The relative impact of brief treatment versus brief intervention in primary health-care screening programs for substance use disorders

By: Arnie Aldridge, William Dowd, and Jeremy W. Bray

This is the peer reviewed version of the following article:

Aldridge, A., Dowd, W., Bray, J. (2017). The relative impact of brief treatment versus brief intervention in primary health-care screening programs for substance use disorders. *Addiction*, 112: 54– 64. doi: 10.1111/add.13653.

which has been published in final form at https://doi.org/10.1111/add.13653.

Abstract:

Aims: To assess the relative impact of brief treatment (BT) compared with brief intervention (BI) on changes in substance use behavior in primary care screening programs for substance use disorders, overall and by patient severity. **Design and participants:** A total of 9029 patients with both baseline and follow-up interviews were identified in the US Government Performance and Results Act (GPRA) data from October 2004 and February 2008. Using a propensity score framework, multiple generalized linear mixed models and a local linear matching method with a difference in difference estimator, patients from the BI group that resemble BT patients were used to determine the relative treatment effect of BT. A total of 3218 of these US patients with baseline and follow-up interviews were used in the final analysis sample after the propensity score-matching procedure (1448 patients assigned to a BI service category and 1770 assigned to a BT service category). **Setting:** United States. **Measurements:** Dependent variables were the number of days of use in the past 30 days of any alcohol, alcohol to intoxication, illicit drugs and marijuana. **Findings:** The relative impact of BT was not significant for alcohol (0.269; $P > 0.1$) or alcohol to intoxication (0.462; $P > 0.1$). BT was found to reduce the frequency of use of illicit drugs at follow-up by 0.634 days more than BI ($P < 0.05$). Marijuana days were not affected significantly by assignment to BT (−0.128; $P > 0.1$). Higher severity patients assigned to BT had a decrease in days of illicit drug use of 1.765 ($P < 0.05$). **Conclusions:** In the United States, brief treatment appears to have a stronger impact on reducing illicit drug use than brief intervention but is similar to brief intervention for reducing alcohol use, alcohol to intoxication and marijuana use alone.

Keywords: brief intervention | brief therapy | brief treatment | illicit drugs | propensity score | quasi-experimental | SBI | SBIRT

Article:

Introduction

Brief treatment (BT) is qualitatively different from brief intervention (BI), and is also more intensive in terms of quantity and length of therapy sessions. Although elements of brief treatment have been studied, there has been little study of its effectiveness when implemented as an additional standalone service helping to fill in gaps in the continuum of care for substance abuse. BT programs are designed to be less intensive than traditional therapy and often use a different clinical approach. For screening and early intervention programs delivered in health-care and other community settings, BT is intended to fill three roles in addressing patients with different levels of risky substance use. BT is a stronger 'dose' of service than BIs. BT offers a more flexible, accessible, less expensive and less stigmatized service than traditional therapy (to which patients might still be referred as they interact with the BT provider). BT can be used to support those patients who have been referred to specialty treatment [1, 2]. BT practitioners often have more clinical training than providers of BIs, which are typically delivered by non-specialists.

Clinically, BIs are generally designed to motivate a patient to think differently about their substance use and to ultimately change risky behaviors. BT is intended for higher levels of risk with stronger objectives for the patients, such as abstinence and managing appropriately factors that exacerbate their substance use. BT often uses cognitive–behavioral therapy (CBT) or motivational enhancement therapy (MET) as clinical models. The distinct clinical approach for BT and its potential value as an additional service option warrant a study of its comparative effectiveness relative to BI. Early literature suggests that BT is effective with a wide range of clients who abuse substances [1, 3]. Length, number of sessions and the use of more highly trained providers may all influence the impact of BT regardless of clinical distinctness. More importantly, they also influence how models of Screening, Brief Intervention and Referral to Treatment (SBIRT) might be implemented in different health-care settings and how much they cost. For example, in the United States, the current reimbursement codes for BT demarcate the rate for reimbursement between BI and BT by the 30-minute mark, with no reference to clinical differences.

In this study, we used data from a large US government-funded implementation of SBIRT in US-based health-care settings [4]. The programs were implemented in community health-care settings to provide SBIRT services and were not designed as research studies. Thus, they provide an opportunity to expand the body of clinical trial research on early interventions to comparative effectiveness studies of different interventions delivered to a large and diverse population [5, 6]. We hypothesized that patients assigned to BT within these SBIRT programs would have greater improvement in substance use outcomes than those assigned to BI. We adopted a propensity score framework commonly used in comparative effectiveness research to examine the intent-to-treat effect of BT on alcohol use, alcohol use to intoxication, illicit drug use and marijuana use relative to being recommended to BI [7, 8] and examined whether this effect varied by baseline substance use intensity.

Methods

Research design

The design is modeled upon regression discontinuity studies which identify quasi-experimental comparison groups using arbitrary cut-offs in continuous measures of eligibility into a 'treatment' group. In this study, we rely upon the fact that a considerable number of SBIRT patients who were recommended to BI services (henceforth referred to as BI patients) were very similar in their observed characteristics to those recommended to BT (henceforth referred to as BT patients). In support of this approach, we note measurement error in screening instruments as well as substantial variation in instruments, staff and other SBIRT program characteristics over time and across grantees and performance sites. These facts imply that some BI patients might have been recommended to BT had they been screened at a different time and place and are therefore eligible comparison candidates for BT patients. We use propensity score-matching methods to create the supports for the regression discontinuity.

