SUBSTANCE USE & MISUSE
2017, VOL. 52, NO. 7, 848–857
http://dx.doi.org/10.1080/10826084.2016.1264970

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

Predicting Change in an Integrated Dual Diagnosis Substance Abuse Intensive Outpatient Program

Edward A. Wise [a], David. L. Streiner[b], and Robert J. Gallop[c]

[a]Mental Health Resources, PLLC, Memphis, Tennessee, USA; [b]Department of Psychiatry, McMaster University, Hamilton, Canada; [c]Applied Statistics Program, West Chester University, West Chester, Pennsylvania, USA

ABSTRACT

Research on routine outcome monitoring in psychotherapy settings is plentiful but not without implementation obstacles. In fact, there is a relative dearth of real-time outcome monitoring in substance use treatment settings. Numerous barriers to the development and implementation of clinical decision support tools and outcome monitoring of substance use patients, including the need to establish expected trajectories of change and use of reliable change indices have been identified (Goodman, McKay, & DePhilippis, 2013). The current study was undertaken to develop expected trajectories of change and to demonstrate the treatment effectiveness of a dual diagnosis intensive outpatient program. The expected trajectories of change for days of substance use and depression scores were developed using predictive equation models from derivation samples and then applied to cross-validation samples. Predictive equations to monitor substance use were developed and validated for all patients and for only patients who were actively using substance at the time of admission, as well as to monitor severity of their depression symptom on a weekly basis. Validation of the equations was assessed through the use of Cohen's kappa (κ), receiver operating characteristic curves, reliable change index, and percentage improvement. Large effect sizes for reductions in substance use (Cohen's $d = .76$) and depressive symptoms ($d = 1.10$) are reported. The best predictive models we developed had absolute accuracy rates ranging from 95 to 100%. The findings from this study indicate that predictive equations for depressive symptoms and days of substance use can be derived and validated on dual diagnosis samples.

KEYWORDS

Substance use outcomes monitoring; predicting substance use change; intensive outpatient outcomes; dual diagnosis outcomes; expected treatment response; predicting substance use treatment outcomes

In spite of the prevalence of dual diagnosis or co-occurring disorders (Grant et al., 2004; National Institute of Drug Abuse [NIDA], 2010; Substance Abuse Mental Health Services Administration [SAMHSA], 2015), the literature related to treatment outcomes for dual diagnosis or co-occurring psychiatric and substance use disorders is sparse (e.g., Drake et al., 2008; Granholm, Anthenelli, Monteiro, Sevcik, & Stoler, 2003; Hesse, 2009; SAMHSA, 2005; Tiet & Mausbach, 2007; Wise, 2010). Furthermore, moving beyond patient outcomes, measurement of routine outcome monitoring and clinical decision support tools to monitor and predict outcomes in substance use practice settings has proved to be challenging (Goodman, McKay, & DePhilippis, 2013). Lambert et al.'s efforts to apply outcome feedback and identify patients at risk for deterioration have been shown to improve outcomes and prevent deterioration (Lambert & Shimokawa, 2011; Shimokawa, Lambert, & Smart, 2010). However, the model used by Lambert, Hansen, and Finch (2001) is based on a derivation sample of 10,000 individuals whose data were used to produce expected recovery curves that were dependent on each individual's baseline severity. This resulted in 50 expected recovery curves consisting of 200 individuals per increment. The challenge of developing similar actuarially derived prediction models based on smaller sample sizes, which could be derived in individual clinics and other settings, including those used in substance use treatments, has thus far been insurmountable.

In fact, Goodman et al. (2013, p. 241) found only one "rigorous study to date" that studied the effects of progress monitoring and feedback to clinicians on symptom change with substance using patients. Crits-Christoph et al. (2012) developed an adapted version of the Outcome Questionnaire-45 (OQ-45; Lambert et al., 1996) for patients with alcohol and drug use problems to monitor expected treatment response under feedback and no feedback conditions. Consistent with the outcomes' monitoring with psychotherapy patients, Crits-Cristoph et al. (2012) reported that "off-track" clients in individual therapy whose therapists received feedback through the clinical decision support tool significantly reduced substance use compared with "off-track" clients whose therapists did not receive this feedback. This study

CONTACT Edward A. Wise Wise@MHRMemphis.com Mental Health Resources, PLLC, 1037 Cresthaven Road, Memphis, TN 38119, USA.

highlights the potential utility of feedback systems to substance use populations in individual therapy. Unfortunately, this approach is limited to extremely large sample sizes such as those used in the development sample of the signal alarm parameters using the OQ-45 (Lambert et al., 1996). Goodman et al. (2013) noted numerous limitations of other studies related to the development of clinical decision support tools and outcome monitoring of substance use patients that were potentially relevant. These included the lack of available measures; the lack of repeated measures; the over sampling of predominantly Caucasian, female, mildly disturbed individuals; the need to establish trajectories of change; and the lack of the use of reliable change indices. Other barriers to the use of these and similar outcome monitoring systems relate to their lack of generalizability to specific types of patients in specific settings.

We previously demonstrated the derivation and validation of predictive equation models based on a relatively small sample of depressed intensive outpatient patients (Wise, Streiner, & Gallop, 2016). However, that predictive equation did not generalize to patients in the dual diagnosis intensive outpatient program (IOP), who were also diagnosed with a depressive condition and co-occurring substance used disorders (SUD). Consequently, this naturalistic study in a real world treatment setting aims to demonstrate a method to develop expected trajectory of change scores for dually diagnosed substance use patients in an IOP. The practical questions we are interested in are directly related to the problems identified by Goodman et al. (2013). That is, can we adopt available measures, administered weekly, to acutely distressed, dual diagnosis patients of mixed race and gender, to assess client improvement and accurately predict the trajectory of change for days of substance use and depressive symptoms? We believe that this is an emerging problem faced by many programs and that a methodologically sound, practical, and cost-effective solution would represent a significant contribution to monitoring treatment progress, providing real-time feedback as a clinical decision support tool, and improving treatment outcomes for specific populations, such as substance use and dual diagnosis patients.

