



A complier average causal effect analysis of the Stimulant Reduction Intervention using dosed exercise study



Thomas Carmody, Tracy L. Greer, Robrina Walker, Chad D. Rethorst, Madhukar H. Trivedi*

University of Texas, Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX, 75390, USA

ARTICLE INFO

Keywords:

Complier average causal effects
Exercise intervention
Health education
Stimulant abuse or dependence
Clinical trials network

ABSTRACT

Objective: Exercise is a promising treatment for substance use disorders, yet an intention-to-treat analysis of a large, multi-site study found no reduction in stimulant use for exercise versus health education. Exercise adherence was sub-optimal; therefore, secondary post-hoc complier average causal effects (CACE) analysis was conducted to determine the potential effectiveness of adequately dosed exercise.

Method: The STimulant use Reduction Intervention using Dosed Exercise study was a randomized controlled trial comparing a 12 kcal/kg/week (KKW) exercise dose versus a health education control conducted at nine residential substance use treatment settings across the U.S. that are affiliated with the National Drug Abuse Treatment Clinical Trials Network. Participants were sedentary but medically approved for exercise, used stimulants within 30 days prior to study entry, and received a DSM-IV stimulant abuse or dependence diagnosis within the past year. A CACE analysis adjusted to include only participants with a minimum threshold of adherence (at least 8.3 KKW) and using a negative-binomial hurdle model focused on 218 participants who were 36.2% female, mean age 39.4 years ($SD = 11.1$), and averaged 13.0 ($SD = 9.2$) stimulant use days in the 30 days before residential treatment. The outcome was days of stimulant use as assessed by the self-reported TimeLine Follow Back and urine drug screen results.

Results: The CACE-adjusted analysis found a significantly lower probability of relapse to stimulant use in the exercise group versus the health education group (41.0% vs. 55.7%, $p < .01$) and significantly lower days of stimulant use among those who relapsed (5.0 days vs. 9.9 days, $p < .01$).

Conclusions: The CACE adjustment revealed significant, positive effects for exercise. Further research is warranted to develop strategies for exercise adherence that can ensure achievement of an exercise dose sufficient to produce a significant treatment effect.

Public health significance

This secondary analysis of the Stimulant use Reduction Intervention using Dosed Exercise study suggests that an exercise level of more than 8.3 kcal/kg/week may reduce relapse to stimulant use and reduce the days of stimulant use for those who relapse. This analysis also demonstrates the importance of ensuring adherence to exercise interventions and accounting for adherence in the interpretation of results, and that statistically rigorous adjustment for post-baseline measures such as exercise dose is possible.

1. Introduction

Currently available treatments for substance use disorders (SUD) are insufficient to achieve abstinence or large reductions in substance use for many treatment-seeking individuals [1,2]. Therefore, the development of new treatments for SUD is an important research goal. Preliminary studies show that exercise has potential as an innovative treatment for SUD [3–6]. Furthermore, exercise acts on a variety of psychological (anxiety [7], depression [8]), and neurobiological mechanisms [9,10] that suggest exercise may be effective as a treatment for SUD [11,12].

The STimulant Reduction Intervention using Dosed Exercise (STRIDE) study evaluated stimulant use outcomes following a dosed

Abbreviations: CACE, Complier Average Causal Effect; ITT, Intention-to-Treat; KKW, kilocalories/kilogram/week; RTP, Residential Treatment Program; STRIDE, STimulant Reduction Intervention using Dosed Exercise; SUD, Substance Use Disorders; TLFB, Timeline Follow Back; UDS, Urine Drug Screens

* Corresponding author. Julie K. Hersh Chair for Depression Research and Clinical Care, Betty Jo Hay Distinguished Chair in Mental Health, Director, Center for Depression Research and Clinical Care, UT Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75390-9119, USA.

E-mail address: Madhukar.Trivedi@utsouthwestern.edu (M.H. Trivedi).

<https://doi.org/10.1016/j.conctc.2018.02.001>

Received 13 June 2017; Received in revised form 8 January 2018; Accepted 12 February 2018

Available online 15 February 2018

2451-8654/ © 2018 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

exercise intervention versus a health education intervention, both of which were provided as augmentation to treatment as usual. The *a priori* primary analysis [13] was based on the intention-to-treat (ITT) principle in which all participants were analyzed according to the groups to which they were randomly assigned [14]. The primary outcome of percent stimulant abstinent days in the exercise and health education groups was compared after (1) imputing missing data as days of drug use and (2) employing a novel method to reconcile discrepancies between subjective and objective measures of drug use [15]. The ITT analysis revealed no treatment effect as percentage of days abstinent were 75.6% (SD = 27.4) for those in exercise and 77.3% (SD = 25.1) for those in health education ($p = 0.60$) [13].

However, the ITT analysis is not by itself sufficient to assess the viability of exercise as an augmentation to SUD treatment because many participants did not exercise at the prescribed dose. The median exercise dose of 8.3 kcal/kg/week (KKW) (interquartile range: 4.2 to 10.6 KKW) in this study was approximately two-thirds of the prescribed dose of 12 KKW. This suboptimal adherence to the prescribed dose confounds our ability to interpret the results of the trial because even an effective treatment may produce small treatment effects in people who do not fully participate in the treatment. In order to assess the viability of exercise, we must answer the following question: Is there an exercise dose that will produce a clinically meaningful exercise effect?

For this analysis, an exercise dose greater than or equal to the median exercise dose (8.3 KKW) exhibited by study participants will be subsequently referred to as an “adequate dose.” An estimate of the exercise effect among participants who exercised at or above this adequate dose provides two major advantages. First, to determine the most appropriate treatment recommendation for a patient, the clinician must consider the size of the treatment effect for exercise versus other possible treatments. If the clinician believes the patient would be adherent to an assigned exercise dose greater than the median 8.3 KKW observed in STRIDE, the STRIDE *a priori* primary analysis results provide no guidance as to treatment effect size because the effect size is influenced by those who exercised less than the median dose [16]. Second, without understanding the efficacy of exercise, it is unclear how to proceed with future research. If exercise for stimulant users is truly ineffective, additional research as a potential treatment option is unwarranted. If, however, exercise is ineffective due to poor adherence, it may be beneficial to continue pursuing exercise as a treatment option while developing interventions to optimize exercise adherence [16].

Per-protocol and as-treated analyses are sometimes used in an attempt to adjust for an observed dose that is less than the prescribed treatment dose, but these approaches are statistically biased. Per-protocol analysis for STRIDE would compare those in the exercise group who were adherent to exercise versus those in the health education group who were adherent to health education. However, those subgroups could differ substantially in important covariates. As-treated analysis would require exercise participants who did not exercise to be considered as belonging to the health education group for the purpose of analysis, thereby creating non-comparable groups [16]. To address these statistical challenges we employed a complier average causal effect (CACE) analysis. This enabled us to make a statistically rigorous estimate of exercise treatment effects based on the majority of participants' exercise dose (ranged from 8.3 to 11.5 KKW) rather than the dose observed in the intention-to-treat STRIDE sample which included those who did not exercise at all (ranged from 0 to 11.5 KKW). CACE analysis has been used in trials of behavioral interventions [17,18], including substance abuse research [19,20]. If the assumptions of the CACE analysis are fulfilled, we can determine in a statistically rigorous manner that an effective range of exercise dose exists and, hence, that exercise is worthy of further research.