The BI models for all projects incorporated an MI approach [9, 10], typically including the six elements of the 'FRAMES' model: feedback regarding personal risk or impairment, emphasis on personal responsibility for change, clear advice to change, providing a menu of change options, empathy as a counseling style, and enhancement of self-efficacy or optimism (FRAMES) [11]. Most of the projects also incorporated information and feedback based on National Institute on Alcohol Abuse and Alcoholism (NIAAA) guidelines [12]. In their BT models, all seven projects reported the use of MI techniques and the Transtheoretical Model of Change [13]. Two projects used cognitive behavioral therapy (CBT) and two used a community reinforcement approach (CRA). More than 75% of practitioners providing BT services were currently or previously certified in addiction treatment (compared to approximately 50% of those providing BI services). Sixty-seven per cent had graduate-level degrees (compared to 34% for BI providers). All seven sites' BT protocol were for at least two sessions and BT session lengths were typically 50 minutes. To improve BT attendance rates, most grantees offered onsite or telephonic services when feasible at the clinical site, especially for the first BT session.

Data

Our data are drawn from SBIRT projects that were implemented by seven US states. Funding to implement SBIRT services in out-patient, emergency department and in-patient settings was provided to the states in 2005 by the Substance Abuse and Mental Health Services Administration (SAMHSA), the US federal agency concerned with assessing and improving the public health and treatment systems for substance abuse and mental health. A stipulation of the funding was that patient data were collected on patients receiving SBIRT services at the time they received services and 6 months later. The standardized patient interview instrument and interview procedures were the same across all projects, in accordance with the federal Government Performance and Results Act (GPRA). SAMHSA's GPRA data coordinating center provided us with access to these patient data from which we constructed an analysis data set of patients with both baseline and follow-up interviews.

The GPRA instrument collected basic socio-demographic data, such as gender, age, race/ethnicity, income and educational attainment, as well as substance use, criminal justice status, employment status, housing status and measures of mental health and social connectedness, all during the past 30 days. However, these data were collected using a tiered structure under which all data were collected on BT and referral to treatment patients, but only a subset of the data—gender, age, race/ethnicity and frequency of substance use in the past 30 days—were collected for BI patients. Therefore, only the variables available on BI patients were utilized in this study.

The necessary data were collected by practitioners delivering SBIRT at the time of screening or immediately after delivery of services. After screening, patients were assigned to a level of service based on the practitioner's recommendation for that patient. This study uses the intent-to-treat categorization, collected in the GPRA survey, to identify BI and BT patients. In addition to collecting GPRA data on all patients at intake, the grantees collected data at discharge and at 6 months after intake. The discharge survey documented which services a patient received and the number of sessions; however, these data were not collected consistently. The 6-month follow-up survey was conducted with a 10% subsample of patients who screened positive and captured the same data elements collected at intake. The analyses in this study rely upon the longitudinal changes in substance use reported at baseline and 6-month follow-up for patients recommended to either BI or BT.

In general, SBIRT service recommendations increased in intensity with severity of patient substance involvement. GPRA data showed that once they screened positive, most patients were willing to continue SBIRT participation: BI patients, 86%; BT patients, 93%. Not surprisingly, willingness to continue with the program generally was lower among those in the BI group, because individuals at the lower end of the severity continuum often perceived less need for services. Illicit drug use by BT patients was almost double that for BI patients, and is a defining feature of the average difference between the two groups. BI was delivered immediately typically on-site, but referrals for more intensive care required typically that patients enroll in a program, often at a later time (following recovery from physical injury or illness) and in another location. In two types of cases, BI sessions were recorded for BT patients' initial service. First, due to time constraints in clinical settings, patients did not always complete the full-length BT session during their clinical visit. Based on the limited time, as well as the clinical content (e.g. MI used but not enough time to in-depth CBT) delivered in the short window, providers decided whether to count the service as a BI instead of a BT. In some project sites, BT provider staff were not available during the patient's visit. In this case, projects attempted to deliver a BI to the patients. Patients in both service recommendation groups received BIs (87% of BI patients, 51% of BT patients) and BT patients commonly received both a BI and a subsequent BT. Most BT patients (82%) participated in one BT session (in addition to any BI).

This study uses GPRA data from October 2004 to February 2008. From these data, we identified 12 886 patients who were eligible for follow-up interviews and who were identified as BI or BT patients according to the SBIRT class code. Of these, 9029 had completed 6-month follow-up data; 6983 were classified as BI patients, while the remaining 2046 were classified as BT patients according to class code, representing follow-up rates of 69 and 74%, respectively. In our final analysis sample, we excluded 1141 BI patients and 257 BT patients who reported no

substance use (neither alcohol nor illicit drugs) at both baseline and follow-up. We were unable to determine whether there were alternative reasons for why they screened positive or if their substance use was recorded incorrectly. However, we have anecdotal evidence that practitioners screened patients as positive based on: past substance use history independent of current use, particularly in the case of current cravings or precipitating life events that might lead to relapse; tobacco use; and other mental health considerations. Patients received services in specific performance sites (health-care facilities) which are identified in these data. Across the seven state-level projects, there were 15, four, four, three, three and two performance sites.