Method

Clinical setting

The study was conducted in a naturalistic, real world treatment setting. An IOP is a state-licensed facility-based program that meets up to 3 hr per day that patients can attend up to five days per week. It is a group-based program that may also include individual or family therapy as well as psychiatric management. The treatment team is composed of Master's level therapists and psychiatrists. The treatment programs are guided by treatment manuals. Broadly speaking, treatment consists of a traditional process group, in which dysfunctional interpersonal relationship themes are addressed, cognitive behavioral groups and skills training groups. The dual diagnosis IOP relies on a motivational interviewing (Miller & Rollnick, 2013) and stages of change in theoretical framework (DiClemente & Prochaska, 1998) that is based on an integrated treatment model and incorporates a harm reduction approach. (The interested reader is referred to Wise, 2010 for further details.)

Participants

All these studies were based on analyses of data gathered for clinical purposes at a multi-disciplinary private practice that operates a dual diagnosis IOP. All study participants provided signed informed consent to use their de-identified data for research purposes. Patients are typically employed or are a family member of an employee and are referred by employee assistance programs, employers, primary care physicians, and behavioral health providers. A semi-structured interview of approximately 90 min is used as the primary intake assessment tool to ensure consistent and reliable collection of information by Master's level clinicians. In addition, all patients complete a battery of self-administered assessment tools prior to being seen by the clinician, including various standardized symptom and functional rating scales. All patients are commercially insured and must pass the pre-authorization process instituted by their respective insurance plans to access this higher level of care. Patients who are imminently suicidal, homicidal, medically unstable, or psychotic are not appropriate for this level of care.

As can be seen in Table 1, these individuals were diagnosed with substance abuse (100%) and primarily depressive disorders (72%), 53% were males, 29% African Americans, and 70% Caucasians, who on average had about 14 years of education. These sample demographics are similar to those reported previously (Wise, 2010) and represent reasonably diverse gender and racial composition similar to our past treatment-seeking populations. Also consistent with the previous article (Wise, 2010), the symptom severity of our dual diagnosis IOP patients was comparable with the SCL-90-R (Derogatis, 1994) inpatient normative sample on 10 of the 12-symptom scales, including all three measures of global distress. In terms of symptom severity, the intensive outpatient sample was not significantly different from the inpatient normative sample on the somatization, obsessive

Table 1. Demographic information.

Variable	Dual diagnosis total IOP sample ($N = 421$)	
Race (%)		
Caucasian	70	
African American	29	
Others	1	
Age (years) (SD)	38.47 (12.73)	
Gender (Males)	53%	
Years education (SD)	14.23 (4.24)	
Number of depressed[a] (%)	303 (72)	
Number of substance abuse[a] (%)	411 (100)	
Complete records for SUD	$N = 380$	
M days used in past week	3.05 (SD = 3.62)	Mdn = 2; range: 0–7
M treatment weeks	4.96 (SD = 1.49)	Mdn = 5; range: 2–7
Actively using at admission	$N = 242$ (64%)	
M days used in past week	4.77 (SD = 3.49)	Mdn = 4; range: 1–7
M treatment weeks	4.97 (SD = 1.47)	Mdn = 5; Range: 2–7
Primary problem at admission		
Alcohol	212 (56%)	
Marijuana	46 (12%)	
Opiates	39 (10%)	
Polysubstance	38 (10%)	
Cocaine	15 (4%)	
Others	30 (8%)	

Notes: IOP = intensive outpatient program, SUD = substance use disorder.
[a]Primary or secondary diagnosis.

compulsive, interpersonal sensitivity, depression, anxiety, hostility, or paranoid ideation scales. Similarly, the indices of general level of distress for the dual diagnosis IOP sample, as measured by the Global Severity Index, Positive Symptom Total and Positive Symptom Distress Indices, were comparable with the inpatient population. Consistent with the above-noted report, 33% of these patients also presented with suicidal ideation and an additional 10% presented with both suicidal and homicidal ideation.

Derivation and validation samples

We had an access to a total of 616 patients ($N = 616$). Of these, 205 were excluded (133 were seen only for intake and never started; 40 were considered dropouts [attended $\leq$3 visits]; and 32 had $<$3 data points), resulting in a total of 411 patients available for the study from the dual diagnosis IOP. Of these, 31 (8%) patients had incomplete substance use data and were omitted from the substance use data analyses, resulting in 380 patients ($N = 380$) for the substance use studies. Table 1 further describes the treatment sample and differences between the subsamples, particularly those who actively engaged in substance use at admission. Since it is common for patients beginning outpatient treatment for substance use to include a significant portion of people who have stopped use prior to treatment engagement (Crits-Christoph et al., 2015), we included these patients in our analyses, even though they could not improve by reducing substance use. Table 2 shows the primary reasons associated with not using alcohol or drugs (AOD) in the week prior to admission ($N = 138$; 36%).

Table 2. Reasons for admission of patients with no AOD use within seven days prior to admission ($N = 138$).

