samples and populations. *J Educ Behav Stat*. 2014;39(6):478–501.
38. Rubin DB. Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Serv Outcomes Res Methodol*. 2001;2:169–188.
39. Khan A, Kolts RL, Thase ME, et al. Research design features and patient characteristics associated with the outcome of antidepressant clinical trials. *Am J Psychiatry*. 2004;161(11):2045–2049.
40. Wisniewski SR, Rush AJ, Nierenberg AA, et al. Can phase III trial results of antidepressant medications be generalized to clinical practice? a STAR*D report. *Am J Psychiatry*. 2009;166(5):599–607.
41. Khan A, Brodhead AE, Kolts RL, et al. Severity of depressive symptoms and response to antidepressants and placebo in antidepressant trials. *J Psychiatr Res*. 2005;39(2):145–150.
42. Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. *Nature*. 2014;505(7485):612–613.
43. Hoertel N, de Maricourt P, Gorwood P. Novel routes to bipolar disorder drug discovery. *Expert Opin Drug Discov*. 2013;8(8):907–918.
44. Lan KKG, Demets DL. Discrete sequential boundaries for clinical-trials. *Biometrika*. 1983;70(3):659–663.
45. Pogue JM, Yusuf S. Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. *Control Clin Trials*. 1997;18(6):580–593, discussion 661–666.
46. Blanco C, Schneier FR, Schmidt A, et al. Pharmacological treatment of social anxiety disorder: a meta-analysis. *Depress Anxiety*. 2003;18(1):29–40.

Supplementary material follows this article.

THE JOURNAL OF
CLINICAL PSYCHIATRY
THE OFFICIAL JOURNAL OF THE AMERICAN SOCIETY OF CLINICAL PSYCHOPHARMACOLOGY

Supplementary Material

Article Title: Toward National Estimates of Effectiveness of Treatment for Substance Use

Authors: Carlos Blanco, MD, PhD; Aimee N. Campbell, PhD; Melanie M. Wall, PhD; Mark Olfson, MD, MPH; Shuai Wang, PhD; and Edward V. Nunes, MD

DOI Number: 10.4088/JCP.15m10333

List of Supplementary Material for the article

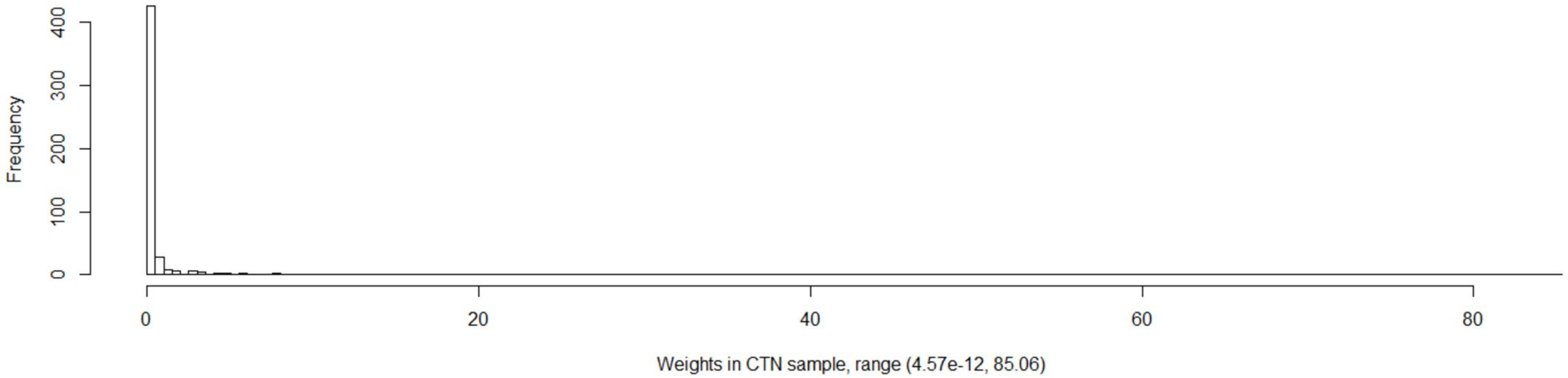
1. eTable 1 — Interactions of background characteristics with study treatment group effects on abstinence outcome
2. eFigure 1 — Comparison of distributions of weights of CTN subjects and NESARC respondents with substance use
3. eFigure 2 — Comparison of distribution of weights of CTN subjects and NESARC respondents with substance use treatment