Measures

Substance use status was measured as the number of days of use in the past 30, and was collected for alcohol, alcohol to intoxication, marijuana, cocaine, amphetamines, inhalants, sedatives, hallucinogens, opiates and other drugs taken illicitly. Our main outcomes were the number of days in the past 30 that patients reported any alcohol use, any alcohol to intoxication and any illicit drug use. We also examined the impact of BT relative to BI on days of marijuana use. For moderator analyses described below, we further constructed indicator variables of zero, low and high baseline substance use days. As noted, screener scores themselves were reported inconsistently and therefore we relied entirely upon the frequency of use variables to characterize risk. The threshold for low versus high was specific to each substance use variable and reflected the median days of use. High was 8 or more alcohol days, 10 or more alcohol intoxication days, 8 or more illicit drug days and 4 or more marijuana days.

Design

We assess the intent-to-treat effect of assignment to BT rather than BI on substance use using a quasi-experimental design known as regression discontinuity, in which BI patients who are similar to BT patients are used as potential comparisons. The problem with simply comparing BI and BT patients is that BT patients are expected to have a greater substance use risk or more severe substance use problems. Regression discontinuity designs rely typically upon the fact that patients are assigned to a ‘treatment’ group based on having a continuous measure or score that is above an arbitrary ‘cut-off’. Patients whose scores are immediately below or immediately above the cut-off are potentially comparable in characteristics other than being assigned to the ‘treatment’. In the case of SBIRT, assignment to BI or BT is based in large part on a substance use screening completed by SBIRT practitioners and other observed patient characteristics. However, several factors influence assignments that are not correlated perfectly with a patient's substance use or other patient characteristics, and thus provide experimental variation that justifies our design. The screening tools varied across grantees and over time within some grantees. The distribution of screening tools is shown in Table 1. The Alcohol Use Disorders Identification Test (AUDIT) and Drug Abuse Screening Test (DAST) were the most common screening instruments used for alcohol and drug use, respectively. Notably, many grantees began to use both validated and non-validated 1–3-question pre-screens.

Table 1. Unadjusted and propensity score weighted means comparison of baseline characteristics between BI and BT patients.

Alcohol		**Illicit drugs**	
Pre-screen			
AUDIT-C	3 Grantees	None	3.5 Grantees
None	2.5 Grantees	Single drug use question	3 Grantees
Single alcohol question	1 Grantee	Three drug use questions	0.5 Grantees
3 NIAAA questions	0.5 Grantees		
Full screen			
AUDIT[a]	5.5 Grantees	DAST-10[a]	3.5 Grantees
TCU screener	1 Grantee	Three drug use questions	1.5 Grantees
AUDIT-C	0.5 Grantee	Drug screen II	1 Grantee
		None (pre-screen only)	1 Grantee

Half grantees listed in the table due to one grantee using different instruments at different sites.
[a] One grantee switched during the SBIRT grant from these instruments to the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST) for both alcohol and drugs. BI = brief intervention; BT = brief treatment; SBIRT = Screening, Brief Intervention and Referral to Treatment; AUDIT-C = Alcohol Use Disorders Identification Test; TCU = Texas Christian University.

Even when using validated screening instruments, measurement error leads to inconsistent service recommendation for patients with similar risk levels. Confounding this naturally occurring variation are several other factors. First, practitioners varied over time and across grantees in several ways. Practitioners were hired based on different criteria across grantees including credentials, experience and behavioral health area of expertise and were different in terms of demographics, education and substance use history or perspective on prevention and treatment. Secondly, in several grantees, medical staff completed either the pre-screen or screen instead of staff with behavioral health backgrounds. Thirdly, staff turnover was cited as a major challenge while all staff evolved in their screening accuracy, as they received ongoing training and gained insight as a result of conducting the BI and BT portion of service delivery for patients with varying substance use risk. Finally, logistics or other constraints such as staff availability or the patient's expected time in the facility were also a factor in service recommendation and changed from day to day and across grantees. There were probably other unobserved factors that varied across sites, such as program culture and what clinical characteristics were emphasized.

Propensity score-matching

Unfortunately, we do not observe a unified measure of all the factors that lead to assignment to BI versus BT. Instead we derive propensity scores for patients which reflect their a priori likelihood of being assigned to BT rather than BI based on observable characteristics. We make the assumption that is motivated by the regression discontinuity design, that BI patients with similar propensity scores to BT patients are appropriate comparisons. The goal of this approach is to determine the potential outcome of an average BI patient had he or she been assigned to BT instead, and compare that outcome to the outcome that we observe—the outcome of an average BI patient who receives a BI. One could conceive of this method equivalently as determining the unobserved, counterfactual outcome of an average BT patient who received a BI and comparing it to the observed outcome of an average BT patient who received a BT. As we only observe an individual in one of the two groups, and we cannot actually subject a patient to an alternative group after the fact, we instead use propensity score-matching to diminish any observed

differences that exist between BI and BT patients prior to their assignment that may impact outcomes. Although we cannot be assured of comparability as in a randomized controlled environment, by controlling for observable factors that may impact upon the practitioner's decision to refer someone to BI or BT we can increase the comparability of our two groups and reduce the treatment-selection bias inherent in the design [14]. The propensity score, determined by probit regression of a treatment indicator on a set of observed covariates, is used to predict each person's probability of receiving treatment.

The key assumption of our approach is that patient recommendation to BI or BT must be independent of the potential outcomes of the patient with respect to observed covariates [15]. In other words, any differences between the control and treatment groups that would have an impact upon outcomes must be controlled for in the matching procedure. Propensity scores are determined for each individual using a probit model with a binary indicator for treatment as the dependent variable and a set of patient characteristics as covariates. The initial probit model has as covariates gender, age, race, ethnicity, state where exposed to SBIRT, ordinal time (e.g. half a calendar year), baseline substance use and interactions of state, time and substance use. We also included quadratic terms of age and days of substance use in the past 30 days.