Stepped down from higher level of care (%)	49 (36)
Relapse prevention	38 (28%)
Employer mandate	34 (25%)
Behavior addictions (eating and gambling)	14 (10%)
Others	3 (2%)

The combined samples (samples 1 and 2) were composed of random subsets of all patients, including those who reported using and not using substance in the week prior to admission ($N = 380$); the actively using samples (samples 3 and 4) were composed of random subsets of only those who were actively using substance in the week prior to admission ($N = 242$). In an effort to examine the effects of these different samples on the equations, we also analyzed their data together (e.g., samples 1 and 4). Finally, we analyzed all patients who completed our depression measure ($N = 396$) in an attempt to derive and validate a predictive equation for their co-occurring depressive symptoms (samples 5 and 6). We randomly assigned all patients to derivation and validation samples; each sample served as both derivation and validation sample. This resulted in a total of 12 series of data analyses for six sample combinations.

Measures

Weekly self-report of substance use and symptom data were obtained prior to the initial intake assessment and during the course of treatment from all patients. These measures were collected prior to intake and then on a weekly basis during a regularly scheduled weekly goal group. If clients were absent for their goal group, the scores from that week were considered missing. Each patient answered items related to the frequency of use from the Substance Use subscale of the Maudsley Addiction Profile Self-Completion form (MAP-sc; Luty, Perry, Umoh, & Gormer, 2006). However, several alterations were necessary to adapt this instrument to our practice. Since we were interested in monitoring weekly changes, we altered the time frame from "month" to "week," due to the infrequency with which it was anticipated that heroin users would be treated, instead of using the item related to the number of days heroin was used, we inserted the word "marijuana." Similarly, to ease respondent burden, the items related to cocaine, crack cocaine, benzodiazepines, amphetamines, methadone, hallucinogens, inhalants, or

other drugs, which we believed to be infrequent in our sample, were combined into a single composite item for the number of days any of these drugs were used. The MAP (Marsden, Gossop, Stewart, Best, Farrell, & Strang, 1998) uses a one-month retrospective recall method; it is in the public domain, and has been adapted in a self-administered format with acceptable psychometric properties. The MAP is a widely used tool designed for repeated administrations and has demonstrated a high test-retest reliability ($M = .88$; Marsden et al, 1998). Of particular relevance, the kappa (κ) coefficients for the MAP drug use items with urinalysis screens ranged from $\kappa = .65$–$.79$, all within the "substantial" range (Landis & Koch, 1977). Luty et al. (2006) reported test-retest $r = .86$ for alcohol days used between MAP and MAP-sc. Additional evidence for the validity of our results, the MAP-sc was demonstrated in a previous study which reported that our IOP patients reported significantly more days used in the week prior to admission than their significant other collateral contacts ($t(31) = 2.55, p < .02$; Wise, 2010). As recommended by Marsden et al. (2011), we calculated Reliable Change Indices (RCI; Jacobson & Truax, 1991) for number of days of substance use as a continuous outcome classification to further reduce the threats to reliability due to measurement error.

All patients also completed a weekly self-report depression measure that is a parallel form of the Brief Symptom Inventory-18 Depression Scale (BSI-18; Derogatis, 2001) and is more fully described in Wise et al. (2016). As Wise (2005) has previously reported, the correlations obtained from the BSI-18 and the six-item self-report depression measure used in this study were .76 for the pre-treatment group ($N = 30$) and .79 for the post-treatment group ($N = 50$), demonstrating reasonable equivalence, and sensitivity to change, between the two forms, particularly in light of the few items (Cronbach's alpha = .87). When the pre-treatment reliability study was replicated ($N = 100$), we obtained test-retest $r = .76$. Our depression measure performs as a parallel version of BSI Depression Scale, and in keeping with SCL-90-R (Derogatis, 1983) and BSI (Derogatis, 1993), the item scores are averaged. In calculating RCI for the depression measure (Jacobson & Truax, 1991), we used the pre- and post-depression scores and a test-retest $r = .76$. For the number of days of substance use and depression measure we derived a predictive equation for each sample and then applied it to the relevant cross-validation samples.

Statistical approach

Our initial methodological goal was the assessment of an individual being on-track or off-track with respect to their substance use and psychiatric symptoms during the course of treatment, as demonstrated in the Lambert et al. (2001) study but using a significantly smaller sample. We derived a predictive interval equation as a function of the observed baseline measure and time using the formula in Neter, Wasserman, and Kutner (1989, p. 246) for regression models. Similar to Lambert et al. (2001), we believe that it is desirable for the recovery trajectory to be dependent on an individual's intake score or baseline measure, and to vary over the range of the scale, including missing data at different intervals with varying lengths of treatment as well as length of time between sequential assessments. To accommodate these two features, we fit the following within-subject regression equation:

$$Y_{\Delta i} = b_1\ \text{Week}_i + b_2\ \text{Week}_i \times \text{Baseline}_i + e_i,$$

where $Y_{\Delta i}$ is the change score in the outcome from baseline to post for the ith person, Week_i is the elapsed time in weeks as described above for the ith person, and e_i is the error in fit of the regression model, which is assumed to be normally distributed with a mean of 0 and variance of σ^2. The b_2 term corresponds to the regression coefficient for the interaction of baseline measure and time, which will accommodate the two features of dependency on an individual's baseline and allowing for variability over the range of the scale. The regression line has no intercept term which refers to a regression line through the origin, corresponding to 0 change based on no change in time from baseline. A separate fitted line and error estimate was calculated for every patient, which accounted for missing data and allowed for the comparison of individuals, irrespective of the number of sessions, and time between sessions and varying lengths of treatment. Confidence intervals provide predictions with regard to the average change in score over time, dependent on baseline score, similar to Lambert's "tolerance intervals" (Lambert, 2010, p. 93).

In developing recovery curves, validation is a necessity to ensure that our recovery model is reproducible. Consequently, we derived our equations on random samples from our population and cross-validated these on the validation samples, as described previously. We took three additional validation steps: (1) we calculated reliable change indices (RCI; Jacobson & Truax, 1991) and calculated receiver operating characteristic (ROC) curves based on the on-track flag created by the predictive equation; and (2) we reported the absolute agreement obtained based on the RCI cut scores. In light of the benefits of reporting percentage improvement (PI) figures (Hiller, Shindler, & Lambert, 2012), we repeated these two steps based on the PI method.