2. Methods

The design and methodology of the STRIDE study have been

previously described [21–26]. Below, we briefly describe study procedures relevant to the analysis presented. The study was conducted in accordance with the Declaration of Helsinki and was approved by the Institutional Review Boards associated with the participating treatment programs. Written informed consent was obtained from all participants prior to beginning study procedures.

2.1. Participants

STRIDE enrolled adult stimulant users, aged 18–65 years, who were admitted to residential substance abuse treatment, had used stimulants within 30 days prior to enrollment, and met DSM-IV criteria for stimulant abuse or dependence within the last 12 months. Participants also had to be medically clear to exercise via a protocol-defined stress test. Exclusion criteria included: opioid dependence within the last 12 months; evidence of a general medical condition, medication, or psychiatric condition that contraindicated study participation; pregnancy; or significant physical activity, defined as aerobic exercise more than 3 times per week for 20 min or more, completed consistently for the three months prior to study enrollment.

2.2. Study procedures and interventions

Randomly assigned participants received either the Dosed Exercise Intervention or the Health Education Intervention. Drug abuse treatment as usual was provided to both groups, beginning with residential treatment (median 17 days, interquartile range 12–22 days) and followed by outpatient treatment. The prescribed exercise dose for the exercise intervention was 12 KKW provided during three one-on-one supervised sessions per week. Twelve KKW is equivalent to 150 min of moderate exercise per week at an exercise intensity of 70–85% of maximal heart rate, and is within public health dose guidelines (<http://www.health.gov/paguidelines>). Health education was also provided during thrice-weekly one-on-one supervised sessions designed to last as long as the exercise sessions and so ensure equivalent staff contact between groups during the 12-week acute phase. These sessions provided information on health-related topics (e.g., diet, mental health, and sleep) via didactics; Websites; and audio, video, and written materials. To reduce any psychosocial effects of health education, no specific goals were set for participants to achieve during the sessions [26]. Health education has been established as a valid control condition in other exercise studies [27–29] (Marcus et al., 1999; Pahor et al., 2006; Rejeski et al., 2005). Trained facilitators implemented both interventions.

2.3. Measures

2.3.1. Drug use

Self-reported drug use was assessed using the TimeLine Follow Back (TLFB), a semi-structured interview that uses calendar prompts to retrospectively recall daily drug use over a specified period of time [30]. Qualitative urine drug screens (UDS) were collected 3 times per week and assessed stimulants (cocaine, amphetamine, methamphetamine) as well as opiates, marijuana, benzodiazepines, barbiturates, methadone, methylenedioxymethamphetamine (ecstasy), and oxycodone. The daily TLFB was compared with the 3 times per week UDS, and contradictions between the two were resolved using the Eliminate Contradiction algorithm [15] as follows: when the UDS was positive but the prior 3 days were all negative according to the TLFB, then the TLFB for the last day in the window was changed from negative to positive. Drug use was assessed during the post-residential treatment program (RTP) period from the day after discharge to 84 days after randomization.

2.3.2. Adequate exercise dose

Participant exercise dose was defined as energy expended per week computed in KKW averaged over the entire acute phase from

randomization until day 84. Weeks in which there were no exercise visits were set to zero KKW for computing the average. STRIDE participants achieved an exercise dose ranging from 0 KKW to 11.5 KKW with a median of 8.3 KKW. For the purpose of this analysis, exercise dose was defined to be “adequate” if it was equal to or greater than the median dose of 8.3 KKW. This dose was selected for several reasons. First, using a sample-based determination of an adequate dose increases the likelihood that this is an achievable dose for this population. Second, the CACE analysis (see Statistical Methods, section 2.4.1) requires the creation of a model to predict a binary outcome (i.e., adequate dose/inadequate dose). To maximize the predictive ability of this model, we chose to define an adequate dose of exercise by the median dose of 8.3 KKW. Using this definition implies that the results of the CACE analysis would generalize to a sample with a range of exercise dose from 8.3 KKW to 11.5 KKW with a median of 10.6 KKW.

2.3.3. Other measures

The predictors used in the analyses described below were derived from several measures fully described in Trivedi et al. [13]. Briefly, demographic information (e.g., gender, marital status) was collected at screening. A Maximal Exercise Test was conducted during the screening process to assess fitness for exercise. Self-report measures collected at baseline included symptom assessments of stimulant abstinence (Stimulant Selective Severity Assessment [31]), depressive symptoms (16-item Quick Inventory of Depressive Symptomatology - Clinician rated version [32]), suicidal thoughts and behaviors (Concise Associated Symptoms Tracking- Self-Report [33]), and the 36-item Short Form Health Survey [34]. Physical and cognitive functioning was assessed using the Massachusetts General Hospital Cognitive and Physical Functioning Questionnaire [35]. Common problems associated with drug use were assessed using the Addiction Severity Index-Lite [36]. Attendance of addiction treatment as usual during the week prior to randomization was assessed using a treatment tracking form created for the study.

2.4. Statistical methods

2.4.1. Complier average causal effect (CACE) analysis

As discussed above, per-protocol and as-treated approaches produce biased estimates of treatment effects. Conceptually, this bias can be eliminated by considering the treatment effect within a subgroup, referred to as a principal strata [37]. Specifically, we consider the strata of participants who would have had an adequate dose (defined to be 8.3 KKW or more) of exercise if they had been assigned to the exercise intervention. Within the principal stratification framework, the strata is conceptualized to exist prior to and independent of randomization. Randomization tends to divide the sample as a whole and any subset of the sample into comparable groups (i.e., groups that are balanced with regard to important covariates). Participants randomized to exercise and health education should therefore be comparable within the principal strata of those who would have achieved an adequate exercise dose had they been assigned to the exercise group. We could in theory, therefore, make a valid unbiased estimate of exercise effects within this principal strata [16].

The principal stratification approach cannot be directly implemented in practice because we cannot observe exercise dose in participants who were assigned to health education. CACE analysis [37,38] can be used to overcome this obstacle. Typically, CACE analysis is implemented with the propensity score approach used here or the instrumental variables approach. The latter approach relies on the assumption of a zero treatment effect in participants who did not receive an adequate exercise dose (i.e., exclusion restriction [39]). Because a participant can exercise on average as much as 8.3 KKW and still be considered to have an inadequate dose, this assumption does not appear to be reasonable for STRIDE. The propensity score approach has the advantage that the weights (described below) can be used with any

model, such as the hurdle model which we describe below as an appropriate analytical strategy for STRIDE. Therefore, we chose the two-step propensity score CACE approach [40].

2.4.2. Propensity score weights

The first step in the propensity score approach is to create a model using appropriate covariates to predict adequate exercise dose for the exercise group (where exercise dose is observed), then apply that model to the health education group to obtain a probability of receiving an adequate exercise dose for each health education participant. To illustrate, consider a specific predictor such as age. If age is a strong predictor of observed exercise dose in the exercise group we can assume, due to randomization, that age will also be a strong predictor of unobserved exercise dose in the health education group. Therefore, the model to predict exercise dose can be applied to the health education group, resulting in an estimated probability of achieving an adequate exercise dose for each member of the health education group. These probabilities are referred to as principal propensity scores and are the basis for propensity score weights. When these weights are applied to the health education participants, the weighted health education group will resemble the exercise-adherent participants in all measured covariates [40]. In the second step, these weights are incorporated into the desired outcome analysis.