As the probit model of treatment participation is atheoretical, we reduced the number of covariates using a stepwise procedure that removed any covariates with *P*-values lower than 0.3. The resulting model was our initial predictive model, which yielded an initial propensity score for each patient. We report a C-statistic to measure the extent to which the probit model assigned patients accurately to the correct group based on the propensity score. A common rule of thumb is that the C-statistic should be close to 0.8, which represents a high probability of correct prediction into patients' actual service category but is not so close to 1 that observable characteristics are predicting service category perfectly.

After weighting, balance is evaluated between the treatment group and the control group by comparing the means of the covariates within groups both before and after weighting. We assess balance on the first order value of each substance use measure and demographic characteristic used in the predictive model. There is some disagreement over the correct method to assess balance [16]. We use the standardized difference method to assess balance in our analysis [17]. Balance is considered to have improved if the standardized difference between groups diminishes after weighting. In addition, it has been suggested that in order for a covariate to be balanced across groups, it should have a standardized difference of no more than 10 [18]. Balancing after propensity score-matching can often result in descriptive statistics that resemble those of a randomized experiment—with little to no difference across groups [14].

Statistical models

After determining the propensity scores and support sample for the regression discontinuity design, we apply several different models of outcomes. There are many methods available to utilize propensity score adjustments and there are trade-offs between estimate bias and efficiency among the different approaches [16, 19]. The first set of models that we apply are multiple generalized linear mixed models (GLMM). We estimate these models with and without inverse propensity score weights. In the latter case, the use of both the weights (based on observable

variables) and the observables themselves as covariates is a form of what are termed ‘doubly robust propensity score models’ [19]. Within the GLMM, we estimated individual-level random effects (i.e. repeated measures for each patient over the two time-periods). Furthermore, these random effects were nested with performance site to account for that level of the error structure and intracluster correlation. The data did not support nesting sites within the seven projects; these were modelled instead as fixed effects by including indicator variables for six projects as independent covariates.

The key independent variables were indicators of whether an observation was baseline or follow-up (post) and whether a patient was in the BI or BT category (BT). The interaction of these two variables (BT × post) is a differences-in-differences (DiD) estimator and its coefficient represents the treatment effect of BT. Other covariates were days of baseline substance use of the substances not being modelled (e.g. when days of alcohol use is the dependent variable, baseline days of alcohol to intoxication, illegal drug use and marijuana were included as dependent variables), age, gender, race and ethnicity.

One advantage of this framework was the ability to test for the moderating effects of the baseline severity of substance use. To test whether any effect of BT is greater for patients with higher baseline use estimated two additional models that included interaction terms. First, we interacted continuous days of baseline substance use being modeled with the key independent variables (i.e. BT, post, and BT × post). The coefficient on the interaction of BT × post with days reflects the moderating effect of severity (as proxied for by days) on the main BT effect. In a second set of models we used the two indicators of baseline substance use, low and high, instead of days for each substance. Low and high days were included as main effects and as interactions with BT, post and BT × post. Similar to the continuous days moderator, the coefficients on BT × post interacted with low and BT × post interacted with high reflected their respective moderating effects on the impact of BT relative to zero days of use at baseline.

We chose to use one additional outcome model that fits our study design in order to corroborate our GLMM results; specifically, local linear matching (LLM) combined with the estimator. The DiD estimator was first recommended by Heckman *et al*., and its use with LLM has been shown to have several advantages [20, 21]. The main advantage of the LLM (and similar kernel matching methods) is its well-studied semi-parametric properties that do not require strong distributional assumptions to estimate treatment effects. Specifically, it compares each treated subject's outcome with a propensity score distance-weighted average of the outcomes of comparison subjects whose propensity scores lie within a certain range of the treated subject's score. The use of additional (weighted) comparison subjects increases the efficiency of the LLM estimate, but adds possible bias from less comparable matches. However, this possible increase in bias is lower than what might be produced by simply increasing the number of comparison matches, such as in nearest-neighbor matching. Limiting the range within which a comparison subject is included in the estimation is conceptually consistent with our quasi-experimental scenario. We contend that for each BT patient there may exist similar BI patients who are reasonable comparators (after controlling for observed characteristics). In practice, the size of the range does not appear to have a strong impact on estimates [21]. However, we estimate the model for several common ranges (0.01, 0.06 0.30) and find little difference in our estimates. Also consistent with our design is the need for common support or, in other words, available BI

patient comparisons for each BT patient, as we are focused on the average effect of BT on those who received BT. We used the common min–max approach of eliminating all BT patients whose propensity score was higher than the score of the highest BI patient. Equivalently, we eliminated all BI patients whose propensity score was lower than the score of the lowest BT patient. This augments our confidence in the internal validity of our experiment, but leaves us with fewer patients in our analysis, which reduces the extent to which our estimates are representative of the overall effect of brief treatment. To provide additional evidence of appropriate matching, we conduct a Kolmogorov–Smirnov test to assess the equality of the propensity score distributions of BI and BT patients who remain in our analysis sample after matching is finalized.

