

Development and preliminary validation of the Opioid Abuse Risk Screener

Patricia Henrie-Barrus^{1,*}, Lynnette A Averill^{2,3,*},
Richard R Sudweeks⁴, Christopher L Averill^{2,3}
and Natalie Mota⁵

Health Psychology Open
January-June 2016: 1–12
© The Author(s) 2016
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/2055102916648995
hpo.sagepub.com



Abstract

Prescription opioid drug abuse has reached epidemic proportions. Individuals with chronic pain represent a large population at considerable risk of abusing opioids. The Opioid Abuse Risk Screener was developed as a comprehensive self-administered measure of potential risk that includes a wide range of critical elements noted in the literature to be relevant to opioid risk. The creation, refinement, and preliminary modeling of the item pool, establishment of preliminary concurrent validity, and the determination of the factor structure are presented. The initial development and validation of the Opioid Abuse Risk Screener shows promise for effective risk stratification.

Keywords

opioid abuse, pain, risk stratification, scale, validation

Introduction

Prescription opioid drug abuse has received unprecedented clinical, research, and public policy attention in the past decade, as it is the fastest growing drug problem in the United States and a significant problem worldwide (Centers for Disease Control and Prevention (CDC), 2012a). In the United States, over 5000 new individuals begin misusing prescription opioids and more than 100 die from opioid-related overdose every day (Bohnert et al., 2011). Furthermore, drug overdose is now the leading cause of injury deaths among US adults, with those resulting from opioid overdose exceeding the death rates from all other illicit drugs combined (Chen et al., 2014). The CDC has declared opioid abuse and diversion a public health crisis of epidemic proportions (CDC, 2011). There are many established definitions used to describe problematic opioid (and other substance) use including abuse, dependence, addiction, disorder, misuse, and non-medical or extra-medical use. The authors refer to the National Institute on Drug Abuse (NIDA) definition when discussing opioid abuse in this article. NIDA defines prescription opioid drug abuse as any “non-medical use” or use different than the exact regimen in which it was prescribed (e.g. in higher doses or increased frequency, using opioid medications that were

not prescribed to you) or for reasons other than why it was prescribed (e.g. to get high, to self-medicate psychiatric symptoms) (Alford and Livingston, 2013).

Individuals with chronic pain compose a large population at considerable risk of abusing prescription opioids given the likelihood that they would be prescribed this class of medication for pain management. More than 80 percent of all physician consults in the United States are pain related, and nearly one-third of all Americans suffer from chronic pain (Bresler, 1979; Institute of Medicine (IOM), 2011; Salovey, 1992). Data from the CDC suggest that approximately 40 percent of the over 100 million individuals

¹The University of Utah, USA

²Yale University, USA

³VA Connecticut Healthcare System, USA

⁴Brigham Young University, USA

⁵The University of Winnipeg, Canada

*Both the authors contributed equally for this article

Corresponding author:

Lynnette A Averill, Department of Psychiatry, Yale School of Medicine, Yale University, 300 George Street, Suite 901, New Haven, CT 06516, USA.

Email: lynette.averill@yale.edu



in the United States with chronic pain will actively seek medical help for pain symptoms, and although prevalence rates may vary across population and clinic settings, as many as 20 percent of these people will become addicted to opioid analgesics during treatment (CDC, 2012a). Each year, insurers pay more than 72.5 billion dollars to cover direct healthcare costs necessitated by the abuse of prescription opioids (CDC, 2012b; White et al., 2005). Adding to this the estimated 53 billion in economic costs, the total annual societal burden is conservatively at US\$125 billion (CDC, 2012b). These figures do not take into account the many individual, interpersonal, and relational costs often associated with prescription opioid abuse.