The unbiased estimation of treatment effects through CACE analysis depends critically on the assumption that the model used to generate the propensity score weights includes all important predictors of adequate exercise dose. Given this critical need, the median split in average KKW expended per week (8.3 KKW) was used to define an adequate exercise dose, as this choice will maximize the power to detect predictors. The concept of the principal strata as existing prior to randomization requires that the model use only pre-randomization variables. To improve our likelihood to capture all the relevant predictors, we started with approximately 200 potential exercise dose predictors covering demographics, fitness to exercise, general health, motivation, severity of drug use, and effects of drug use. This list was reduced by retaining only those predictors for which the difference between adequate and inadequate exercise dose participants was significant at $p < 0.20$. To reduce the possibility of including spurious predictors, 50 bootstrap samples were created and the search was conducted in each sample. Only predictors with $p < 0.20$ in a majority of bootstrap samples were retained. Next, predictors which were highly correlated as defined by a correlation coefficient above 0.9 were removed, resulting in a final list of 32 predictors (Table 1).

The model used to predict adequate exercise dose status in the exercise group was created using a machine-learning algorithm known as extreme generalized boosted regression modeling [41], implemented using the ‘xgboost’ package in R software [42]. Xgboost combines a large number of simple regression trees to obtain a prediction [43] using a boosting algorithm while also allowing for random sampling of predictors as in the random forest approach. Simulation studies by Lee et al. [44] show that the boosting and random forest approaches provide superior covariate balance in propensity score estimation. The xgboost algorithm tuning parameters were chosen by cross-validation. The ability of the model to achieve balance between the adequate exercise dose participants and the weighted health education participants was assessed using the average standardized absolute mean difference across the 32 predictors. The average standardized absolute mean difference was reduced by 28% from 0.147 to 0.106 between the adequate dose exercise participants versus the unweighted and weighted health education groups, respectively. A measure of the relative importance of each predictor based on the gain in predictive ability averaged over all trees in the model was obtained from the final prediction model (Table 1).

The covariate values of each health education participant were applied to this xgboost model to compute the probability of achieving an adequate exercise dose. Following Stuart and Jo [40], we used the

Table 1
Effect of propensity score weights on the difference between exercise adherers and health education as measured by effect size.

Predictor	Predictor Relative Influence	DEI Adherent Mean	Unweighted HEI Mean	Unweighted Effect Size	Weighted HEI Mean	Weighted Effect Size
ASI: Employment Status Subscale	11.00	0.68	0.72	0.15	0.69	0.04
% Attendance at Pre-Study TAU Sessions	7.33	101.46	89.17	0.34	91.75	0.27
Gender	7.20	0.32	0.39	0.15	0.27	0.11
ASI:# Cocaine Use Days	7.10	5.51	5.55	0.01	4.30	0.18
ASI: Alcohol Use Subscale Score	6.16	0.18	0.21	0.14	0.20	0.09
SSSA: Hyperphagia Item	5.78	0.55	1.14	0.42	0.90	0.25
ASI: Legal Status Subscale Score	5.47	0.11	0.15	0.22	0.11	0.04
Study Conducted at Site A (y/n)	4.85	0.04	0.11	0.35	0.07	0.18
ASI: Drug Use Subscale Score	4.65	0.16	0.16	0.02	0.15	0.16
SSSA: Hypersomnia Item	4.64	0.32	0.36	0.04	0.40	0.07
CPFQ: Total Score	4.36	18.36	16.97	0.21	17.05	0.19
SF-36: General Health Subscale Score	3.91	51.40	52.56	0.14	51.58	0.02
QIDS: Sum of Insomnia Items	3.87	3.17	2.50	0.28	3.01	0.07
CAST: Sleep Item (Slept Well)	3.65	0.60	0.73	0.26	0.67	0.14
TLFB:# Cocaine Use Days	3.27	8.19	8.89	0.08	7.18	0.12
Study Conducted at Site B (y/n)	2.39	0.24	0.18	0.13	0.16	0.19
ASI:# Illegal Activity Days	2.37	2.49	2.02	0.08	1.34	0.20
ASI: Legal Problems Seriousness*Motivation	2.27	1.29	2.05	0.22	1.18	0.03
ASI:# Days Incarcerated	2.00	0.83	0.40	0.11	0.33	0.13
ASI:# Drugs Used Past 30 Days	1.99	0.89	1.01	0.11	0.91	0.02
MET: Resting Diastolic Blood Pressure	1.79	74.09	72.76	0.12	73.52	0.05
SF-36: Energy Subscale Score	1.16	50.43	52.02	0.14	51.17	0.06
Marital Status: Divorced/Separated	0.82	0.27	0.27	0.00	0.26	0.01
ASI: Spend free time alone (y/n)	0.68	0.69	0.67	0.05	0.66	0.08
ASI:# Lifetime Hospitalizations	0.67	2.20	1.70	0.22	1.78	0.18
QIDS: Hypersomnia Item	0.62	0.09	0.12	0.08	0.12	0.07
ASI: Satisfied how Time Spent	0.00	0.73	0.73	0.01	0.71	0.04
CPFQ: Interest Past Month	0.00	0.21	0.17	0.10	0.19	0.07
ASI: Lives in Controlled Environment	0.00	0.07	0.01	0.21	0.01	0.21
Marital Status: Widowed	0.00	0.05	0.03	0.08	0.03	0.08
Study Conducted at Site C (y/n)	0.00	0.01	0.03	0.18	0.01	0.02
Study Conducted at Site D (y/n)	0.00	0.13	0.10	0.09	0.14	0.03

*During the 30 days prior to randomization.

Abbreviations: ASI: Addiction Severity Index; CAST: Concise Associated Symptoms Tracking; CPFQ: Cognitive and Physical Functioning Questionnaire; DEI: Dosed Exercise Intervention; HEI: Health Education Intervention; MET: Maximal Exercise Test; QIDS: Quick Inventory of Depressive Symptomatology; SF-36: 36-item Short Form Health Survey; SSSA: Stimulant Selective Severity Assessment; TAU: Treatment As Usual; TLFB: TimeLine Follow Back.

probability of an adequate exercise dose as the propensity score weight in the health education group, where exercise adherence is not observed. In the exercise group, where exercise dose is observed, adequate exercise dose participants were assigned a weight of 1 and the others were assigned a weight of 0.

The weighted sample attains the goal of principal stratification: identification of exercise participants who received an adequate exercise dose and, through weighting, a sample of health education participants who are comparable to them so that an unbiased estimate of treatment effect can be obtained. Note that no post-baseline information from the health education group was used and no post-baseline predictors in the exercise group were used, which eliminates potential bias caused by using post-baseline information (e.g., reverse causality).

2.4.3. *The hurdle model*

The outcome chosen for this analysis, standardized number of days of stimulant use during the post-RTP period (described below), consists of count data which typically follow a Poisson or Negative Binomial distribution. However, as is often the case with substance use data, STRIDE data contains more zeros than expected under these distributions (44.1% of the 295 participants with post-RTP data reported zero days of stimulant use, 43.2% in the health education group and 45.0% in the exercise group). Two main approaches have been developed to deal with excess zeros in count data. A zero-inflated model assumes that some participants have the potential to use drugs and that the amount of their use can be modeled with the Poisson or Negative Binomial distribution, while other participants have no potential for drug use, resulting in an excess of participants with zero use. A hurdle model

assumes that all participants have the potential to use drugs but “resistance” to drug use, or a hurdle, must be overcome before drugs are used. The existence of the hurdle results in an excess of participants with zero use [45]. The assumptions of the hurdle model are more appropriate for STRIDE given that all participants have the potential to resume using stimulants, but resistance to use is present due to the fact that participants chose to enter treatment and that we expect a beneficial effect of treatment as usual. A hurdle model provides estimates of two aspects of drug use: the probability of use (i.e., relapse) and the amount of use among those who used.