The DiD estimator takes advantage of the longitudinal nature of our data by using the difference within a patient's outcomes between follow-up and baseline as the dependent variable. In our case, we calculate the difference in days of use of each of our outcomes of interest. While the LLM controls for selection into treatment based on observable characteristics, the DiD estimator controls additionally for unobserved patient characteristics that do not change over time. The differenced days of use variables yield a fairly symmetrical distribution. As an additional check for robustness, we also estimated LLM models of the follow-up days of use as dependent variables. For all LLM-DiD results, standard errors were estimated by bootstrapping. All models were estimated using the psmatch2 routine using Stata version 11 [22].

Results

The C-statistic for the probit model was 0.78, indicating an appropriate level of accuracy in prediction of group assignment conditional on observable characteristics. The number of unique comparison BI patients who were identified was 1448, with 332 of them serving as comparisons for more than one BT patient. Only 19 BT patients were eliminated to maintain common support, which strengthens our confidence that our estimates are representative of the average effect of brief treatment on BTs served by the program as a whole. This led to a BT sample of 1770 and a total analysis sample of 3218. We compared the distribution of propensity scores of the BI patients who remained in our support with the scores of the BT patients in the support using a Kolmogorov–Smirnov test and failed to reject the null hypothesis of equal distributions ($P = 0.89$), providing evidence that the BI group is similar enough to the BT group to provide adequate comparability.

Balance was assessed by comparing the between-group standardized differences of unweighted and weighted means. Although balance was examined on all covariates used in the analysis, we report only the balance for substance use variables and demographics in Table 2. In general, the balance assessment suggests that the adjustment eliminates much of the difference between the BI and BT groups and reduces the impact of potential confounders. The sum of the standardized differences decreases from 133.88 on the unweighted means to 54.32 on the weighted means. Seven standardized differences on the unweighted means were greater than 10. None of the adjusted standard differences exceeded 10. Although the weighted standard difference for the conditional days of eliminated marijuana use is less than 10, its size suggests a persistent difference between the groups that was not completely by the adjustment. This difference persisted even after an iterative process of including higher-order interactions of marijuana use with grantee, time and gender variables.

Table 2. Unadjusted and propensity score weighted means comparison of baseline characteristics between brief intervention (BI) and brief treatment (BT) patients.

	Unweighted			Weighted		
Variable	**BI (*n* = 5842)**	**BT (*n* = 1789)**	**SD**	**BI (*n* = 1448)**	**BT (*n* = 1770)**	**SD**
Alcohol						
Proportion using in past 30 days	0.83	0.77	10.06	0.82	0.77	8.59
Average conditional days of use	10.52	11.18	6.75	11.37	11.18	1.44
Alcohol to intoxication						
Proportion using in past 30 days	0.63	0.60	5.81	0.61	0.60	3.01
Average conditional days of use	14.31	16.95	13.56	16.81	16.08	5.37
Illicit drugs						
Proportion using in past 30 days	0.55	0.64	15.07	0.64	0.64	0.00
Average conditional days of use	12.69	13.93	10.92	14.28	13.93	3.09
Marijuana						
Proportion using in past 30 days	0.33	0.35	2.30	0.34	0.35	1.11
Average conditional days of use	13.37	12.52	7.44	13.97	12.82	9.1
Cocaine						
Proportion using in past 30 days	0.21	0.27	10.53	0.29	0.27	3.80
Average conditional days of use	7.01	8.45	16.15	8.07	8.45	4.06
Opiates						
Proportion using in past 30 days	0.08	0.13	9.21	0.13	0.13	0.19
Average conditional days of use	13.22	15.08	15.61	15.17	15.08	0.72
Age						
Average	37.35	37.14	1.62	37.72	37.14	4.70
Gender						
Female	0.35	0.35	0.05	0.34	0.35	1.57
Race/ethnicity						
White, non-Hispanic	0.32	0.32	1.68	0.33	0.32	0.48
Black, non-Hispanic	0.24	0.24	1.81	0.24	0.24	1.11
Other, non-Hispanic	0.22	0.21	2.69	0.22	0.21	2.24
Hispanic	0.22	0.23	2.62	0.21	0.23	3.74

SD = standard deviation.

Table 3. Mean unconditional days of use over time by service recommendation.

	BI (*n* = 1448)			BT (*n* = 1770)		
Substance	**Baseline**	**6-month follow-up**	**Difference**	**Baseline**	**6-month follow-up**	**Difference**
Alcohol	9.304	3.813	–5.491	8.610	3.275	–5.335
	(0.303)	(0.208)	(0.297)	(0.234)	(0.16)	(0.236)
Alcohol to Intoxication	6.623	2.502	–4.121	6.060	2.251	–3.809
	(0.276)	(0.172)	(0.278)	(0.208)	(0.132)	(0.218)
Illicit drugs	9.120	2.890	–6.229	8.894	2.068	–6.826
	(0.337)	(0.228)	(0.341)	(0.27)	(0.151)	(0.271)
Marijuana	4.777	2.007	–2.771	4.364	1.429	–2.935
	(0.279)	(0.194)	(0.260)	(0.214)	(0.128)	(0.206)

Propensity score weighted means reported with standard errors in parentheses. BI = brief intervention; BT = brief intervention.

Table 4. Estimates of the effect of brief intervention (BT): overall and by baseline severity.

