ORIGINAL RESEARCH

Measuring recovery in opioid use disorder: clinical utility and psychometric properties of the Treatment Effectiveness Assessment

Walter Ling[1]
Vijay R Nadipelli[2]
Caitlyn T Solem[3]
David Farabee[1,4]
Naoko A Ronquest[2]
Brian Perrochet[1,5]
Susan M Learned[6]
Chinmay G Deshpande[3]
Christian Heidbreder[7]

[1]UCLA Department of Family Medicine, Center for Behavioral and Addiction Medicine, Los Angeles, CA, USA; [2]Global Health Economics and Outcomes Research, Indivior Inc, Richmond, VA, USA; [3]Patient-centered outcomes Center of Excellence, Pharmerit International, Bethesda, ML, USA; [4]Department of Population Health, School of Medicine,, New York, NY, USA; [5]Marron Institute of Urban Management, Marron Institute, New York University, New York, NY, USA; [6]Global Medicines Development, Indivior Inc, Richmond, VA, USA; [7]Global Research and Development, Indivior Inc, Richmond, VA, USA

Purpose: The Treatment Effectiveness Assessment (TEA) is a patient-centered instrument for evaluating treatment progress and recovery from substance use disorders, including opioid use disorder (OUD). We assessed the TEA's reliability and validity and determined minimal clinically important differences (MIDs) in participants with moderate to severe OUD.

Patients and methods: The TEA measures change in four single-item domains (substance use, health, lifestyle, community involvement) from treatment initiation across the duration of a treatment program. Self-reported responses range from 1 ("none or not much") to 10 ("much better") with items summed to a total score ranging from 4–40. We assessed floor and ceiling effects, internal consistency, test-retest reliability, known-groups validity (ANOVA stratified by current health status [36-Item Short Form Health Survey item 1]), convergent/divergent validity, and MIDs using data from a phase 3, open-label clinical trial of buprenorphine extended-release monthly injection for subcutaneous use (BUP-XR). Participants with OUD completed the TEA at screening and before monthly injections for up to 12 months.

Results: Among 410 participants (mean age 38 years; 64% male), the mean baseline (pre-injection 1) TEA total score was 25.4 (SD 9.7), with <10% of participants at the measure floor and 10%–20% at the ceiling across domains. Internal consistency was high (Cronbach's α=0.90), with marginal test-retest reliability (intraclass correlation coefficient =0.69). Mean TEA total score consistently increased from baseline (n=410; mean 25.4 [SD 9.7]) to end of study (n=337; 35.0 [6.7]) and differentiated between current health status groups ($P<0.001$); it was weakly correlated with other measures of health-related quality of life/severity. MIDs ranged from 5–8 for the TEA total score across anchor- and distribution-based approaches.

Conclusion: The TEA exhibited acceptable reliability and validity in a cohort of participants with moderate to severe OUD treated with BUP-XR. Given its brevity and psychometric properties, the TEA is a promising tool for use in clinical practice and research.

Keywords: Treatment Effectiveness Assessment, TEA, opioid use disorder, patient-reported outcomes, addiction, psychometrics

Introduction

Most studies evaluating the treatment of substance use disorders (SUDs), including opioid use disorder (OUD), focus on measuring retention and abstinence, paying little attention to recovery. Retention and abstinence alone may not adequately describe a person's recovery journey toward regaining and maintaining a meaningful and useful life.[1,2] Tools clinicians can use to easily measure gains in life

Correspondence: Walter Ling
UCLA Department of Family Medicine, Center for Behavioral & Addiction Medicine, 16556 Park Lane Circle, Los Angeles, CA 90049, USA
Tel +1 310 993 8111; +1 310 476 6940
Email lwalter@ucla.edu

activities that promote and help sustain comprehensive recovery in people with SUDs are lacking.

The Treatment Effectiveness Assessment (TEA),[3] designed to address this need, is a brief instrument used to assess recovery outcomes in people with SUDs receiving treatment. The TEA assesses outcomes through the use of questions within four major dimensions involved in recovery: substance use (eg, drugs, alcohol, tobacco), health (eg, physical and emotional health), lifestyle (eg, housing or living situation, family, employment, relationships), and community (eg, obeying laws and becoming a responsible member of society).