The hurdle model in this analysis contained a random site effect, fixed effects for treatment group, three pre-specified covariates (days of stimulant use in the 30 days prior to RTP, age, and gender), and additional covariates (described below) included to improve balance between the exercise and health education groups. The propensity score weights were incorporated into the hurdle model. Mplus software Version 7.3 was used to implement a weighted hurdle model based on the Negative Binomial Distribution, which is a more flexible variant of the Poisson distribution [46]. Effect sizes were provided for the estimated probability of use and days of use among those who used, where the effect size is defined as the hurdle model coefficient divided by the standard deviation of the coefficient which is computed as the square root of the sample size multiplied by the standard error and the sample size is the sum of the propensity score weights.

2.4.4. *Standardization of counts to a common time period*

To be meaningful, the count of days of use should reflect a standard time frame. A count of 5 days of use out of 50 days is clearly not

Table 2
Results for Use vs. No Use.

Effect	Estimate	Standard Error	Z-Value	P-Value
Intercept	0.230	0.23	1.0	0.308
Treatment Group	-0.594	0.19	-3.1	0.002
Age	-0.021	0.02	-1.0	0.311
Gender	-0.998	0.86	-1.2	0.247
TLFB:# Stimulant Use Days	0.074	0.02	4.4	< 0.001
% Attendance at Pre-Study TAU Sessions	0.792	0.48	1.7	0.098
ASI:# Cocaine Use Days	-0.021	0.05	-0.4	0.683
SSSA: Hyperphagia Item	0.125	0.10	1.3	0.199
Study Conducted at Site A (y/n)	1.588	0.27	5.9	< 0.001
ASI: Drug Use Subscale Score	2.240	2.61	0.9	0.391
CPFQ: Total Score	-0.024	0.04	-0.6	0.567
CAST: Sleep Item (Slept Well)	-0.674	0.23	-2.9	0.004
TLFB:# Cocaine Use Days	-0.022	0.04	-0.6	0.538
Study Conducted at Site B (y/n)	0.495	0.73	0.7	0.500
ASI:# Illegal Activity Days	-0.005	0.04	-0.1	0.888
ASI:# Days Incarcerated	0.110	0.08	1.4	0.162

Abbreviations: ASI: Addiction Severity Index; CAST: Concise Associated Symptoms Tracking; CPFQ: Cognitive and Physical Functioning Questionnaire; SSSA: Stimulant Selective Severity Assessment; TAU: Treatment As Usual; TLFB: TimeLine Follow Back.

equivalent to a count of 5 days of use out of 70 days. Thus were the days of use standardized by the following procedure. The proportion of stimulant use days were first computed using available post-RTP TLFB data and then adjusted by multiplying by 63 to provide the days of use in a 63-day period based on the *a priori* assumption of the post-RTP period lasting from day 22 to day 84, as was also assumed in the primary analysis [13].

In participants with missing TLFB data, this standardization has the effect of imputing the same rate of use for missing days as for observed days. Only 15 of the 218 (6.9%) participants used in the CACE analysis had incomplete TLFB data. Reasons for withdrawal from the study were: jail (1 health education), withdrew consent (1 health education), lost to follow-up (8 health education, 2 exercise), moved (1 health education, 1 exercise), and other (1 exercise). The rate of missing TLFB days in the health education and exercise groups was 4.4% and 4.9%, respectively. Given the small amount of missing TLFB data and the similar levels of missing data between groups, a more sophisticated multiple imputation approach was not considered necessary [47].

2.4.5. Hurdle model covariates

Covariate balance between treatment groups is achieved primarily through the propensity score weight. However, as can be seen in the last column of Table 1, not all covariates achieved the same degree of balance as measured by the effect size. Harder et al. [48] proposed a less strict (effect size < 0.25) and more strict (effect size < 0.10) rule of thumb for assessing whether a covariate is in balance, and they suggested that residual imbalances may be reduced by including unbalanced covariates in the final model. Based on these considerations, we include as covariates in the hurdle model all propensity score predictors with a weighted effect size greater than 0.10 and a relative influence greater than 1%. Table 1 shows the 11 additional covariates which fit this criteria and were added to the hurdle model along with the three pre-specified covariates.

3. Results

Altogether, 295 participants (149 exercise group, 146 health education group) provided some post-RTP data. In the CACE propensity score weighted analysis, 74 exercise participants who completed an inadequate exercise dose received a weight of zero (i.e., were not used) and 3 health education participants were not used in the hurdle model due to missing covariates; therefore the analytic sample size was 218

participants (75 exercise group, 143 health education group). The mean age of this sample was 39.4 years (SD = 11.1) and the mean days of stimulant use in the 30 days prior to RTP was 13.0 days (SD = 9.2). Mean time from randomization to discharge from RTP was 18.9 days (SD = 10.3) and did not differ between groups with 20.4 days (SD = 10.8) in RTP in the exercise group and 18.0 days (SD = 10.0) in the health education group (t = 1.63, df = 216, p = 0.106). Females accounted for 36.2% of the sample. Other demographic and clinical characteristics of the sample are presented in Table 1, which provides the list of 32 predictors used in the propensity score model and the effect of the propensity score weighting for each predictor. For example, the mean Addiction Severity Index employment subscale among adequate dose exercise participants was 0.68 and the un-weighted mean in the health education group was 0.72 (effect size = 0.15). After applying the propensity score weights, the weighted health education mean decreased to 0.69, which is closer to the exercise mean of 0.68 and provided a lower effect size of 0.04. Table 1 also includes a measure of the relative importance of each predictor in its contribution to the prediction of adequate exercise dose.

3.1. CACE adjusted hurdle model of stimulant use

The CACE adjusted hurdle model results for use of stimulants among participants who would have achieved an adequate exercise dose if assigned to exercise (Table 2) found a significant group effect for the probability of use (effect size = 0.26, z = -3.1, p = 0.002). The estimated probability of stimulant use for a hypothetical participant with all covariates set to adequate exercise dose group mean levels was 55.7% in the health education group and 41.0% in the exercise group. A significant group effect also emerged for the days of use among those who used (defined conservatively as anyone who used once) (effect size = 0.22, z = -2.7, p = 0.007) (Table 3). Estimated days of use among those who used was 9.9 days in the health education group and 5.0 days in exercise group. Thus, exercise participants experienced a significantly lower relapse rate than did health education participants and once a participant relapsed, the days of use were significantly lower in the exercise group compared to the health education group by an estimated 4.9 days (over a post-RTP period of 63 days).

3.2. Validation of CACE adjustment model

An analysis was done to determine the sensitivity of these results to

Table 3
Results for days of use among those who used.