		Generalized linear mixed models						
		Unweighted			Weighted			
Substance	Covariate	Base model	Days moderator	Levels moderator	Base	Days moderator	Levels moderator	Weighted local linear regression
Alcohol	BT indicator × post	–0.026	–0.012	0.340	–0.102	0.129	0.418	0.269
		(0.250)	(0.350)	(0.289)	(0.310)	(0.366)	(0.270)	(0.373)
	BT indicator × post × baseline days of use		–0.002			–0.026		
			(0.041)			(0.042)		
	BT indicator × post × low days of use indicator			–0.398			–0.420	
				(0.368)			(0.317)	
	BT indicator × post × high days of use indicator			–0.497			–0.876	
				(0.705)			(0.701)	
Alcohol to intoxication	BT indicator × post	0.337	0.601	0.275	–0.027	0.797		0.462
		(0.378)	(0.455)	(0.798	(0.470)	(0.494)		(0.378)
	BT indicator × post × baseline days of use		–0.020			–0.057		
			(0.037)			(0.043)		
	BT indicator × post × low days of use indicator			0.471			0.320	
				(0.791)			(0.832)	
	BT indicator × post × high days of use indicator			–0.357			–0.978	
				(1.227)			(1.351)	
Illicit drugs	BT indicator × post	–0.634**	–0.282**	–0.218	–0.878**	–0.355**	–0.154	–0.651**
		(0.274)	(0.141)	(0.157)	(0.378)	(0.146)	(0.164)	(0.328)
	BT indicator × post × baseline days of use		–1.260*			–1.626*		
			(0.701)			(0.929)		
	BT indicator × post × low days of use indicator			–0.239			–0.508	
				(0.474)			(0.694)	
	BT indicator × post × high days of use indicator			–1.334*			–1.765**	
				(0.771)			(0.897)	
Marijuana	BT indicator × post	–0.379	–0.217	0.039	–0.458*	–0.205	0.072	–0.128
		(0.248)	(0.147)	(0.108)	(0.271)	(0.154)	(0.119)	(0.258)
	BT indicator × post × baseline days of use		–0.770			–1.190		
			(0.687)			(0.899)		
	BT indicator × post × low days of use indicator			–0.882			–0.571	
				(0.581)			(0.488)	
	BT indicator × post × high days of use indicator			–1.477			–2.117	
				(1.009)			(1.292)	

Weights are inverse propensity scores. Multivariate models included age, indicators for being female, African American, another race category and Hispanic, indicators of project and baseline counts of days of alcohol use, alcohol to intoxication, illegal drug use and marijuana use. Coefficients are presented for the indicators of being a BT patient and any interactions with that indicator. The low and high level moderator models included indicators of whether a patient's baseline days of use for that substance was greater than zero and less than the median days of use in the sample or greater than the median days of use in the sample; zero days of use was the reference category. The error structure included random effects for the performance site a patient was in and adjusted standard errors for clustering. Robust standard errors in parentheses. *** $P < 0.01$; ** $P < 0.05$; * $P < 0.1$.

Table 3 presents inverse propensity score-weighted mean unconditional number of days of use of each assessed substance at baseline and follow-up. These estimates provide context for the models that follow and are no statistical testing is applied. For both BI and BT patients and for all substances, days of use decreased after recommendation to SBIRT services. For both BI and BT patients and for all substances, days of use decreased after recommendation to SBIRT services. The largest decreases are for illicit drugs at 6.2 days for BI and 6.8 days for BT, with an unadjusted difference-in-differences of a little more than half a day of use [6].

Table 4 presents the primary and moderator results of the GLMMs as well as the results of the propensity score LLR. Unweighted and weighted GLMM estimates are presented. The overall effect estimates for BT are presented in the base model columns. We find no effect for alcohol days in either the unweighted or the weighted models. The estimated decreases for alcohol to intoxication are somewhat larger than for alcohol days, but were not significant. We found significant effects for illicit drug days in both the unweighted (−0.634, $P < 0.05$) and weighted (−0.878, $P < 0.05$) models. The weighted result for marijuana days was marginally statistically significant (−0.458, $P < 0.1$).

In the continuous days moderator models, only illicit drug days had statistically significant findings. The main effect estimates for BT remained statistically significant at the 0.05 level but were smaller in magnitude (e.g. −0.355 versus −0.878 for the weighted models). The interaction terms were large, but significant only at the 0.1 level. The low/high models did yield statistically stronger moderating effects in the weighted models. The high-category BT patients had a 1.765 ($P < 0.05$) decrease in days of use, while the zero (i.e. the main BT × post effect) and low BT patients, although negative, were not statistically significant.

The final column displays the estimated effect of brief treatment on BT patients' difference in days of use using the LLR-DiD method. For both alcohol and alcohol to intoxication, BI comparison patients demonstrated slightly larger decreases in days (0.269 and 0.462 days, respectively). These differences were not statistically significant. BT patients decreased their days of illicit drug use by 0.651 days more than BI patients, a statistically significant result ($P < 0.05$). The effect on days of marijuana use was also higher for BT patients, but the result was smaller in magnitude than for the composite illicit drug days and was not statistically significant.

Discussion

Within the emerging set of early intervention services directed at substance users in primary health-care and other settings, brief treatment is a distinct option from brief interventions (BIs) In this study we have employed a quasi-experimental design to analyze the comparative effectiveness of brief treatment. By taking advantage of natural variation in whether screened patients were recommended to brief intervention or brief treatment, we used a propensity score approach to develop a comparison sample of BI patients to estimate the relative impact of brief treatment. Our assessment of balance between our BI and BT groups demonstrated that the comparison sample of BI patients were very similar to BT patients in their substance use and other observable characteristics prior to exposure to SBIRT. Days of use of all substances decreased substantially for both our BI comparison group and the BT group. Although we found

no differential effect of recommendation to BT in days of using alcohol, we found a statistically significant reduction in days of illicit drug use for those in the BT group of over a half day of use in the prior 30 days. We note that days of illicit drug use is a meaningful outcome, as it combines all illicit drug use and represents an upper bound on number of days of abuse. It also manages any substitutions among drugs that may lead to spurious estimates on individual drugs.