Table 3. Predicting number of days of substance use and symptom severity.

Comparison	*n*	Kappa	95% CI	% absolute agreement	% on-track agreement
Predicting number of days of substance use					
Study 1					
1 vs 2	380	.98	.94–1.00	99	84
2 vs 1	380	.71	.49–.93	96	93
Study 2					
3 vs 4	242	.80	.64–.95	95	84
4 vs 3	242	.91	.81–1.00	98	82
Study 3					
1 vs 3	242	.94	.85–1.00	98	84
2 vs 3	242	1.00	1.00–1.00	100	82
1 vs 4	242	1.00	1.00–1.00	100	82
2 vs 4	242	.89	.78–1.00	97	81
4 vs 1	242	.91	.86–.96	98	82
4 vs 2	242	.89	.83–.94	98	82
Predicting depression symptom severity					
Study 4					
5 vs 6	396	.81	.69–.93	95	84
6 vs 5	396	.77	.56–.98	92	77

Results

Study 1: Predicting days of substance use based on combined samples

The initial predictive equation was derived using the number of days of substance use in the preceding week from half of the dual diagnosis sample 1 ($N = 190$; both who used substance in the week prior to admission, and who did not use) and was applied to the patients in sample 2 ($N = 190$). Applying the equation derived from sample 1 to the participants in sample 2 resulted in Cohen's $\kappa = .98$, and 95% Confidence Interval (CI) [.94, 1.00] (Cohen, 1960), falling within the "near perfect" range defined by Landis and Koch (1977). This model classified 30 (16%) patients as "off-track" with only one classification disagreement. For the purpose of illustration, the equation based on sample 2 was then cross validated with sample 1, which resulted in Cohen's $\kappa = .71$, and 95% CI [.49, .93]. In this model, eight (4%) patients were classified as "off-track" and six (3%) did not reach agreement. The equation derived from sample 1 to predict number of days of substance use showed near perfect agreement and outperformed the equation derived from sample 2 (see Table 3, study 1). The same method was followed for each equation, but for the sake of brevity, only the most accurate predictive equation from their respective samples (e.g., sample 1) will be reported in the following text; the remaining summary statistics for the eight additional predictive equations and their respective cross validation results are presented in Table 3.

Study 2: Predicting days of substance use based on active users

There were 138 patients included in the previous results who had not reported any substance use in the week prior to admission in spite of having been diagnosed with a substance abuse disorder. These patients were removed from the dual diagnosis sample, leaving 242 patients ($N = 242$) who had all reported active substance use in the week prior to admission. These patients were randomly divided into two groups, samples 3 and 4 ($N = 141$ each) to determine recovery curves and predictive accuracy of the equations for patients who were actively using substance at the time of admission to IOP. The most accurate equation was derived from sample 4 and applied to sample 3, resulting in Cohen's $\kappa = .91$, and 95% CI [.81, 1.00] with 19 (16%) classified as "off-track" and three (2%) did not reach agreement.

Study 3: Predicting days of substance use based on combined and active users' samples

The equation derived from sample 1 (all patients) applied to the patients in sample 4, composed of only patients who reported use in the week prior to admission, resulted in perfect agreement ($\kappa = 1.00$, 95% CI [1.00, 1.00]), with 22 (18%) patients classified as "off-track" and no misclassifications. To further test the hypothesis that the predictive equation derived from those who were actively using substance on admission might add additional information, we reversed the procedure, applying the equation derived from sample 4 (active users) to sample 1 (combined group) and also obtained a near perfect agreement ($\kappa = .91$, and 95% CI [.86, .96]). In this analysis, 19 (16%) patients were classified as off-track and three (2%) were misclassified.

Study 4: Predicting depression scores

Since the majority of these patients also experienced co-occurring depression, we were also interested in developing predictive equations to monitor recovery from depression. Hence, we obtained two random samples (samples 5 and 6) from the 198 patients with depression data and derived two additional predictive equations to determine patients who were "on-track" and "not on-track" with respect to their expected trajectory of change in depressive symptoms. We again cross validated the prediction equations by applying them to the other sample, but only the results for the more accurate equation are reported here. In this analysis, when the equation from sample 5 was run with the data from sample 6, Cohen's $\kappa = .81$, and 95% CI [.690–.930]. The absolute on-track/off-track agreement rate was 95% in this cross validation. That is, 23 (12%) were flagged as "off-track" and in nine (5%) cases agreement was not reached.

Table 4 shows a summary of the descriptive pre- and post-outcome measures for the combined, actively using,

Table 4. Pre- and post-outcome measures by sample.

	Sample		
	Days used		Depression
	1	4	5
N	190	121	198
Pre-treatment	3.04 (3.56)	4.71 (3.60)	1.92 (1.24)
Post-treatment	1.62 (2.54)	2.24 (2.90)	.76 (.84)
Cohen's *d* effect size	.46	.76	1.10
RCI number (%)			*N* = 173
≥1.28*	67 (35)	45 (37)	92 (53)
≤−1.28	18 (10)	4 (3)	1 (1)
Improvement number (%)			
Remission (100%)	57 (30)	53 (44)	55 (32)
Response (≥50%)	81 (43)	29 (24)	111 (64)

Note: RCI ± 1.28 corresponds to an 80% CI of improvement and/or deterioration.

and depression outcomes (samples 1, 4, and 5, respectively). Samples 1 and 4 showed different pre-treatment severity and PI classification rates. Cohen's *d* for samples 1, 4, and 5 were .46, .76, and 1.10, respectively (where .50 is considered as a medium effect and .80 a large effect; Cohen, 1988). It should be noted that the discrepancy between the effect sizes for the number of days used in samples 1 and 4 is due to the inclusion of patients who were abstinent throughout treatment. Hence, the large effect size with Sample 4 (.76) represented the outcomes of those who were actively using substance at the time of admission.