The TEA was included as a patient-centered measure of recovery from OUD in a 12-month open-label safety study (NCT02510014)[4] evaluating buprenorphine extended-release for monthly subcutaneous injection (CIII) (BUP-XR, SUBLOCADE™, Indivior Inc), which was shown to be efficacious and safe for the treatment of moderate or severe OUD.[5] As this was the first use of the TEA in an OUD phase 3 clinical trial to our knowledge, we assessed the reliability, validity, and responsiveness of the TEA in people receiving BUP-XR for moderate to severe OUD. Additionally, we determined the minimal clinically important differences (MIDs) in TEA total scores for reference in future studies.

Patients and methods
Study design
This multicenter, phase 3, open-label, long-term safety study (NCT02510014)[4] was conducted in accordance with International Council for Harmonisation Good Clinical Practice guidelines and US Food and Drug Administration (FDA) regulations governing clinical study conduct. An institutional review board reviewed and approved the study protocol, amendments, informed consent form, and all other appropriate study-related information. All patients provided written informed consent.

Following study enrollment (screening), participants began a 3-day induction period with buprenorphine/naloxone sublingual film (2–24 mg); buprenorphine/naloxone doses were adjusted for up to 11 days, followed by a 48-week BUP-XR treatment period. On day 1 of the open-label study, participants received BUP-XR 300 mg. Subsequent doses (300 mg or 100 mg at investigator's discretion) were administered every 28 (−2/+4) days. All participants also received manual-guided individual behavioral therapy. While the long-term safety study itself included participants who rolled over from the randomized placebo-controlled trial of BUP-XR, the current analysis only focused on those participants who were newly enrolled (de novo) within the safety trial.

Measures
The TEA measures improvement associated with treatment along four single-item domains: substance use, health, lifestyle, and community. Scores range from 1 ("none or not much") to 10 ("much better"), resulting in a TEA total score ranging from 4–40. The TEA allows respondents to consider the characteristics within the domains that are most important to them in evaluating how their lives have changed since initiating treatment. Within the trial, the TEA was completed at initial screening and monthly before BUP-XR injections 1 (ie, BUP-XR baseline) through 12 and at the end-of-study visit. As the majority of analyses required that the participant completed the baseline assessment, only participants who completed the TEA at baseline were included in the TEA analysis for psychometric assessment.

Other key measures used in the study to monitor clinical status and symptoms included the opioid craving (OC) visual analog scale (VAS) and the Addiction Severity Index-Lite (ASI-Lite).[6–9] Measures of health status and health-related quality of life (HRQoL) included the 5-level EuroQol 5 dimensions instrument (EQ-5D-5L),[10] which is a generic measure of health status, and the 36-Item Short Form Health Survey, Version 2 (SF-36).[11]

Statistical analysis
Statistical analyses were undertaken to assess the psychometric properties of the TEA. The proportions of participants indicating each level of response were calculated to assess floor and ceiling effects[12] at baseline for TEA single-item domains.

Tests of reliability included an assessment of internal consistency (the extent to which items constituting a scale measure the same concept) as well as test-retest reliability (a measure of whether the instrument would give consistent scores over time). Internal consistency of the TEA total score was assessed at baseline using Cronbach's alpha. Test-retest reliability for TEA single-item domains and total score was examined using intraclass correlations between responses at BUP-XR injection 12 and end of study (approximately 1 month apart) among people who had stable urine drug screen results between the two visits (ie, either both opioid positive or both opioid negative).

These time point visits were chosen based on the expectation that most improvement related to pharmacotherapy should occur early on and then remain more stable. A threshold of 0.7 for Cronbach's alpha[13] and intraclass correlation coefficients (ICCs)[14] was used to evaluate reliability.