The finding that recommendation to brief treatment is effective in reducing illicit drug use is important for several reasons. First, the effect of recommendation to brief treatment was not estimated relative to 'no recommendation to services' or to 'care as usual', but instead to recommendation to brief intervention, which has been shown to have a substantial impact on substance use in and of itself. The relative impacts of 0.652–0.878 days (depending on the model) represent a 20–23% reduction in days of illicit drug use at follow-up relative to brief intervention for the average patient. Secondly, recommendation to a service does not guarantee receipt of the service for each study patient. Some patients recommended to SBIRT services declined to receive any services and others chose to receive a less intensive service, often for practical reasons. Rarely did a patient recommended to a less intensive service request to receive a higher-intensity service. These factors suggest that our findings may indeed represent a lower bound of the effectiveness of BT relative to BI.

We note that only illicit drug days out of four substance use outcomes had a strong finding. However, this outcome is arguably more important than the other three. Illicit drug use is a global measure—the maximum of days of all substances. Marijuana is only a single component of illicit days, and is thus a partial (albeit prominent) measure. Similarly, alcohol and alcohol to intoxication are specific substance measures. It must be kept in mind that the BT following a screen is likely to prioritize specific substances. The broadest measure of what is more likely to be targeted for BT level patients (illicit drug days) therefore represents more of the overall sample bringing higher statistical power and opportunity of change in the measure over time. A final point of context is that BT is not being compared against placebo or a low-content care as usual, but against BI. Inasmuch as BI is both effective for and may more often emphasize alcohol outcomes, it may not be inconsistent that BT was found significant only for illicit drug use. In this interpretation, BT may have had a qualitatively different effect on illicit drug use than BI, which is consistent with the literature on BI's primary efficacy being for alcohol.

Our moderator analyses provide additional insight into BT's relative impact. Specifically, the overall effects appear to be explained by relative improvement among more severe patients. Focusing upon the weighted GLMM result, the 0.878 overall effect is arguably driven by the 1.765-day decrease among high users (8 or more days reported at baseline). In the case of illicit drugs, high users accounted for 26% of the BI sample and 30% of the BT sample. The implication of these findings is that when interventionists identify more severe substance-using patients, BT is likely to be a better service alternative than BI. None the less, it is important to keep in mind that this does not preclude a role for BI as an option when BT (or any related service that is more intense than BI) is not available. Moreover, as noted in the Data section, BI services still played a substantial role in the average set of services received by patients assigned to the BT service category.

One consideration in interpreting these results is the two distinct types of models being used: the inverse propensity score weighted GLMM that conditioned on a robust set of control variables and the LLR which semi-parametrically estimates 'local' effects for treatment group patients relative to comparison patients whose propensity scores fall within a small radius. The GLMM estimate (holding aside the covariate conditioning) is analogous to comparisons of weighted population averages, while the LLR estimate is more akin to the average of many local outcome differences. It is therefore possible that different estimates might be generated based simply upon having imperfect (not normal or even symmetric) distributions. There are limited simulation studies of how these two approaches perform under different scenarios [14]. We believe that the LLR more effectively fits the study design conceptually, especially given the distribution of severity seen at baseline. Moreover, LLR is notably less efficient than GLMM, and thus stacks the deck against finding statistically significant results. However, the use of covariates in the GLMM and the results of the moderator models which explicit partition the treatment effect by severity arguably improve their appropriateness within this design. We therefore include results from both models, noting that their results are overall consistent with each other. We note several limitations to our analyses. Our design is quasi-experimental, and our propensity score approach only helps to eliminate bias from the correlation of observed patient characteristics with assignment to treatment or potential outcomes. Although these observable characteristics include substance use, arguably the most important indicator of risk, there may be other patient factors that lead practitioners to make certain treatment recommendations systematically, or affect a patient's response to a certain treatment. A second limitation is that the validity of our design applies only to a specific group of higher-risk patients. Related to both of these limitations is the fact that we do not know the exact substance that was being addressed by the intervention or therapy. We note again that our design is intent-to-treat and thus, despite evidence that a substantial amount of services was delivered, we cannot make strong conclusions about the extent to which dosage differences or qualitative differences in the type of service are the cause of the findings. Our results do not provide information on how lower-risk BIs would respond to brief treatment. Our follow-up rates were not higher than the conventional 80% threshold, which should be taken into account when interpreting the generalizability of our results. Finally, our analyses are based on self-report administrative data. A particular challenge in this case is the possibility that patients who were engaged more strongly in services (BTs) might simply answer substance use questions differently at follow-up in order to avoid perceived social stigma.

Despite these limitations, we note the uniqueness and benefit of studies like this one. It is difficult to design and implement more rigorous effectiveness studies in the natural conditions of SAMHSA's SBIRT projects. Human subjects protections, reluctance of provider sites to participate fully and cost are all up-front challenges. Compromises to overcome these challenges, as well as inclusion and exclusion criteria, can lead to samples with less external validity. Moreover, estimates from these types of studies reflect effects of interventions that are based on clinical research protocols and with their requisite fidelity and quality control. The utility of such estimates for justifying services in natural settings or informing policy around them is limited. In our case, our experiment is embedded in the natural setting and subsumes all the variation in provider quality, intervention quality and setting and patient characteristics. In other words, following an efficacy study, the common question is: 'does it work in the real world?'. Our study is a rare opportunity to provide an answer to that question.