In an effort to mitigate the opioid crisis, the Food and Drug Administration (FDA) now mandates compliance to the strict standards of their risk evaluation and mitigation strategy (REMS) for opioid analgesics, which requires comprehensive screening and documentation of assessment (U.S. Food and Drug Administration (FDA), 2012). The American Pain Society (APS) and American Academy of Pain Medicine (AAPM) have developed evidence-based practice guidelines for opioid therapy (Chou et al., 2009a, 2009b, 2009c). The first major recommendation of the APS/AAPM guidelines is for careful patient selection and risk stratification (Chou et al., 2009b). These guidelines, as well as those developed by the US Department of Veterans Affairs/Department of Defense (Department of Veterans Affairs and Department of Defense (VA/DoD), 2010) and by other countries, highlight the need for comprehensive screening and suggest that proper patient selection can minimize potential risks and increase potential benefits of opioid analgesics in the treatment of chronic pain (Graziotti and Goucke, 1997; Jovey et al., 2003; Kalso et al., 2003; Society, 2010; VA/DoD, 2010).

Appropriate and effective screening and risk stratification as part of a comprehensive evaluation can help to lower the rates of opioid abuse, overdose, and death by providing useful risk information. This can further aid in reduced incidence of inaccurate or missed diagnoses, provide evidence to support appropriate monitoring, and reduce rates of doctor shopping and diversion. Comprehensive screening and risk stratification are associated with decreased costs for patients, providers, and insurers as those prescribing are able to make increasingly well-informed decisions when treatment-planning regarding what to prescribe and how to best monitor patients for safety based on individual risk profiles.

Despite FDA requirements, practice guidelines, and staggering prevalence rates of abuse and overdose, many practitioners are not formally screening for risk, either at the time of initial evaluation and prescribing nor at follow-up visits or are screening patients solely depending on their "gut" or instinct about risk level (Michna et al., 2007; Wasan et al., 2005). In addition, it is suggested that many physicians prescribing opioid pain medications have very

little training in aberrant drug-related behavior and substance abuse; thus, they may have limitations in their ability to effectively and accurately assess risk (Chou et al., 2009a, 2009b, 2009c; Graziotti and Goucke, 1997; Jovey et al., 2003; Kalso et al., 2003; Sehgal et al., 2012; Wasan et al., 2005). Research suggests even the best trained prescribers are not always successful in accurately screening for risk (Wasan et al., 2005) and all too often patients are being prescribed opioids without their prescriber having a clear and well-informed idea of their level of risk of abuse (Sehgal et al., 2012). For example, Wasan et al. (2007) found that prescribers estimated only 13.9 percent of their patients demonstrate aberrant behaviors (ABs), yet 50 percent of those prescribed opioids had positive urine toxicology screens for illicit drugs and 8.7 percent had no opioids in their urine at all. There is a clear need for improved screening that provides comprehensive and accurate evidence of opioid risk stratification, especially as many insurance companies are recommending that risk assessments be given to provide evidence for a coverage determination of medical necessity for urine drug screening (Owen et al., 2012).

There are several validated measures designed to assess risk of misusing opioids including the Opioid Risk Tool (ORT; Webster and Webster, 2005), Attitudes and Behaviors Questionnaire (Passik et al., 2000), and the Screener and Opioid Assessment for Patients with Pain-Revised (SOAPP-R; Butler et al., 2008, 2009). These measures provide useful information and insights into potential risk of abuse, yet are relatively limited in scope regarding biopsychosocial risk including psychiatric variables and history of AB. Furthermore, there is concern that some measures may overestimate risk (Moore et al., 2009) and although it is likely better to over-estimate rather than under-estimate risk in these circumstances, this potential overestimation translates directly to the pain management options an individual may be given and may limit opioid medications that provide effective reductions in pain. Of important note, the authors do not believe that any assessment measure should, in and of itself, be used to determine pain management options or to deny opioid medications. Rather risk assessments should be used in tandem with a clinical assessment, review of medical records, and other available collateral information to develop a well-informed and comprehensive patient profile.

Strong evidence connects a variety of psychiatric and biopsychosocial variables to increase the risk of opioid abuse in individuals with chronic pain (Alford and Livingston, 2013; Ballantyne and Mao, 2003; Edlund et al., 2007; Richardson et al., 2012; Seal et al., 2012; Sehgal et al., 2012; Sullivan et al., 2006); yet, to date, no assessments have been developed to adequately evaluate these often co-occurring and sometimes mutually exacerbating factors. In order to address the limitations with current risk screeners, the Opioid Abuse Risk Screener (OARS) was

Table 1. CFA fit statistics by construct.