Also assessed were known-groups validity (ie, the extent to which scores can distinguish among groups hypothesized a priori to be different), convergent validity (ie, whether two different scores measuring similar concepts were associated), and divergent validity (ie, whether different concepts were not associated). Known-groups validity at baseline was assessed using 1-way ANOVA to determine whether the TEA total score and/or domains were able to distinguish between participants who rated their current health status as excellent, very good, good, fair, or poor based on SF-36 item 1 (current health status). Convergent and divergent validity for baseline TEA domain and total scores were assessed by correlating these scores with EQ-5D-5L index and VAS scores, SF-36 mental and physical component scores, ASI-Lite problem areas, and OC VAS scores at baseline. We anticipated moderate negative correlations between the TEA substance use domain and the ASI-Lite alcohol use and drug use problem areas and moderate positive correlations with OC VAS; moderate positive correlations between the TEA health domain and EQ-5D-5L index and VAS scores and SF-36 scores; and negative correlations between the TEA health domain and ASI-Lite medical and psychiatric status problem areas; moderate negative correlations between the TEA lifestyle domain and ASI-Lite social/family problem areas; and moderate negative correlations between the TEA community domain and the ASI-Lite legal status and community problem areas.

The TEA's ability to detect change was evaluated by comparing change in TEA scores to change in SF-36 item 1 (current health status), evaluated from screening to baseline. The rationale for this window was that previous analyses of the same population[15] had observed the most sizeable treatment effect on the SF-36 in this window, and it was anticipated to be similarly associated with improvement in effectiveness as measured by the TEA. The participants were grouped as those who improved by ≥1 category in their current health status, had no change in their current health status (ie, same health status at screening and baseline), or declined by ≥1 category in their current health status based on SF-36 item 1. The corresponding score on the TEA total scale was evaluated to assess whether the TEA measure would be sensitive to changes in current health status.

Additionally, the MID estimates for the TEA total score were calculated using multiple techniques to understand the range of potentially important differences. Distributional approaches included calculating half the SD at baseline (0.5 SD) and the standard error of measurement (computed as SD × √[1-reliability]) as estimates of the MID. The change from baseline on TEA score was calculated among participants in whom ≥80% of their weekly urine drug screens and self-reports of opioid use were both negative between BUP-XR injection 2 (study week 5) and end of study as an additional indication of MID.[16–18] Treatment success was a key secondary efficacy endpoint of the pivotal phase 3 trial and was defined as ≥80% opioid abstinence during weeks 5 through 24.[5] This endpoint was included because patients can be clinically successful despite occasional lapses in abstinence. The MID was also evaluated using ROC curves that assess what change in TEA total score from baseline to end of study would most discriminate between participants who achieved the same ≥80% threshold of negative opioid use, balancing sensitivity and specificity.[19]

Results

Among 412 de novo participants who were enrolled into the open-label safety study, 410 (99.5%) completed the baseline TEA assessment and formed the cohort for this analysis. The mean age was 38.4 years; 63.7% of participants were male and 71.5% indicated their race as white (Table 1).

Item distribution

The mean (SD) TEA total score before BUP-XR injection 1 (following up to 14 days of sublingual buprenorphine) was 25.4 (9.7). At this time point, less than 10% of participants were at the floor of any domain (ie, responded with 1 ["none or not much"]), and 15%–25% were at the ceiling of the measure (ie, 10 ["much better"]) across domains (Figure 1). All response levels were used in all domains.

Reliability

The TEA exhibited strong internal consistency (Cronbach's α=0.90), which was retained when each item was removed and internal consistency was re-calculated (Table 2). When considering the test-retest reliability among the 177 participants who had stable urine drug screen results at the last 2 visits and a TEA collected at these time points, the ICC for

Table 1 Baseline descriptive characteristics

	Study population (N=410)
Age, mean (SD), years Male, number (%)	38.4 (12.1) 261 (63.7)
Race, number (%) 　White 　African American or Black 　Other	 293 (71.5) 107 (26.1) 10 (2.4)
Tobacco use, number (%) 　Current 　Former	 353 (86.1) 17 (4.2)
Alcohol use, number (%) 　Current 　Former	 192 (46.8) 97 (23.7)
Opioid use at screening, number (%) 　Non-injectable 　Injectable	 217 (52.9) 193 (47.1)

the total score approached 0.7, with CIs overlapping this threshold (ICC 0.69; 95% CI 0.60–0.76; Table 2).