Practical utility: Outcome measurement validation

We used RCI and PI scores with ROC procedures as another validation strategy. We calculated these scores for each patient and then used the ROC procedures to determine the sensitivity and specificity of the predictive equations.

Days of substance use RCI analysis

Based on the equation derived from sample 1 (combined sample), we calculated days of substance used RCI scores for each individual in sample 1. When the continuous RCI results were plotted against the dichotomous on-track/off-track prediction, the area under the curve (AUC) = .78, and 95% CI [.68, .88]. The best cut-point to maximize sensitivity and specificity was 1.45, where sensitivity = .91 and specificity = .55, yielding a value for *J* = .45. We then used these RCI scores to categorize improvement rates. (The RCI analysis relied upon the test-retest $r = .87$ for the alcohol days used self-report; Luty et al., 2006). The RCI results of all patients indicated that 67 patients (35%) could be classified as reliably improved with respect to days of substance use in the preceding week, using on the 1.28 (80%) cut-off (Table 4). Based on the most liberal RCI criterion for deterioration (≤−1.28), 10% of this sample was deteriorated, and 55% (105) was classified as indeterminant (<1.28 and >−1.28). Within the indeterminant category, 41 patients (39%) reported no substance use throughout IOP. In fact, because sample 1 included 68 patients (36%) who could not improve (baseline = 0 days used), the improvement rates from sample 1 are under-estimates.

When the RCI results for days of substance use derived from sample 4 (active users) were plotted against the on-track/off-track prediction flag in the ROC analysis, the AUC = .85, and 95% CI [.73, .96]. The best cut-point to maximize sensitivity and specificity was 1.55, where sensitivity = .92 and specificity = .63, yielding a value for *J* = .55. Using the RCI criteria above, the RCI scores for active users indicated that 45 (37%) would be classified as reliably improved with respect to days of substance use in the preceding week, 4 (3%) deteriorated, and 72 (60%) were classified as indeterminant (Table 4).

Depression RCI analysis

In order to validate the predictive equation derived from sample 5, we conducted an RCI analysis on depression scores. The ROC analysis for the RCI by on-track/off-track prediction resulted in an AUC = .65, and 95% CI [.55, .74]. The best cut-point to maximize sensitivity and specificity was 1.14, where sensitivity = .72 and specificity = .42 and *J* = .14. Eliminating the 25 patients who scored in the asymptomatic range (≤.23; Wise, 2003) from the RCI, left *N* = 173. Based on these RCI cut-off classifications, 53% (92) could be classified as reliably improved on their depression scores utilizing the 1.28 (80%) confidence interval (Table 4, Sample 5). Only one individual deteriorated based on the RCI criteria, and 46% (80) were classified in the indeterminant range.

Percent improvement

Hiller et al. (2012, p. 4) demonstrated that PI rates were valid psychotherapy outcome measures, allowed for between-study comparisons, accounted for baseline severity, achieved convergent validity with clinical judgment, and could be readily understood and communicated. However, rather than using the PI methodology with all patients, the authors argued in favor of only including symptomatic patients and calculating Response rates to require a reduction of ≥ 25% "on the entire range of the instrument." Due to the restricted range of our scales (i.e., 0–7 days; 0–4 depression severity), a 25% reduction in range would arbitrarily exclude improved patients. Hiller et al. (2012, p. 2) also noted, however, that

to be classified as a Responder, the conventional calculation of PI simply requires "≥ 50% reduction" in symptom severity. Hence, we adopted the conventional calculation and identified patients who improved ≥50% as Responders (e.g., 50% reduction in days used). Remission was defined as 100% change, crossing over from prior use to abstinence by the last measure. As a final cross-validation step, we calculated PI rates for the samples with the best predictive equations (samples 1, 4, and 5) and performed ROC analyses to determine the sensitivity and specificity of classification predictions.

Days of substance use PI: Combined sample

We calculated a percentage of change score based on pre- and post-treatment number of days used in the past week and multiplied this by 100. Using this as a continuous measure, we then used the on-track and off-track flags to calculate an ROC (AUC = .82, and 95% CI [.72, .89]) for sample 1. The best cut point to maximize sensitivity and specificity for predicting days used was $J = .71$, where sensitivity = .91 and specificity = .55. We then conservatively defined Remission as 100% change, crossing over from days used into the asymptomatic range of abstinence or no use by the last week of treatment; 30% of the sample achieved Remission (Table 4). Based on the conventional ≥50% symptom improvement rate as the definition of a treatment Response, 43% of the sample achieved a Response. An additional 11% ($N = 20$) of the sample showed reduction of days used, consistent with our harm reduction philosophy, ranging from 20–49%. Sixty-three patients (33%) reported no change in the number of days used during treatment, but this included a subgroup of patients who were abstinent throughout IOP ($N = 41$). In fact, the total abstinence rate reported by sample 1 in the last week of treatment prior to discharge was 57% (108). Thirty-two patients (17%) showed an increase in the number of days used from admission to discharge.

Days of substance use PI: Active users

The predictive equation obtained from sample 4 resulted in Cohen's $\kappa = .91$ for on-track classification agreement with sample 3 and an absolute agreement rate of 98% in predicting change for other active users, second only to the equation derived from sample 1. As above, with days of substance use as a continuous measure for the on-track/off-track flags derived from sample 4, AUC = .94, and 95% CI [.89, .99]. The best cut point to maximize sensitivity and specificity for predicting days used was $J = .81$, where sensitivity = .86 and specificity = .95. Sample 4 obtained a Remission rate of 44%, and 24% were classified as Responders (Table 4). As might be expected, the no-change classification group was reduced to 22%, probably due to the elimination of people who reported no substance use prior to or during IOP.

