Journal of Substance Abuse Treatment 46 (2014) 251–256

Contents lists available at ScienceDirect

Journal of Substance Abuse Treatment

Psychometric Properties of the Adjective Rating Scale for Withdrawal across treatment groups, gender, and over time

Celestina Barbosa-Leiker, Ph.D. [a,b,c,d,*], Sterling McPherson, Ph.D. [a,b,c,d], Mary Rose Mamey, M.A. [b], G. Leonard Burns, Ph.D. [b], John Roll, Ph.D. [a,b,c,d]

[a] Washington State University, College of Nursing, PO Box 1495, Spokane, WA, 99210–1495, USA
[b] Department of Psychology, Washington State University, PO Box 4820, Pullman, WA, 99164–4820, USA
[c] Programs of Excellence in Addictions Research, Washington State University, Pullman, WA 99164–4820, USA
[d] Translational Addictions Research Center, Washington State University, Pullman, WA 99164–4820, USA

ARTICLE INFO

Article history:
Received 21 May 2013
Received in revised form 19 July 2013
Accepted 14 August 2013

Keywords:
Opiate withdrawal
Opiate dependence
Measurement invariance
Psychometrics

ABSTRACT

The Adjective Rating Scale for Withdrawal (ARSW) is commonly used to assess opiate withdrawal in clinical practice and research. The aims of this study were to examine the factor structure of the ARSW, test measurement invariance across gender and treatment groups, and assess longitudinal measurement invariance across the clinical trial. Secondary data analysis of the National Drug Abuse Treatment Clinical Trials Network 000–3, a randomized clinical trial comparing two tapering strategies, was performed. The ARSW was analyzed at baseline, end of taper and 1-month follow-up (N=515 opioid-dependent individuals). A 1-factor model of the ARSW fit the data and demonstrated acceptable reliability. Measurement invariance was supported across gender and taper groups. Longitudinal measurement invariance was not found across the course of the trial, with baseline assessment contributing to the lack of invariance. If change over time is of interest, change from post-treatment through follow-up may offer the most valid comparison.

1. Introduction

Assessment of opioid withdrawal is important to clinically manage opioid dependent individuals (Tompkins et al., 2009). Opioid withdrawal is a function of the severity of physical dependence on opioids and the occupancy of the μ receptor at a specific time (Wesson & Ling, 2003). Withdrawal is assessed in treatment and recovery from opioid addiction in both research and clinical settings. One common scale used to examine withdrawal is the Adjective Rating Scale for Withdrawal (ARSW) (Amass, Kamien, & Mikulich, 2000; Bickel et al., 1988a; Bickel et al., 1988b). As it was developed, the ARSW was one of two 20-item adjective scales that were given to assess opiate withdrawal and opiate effects (Bickel et al., 1988a, 1988b), and was later trimmed to 16 items (Amass et al., 2000) that assessed symptoms such as painful joints, poor appetite, trouble getting to sleep, etc. The ARSW is typically summed to create a total score, and the scores have been used to assess subjective ratings of withdrawal across treatment groups (Amass et al., 2000; Bickel et al., 1988a; Ling et al., 2005; Ling et al., 2009; Ziedonis et al., 2009) and over time (Amass et al., 2000; Bickel et al., 1988a, 1988b; Ling et al., 2005; Ziedonis et al., 2009), typically in combined samples of men and women (Amass et al., 2000; Bickel et al., 1988a; Ling et al., 2005, Ling et al., 2009; Ziedonis et al., 2009).

Although the ARSW is a widely used measure of withdrawal in clinical trials, there is surprisingly little psychometric evaluation of this measure. To our knowledge, there is no published work on the factor structure or reliability of this scale. Additionally, measurement invariance of the ARSW across groups or over time has not yet been tested. The demonstration of measurement invariance is critical in order to effectively compare different groups on a latent construct (Brown, 2006; Chen, Sousa, & West, 2005; Cheung & Rensvold, 2002; Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000). In clinical trials, if measures are not invariant, there is the potential for bias in clinicians' interpretation of patient outcomes, complicating the interpretation of treatment effects (McHorney & Fleishman, 2006). For example, although treatment or taper groups may both have the potential to experience withdrawal similarly, one treatment group may endorse more withdrawal symptoms compared to the other treatment group, even though both groups are experiencing the same amount of withdrawal. This same issue of measurement inequality may occur across gender.

Longitudinal measurement invariance is used to evaluate the temporal change in a construct and, similar to testing measurement invariance across groups, is needed prior to comparing means over time (Brown, 2006; Vandenberg & Lance, 2000). For example, endorsement of withdrawal symptoms may change over time even though actual levels of withdrawal are consistent over time. This

* Corresponding author. Washington State University, College of Nursing, PO Box 1495 Spokane, WA, 99210–1495, USA. Tel.: 1 509 324 7477; fax: +1 509 324 7341.
E-mail address: celestina@wsu.edu (C. Barbosa-Leiker).

0740-5472/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jsat.2013.08.020

concern is highlighted in the clinical management of opioid withdrawal as assessing change in withdrawal over the course of treatment is crucial. Without testing the assumptions related to multiple-group and longitudinal measurement invariance, one cannot know whether 1) detected differences in withdrawal are true differences across groups or due to differences in assessment or structure of withdrawal across the groups, and 2) detected changes in withdrawal over time reflect true changes or are due to changes in assessment or structure of withdrawal.