Consensus defined construct	α	CFI	RMSEA	Chi-square ^a	df
Substance use history	.783	.981	.097	41.826	18
Aberrant medication-related behaviors	.752	.973	.080	15.230	8
Depression	.823	.996	.041	8.708	7
Anxiety	.853	.972	.123	25.157	8
PTSD and traumatic stress	.824	.986	.078	16.780	9
Quality of life	.822	.987	.051	22.004	16

CFA: confirmatory factor analysis; CFI: confirmatory fit index; RMSEA: root mean square error of approximation; PTSD: post traumatic stress disorder.

^aValues estimated using weighted least squares means and variance (WLSMV) cannot be used for chi-square difference tests.

developed as a comprehensive self-administered measure of opioid abuse that includes a wide range of critical elements noted in the literature to be relevant to opioid risk (e.g. depressive and anxiety symptoms, exposure to trauma/post-traumatic stress disorder (PTSD) symptoms, history of abuse/neglect, history of substance abuse, tobacco use, and impulsivity). As noted above, the OARS is not intended to assess and stratify risk of developing substance use disorder based on Diagnostic and Statistical Manual of Mental Disorders–Fifth Edition (DSM-5) criteria specifically, but rather to assess risk of opioid abuse more broadly including self-medication of psychiatric symptoms, diversion and other AB, and recreational use. These behaviors can, of course, lead to the development of a substance use disorder, as defined by the DSM-5, if left untreated.

The OARS is a brief, yet comprehensive screening measure grounded in the evidence regarding risk factors for opioid abuse that meets and surpasses clinical practice guidelines for effective risk stratification. This article describes the creation of the OARS in three steps: (1) the creation, refinement, and preliminary modeling of the item pool to an anonymous sample; (2) testing the refined items in a new sample, establishing preliminary concurrent validity by comparing the measure to a validated opioid risk measure; and (3) testing a revised item set and determining the factor structure in a larger, clinical sample. Each step is presented followed by a brief discussion of areas for future development and research. All studies presented in this article were conducted with the approval of the University of Utah Institutional Review Board.

Methods

Study 1: item development

Expert consensus group. A group of clinical professionals experienced in working with pain patients consulted with the authors throughout item development and the other stages of study described in this article. The group comprised eight psychologists, five physicians, two master-level mental health counselors, and two advance practice registered nurse (APRN) practitioners.

Construct and item development. The authors first performed a review of the pain and substance abuse literature, and existing risk assessment measures, including the SOAPP-R (Butler et al., 2008, 2009) and ORT (Webster and Webster, 2005). This review identified several prospective content domains likely to increase the risk of opioid abuse (i.e. substance use history, depression, and anxiety). These constructs were presented to the consensus group, which was asked to provide feedback on the clinical relevance of each domain, and to identify any spurious or missing constructs. The resulting domains are noted in Table 1.

An initial item pool was generated consisting of 230 items, positively oriented to one of the six accepted content domains. Many redundant items were included to identify the strongest items for each domain (i.e. “I have abused drugs or alcohol” and “My drug or alcohol use has caused me problems in the past”). The consensus group, who accepted, rejected, or requested revisions to item content, then iteratively refined the items, leaving a pool of 162 items deemed “content valid.” The group was also asked to theoretically load each item onto the factor they perceived was best being measured by the item. In all, 26 items that, by group consensus, could not be theoretically loaded onto a single factor or were thought to load on multiple factors were omitted. Additionally, five volunteer patients with chronic, non-cancer pain (aged 28–51 years; three males and two females) were asked to identify items that were unclear, confusing, or difficult to endorse. These 31 items were removed, leaving 105 prospective items total.