Depression scores PI

We ran a similar analysis using a percentage of change score based on the depression scale and included only those patients who scored in the symptomatic range at baseline ($N = 173$). On the depression measure, AUC = .89, and 95% CI [.84, .93]; the best cut point to maximize sensitivity was $J = .43$, where sensitivity = .98 and specificity = .45. Based on the depression scale, 32% (55) were classified in the Remission category, 64% (111) achieved a Response, and 4% (6) reported no change in depressive symptoms (Table 4).

Discussion

The best predictive models that we developed for the three samples of interest (actively using at admission, all users, and depressed) had absolute accuracy rates ranging from 95 to 100%. Our method demonstrates that trajectories of change can be modeled on smaller samples than previously demonstrated and used to monitor "off-track" status quite accurately. In light of the prevalence of IOPs delivering substance use treatment, the development of routine outcomes monitoring and the prediction of "off-track" individuals has the potential to identify individuals at risk of deterioration and thereby improve treatment engagement, retention, and outcomes. This approach addresses the significant barriers raised by Goodman et al. (2013) related to outcomes monitoring in substance use treatment settings by utilizing available measures; with repeated measures; on more diverse and severely distressed individuals; developing and validating expected trajectories of change; and using RCI and PI measures to validate outcomes.

The use of routine outcomes monitoring in higher level of care settings has numerous practical implications. The most important implication is that these studies demonstrate routine outcome monitoring and clinical decision support tools can be derived from relatively small samples, utilizing available measures and adapted for specific populations. Based on expected trajectories of change, the identification of an off-track individual could trigger a review of circumstances, potentially contributing to this event. In the current dual diagnosis IOP, it would be possible for an individual to remain on-track with one measure (depression) and off-track with the other (days used). Following the identification of an off-track score, such a review might reveal that a recent interpersonal crisis has occurred and has resulted in an off-track flag on the

depression measure, suggesting that more support and days of treatment might be indicated. On the other hand, the identification of an off-track alert for substance use could signal an early lapse and provide an opportunity for a reassessment of factors contributing to the lapse and result in a redesigned treatment plan.

Within this group of dually diagnosed IOP patients, an easily identifiable sub-group comprised those who were actively using substance at admission, compared with a larger group that included all patients admitted. Compared with sample 1, sample 4 demonstrated a relatively higher average number of days used in the week prior to admission (4.77), a larger effect size (.76), and higher Response and Remission rates (44% and 68% respectively). The existence of relatively homogeneous sub-groups is consistent with the research of Zheng, Cleveland, Molenaar, and Harris (2015), who demonstrated that sub-groups of individuals in recovery cope with daily relapse risk and negative affect through different methods of coping, and hence demonstrate different treatment outcomes.

The differences in treatment outcomes between samples 1 and 4 may also have been due in part to the methodological problems associated with the no-change or indeterminant groups, which included patients who were abstinent throughout the treatment (sample 1, $N = 41$), and hence unable to change. While maintaining abstinence would be considered a successful treatment outcome in the real world, these individuals were classified as indeterminant or unchanged in the outcome analysis. The overall abstinence rate at discharge for Sample 1 was 57% and for Sample 4, it was 44%. Furthermore, 65% of those from sample 1 who were classified in the no-change group reportedly were abstinent when they entered IOP and remained abstinent throughout IOP.

In spite of differences between samples 1 and 4 (substance use in the week prior to admission, pre-treatment severity, Cohen's *d*, and Response rates), the equation derived from sample 1 (all patients) applied very well (average Cohen's $\kappa = .97$) to both samples and appears to predict an expected trajectory of change for those who report substance use in the week prior to admission as well as those who do not report substance use in the week prior to admission. The ROC analysis further validated the predictive equation for days of substance use. The RCI scores from samples 1 and 4 showed comparable levels of improvement and their deterioration rates were within the range reported in the literature (Lambert, 2010). The RCI scores of active users from a previously published study (Wise, 2010) found a 6% deterioration rate, comparable with the 4% reported here. Compared with RCI, the PI method was more liberal in classifying patients who improved, for both days of substance use and depression scores. Irrespective of the method of measuring outcomes one uses (PI or RCI), the predictive equations derived from samples 1 and 4 appear to provide the foundation for a clinical decision support tool to identify which patients using substances are "on-track" and "off-track" of their expected recovery curves as related to their final outcome classification. If one were interested in an equation based on all patients, including those who were abstinent at admission, Equation 1 would be preferred. On the other hand, if it was deemed more desirable to monitor only those who could reduce their number of days of substance use, the equation derived from sample 4 (active users) would be preferred.

With respect to depressive symptoms, 53% of these dually diagnosed patients achieved reliable change and 64% achieved a Response, while only 1% deteriorated. This 1% deterioration rate in depressive symptoms is comparable with the 3% deterioration rate reported previously by Wise (2010). Based on Cohen's *d*, their response to treatment on the depression measure was quite large ($d = 1.10$). It is conceivable that this sample showed a stronger and perhaps more rapid response to their depression treatment, suggesting multiple interactions and patterns of underlying change. This finding is consistent with a recent meta-analysis, which found that the response for depressive symptoms was achieved sooner than a reduction in alcohol use (Riper et al., 2014) and a reduction in depression has also been associated with a later reduction in substance use (Hunter et al., 2012; Worley et al., 2012). In light of the absolute agreement and on-track agreement rates (95% and 84% respectively) found with the equation derived from sample 5, this equation should be of considerable assistance in identifying those individuals who are either indeterminant or off-track from their expected depression trajectory of change and therefore not responding as expected to IOP.