Effect	Estimate	Standard Error	Z-Value	P-Value
Intercept	2.296	0.14	16.5	0.000
Treatment Group	-0.685	0.25	-2.7	0.007
Age	-0.009	0.01	-1.3	0.207
Gender	-0.789	0.29	-2.7	0.006
TLFB:# Stimulant Use Days	0.018	0.02	1.1	0.278
% Attendance at Pre-Study TAU Sessions	-0.470	0.24	-1.9	0.052
ASI:# Cocaine Use Days	-0.015	0.01	-1.1	0.284
SSSA: Hyperphagia Item	-0.002	0.04	-0.0	0.965
Study Conducted at Site A (y/n)	-0.062	0.20	-0.3	0.756
ASI: Drug Use Subscale Score	-0.044	0.83	-0.1	0.958
CPFQ: Total Score	0.006	0.02	0.4	0.720
CAST: Sleep Item (Slept Well)	-0.192	0.26	-0.7	0.466
TLFB:# Cocaine Use Days	0.011	0.02	0.7	0.498
Study Conducted at Site B (y/n)	0.736	0.19	4.0	0.000
ASI:# Illegal Activity Days	0.026	0.01	1.9	0.061
ASI:# Days Incarcerated	-0.048	0.02	-2.5	0.014

Abbreviations: ASI: Addiction Severity Index; CAST: Concise Associated Symptoms Tracking; CPFQ: Cognitive and Physical Functioning Questionnaire; SSSA: Stimulant Selective Severity Assessment; TAU: Treatment As Usual; TLFB: TimeLine Follow Back.

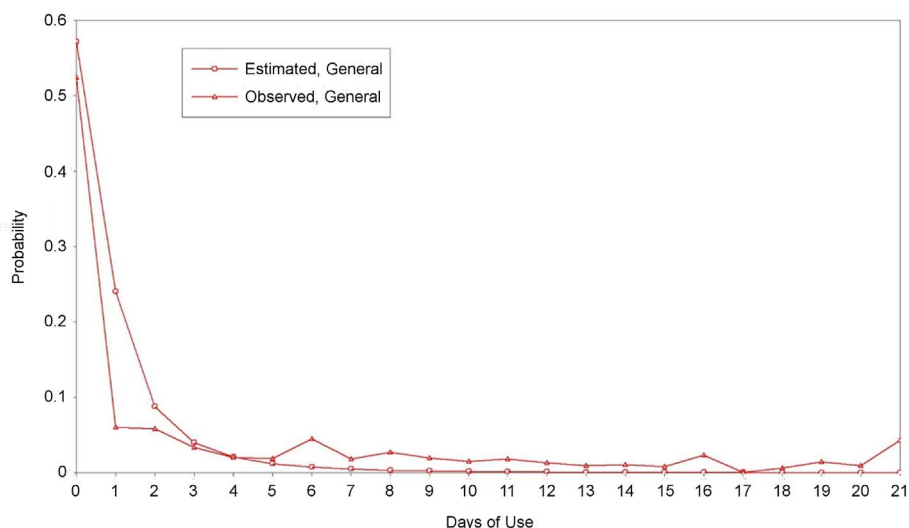


Fig. 1. Observed and expected probability by number of days of use.

violation of the critical assumption that each important predictor of adequate exercise dose had been incorporated into the model used to generate the propensity score weights. The sensitivity analysis required repeating the hurdle model but with modified propensity scores weights and an additional parameter added to the weighting scheme which quantified the degree of departure from this critical assumption, where a value of 1 implies no departure [49]. The hurdle model was rerun with sensitivity parameter values between 0.5 and 2 as recommended in Ding and Lu [49]. The sensitivity analysis found that the results for probability of use were quite robust to departures from model adequacy with very little change in effect sizes or significance levels over the range of sensitivity parameters tested. This was not the case for the results for days of use among users where significance was not obtained for any test of departure from the critical assumption.

As a check on the adequacy of the hurdle model, the distribution of days of use estimated from the model was compared to the observed distribution (Fig. 1).

As another check on the reasonableness of the propensity score weighting procedure, an analysis was done in which propensity score weights were computed to achieve balance between the health education group and the inadequate exercise dose participants [50]. It is expected that the absolute value of the treatment effect would be smaller for inadequate exercise dose participants than for adequate exercise dose participants. The hurdle model estimated probability of stimulant use and days of use were 61.1% and 9.4 days in the health education group and 71.3% and 12.0 days in the exercise group, respectively. In fact, these treatment effects were smaller than those found for the adequate dose analysis above, as expected, and not significant (probability of use: effect size = 0.12, $z = 0.9$, $p = 0.373$, and days of use: effect size = 0.07, $z = 1.4$, $p = 0.145$).

4. Discussion

The CACE efficacy analysis was conducted for the STRIDE study to account for exercise dose in the evaluation of exercise as a potential treatment for stimulant use disorders. This analysis demonstrated statistically significant differences for the probability of stimulant use, such that those who would achieve an adequate exercise dose (defined to be an average of 8.3 KKW or more) have an estimated lower probability of relapsing to any stimulant use. Analyses also demonstrated that, even among those who relapsed, the amount of estimated stimulant use was significantly less among those who would achieve an adequate exercise dose. Together, these results suggest a beneficial effect of exercise in the treatment of stimulant abuse.

The primary strength of this secondary analysis of STRIDE is the statistically rigorous method of data analysis. The CACE-adjusted analysis enabled us to estimate the impact of exercise on stimulant use given participants who exercised in a range between 8.3 and 11.5 KKW, which is higher than range of 0–11.5 KKW actually observed during the study. An additional strength is that our definition of adequate exercise dose is based on the observed level of exercise in our sample. Thus, although those participants defined as having an adequate dose did not necessarily meet the targeted exercise dose of 12 KKW, the median exercise dose (8.3 KKW or greater) used to define adequacy may be more indicative of exercise levels that can be expected in this population.

The primary weakness of the CACE method is that the unbiased comparison of exercise and health education is dependent on the inclusion of all important predictors of adequate exercise dose in the propensity score model. If important predictors of exercise dose have been omitted from the model, then the comparison may be biased. Although it is not possible to know if all important predictors have been included, the sensitivity analysis provided evidence that the lower probability of relapse found in the exercise group compared to the health education group was robust with respect to the exclusion of important predictors from the propensity score model. The fewer days of use among users in the exercise group compared to the health education group was not found to be robust and should be interpreted more cautiously. As such, while our results suggest a potential benefit for exercise, additional studies are needed to prospectively evaluate and replicate this finding using standard intention-to-treat analyses.

Our results indicate that further research in the use of exercise as an intervention for stimulant users is warranted. The use of a median split to define an adequate exercise dose resulted in a dose that was at least two-thirds of the assigned dose. This rate is comparable to other studies of exercise in SUD [51,52]. Given the difficulties experienced by participants in STRIDE regarding adherence to the exercise program, future research is also needed to examine strategies to improve exercise adherence. A survey of individuals receiving treatment for an SUD indicated interest in exercise programs designed specifically for the SUD population [53]. They also indicated an interest in engaging in a variety of exercise activities. Participants in STRIDE were limited to aerobic exercise on a treadmill. Offering alternative exercise activities, such as group exercise classes, walking/running groups, or resistance training, would allow the intervention to match patient preference for exercise activity. Finally, these patients indicated an interest in being able to self-monitor their activity using a pedometer or activity tracker (e.g., Fitbit). Self-monitoring is an established strategy to increase physical

activity [54,55]. Though participants in STRIDE were provided a pedometer, its purpose was to provide an assessment of difference in activity between the two treatment groups and it was not emphasized as a tool to support exercise adherence.