Overall, these findings suggest that recommendation to brief treatment can have a larger impact upon substance use, or is at least as effective as recommendation to brief intervention, for a certain group of higher-risk patients. The mechanism through which brief treatment is more effective is uncertain and requires future study. A previous trial of BT for chronic marijuana users found that assignment to more BT sessions (nine versus two) led to better outcomes [23]. They note, however, that the relative improvement of the nine-session patients began around the second session. This suggests that there may be other elements in addition to *per se* dosage that make a difference. For example, the expectation of the greater commitment may have a significant factor. From the patient perspective, simply being labeled as 'higher risk' may have its own intervention effect and lead the patient to change or to more thoroughly engage in services.

From the perspective of the practitioner, even if he or she knows the 'borderline' risk level of the patient initially, by recommending to BT that practitioner or subsequent practitioners may begin to think of that patient differently by, for example, engaging them more intensively in follow-up services. As brief treatment may represent a key component in the continuum of care between lower-risk BI patients and patients who require referral to specialty treatment, ongoing study of its effectiveness is justified.

References

1. Moyer A., Finney J. W., Swearingen C. E., Vergun P. Brief interventions for alcohol problems: a meta-analytic review of controlled investigations in treatment-seeking and non-treatment-seeking populations. *Addiction* 2002; **97**: 279– 292.
2. Krupski A., Sears J. M., Joesch J. M., Estee S., He L., Dunn C. *et al.* Impact of brief interventions and brief treatment on admissions to chemical dependency treatment. *Drug Alcohol Depend* 2010; **110**: 126– 136.
3. Stephens R. S., Roffman R. A., Curtin L. Comparison of extended versus brief treatments for marijuana use. *J Consult Clin Psychol* 2000; **68**: 898.
4. Madras B. K., Compton W. M., Avula D., Stegbauer T., Stein J. B., Clark H. W. Screening, brief interventions, referral to treatment (SBIRT) for illicit drug and alcohol use at multiple healthcare sites: comparison at intake and 6 months later. *Drug Alcohol Depend* 2009; **99**: 280– 295.
5. Bernstein E., Bernstein J. A., Stein J. B., Saitz R. SBIRT in emergency care settings: are we ready to take it to scale. *Acad Emerg Med* 2009; **16**: 1072– 1077.
6. Kaner, E. F., Beyer, F., Dickinson, H. O., Pienaar E., Campbell F., Schlesinger C. *et al*. Effectiveness of brief alcohol interventions in primary care populations. *Cochrane Database Syst Rev* 2007; Issue **2**. Art. No.: CD004148. DOI: 10.1002/14651858.CD004148.pub3.
7. Dehejia R. H., Wahba S. Causal effects in non-experimental studies: re-evaluating the evaluation of training programs. *J Am Assoc* 1998; **94**: 1.053– 1.062.

8. Johnson M. L., Crown W., Martin B. C., Dormuth C. R., Siebert U. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—Part III. *Value Health* 2009; **12**: 1062.

9. Miller W. R. Motivational interviewing with problem drinkers. *Behav Psychother* 1983; **11**: 147– 172.

10. Miller W. R., Rollnick S. *Motivational Interviewing: Preparing People to Change Addictive Behavior*. New York: Guilford Press; 1991.

11. Bien T. H., Miller W. R., Tonigan J. S. Brief interventions for alcohol problems: a review. *Addiction* 1993; **88**: 315.

12. National Institute on Alcohol Abuse and Alcoholism. *Helping patients who drink too much: a clinician's guide*. Updated 2005 edn. NIH Publication no. 07-3769. Rockville, MD: Author; 2007.

13. Prochaska J. O., DiClemente C. C. Toward a comprehensive model of change. In: M. W. R., H. N., editors. *Treating Addictive Behaviors: Processes of Change*. New York: Plenum Press; 1986, pp. 3– 27.

14. Rosenbaum, P. R., Rubin, D. B. Reducing bias in observational studies using subclassification on the propensity score. J Am Stat Assoc 1984; **79**: 516– 524.

15. Rubin D. B. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol* 1974; **66**: 688.

16. Caliendo M., Kopeinig S. Some practical guidance for the implementation of propensity score matching. *J Econ Surv* 2008; **22**: 31– 72.

17. Ho D. E., Imai K., King G., Stuart E. A. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Polit Anal* 2007; **15**: 199– 236.

18. Austin P. C. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. *Commun Stat Stim Comput* 2009; **38**: 1228– 1234.

19. Baser O. Too much ado about propensity score models? Comparing methods of propensity score matching. *Value Health* 2006; **9**: 377– 385.

20. Heckman J. J., Ichimura H., Todd P. Matching as an econometric evaluation estimator. *Rev Econ Stud* 1998; **65**: 261– 294.

21. A. S. J., Todd P. E. Does matching overcome LaLonde's critique of nonexperimental estimators? *J Econometr* 2005; **125**: 305– 353.

22. Leuven, E., Sianesi, B. *PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing.* Boston College of Econonmetrics; Statistical Software Components: 2003.

23. Babor T. F. Brief treatments for cannabis dependence: findings from a randomized multisite trial. *J Consult Clin Psychol* 2004; **72**: 455.