Validity

The TEA total score demonstrated differentiation between known groups as described by current health status (SF-36 item 1; ANOVA-based $P<0.001$; Table 2). For convergent and divergent validity, TEA correlations with ASI-Lite domains, SF-36 physical and mental component scores, and EQ-5D-5L index and VAS scores were uncorrelated or weakly correlated, but in the anticipated directions (Table 2).

Ability to detect change

The TEA instrument was sensitive to change in current health status (SF-36 item 1; Table 2). Participants whose health status had improved had a 13.5-point increase in TEA total score compared to an 11.5-point increase in those indicating no change in health status and an 8.7-point increase in those who reported a decline in health status. Similar trends were observed at a domain level.
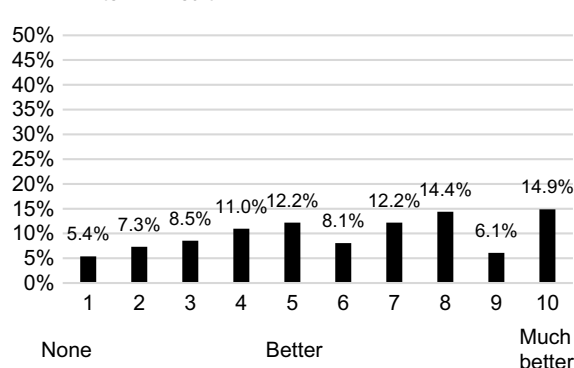
MIDs

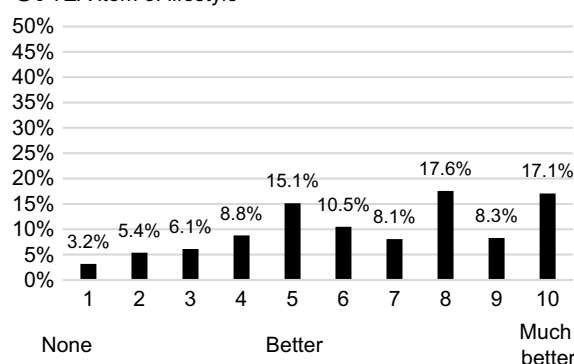MIDs for the TEA total score were highly dependent on the method used. When using the distribution-based method and
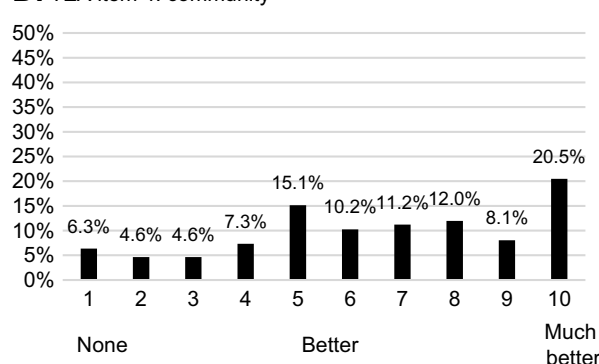

Figure 1 Item distribution and floor-ceiling effects at baseline[a]. (**A**) TEA Item 1: substance use. (**B**) TEA Item 2: health. (**C**) TEA Item 3: lifestyle. (**D**) TEA Item 4: community.
Note: [a]Given that there are ten response options for each item, an item may be subject to a floor or ceiling effect if more than 10% (=100/the number of response options) of responses are in the highest or lowest response categories.
Abbreviation: TEA, Treatment Effectiveness Assessment.
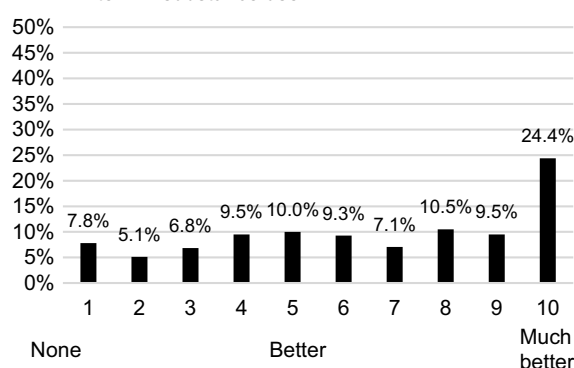
Table 2 Measures of Treatment Effectiveness Assessment (TEA) reliability and validity