Our results also emphasize the importance of factoring adherence into the interpretation of trial results. Recent clinical trials have also highlighted this issue [56,57], underscoring the potential misinterpretation that may occur if adherence is not accounted for in trials with differential treatment adherence between groups and/or significant non-adherence to treatment. Traditional methods to account for non-adherence may have important limitations and biases, and therefore adequate approaches to statistical adjustment for non-adherence should be carefully considered and applied, when possible, to reduce the likelihood of abandoning the investigation or clinical utilization of potentially effective treatments.

5. Conclusions

Our results demonstrate that an adequate dose of exercise has the significant positive effect of reducing the probability of relapse to stimulant use and reducing the days of stimulant use in those who relapse. Further research is needed to develop adherence strategies that can ensure patients receive an exercise dose sufficient to produce a significant positive effect on treatment.

Funding support and role of funder/sponsor

This work was supported by the National Institute on Drug Abuse of the National Institutes of Health [Award Number U10DA020024 and UG1DA020024 (PI: Trivedi)]. Additional grant support was provided by NIMH [K01 MH097847 (PI: Rethorst)]. The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health. Funders/sponsors had no role in study design; the collection, analysis and interpretation of data; the writing of the report; or the decision to submit the article for publication.

Acknowledgements

The authors are very grateful to Booil Jo, Ph.D., co-developer of the CACE method used in this paper, whose comments on an earlier draft led to substantial improvement in the final version. We are very grateful to our STRIDE participants, as well as all of our participating residential treatment program sites: Wave 1: Arapahoe House, Gateway Community Services Inc., Gibson Recovery Center, Nexus Recovery Center; Wave 2: Charleston VAMC, Memorial Hermann PaRC, Morris Village, Penn Presbyterian Medical Center, and St. Luke's-Roosevelt Hospital. Appreciation also goes to all participating National Drug Abuse Treatment Clinical Trials Network Nodes: Delaware Valley Node, Florida Node Alliance, Greater New York Node, Ohio Valley Node, Texas Node, and Southern Consortium Node. We also greatly appreciate the expertise and guidance of the STRIDE Executive Committee, which includes Colleen Allen, M.P.H., CCRA; Steven N. Blair, PED, FASCM; Jack Chally, M.B.A.; Timothy Church, M.D., Ph.D., MPH; Becca Crowell, M.Ed., Ed.S.; Eve Jelstrom, CRNA, M.B.A.; Robert Lindblad, M.D.; Tiffany L. Linkovich Kyle, Ph.D.; David Liu, M.D.; Bess H. Marcus, Ph.D.; Edward Nunes, M.D.; John P. Rotrosen, M.D.; Eugene Somoza, M.D., Ph.D.; James L. Sorensen, Ph.D.; Michele Straus, R.Ph., MS; Paul Van Veldhuisen, Ph.D.; Diane Warden, Ph.D.; and Jeremy Wolff, B.A. We thank the site PIs: Trey Causey, M.D.; Candace Hodgkins, Ph.D., LPC, LMHC, NCC; Lee Love, M.A., LPC; Hugh Myrick, M.D.; Thomas Northrup, Ph.D.; Paul Rinaldi, Ph.D.; Cindy Seamans, Ph.D.; Meredith Silverstein, Ph.D.; and Regina P. Szucs-Reed, M.D. We are very grateful to Bruce D. Grannemann, M.A., Neal Oden, Ph.D., Kolette Ring, B.A., Angela Stotts, Ph.D., Mark Stoutenberg, Ph.D., and Kathy Shores-Wilson, Ph.D., for their assistance with the project. We thank the EMMES Corporation for serving as the Data and Statistics Center and

the Clinical Coordinating Center. We also thank Carol A. Tamminga, M.D., Communities Foundation of Texas, Inc. Chair in Brain Science, and Chair, Department of Psychiatry, University of Texas Southwestern Medical Center, Savitha Kalidas, Ph.D., and Jennifer Furman, Ph.D. (Dallas, TX) and Jeremy A. Kee, M.A. (Dallas, TX) for their administrative support; and Jon Kilner, MS, MA (Pittsburgh, PA) for his editorial support.