There are a number of limitations with this study. First, the sample came from only one treatment setting; and second, the sample size was relatively small, especially in comparison to that used by Lambert et al. (2001). However, these weaknesses can also be seen in a different light. This paper demonstrates that small outpatient clinics and practices delivering treatment in naturalistic real world clinical settings can derive prediction equations specific to their populations of clinical interest, using statistical techniques that are known and available to most practitioner-researchers. Unlike other methods relying on very large sample sizes, this method does not adjust the expected trajectory of change based on the most recent score and these specific curves are not expected to generalize to dissimilar patients or treatment settings. For example, in light of harm reduction, stages of change, and motivational interviewing philosophy of this particular IOP, it is

not clear whether the expected recovery would generalize to 12-step abstinence based IOPs. The method does, however, create empirically derived expected recovery curves based on average expected treatment responses utilizing relatively smaller sample sizes, and the method has been proven with various treatment seeking populations and could easily be applied to varied settings, including abstinence-based programs. Another limitation relates to the inclusion of patients who did not use any substances in the week prior to admission. Patients who do not use substances prior to admission cannot improve and can only increase days used, and therefore can only dilute or negatively skew outcome results. While excluding these patients from data analysis does show improved treatment effectiveness, excluding these patients entirely from the data analysis would pose a threat to the validity of this naturalistic treatment study. For example, Crits-Christoph et al. (2012, 2015) have pointed out that the majority of their patients, from a four-site substance use study, started treatment with no use of substances in the week prior to admission. They also included these patients in their analyses because their results showed this was a common occurrence in this treatment seeking population in their real world treatment settings.

The future studies interested in replicating this predictive model could circumscribe comorbid psychiatric conditions to more homogeneous diagnostic categories, particularly when using small sample sizes. Another alternative would be to use a broad band measure designed to capture a broader range of general distress, negative affect, or the addition of items that match the population under study (e.g., anxiety). Recovery curves for specific drugs of choice (e.g., cocaine, heroin, alcohol, marijuana, etc.) might also be of interest. Other baseline predictor variables might be identified that mediate outcomes. The current study also shows that predicting and monitoring co-occurring depression and substance use trajectories of change can be achieved with dual diagnosis IOP patients and can provide foundation for a clinical support decision tool.

Conclusions

The findings of this study indicate that predictive equations for depressive symptoms and days of substance use can be derived and validated on dual diagnosis samples. These findings also confirm our previous report that by using this method (1) it is possible to derive equations for relatively homogeneous patient groups that can identify patients who are off-track; (2) it can be done using relatively small sample sizes; and (3) relatively homogeneous sub-groups may benefit from equations derived from their unique sub-samples in some cases (e.g., major depression vs. co-occurring substance abuse and major depression). This study is the first report utilizing this method to derive and validate expected treatment response curves using dual diagnosis IOP patients and demonstrates that this method can be applied in substance use treatment settings to derive predictive equations for expected treatment recovery curves and management of routine outcomes.

Declaration of interest

The authors declare that they have no conflicts of interest. The authors alone are responsible for the content and writing of the article.

ORCID

Edward A. Wise http://orcid.org/0000-0003-3682-790X

References

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Crits-Christoph, P., Markell, H. M., Gallop, R., Gibbons, M. B., McClure, B., & Rotrosen, J. (2015). Predicting outcome of substance abuse treatment in a feedback study: Can recovery curves be improved upon? *Psychotherapy Research*, *25*, 694–704.

Crits-Christoph, P., Ring-Kurtz, S., Hamilton, J. L., Lambert, M. J., Gallop, R., McClure, B., … Rotrosen, J. (2012). A preliminary study of the effects of individual patient-level feedback in outpatient substance abuse treatment programs. *Journal of Substance Abuse Treatment*, *42*, 301–309.

Derogatis, L. R. (1983). *SCL-90-R: Administration, scoring and procedures manual-II*. Towson, MD: Clinical Psychometric Research.

Derogatis, L. R. (1993). *Brief Symptom Inventory (BSI): Administration, scoring and procedures manual* (4th ed.). Minneapolis, MN: NCS Pearson.

Derogatis, L. (1994). *SCL-90-R: Administration, scoring and procedures manual-II*. Minneapolis, MN: NCS Pearson.

Derogatis, L. (2001). *Brief symptom inventory: Administration, scoring and procedures manual*. Minneapolis, MN: NCS Pearson.

DiClemente, C. C., & Prochaska, J. (1998). Toward a comprehensive, transtheoretical model of change: Stages of change and addictive behaviors. In W. R. Miller & N. Heather (Eds.), *Treating addictive behaviors* (2nd ed., pp. 3–24). New York, NY: Plenum Press.

Drake, R. E., Mercer-McFadden, C., Mueser, K. T., McHugo, G. J., & Bond, G. R. (1998). Review of integrated mental health and substance abuse treatment for patients with dual disorders. *Schizophrenia Bulletin*, *24*, 589–608.

Drake, R. E., O'Neal, E., & Wallach, M. (2008). A systematic review of psychosocial research on psychosocial interventions for people with co-occurring severe mental and substance abuse disorders. *Journal of Substance Abuse Treatment*, *34*, 123–138.

Finch, A. E., Lambert, M. J., & Schaalje, B. G. (2001). Psychotherapy quality control: The statistical generation of expected recovery curves for integration into an early warning system. *Clinical Psychology and Psychotherapy*, *8*, 231–242.