	Number of participants	TEA total score	TEA domains			
			Substance use	Health	Lifestyle	Community
Reliability						
Internal consistency, Cronbach's alpha	410	0.90	0.93[b]	0.84[b]	0.84[b]	0.86[b]
Test-retest reliability, ICC (95% CI)[a]	177	0.69 (0.60–0.76)	0.65 (0.55–0.72)	0.71(0.63–0.77)	0.65 (0.56–0.73)	0.61(0.51–0.69)
Known groups validity, mean (SD) TEA score						
Current health status (SF-36 Item 1)						
Excellent	56	29.4 (9.8)	7.3 (2.9)	7.1 (2.9)	7.7 (2.6)	7.4 (2.8)
Very good	167	26.5 (9.7)	6.5 (3.1)	6.5 (2.7)	6.7 (2.6)	6.8 (2.7)
Good	140	24.0 (9.1)	6.3 (3.0)	5.6 (2.5)	6.1 (2.4)	6.1 (2.5)
Fair/Poor	47	21.1 (9.0)	5.9 (2.9)	4.5 (2.5)	5.3 (2.6)	5.3 (3.1)
P-value[c]		<0.001	0.097	<0.001	<0.001	<0.001
Convergent/divergent validity, Pearson correlations						
OC VAS	410	−0.13**	−0.07	−0.15**	−0.14**	−0.11*
ASI-Lite						
Medical status	329	−0.04	0.00	−0.09	−0.07	0.00
Employment status	298	−0.05	−0.10	−0.03	−0.01	−0.04
Alcohol use	296	0.00	0.03	−0.07	0.00	0.04
Drug use	281	−0.12	−0.12*	−0.11	−0.08	−0.08
Legal status	288	−0.08	−0.10	−0.06	−0.08	−0.06
Social/family status	313	−0.05	0.00	−0.09	−0.06	−0.04
Psychiatric status	309	0.06	0.06	−0.01	0.07	0.08
SF-36						
PCS	410	0.11*	0.04	0.16**	0.11*	0.07
MCS	410	0.25***	0.06	0.30***	0.26***	0.26***
EQ-5D-5L						
Utility index	410	0.30***	0.18***	0.34***	0.28***	0.26***
VAS	410	0.17***	0.05	0.20***	0.16**	0.17***
Ability to detect change in current health status[d], mean (SD) TEA score						
Improvement	189	13.5 (10.9)	3.4 (3.9)	3.5 (2.9)	3.4 (3.0)	3.3 (3.1)
No change	177	11.5 (10.6)	3.2 (3.8)	2.6 (3.0)	2.9 (3.0)	2.8 (3.1)
Decline	44	8.7 (9.7)	2.3 (3.2)	2.0 (3.0)	2.4 (2.5)	2.0 (3.0)

Notes: [a]Stable participants were defined based on last two visits in the open-label safety study and limited to those with stable urine drug screen results between these last visits. [b]Alpha if item deleted. [c]P-value is based on the overall ANOVA. [d]"Improved" defined as a ≥1-point increase between screening and baseline, "no change" defined as the same score at screening and baseline, and "declined" defined as a ≥1-point decline between screening and baseline. *P<0.05; **p<0.01; ***P<0.001.
Abbreviations: ASI-Lite, Addiction Severity Index-Lite; EQ-5D-5L, 5 level EuroQol 5 dimensions instrument; ICC, intraclass correlation coefficient; MCS, mental component summary, OC, opioid craving; PCS, physical component summary; SF-36, 36-Item Short Form Health Survey; VAS, visual analog scale.