Reducing the item count. Data were gathered from 142 fully anonymous volunteers at a pain clinic in Utah, and statistical analyses were performed to reduce spurious and confounding items, and to add some additional evidence supporting the content validity established by the consensus group. First, items were purged based on reliability and corrected item-total correlation analyses. A confirmatory factor analysis (CFA) was then conducted separately for each of the content domains, using the “majority rules” item-to-factor assignments obtained from the consensus group. Since the items were rated on a 4-point Likert-type

Table 2. Fit statistics by study and model.

Model	CFI	TLI	RMSEA	Chi-square ^a	df
1A: one-factor	.927	.923	.079	1668.007	593
1B: two-factor	.899	.893	.093	2085.105	594
2A: two-factor	.908	.958	.128	335.829	66
2B: one-factor	.851	.929	.167	498.658	63
3A: five-factor	.959	.954	.070	1874.035	340
3B: two second-order factors	.958	.954	.071	1912.260	344
3C: one second-order Factor	.950	.945	.077	2205.407	345
3D: Bifactor	.964	.958	.068	1663.328	322

TLI: Tucker-Lewis index; CFI: confirmatory fit index; RMSEA: root mean square error of approximation; WLSMV: weighted least squares means and variance.

^aValues estimated using WLSMV cannot be used for chi-square difference tests.

scale, the default maximum likelihood (ML) estimator, which assumes continuous data conforming to a multivariate normal distribution, could not be used (Brown, 2015; Edwards et al., 2012). Instead, categorical assumptions using WLSMV adjusted estimation was applied as in Muthen (2009). Items with unacceptably high modification indices (MIs) or low loadings were removed and the CFA was repeated. The results of the six, single factor, CFAs are presented alongside the representative content domains established by the consensus group in Table 1.

The results of these analyses were provided to the same panel of experts who recommended several revisions (e.g. word choice, number of items) and supported the notion that the remaining items were still content valid to the defined constructs. Minor wording revisions were made to 6 items and 38 were omitted before continuing data collection. Data for 61 items were then gathered from 289 fully anonymous volunteers at a multi-site pain clinic in Utah. Poorly functioning items were removed following classical item analysis and exploratory factor analysis (EFA), from which evidence of either one or two factors was obtained. CFA was then conducted to test both the one-factor (1A) and the two-factor model (1B); the results of these analyses are presented in Table 2. Neither model yielded robust fit statistics; however, with CFI = .927 (trending close to .95, the gold standard), the two-factor model (with emotional and psychological items loading on one factor, and behavioral items loading on the other) emerged as the stronger of the two. After removing items with high poor factor loadings or large correlated errors, 38 items were retained.

Study 2: preliminary modeling and validity

Participants. Data were gathered from a sample of 267 adults who presented for consultation at one of several outpatient, community-based pain clinics in Utah, Idaho, and Nevada that included the OARS items in standard preliminary screening paperwork. In all, 249 patients with no missing data were included in analysis, while 18

incomplete records were omitted. The completing and non-completing populations appear similar: 55 versus 50 percent were females, 53 versus 55 percent were unemployed, 63 versus 61 percent were married or partnered, 85 versus 95 percent were White, and 50 versus 44 percent reported currently taking opioid analgesics for pain.

Methods. As in Study 1, EFA and CFA were each conducted on the full sample, and WLSMV rather than ML estimation was applied. EFA again showed strong support for one or two factors and identified six items that did not load strongly on any factor or that loaded very strongly on multiple factors. The consensus group reviewed these items and determined that they were either redundant or not critical. A CFA was performed on 32 items, testing multiple alternate models of factor structure: a two-factor congeneric model (2A) in which the items were specified to load on two separate first-order factors; a one-factor model (2B); and a hierarchical model in which each of the two first-order factors in 2A were specified to load on an underlying second-order factor (2C).

Preliminary evaluation of convergent and discriminant validity was performed on model 2A, which demonstrated the best fit, using the recommendations of Fornell and Larcker (1981) and Hair et al. (2006). Convergence is indicated by the degree to which the average of the squared loadings (R^2) exceeds .50 for a given factor. Discriminant validity is evidenced when the average R^2 for each factor exceeds the squared correlation between factors. A preliminary test of concurrent validity was performed by comparing the degree to which raw scores obtained from each of the factors in model 2A were correlated with each other and with raw scores from the SOAPP-R. Although the authors feel the SOAPP-R is limited in biopsychosocial comprehensive assessment, this measure was selected as the comparator given its wide spread use as a measure of risk of opioid abuse and related ABs, and the scientific evidence supporting its validity and reliability.