Goodman, J. D., McKay, J. R., & DePhilippis, D. (2013). Progress monitoring in mental health and addiction treatment: A means of improving care. *Professional Psychology: Research and Practice*, *44*, 231–246.

Granholm, E., Anthenelli, R., Monteiro, R., Sevcik, J., & Stolar, M. (2003). Brief integrated outpatient dual diagnosis treatment reduces psychiatric hospitalizations. *American Journal of Addictions*, *12*, 306–313.

Grant, B. F., Stinson, F. S., Dawson, D. A., Chou, S. P., Dufour, M. C., Compton, W. … Kaplan, K. (2004). Prevalence and co-occurrence of substance use disorders and independent mood and anxiety disorders: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Archives of General Psychiatry*, *61*, 807–816.

Hesse, M. (2009). Integrated psychological treatment for substance use and comorbid anxiety or depression vs. treatment for substance use alone. A systematic review of the published literature. *BMC Psychiatry*, *9*, 6. doi:10.1186/1471-244X-9-6

Hiller, W., Schindler, A. C., & Lambert, M. J. (2012). Defining response and remission in psychotherapy research: A comparison of the RCI and the method of percent improvement. *Psychotherapy Research*, *22*, 1–11.

Hunter, S. B., Watkins, K. E., Hepner, K. A., Paddock, S. M., Ewing, B. A., Osilla, K. C., & Perry, S. (2012). Treating depression and substance use: A randomized controlled trial. *Journal of Substance Abuse Treatment*, *43*, 137–151.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12–19.

Lambert, M. J. (2010). *Prevention of treatment failure*. Washington, DC: American Psychological Association.

Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology*, *69*, 159–172.

Lambert, M. J., Hansen, N. B., Umphress, V., Lunnen, K., Okiishi, J., & Burlingame, G. M. (1996). *Administration and scoring manual for the OQ-45.2*. Wilmington, DE: American Professional Credentialing Services.

Lambert, M. J., & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy*, *48*, 72–79.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.

Luty, J., Perry, V., Umoh, O., & Gormer, D. (2006). Validation and development of a self-report outcome measures (MAP-sc) in opiate addiction. *Psychiatric Bulletin*, *30*, 134–139.

Marsden, J., Eastwood, B., Wright, C., Bradbury, C., Knight, J., & Hammond, P. (2011). How best to measure change in evaluations of treatment for substance use disorder. *Addiction*, *106*, 294–302.

Marsden, J., Gossop, M., Stewart, D., Best, D., Farrell, M., & Strang, J. (1998). *The Maudsley addiction profile: A brief instrument for treatment outcome research. Development and user manual*. London, UK: National Addiction Centre, Institute of Psychiatry.

McCarty, D., Braude, L., Lyman, D. R., Dougherty, R. H., Daniels, A. S., Ghosem, S. S., & Delphin-Rittmon, M. E. (2014). Substance abuse intensive outpatient programs: Assessing the evidence. *Psychiatric Service*, *65*, 718–726.

Miller, W. R., & Rollnick, S. (2013). *Motivational interviewing: Helping people change* (3rd ed.). New York, NY: Guilford Press.

National Institute of Drug Abuse (NIDA). (2010). *Comorbidity: Addiction and other mental disorders*. Retrieved from http://www.drugabuse.gov/publications/drugfacts/comorbidity-addiction-other-mental-disorders

Neter, J., Wasserman, W., & Kutner, M. H. (1989). *Applied linear regression models*. Homewood, IL: Irwin.

Riper, H., Andersson, G., Hunter, S. B., de Wit, J., Berking, M., & Cuijpers, P. (2014). Treatment of comorbid alcohol use disorders and depression with cognitive-behavioural therapy and motivational interviewing: A meta-analysis. *Addiction*, *109*, 394–406.

Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology*, *78*, 298–311.

Substance Abuse and Mental Health Services Administration (SAMHSA). (2005). *Substance abuse treatment for persons with co-occurring disorders. Treatment Improvement Protocol (TIP) Series, No. 42. HHS publication No. (SMA) 133992*. Rockville, MD: Author.

Substance Abuse and Mental Health Services Administration (SAMHSA). (2015). *Suicidal thoughts and behavior among adults: Results from the 2014 national survey on drug use and health*. Rockville, MD: Center for Behavioral Health Statistics and Quality.

Tiet, Q., & Mausbach, B. (2007). Treatments for patients with dual diagnosis: A review. *Alcoholism: Clinical and Experimental Research*, *31*, 513–536.

Wise, E. A. (2003). Psychotherapy outcome and satisfaction methods applied to intensive outpatient programming in a private practice setting. *Psychotherapy: Theory, Research and Practice*, *40*, 203–214.

Wise, E. A. (2005). Effectiveness of intensive outpatient programming in private practice: Integrating practice, outcomes and business. *American Psychologist*, *60*, 885–895.

Wise, E. A. (2010). Evidence-based effectiveness of a private practice intensive outpatient program with dual diagnosis patients. *Journal of Dual Diagnosis*, *6*, 25–45.

Wise, E. A., Streiner, D. L., & Gallop, R. (2016). Predicting individual change during the course of treatment. *Psychotherapy Research*, *26*, 623–631.

Worley, M. J., Trim, R. S., Roesch, S. C., Mrnak-Meyer, J., Tate, S. R., & Brown, S. A. (2012). Comorbid depression and substance use disorder: Longitudinal associations between symptoms in a controlled trial. *Journal of Substance Abuse Treatment*, *43*, 291–302.

Zheng, Y., Cleveland, H. H., Molenaar, P. C., & Harris, K. S. (2015). An alternative framework to investigating and understanding intraindividual processes in substance abuse recovery: An idiographic approach and demonstration. *Evaluation Review*, *39*, 229–354. doi:10.1177/0193841X14567313