Table 3 Summary of MID estimates for TEA total score

	Method/Approach			
	0.5 SD[a]	SEM[b]	Anchor-based[c]	ROC curve
Total score	4.8	5.6	6.8	8.0

Notes: [a]0.5 SD calculated as 1/2× SD at baseline. [b]SEM is calculated as SD × √(1-reliability) at baseline. [c]Anchored on participants with ≥80% negative urine drug screens with simultaneous report of no opioid use between week 5 and end of study.
Abbreviations: MID, minimal clinically important difference; SEM, standard error of measurement; TEA, Treatment Effectiveness Assessment.

ROC analysis, MIDs ranged from 4.8–8.0. Among participants who reached 80% or more abstinence as an anchor, the mean MID for the TEA total score was 6.8 (Table 3).

Discussion

The US FDA supports including the patient voice within medication development, as evidenced by the inclusion of patient-reported outcomes in a 2018 draft guidance document describing endpoints demonstrating effectiveness in people receiving medication-assisted treatment. The TEA is an easy-to-use clinical tool that clinicians can use to monitor a patient's recovery during treatment for an SUD. Results from our analysis demonstrate that the TEA has acceptable levels of reliability, validity, and ability to detect changes in a cohort of people with moderate to severe OUD treated with BUP-XR. While we did not do a formal testing of factor structure in this particular analysis, the relatively high Cronbach's alpha across the items would point to the high inter-relationship between the single-item domains. The purpose of the TEA is to provide a simple tool in which each area can be measured over time. Additionally, we determined the MID for change in the TEA total score, further informing health care providers on what constitutes a clinically meaningful change for this instrument.

Our results broadly support using the TEA to evaluate improvements associated with OUD treatment. Floor and ceiling effects were consistent with expectations; items exhibited strong internal consistency, illustrating that these items work together; and scores were able to discriminate between participants self-reporting excellent, very good, good, or fair/poor health. Furthermore, we observed that participants self-reporting improvement in health status showed greater improvements in TEA total scores as compared to participants who had no change in their health status or with declined health status. However, in interpreting these results it should be noted that while the SF-36 current health status variable would be expected to be associated with changes in the TEA health domain (and therefore TEA total score as well), this anchor does not capture global improvement as it may relate to treatment effectiveness in SUDs.

A challenge in evaluating measurement properties within this cohort was the lack of a clear anchor for change. This affected our ability to anchor measures of meaningful change and identify a group of clinically stable participants for test-retest reliability. Because of this, we used change in health status (during a window when health change was expected) as a proxy for assessing the ability to detect meaningful change. However, this anchor was imperfect as treatment effectiveness and recovery from OUD go far beyond health. This was evidenced by an improvement in participants' TEA total scores, including improvement in participants with declining health status, consistent with expectations, as this measurement occurred when participants were first inducted onto pharmacotherapy treatment and improvement was expected.

Our test-retest reliability results overlapped with the 0.7 threshold for reliability; however, the point estimate ranged from 0.609 (community domain) to 0.708 (health domain) despite selecting a time period when participants would be expected to be more stable and limiting measurement to those with stable urine drug screens between the two visits. For assessing MIDs, data obtained from participants achieving at least 80% abstinence were used as an anchor; this may be a high threshold for establishing meaningful change in the TEA and changes in TEA total scores under 7 points may still be meaningful. Confirming these results in future studies with a global indicator of change is warranted.

When assessing convergent and divergent validity, the direction of correlations between the TEA and other outcomes was consistent with expectations, but the correlations were low. It is likely that the TEA is capturing different constructs than measures of the EQ-5D-5L and SF-36, which focus on HRQoL. The TEA may also differ from the ASI-Lite, which focuses on disease severity, whereas the TEA measures person-specific measures of treatment effectiveness. These differences should be considered when weighing the comparative evidence for the

TEA's validity. Future research may also be warranted to explore correlations with other patient-centered measures in SUD, particularly those that move beyond a simple abstinence-based definition and encompass a more complete picture of the recovery process.[20–22] The TEA was developed to implicitly incorporate patient perspectives, as patients are solely responsible for determining which characteristics within the domains are most important to them in evaluating how their lives have been affected by treatment. However, if administered during a first treatment visit, the TEA results must be judiciously interpreted because the instrument asks about changes for the better since beginning a treatment program. Similarly, care should be taken when the possibility exists that a given program or treatment modality may cause decline instead of improvement, as the tool is worded to measure changes for the better. Future studies may be needed to explore the effects of widening this scale to include a version for assessment of baseline status, and/or providing wording specific to measuring change from a fixed time point.