References

- [1] M.S. de Lima, B.G. de Oliveira Soares, A.A. Reisser, M. Farrell, Pharmacological treatment of cocaine dependence: a systematic review, *Addiction* 97 (8) (2002) 931–949.
- [2] W.P. Knapp, B.G. Soares, M. Farrel, M.S. Lima, Psychosocial interventions for cocaine and psychostimulant amphetamines related disorders, *Cochrane Database Syst. Rev.* 3 (2007), <http://dx.doi.org/10.1002/14651858.CD003023.pub2> Cd003023.
- [3] M.S. Buchowski, N.N. Meade, E. Charboneau, S. Park, M.S. Dietrich, R.L. Cowan, P.R. Martin, Aerobic exercise training reduces cannabis craving and use in non-treatment seeking cannabis-dependent adults, *PLoS One* 6 (3) (2011) e17465, <http://dx.doi.org/10.1371/journal.pone.0017465>.
- [4] T.R. Collingwood, R. Reynolds, H.W. Kohl, W. Smith, S. Sloan, Physical fitness effects on substance abuse risk factors and use patterns, *J. Drug Educ.* 21 (1) (1991) 73–84.
- [5] D. Sinyor, T. Brown, L. Rostant, P. Seraganian, The role of a physical fitness program in the treatment of alcoholism, *J. Stud. Alcohol Drugs* 43 (3) (1982) 380–386.
- [6] J. Weinstock, D. Barry, N.M. Petry, Exercise-related activities are associated with positive outcome in contingency management treatment for substance use disorders, *Addict. Behav.* 33 (8) (2008) 1072–1075, <http://dx.doi.org/10.1016/j.addbeh.2008.03.011>.
- [7] B.M. Wipfli, C.D. Rethorst, D.M. Landers, The anxiolytic effects of exercise: a meta-analysis of randomized trials and dose-response analysis, *J. Sport Exerc. Psychol.* 30 (4) (2008) 392–410.
- [8] C.D. Rethorst, B.M. Wipfli, D.M. Landers, The antidepressive effects of exercise: a meta-analysis of randomized trials, *Sports Med.* 39 (6) (2009) 491–511.
- [9] R.K. Dishman, H.R. Berthoud, F.W. Booth, C.W. Cotman, V.R. Edgerton, M.R. Fleshner, S.C. Gandevia, F. Gomez-Pinilla, B.N. Greenwood, C.H. Hillman, A.F. Kramer, B.E. Levin, T.H. Moran, A.A. Russo-Neustadt, J.D. Salamone, J.D. Van Hoomissen, C.E. Wade, D.A. York, M.J. Zigmond, *Neurobiology of exercise*, *Obesity (Silver Spring)* 14 (3) (2006) 345–356.
- [10] W.J. Lynch, A.B. Peterson, V. Sanchez, J. Abel, M.A. Smith, Exercise as a novel treatment for drug addiction: a neurobiological and stage-dependent hypothesis, *Neurosci. Biobehav. Rev.* 37 (8) (2013) 1622–1644.
- [11] S.E. Linke, M. Ussher, Exercise-based treatments for substance use disorders: evidence, theory, and practicality, *Am. J. Drug Alcohol Abuse* 41 (1) (2015) 7–15.
- [12] M. Stoutenberg, C.D. Rethorst, O. Lawson, J.P. Read, Exercise training - a beneficial intervention in the treatment of alcohol use disorders? *Drug Alcohol Depend.* 160 (2016) 2–11.
- [13] M.H. Trivedi, T.L. Greer, C.D. Rethorst, T. Carmody, B.D. Grannemann, R. Walker, D. Warden, K. Shores-Wilson, M. Stoutenberg, N. Oden, M. Silverstein, C. Hodgkins, L. Love, C. Seamans, A. Stotts, T. Causey, R.P. Szucs-Reed, P. Rinaldi, H. Myrick, M. Straus, D. Liu, R. Lindblad, T. Church, S.N. Blair, E.V. Nunes, Randomized controlled trial comparing exercise to health education for stimulant use disorder: results from the CTN-0037 Stimulant Reduction Intervention Using Dosed Exercise (STRIDE) study, *J. Clin. Psychiatr.* (2017 Feb 14), <http://dx.doi.org/10.4088/JCP.15m10591> (Epub ahead of print).
- [14] J.M. Lachin, Statistical considerations in the intent-to-treat principle, *Contr. Clin. Trials* 21 (3) (2000) 167–189.
- [15] N.L. Oden, P.C. VanVeldhuisen, P.G. Wakim, M.H. Trivedi, E. Somoza, D. Lewis, Power of automated algorithms for combining time-line follow-back and urine drug screening test results in stimulant-abuse clinical trials, *Am. J. Drug Alcohol Abuse* 37 (5) (2011) 350–357, <http://dx.doi.org/10.3109/00952990.2011.601777>.
- [16] L.B. Sheiner, D.B. Rubin, Intention-to-treat analysis and the goals of clinical trials, *Clin. Pharmacol. Ther.* 57 (1) (1995) 6–15, [http://dx.doi.org/10.1016/0009-9236\(95\)90260-0](http://dx.doi.org/10.1016/0009-9236(95)90260-0).
- [17] C.R. Knox, R. Lall, Z. Hansen, S.E. Lamb, Treatment compliance and effectiveness of a cognitive behavioural intervention for low back pain: a complier average causal effect approach to the BeST data set, *BMC Musculoskel. Disord.* 15 (2014) 17, <http://dx.doi.org/10.1186/1471-2474-15-17>.
- [18] Y. Liang, B.R. Ehler, C.S. Hollenbeck, B.J. Turner, Behavioral support intervention for uncontrolled hypertension: a complier average causal effect (CACE) analysis, *Med. Care* 53 (2) (2015) e9–e15, <http://dx.doi.org/10.1097/MLR.0b013e31827da928>.
- [19] A.M. Connell, Employing complier average causal effect analytic methods to examine effects of randomized encouragement trials, *Am. J. Drug Alcohol Abuse* 35 (4) (2009) 253–259, <http://dx.doi.org/10.1080/00952990903005882>.
- [20] S. Huang, D. Cordova, Y. Estrada, A.M. Brincks, L.S. Asfour, G. Prado, An application of the Complier Average Causal Effect analysis to examine the effects of a family intervention in reducing illicit drug use among high-risk Hispanic adolescents, *Fam. Process* 53 (2) (2014) 336–347, <http://dx.doi.org/10.1111/famp.12068>.
- [21] T.L. Greer, K.M. Ring, D. Warden, B.D. Grannemann, T.S. Church, E. Somoza, S.N. Blair, J. Szapocznik, M. Stoutenberg, C. Rethorst, R. Walker, D.W. Morris,