The aims of this study are to build on previous work that utilized the ARSW by 1) examining the assumed unidimensionality of withdrawal by testing a 1-factor model of the ARSW, 2) provide a measure of ARSW reliability, 3) examine measurement invariance across gender and taper groups, and 4) assess longitudinal measurement invariance at pre-treatment, post-treatment, and at 1-month follow-up in a sample of opioid-dependent individuals enrolled in a clinical trial. If measurement invariance is demonstrated across groups and/or over time, differences in latent means will be examined.

2. Materials and methods

2.1. Clinical Trials Network 000–3

Participants were from the National Drug Abuse Treatment Clinical Trials Network Number 0003 (Ling et al., 2009). This was a randomized, parallel-group, open-label study design consisting of two buprenorphine/naloxone taper periods for opioid-dependent individuals. The participants first completed baseline assessments and were then stabilized on buprenorphine/naloxone. After the stabilization phase, the patients were stratified across the maintenance dose (8, 16 or 24 mg) of buprenorphine/naloxone and then randomized to the 7-day or 28-day taper groups. Follow-up measures were collected at 1-month and 3-month post-taper.

2.2. Measures

The ARSW is a 16-item self-report scale of opiate withdrawal symptoms (e.g., *Muscle cramps, Hot or cold flashes, Runny nose, Tense/jittery*) in which individuals rate the withdrawal adjectives on a 9-point scale from *none* to *severe*, with a maximum score of 144 (Amass et al., 2000; Bickel et al., 1988a; Bickel et al., 1988b).

2.3. Study sample

The final intention-to-treat sample was 516 participants who were potentially available for data collection at the end of the taper. Data were not available for one participant at baseline and was therefore excluded from the current analyses at that phase of the study. The effective sample size for the current investigation was n = 515. Self-report gender, age and ethnicity were assed at screening. This sample consisted of 346 males (67%) and 169 females (33%) (7-day taper group: male = 171, female = 83; 28-day taper group: male = 175, female = 86), with a mean age of 35.91 (SD = 10.45). Seventy-one percent identified as Caucasian, 11% as African American, 7% as Hispanic, 9% as multiple races and 2% as "other." Participants were from 11 out-patient treatment programs from 10 cities in the United States.

2.3. Statistical analyses

As noted in the applied measurement invariance literature (Barbosa-Leiker, in press; Barbosa-Leiker, Wright, Burns, Parks, & Strand, 2011), a series of analyses to examine measurement invariance were conducted after separate confirmatory factor analyses were fit for each group. Configural invariance (Horn & McArdle, 1992) examined if there was an equal factor structure across the groups (Vandenberg & Lance, 2000). This tested the final single-occasion model and examined if the theoretical framework of the ARWS was the same across taper groups and gender. Metric invariance (Horn & McArdle, 1992) tested if the relationships of ARSW items were equivalent for like items across taper groups and gender (i.e., constrained factor loadings). Lastly, scalar invariance (Meredith, 1993) constrained the intercepts of like item to be equal across groups to test whether differences in means of the items were due to differences in the construct (Steenkamp & Baumgartner, 1998). This step evaluated whether observed values of the ARSW items were equivalent across taper groups and gender when there was a constant level of withdrawal. After the demonstration of scalar invariance across taper groups and gender is shown, mean differences across taper groups and gender can be attributed to true differences in the construct (Brown, 2006; Vandenberg & Lance, 2000).

If withdrawal was invariant across groups, longitudinal measurement invariance was conducted following the above steps across baseline, end of taper, and at 1-month post-taper follow-up. Once measurement invariance was demonstrated across taper groups, gender, and over the course of the clinical trial and follow-up, differences across taper group, gender (regardless of taper group), and across the phases of the clinical trial in the withdrawal latent factor means were tested via *Z*-tests. Latent factor means were examined in this way so as to map directly onto the tests of invariance.

Overall model fit was evaluated using the robust comparative fit index (CFI; study criterion ≥ 0.900), the robust root mean square error of approximation (RMSEA; study criterion ≤ 0.080), and the robust standardized root mean square residual (SRMR; study criterion ≤ 0.080). For tests of measurement invariance, a change in CFI between comparison and nested models of ≥ -0.010 in addition to a change in RMSEA of ≥ 0.015 or a change in SRMR of ≥ 0.030 (for loading invariance) and ≥ 0.010 (for intercept invariance) will be used (Chen, 2007). For latent means analyses, differences are considered significant at $p \leq .05$.

All primary statistical testing was conducted in Mplus, Version 6 (Muthén, 1998–2010), using the robust maximum likelihood (MLR) estimator. There was 43% missing data in both taper groups from baseline to 1-month follow-up. Missing data was handled with full information maximum likelihood using all available data (Enders, 2010).

3. Results

3.1. 1-factor model

Mean scale scores, standard deviations, and reliability coefficients over time within each group are presented in Table 1. The 1-factor

Table 1
Mean scores (standard deviation) and reliability coefficients for the Adjective Rating Scale for Withdrawal.

	Total score	Cronbach's alpha
Males		
Baseline (n = 346)	60.61 (32.05)	.93
End of taper (n = 244)	19.51 (24.10)	.95
1-month follow-up (n = 192)	12.49 (18.76)	.94
Females		
Baseline (n = 169)	65.23 (32.26)	.93
End of taper (n = 128)	21.28 (28.10)	.95
1-month follow-up (n = 103)	16.59 (24.46)	.95
7-day taper		
Baseline (n = 254)	62.82 (31.65)	.93
End of taper (n = 201)	22.03 (25.98)	.95
1-month follow-up (n = 146)	14.30 (22.47)	.95
28-day taper		
Baseline (n = 261)	61.45 (32.70)	.93
End of taper (n = 171)	17.88 (24.81)	.95
1-month follow-up (n = 149)	13.56 (19.50)	.95