Additional limitations of this analysis include that it was a secondary analysis of participants within a long-term open-label clinical trial and was not a psychometric study on its own; this may impact the generalizability of stated results. Future comparisons of the TEA with other patient-centered instruments could help elucidate aspects of the instrument's utility in measuring recovery from OUD and other SUDs.

Conclusion

Pragmatic tools such as the TEA are needed to quantify the patient experience in clinical practice and translate it into a format that provides meaningful, evaluable data that can benchmark person-specific improvement over time in clinical research and in real-world treatment settings. Equally important is that participants are reminded that life changes, as concretely assessed by the TEA, are critical steps in their pathway to recovery. Results from this study document the reliability and validity of the TEA in a cohort of participants with moderate to severe OUD and provide preliminary benchmarks for assessing meaningful changes in TEA scores. Further research is needed to evaluate the reliability, validity, and responsiveness of the TEA in real-world settings, and to identify person-centered recovery-related outcomes that may be strongly correlated with TEA scores.

Acknowledgments

Disclosure

References

1. Betty Ford Institute Consensus Panel. What is recovery? A working definition from the Betty Ford Institute. *J Subst Abuse Treat*. 2007;33 (3):221–228. doi:10.1016/j.jsat.2007.06.001

2. Substance Abuse and Mental Health Services Administration. SAMHSA's working definition of recovery. *Substance Abuse and Mental Health Services Administration* 2012; Available from: https:// store.samhsa.gov/shin/content/PEP12-RECDEF/PEP12-RECDEF.pdf. Accessed July 27, 2018.

3. Ling W, Farabee D, Liepa D, Wu LT. The Treatment Effectiveness Assessment (TEA): an efficient, patient-centered instrument for evaluating progress in recovery from addiction. *Subst Abuse Rehabil*. 2012;3(1):129–136. doi:10.2147/SAR.S38902

4. Indivior Inc. Safety and tolerability study of depot buprenorphine in treatment seeking subjects with opioid use disorder. March 29, 2018; Available from: https://clinicaltrials.gov/ct2/show/NCT02510014. Accessed April 16, 2019.

5. Haight BR, Learned SM, Laffont CM, et al. Efficacy and safety of a monthly buprenorphine depot injection for opioid use disorder: a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet*. 2019;393(10173):778–790. doi:10.1016/S0140-6736(18) 32259-1

6. Cacciola JS, Alterman AI, McLellan AT, Lin YT, Lynch KG. Initial evidence for the reliability and validity of a "Lite" version of the Addiction Severity Index. *Drug Alcohol Depend*. 2007;87(2–3):297– 302. doi:10.1016/j.drugalcdep.2006.09.002

7. McMillan DE, Gilmore-Thomas K. Stability of opioid craving over time as measured by visual analog scales. *Drug Alcohol Depend*. 1996;40(3):235–239. doi:10.1016/0376-8716(96)01218-5

8. Tompkins DA, Bigelow GE, Harrison JA, Johnson RE, Fudala PJ, Strain EC. Concurrent validation of the Clinical Opiate Withdrawal Scale (COWS) and single-item indices against the Clinical Institute Narcotic Assessment (CINA) opioid withdrawal instrument. *Drug Alcohol Depend*. 2009;105(1–2):154–159. doi:10.1016/j.drugalcdep.2009.07.001

9. Handelsman L, Cochrane KJ, Aronson MJ, Ness R, Rubinstein KJ, Kanof PD. Two new rating scales for opiate withdrawal. *Am J Drug Alcohol Abuse*. 1987;13(3):293–308. doi:10.3109/00952998709001515

10. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727–1736. doi:10.1007/s11136-011-9903-x

11. Ware JE, Kosinski M, Bjorner JB, Turner-Bowker DM, Gandek B, Maruish ME. *User's Manual for the SF-36v2 Health Survey*. Lincoln, RI: Quality Metric; 2008.