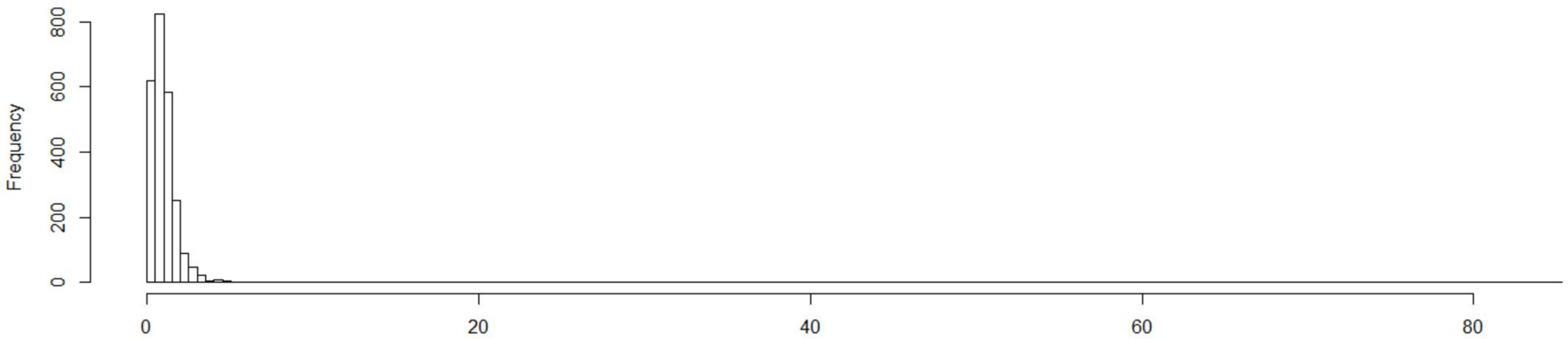
Disclaimer

This Supplementary Material has been provided by the author(s) as an enhancement to the published article. It has been approved by peer review; however, it has undergone neither editing nor formatting by in-house editorial staff. The material is presented in the manner supplied by the author.

Supplemental Table 1: Interactions of background characteristics with study treatment group effects on abstinence outcome		
Background Characteristics	OR (95% CI)	*p*
Age, year		0.65
18-35 (n=295)	1	
36-67 (n=212)	0.84 (0.39, 1.79)	
Sex		0.59
Male (n=314)	1	
Female (n=192)	0.81 (0.37, 1.75)	
Race/Ethnicity		0.72
White (n=284)	1	
Non-white (n=221)	0.87 (0.41, 1.84)	
Hispanic/Latino		0.50
Hispanic (n=55)	1.54 (0.43, 5.48)	
Non-Hispanic (n=451)	1	
Education		0.83
< HS Degree (n=118)	0.91 (0.37, 2.22)	
HS Degree + (n=389)	1	
Marital Status		0.12
Single (n=308)*	1	
Other (n=199)	1.80 (0.86, 3.76)	
Employment		0.69
Underemployed (n=190)	1.17 (0.54, 2.58)	
Employed (n=317)	1	
Alcohol dependence		0.48
Presence (n=224)	1.31 (0.62, 2.78)	
Absent (n=283)	1	
Cocaine dependence		0.15
Presence (n=177)	1.78 (0.82, 3.88)	
Absent (n=330)	1	
Stimulant dependence		0.25
Presence (n=100)	1.71 (0.68, 4.29)	
Absent (n=407)	1	
Cannabis dependence		0.30
Presence (n=146)	1.54 (0.67, 3.53)	
Absent (n=361)	1	
Opiate dependence		0.14
Presence (n=158)	0.54 (0.24, 1.23)	
Absent (n=349)	1	
Other dependence		0.94
Presence (n=41)	1.04 (0.33, 3.30)	
Absent (n=465)	1	
Data from CTN. Results based on the original CTN sample (n=508).		

Supplementary eFigure 1:
Comparison of distributions of weights of CTN subjects and NESARC respondents with substance use (Generalized Index Estimate=0.73)





Supplementary eFigure 2:
Comparison of distribution of weights of CTN subjects and NESARC respondents with substance use treatment (Generalized Index Estimate=0.86)