- A.S. Kosinski, T. Kyle, B. Marcus, B. Crowell, N. Oden, E. Nunes, M.H. Trivedi, Rationale for using exercise in the treatment of stimulant use disorders, *J. Global Drug Policy and Pract.* 6 (1) (2012) pii <http://ctndisseminationlibrary.org/display/825.htm>.
- [22] M. Stoutenberg, C. Rethorst, G. Fuzat, T. Greer, S. Blair, T. Church, B. Marcus, M. Trivedi, Stimulant reduction intervention using dosed exercise (STRIDE) - description of the exercise intervention and behavioral program to ensure adherence, *Ment. Health Phys. Act.* 5 (2) (2012) 175–182, <http://dx.doi.org/10.1016/j.mhpa.2012.08.003>.
- [23] M.H. Trivedi, T.L. Greer, B.D. Grannemann, T.S. Church, E. Somoza, S.N. Blair, J. Szapocznik, M. Stoutenberg, C. Rethorst, D. Warden, K.M. Ring, R. Walker, D.W. Morris, A.S. Kosinski, T. Kyle, B. Marcus, B. Crowell, N. Oden, E. Nunes, Stimulant reduction intervention using dosed exercise (STRIDE) - CTN 0037: study protocol for a randomized controlled trial, *Trials* 12 (2011) 206, <http://dx.doi.org/10.1186/1745-6215-12-206>.
- [24] R. Walker, D.W. Morris, T.L. Greer, M.H. Trivedi, Research staff training in a multisite randomized clinical trial: methods and recommendations from the Stimulant Reduction Intervention using Dosed Exercise (STRIDE) trial, *Addiction Res. Theor.* 22 (5) (2014) 407–415, <http://dx.doi.org/10.3109/16066359.2013.868446>.
- [25] D. Warden, M.H. Trivedi, T.L. Greer, E. Nunes, B.D. Grannemann, V.E. Horigian, E. Somoza, K. Ring, T. Kyle, J. Szapocznik, Rationale and methods for site selection for a trial using a novel intervention to treat stimulant abuse, *Contemp. Clin. Trials* 33 (1) (2012) 29–37, <http://dx.doi.org/10.1016/j.cct.2011.08.011>.
- [26] C.D. Rethorst, T.L. Greer, B. Grannemann, K.M. Ring, B.H. Marcus, M.H. Trivedi, A health education intervention as the control condition in the CTN-0037 STRIDE multi-site exercise trial: rationale and description, *Ment. Health Physical Act.* 7 (1) (2014) 37–41, <http://dx.doi.org/10.1016/j.mhpa.2013.12.001>.
- [27] B.H. Marcus, A.E. Albrecht, T.K. King, A.F. Parisi, B.M. Pinto, M. Roberts, D.B. Abrams, et al., The efficacy of exercise as an aid for smoking cessation in women: a randomized controlled trial, *Arch. Intern. Med.* 159 (11) (1999) 1229–1234.
- [28] M. Pahor, S.N. Blair, M. Espeland, R. Fielding, T.M. Gill, J.M. Guralnik, S. Studenski, et al., Effects of a physical activity intervention on measures of physical performance: results of the lifestyle interventions and independence for elders pilot (LIFE-P) study, *J. Gerontol. Series A, Biol. Sci. Med. Sci.* 61 (11) (2006) 1157–1165.
- [29] W.J. Rejeski, R.A. Fielding, S.N. Blair, J.M. Guralnik, T.M. Gill, E.C. Hadley, A.B. Newman, et al., The lifestyle interventions and independence for elders (LIFE) pilot study: design and methods, *Contemp. Clin. Trials* 26 (2) (2005) 141.
- [30] L. Sobell, M. Sobell, Timeline follow-back, in: R. Litten, J. Allen (Eds.), *Measuring Alcohol Consumption*, Humana Press, New Jersey, 1992, pp. 41–72.
- [31] K.M. Kampman, J.R. Volpicelli, D.E. McGinnis, A.I. Alterman, R.M. Weinrieb, L. D'Angelo, L.E. Epperson, Reliability and validity of the cocaine selective severity assessment, *Addict Behav.* 23 (4) (1998) 449–461.
- [32] A.J. Rush, M.H. Trivedi, H.M. Ibrahim, T.J. Carmody, B. Arnow, D.N. Klein, J.C. Markowitz, P.T. Ninan, S. Kornstein, R. Manber, M.E. Thase, J.H. Kocsis, M.B. Keller, The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression, *Biol. Psychiatr.* 54 (5) (2003) 573–583.
- [33] M.H. Trivedi, S.R. Wisniewski, D.W. Morris, M. Fava, B.T. Kurian, J.K. Gollan, A.A. Nierenberg, D. Warden, B.N. Gaynes, J.F. Luther, A.J. Rush, Concise Associated Symptoms Tracking scale: a brief self-report and clinician rating of symptoms associated with suicidality, *J. Clin. Psychiatr.* 72 (6) (2011) 765–774, <http://dx.doi.org/10.4088/JCP.11m06840>.
- [34] J.E. Ware, K.K. Snow, M. Kosinski, B. Gandek, SF-36 Health Survey Manual and Interpretation Guide, The Health Institute, New England Medical Center, Boston, MA, 1993.
- [35] M. Fava, L.M. Graves, F. Benazzi, M.J. Scalia, D.V. Iosifescu, J.E. Alpert, G.I. Papakostas, A cross-sectional study of the prevalence of cognitive and physical symptoms during long-term antidepressant treatment, *J. Clin. Psychiatr.* 67 (11) (2006) 1754–1759.
- [36] A.T. McLellan, L. Luborsky, G.E. Woody, C.P. O'Brien, An improved diagnostic evaluation instrument for substance abuse patients. The Addiction Severity Index, *J. Nerv. Ment. Dis.* 168 (1) (1980) 26–33.
- [37] C.E. Frangakis, D.B. Rubin, Principal stratification in causal inference, *Biometrics* 58 (1) (2002) 21–29.
- [38] R.J. Little, D.B. Rubin, Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches, *Annu. Rev. Publ. Health* 21 (2000) 121–145, <http://dx.doi.org/10.1146/annurev.publhealth.21.1.121>.
- [39] J.D. Angrist, G.W. Imbens, D.B. Rubin, Identification of causal effects using instrumental variables, *J. Am. Stat. Assoc.* 91 (434) (1996) (Applications and Case Studies).
- [40] E.A. Stuart, B. Jo, Assessing the sensitivity of methods for estimating principal causal effects, *Stat. Meth. Med. Res.* 24 (6) (2015) 657–674, <http://dx.doi.org/10.1177/0962280211421840>.
- [41] T. Chen, C. Guestrin, Xgboost: a scalable tree boosting system, Accessed 4.24.16, 2016. <https://arxiv.org/abs/1603.02754>.
- [42] T. Chen, T. He, Xgboost: extreme gradient boosting (Accessed 8.16.16), 2016. <https://cran.r-project.org/web/packages/xgboost/vignettes/xgboost.pdf>.
- [43] J.H. Friedman, Greedy function approximation: a gradient boosting machine, *Ann. Stat.* 29 (5) (2001) 1189–1232.
- [44] B.K. Lee, J. Lessler, E.A. Stuart, Improving propensity score weighting using machine learning, *Stat. Med.* 29 (3) (2009) 337–346, <http://dx.doi.org/10.1002/sim.3782>.
- [45] A. Agresti, Y. Min, Simple improved confidence intervals for comparing matched proportions, *Stat. Med.* 24 (5) (2005) 729–740, <http://dx.doi.org/10.1002/sim.1781>.
- [46] L.K. Muthén, B.O. Muthén, *Mplus User's Guide*, seventh ed., Muthén & Muthén, Los Angeles, CA, 2012, pp. 1998–2010 7th edition.
- [47] K.F. Schulz, D.A. Grimes, Sample size slippages in randomised trials: exclusions and the lost and wayward, *Lancet* 359 (9308) (2002) 781–785.
- [48] V.S. Harder, E.A. Stuart, J.C. Anthony, Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research, *Psychol. Meth.* 15 (3) (2010) 234–249, <http://dx.doi.org/10.1037/a0019623>.
- [49] P. Ding, J. Lu, Principal stratification analysis using principal scores, *J. Roy. Stat. Soc. B Stat. Meth.* 79 (2017) 757–777.
- [50] B. Jo, E.A. Stuart, On the use of propensity scores in principal causal effect estimation, *Stat. Med.* 28 (23) (2009) 2857–2875, <http://dx.doi.org/10.1002/sim.3669>.
- [51] R.A. Rawson, J. Chudzynski, L. Mooney, R. Gonzales, A. Ang, D. Dickerson, J. Penate, B.A. Salem, B. Dolezal, C.B. Cooper, Impact of an exercise intervention on methamphetamine use outcomes post-residential treatment care, *Drug Alcohol Depend.* 156 (2009) 21–28.
- [52] R.A. Brown, A.M. Abrantes, J.P. Read, B.H. Marcus, J. Jakicic, D.R. Strong, J.R. Oakley, S.E. Ramsey, C.W. Kahler, G.G. Stuart, M.E. Dubreuil, A.A. Gordon, A pilot study of aerobic exercise as an adjunctive treatment for drug dependence, *Ment. Health Phys. Act.* 3 (1) (2010) 27–34.
- [53] A.M. Abrantes, C.L. Battle, D.R. Strong, E. Ing, M.E. Dubreuil, A. Gordon, R.A. Brown, Exercise preferences of patients in substance abuse treatment, *Ment. Health Phys. Act.* 4 (2) (2011) 79–87.
- [54] M.B. Conroy, K. Yang, O.U. Elci, K.P. Gabriel, M.A. Styn, J. Wang, A.M. Kriska, S.M. Sereika, L.E. Burke, Physical activity self-monitoring and weight loss: 6-month results of the SMART trial, *Med. Sci. Sports Exerc.* 43 (8) (2011) 1568–1574.
- [55] J.K. Vallance, K.S. Courneya, R.C. Plotnikoff, Y. Yasui, J.R. Mackey, Randomized controlled trial of the effects of print materials and step pedometers on physical activity and quality of life in breast cancer survivors, *J. Clin. Oncol.* 25 (17) (2007) 2352–2359.
- [56] N.J. Wiles, K. Fischer, P. Cowen, D. Nutt, T.J. Peters, G. Lewis, I.R. White, Allowing for non-adherence to treatment in a randomized controlled trial of two anti-depressants (citalopram versus reboxetine): an example from the GENPOD trial, *Psychol. Med.* 44 (13) (2014) 2855–2866, <http://dx.doi.org/10.1017/S0033291714000221>.
- [57] Y. Kubo, L.R. Sterling, P.S. Parfrey, K. Gill, K.W. Mahaffey, I. Gioni, M.L. Trotman, B. Dehmel, G.M. Chertow, Assessing the treatment effect in a randomized controlled trial with extensive non-adherence: the EVOLVE trial, *Pharmaceut. Stat.* 14 (3) (2015) 242–251, <http://dx.doi.org/10.1002/pst.1680> Epub 2015 Apr 6. Erratum in: *Pharm. Stat.* 14(4) (2015) 368.