12. O'Mahony PG, Rodgers H, Thomson RG, Dobson R, James OF. Is the SF-36 suitable for assessing health status of older stroke patients? *Age Ageing*. 1998;27(1):19–22.

13. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16(3):297–334. doi:10.1007/BF02310555

14. Oliva TA, Oliver RL, MacMillan IC. A catastrophe model for developing service satisfaction strategies. *J Mark*. 1992;56(3):83–95. doi:10.1177/002224299205600306

15. Ling W, Nadipelli VR, Solem CT, et al. Impact of RBP-6000 (Once-Monthly Depot Buprenorphine) on patient-reported outcomes: a long-term study. Poster presented at: The American Society of Addiction Medicine 49th Annual Conference; April 12–15; 2018; San Diego.

16. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407–415.

17. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care*. 2003;41(5):582–592. doi:10.1097/01.MLR.0000062554.74615.4C

18. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61(2):102–109. doi:10.1016/j.jclinepi.2007.03.012

19. Froud R, Abel G. Using ROC curves to choose minimally important change thresholds when sensitivity and specificity are valued equally: the forgotten lesson of pythagoras. Theoretical considerations and an example application of change in health status. *PLoS One*. 2014;9(12):e114468. doi:10.1371/journal.pone.0114468

20. Kelly JF, Abry AW, Milligan CM, Bergman BG, Hoeppner BB. On being "in recovery": a national study of prevalence and correlates of adopting or not adopting a recovery identity among individuals resolving drug and alcohol problems. *Psychol Addict Behav*. 2018;32(6):595–604. doi:10.1037/adb0000386

21. Neale J, Nettleton S, Pickering L. What is the role of harm reduction when drug users say they want abstinence? *Int J Drug Policy*. 2011;22(3):189–193. doi:10.1016/j.drugpo.2010.09.007

22. Neale J, Vitoratou S, Finch E, et al. Development and validation of 'sure': a patient reported outcome measure (PROM) for recovery from drug and alcohol dependence. *Drug Alcohol Depend*. 2016;165:159–167. doi:10.1016/j.drugalcdep.2016.06.006

Supplementary material
Treatment Effectiveness Assessment instrument

Treatment Effectiveness Assessment (TEA)

The TEA asks you to express the extent of changes for the better from your involvement in the program to this point (or how things are if it's your first TEA or baseline) in four areas: substance use, health, lifestyle, and community. For each area, think about how things have become better and circle the results on the scale below: the more you have improved, the higher the number – from 1 (not better at all) to 10 (very much better). In each area write down the one or two changes most important to you in the Remarks section. Feel free to use the back of this page to add details, explain remarks, and make comments.

Substance use: How much better are you with drug and alcohol use? Consider the frequency and amount of use, money spent on drugs, amount of drug craving, time spent being loaded, being sick, in trouble and in other drug-using activities, etc.

None or not much				Better				Much better	
1	2	3	4	5	6	7	8	9	10

Remarks:

Health: Has your health improved? In what way and how much? Think about your physical and mental health: Are you eating and sleeping properly, exercising, taking care of health problems or dental problems, feeling better about yourself, etc?

None or not much				Better				Much better	
1	2	3	4	5	6	7	8	9	10

Remarks:

Lifestyle: How much better are you in taking care of personal responsibilities? Think about your living conditions, family situation, employment, relationships: Are you paying your bills? Following through with your personal or professional commitments?

None or not much				Better				Much better	
1	2	3	4	5	6	7	8	9	10

Remarks:

Community: Are you a better member of the community? Think about things like obeying laws and meeting your responsibilities to society: Do your actions have positive or negative impacts on other people?

No or not much				Better				Much better	
1	2	3	4	5	6	7	8	9	10

Remarks:

Name: _____ Date: _____ First TEA?: [ ]