Table 3. Item-by-item factor loadings for models 2A and 2B.

2A factor	Subdomain	Item	Model 2A		Model 2B	
			Loading	R ²	Loading	R ²
Emotional lability	Anxiety	8	.645	.416	.616	.379
		17	.779	.607	.744	.554
		25	.733	.537	.704	.496
		33	.820	.672	.791	.626
	Depression	4	.704	.496	.675	.456
		13	.667	.445	.636	.404
		21	.831	.691	.798	.637
		29	.750	.563	.717	.514
		38	.769	.591	.744	.554
	Quality of life	15	.791	.626	.766	.587
		23	.754	.569	.719	.517
		31	.763	.582	.735	.540
		30	.763	.736	.542	.736
	PTSD/trauma	5	.741	.716	.513	.716
		14	.794	.768	.590	.768
22		.841	.806	.650	.806	
30		.763	.736	.542	.736	
7		.853	.728	.825	.681	
16		.810	.656	.777	.604	
Aberrant behavior	Substance use history	18	.753	.567	.715	.511
		26	.743	.552	.694	.482
		32	.806	.650	.774	.599
		2	.614	.377	.573	.328
		9	.733	.537	.694	.482
	Medical non-compliance	10	.655	.429	.605	.366
		19	.575	.331	.526	.277
		27	.788	.621	.731	.534
		3	.733	.537	.682	.465
	Non-substance-related behavioral risks	11	.824	.679	.779	.607
		12	.681	.464	.641	.411
		20	.806	.650	.754	.569
28		.795	.632	.721	.520	
35		.832	.692	.782	.612	

PTSD: post traumatic stress disorder.

Results

Model 2A: two first-order factors. Model 2A postulates that 16 of the 32 retained items load separately on each of two first-order factors, as identified in the EFA. Based on item content, the authors and the consensus group labeled the factors “emotional lability” (EL) and “AB.” Table 3 compares the standardized factor loadings and R^2 values for each item in models 2A and 2B. All items had acceptable factor loadings, and 28 items had strong loadings (>7.0). The average R^2 values were .579 and .569 for the EL and AB factors, respectively. The overall fit of this model (and others) is summarized in Table 2. Acceptable fit is indicated by a CFI of .908 and Tucker-Lewis index (TLI) of .958 although RMSEA appears to be undesirably high (.128) based on the range of .05–.08 suggested by Browne et al. (1993). All parameter estimates were significant, and MI >10.0 were limited to four potential cross-loadings between

10.1 and 17.4 and three minor correlated error terms ranging from 12.91 to 24.11.

A relatively high association between factors was observed. The correlation between the two factors in 2A was estimated to be .765. Squaring the correlation indicates that 58 percent of the variance in one of these two factors is explained by variability in the other factor. Finally, Cronbach’s alpha (α) was found to be .932 for factor EL and .895 for factor AB. Together, the 32 items demonstrate acceptable internal reliability and validity ($\alpha = .88$; sensitivity = .81; specificity = .68).

Model 2B: one first-order factor. Comparing the item-by-item loadings in Table 3 reveals that each item has a lower loading on model 2B than 2A, smaller R^2 values, higher chi-square (a measure of misfit), lower CFI and TLI, and higher RMSEA. The one-factor model does not fit the data better than 2A.

Model 2C: a hierarchical model. Given the degree of correlation between factors in 2A, it was reasonable to determine whether a hierarchical model specifying a single second-order factor would account for the estimated correlation between two first-order factors. However, 2C was unidentified unless the loadings of each of the two first-order factors were constrained to be equal. Under this constraint, the fit statistics were identical to those of model 2A—see discussions by Brown (2006) and MacCallum et al. (1993). Model 2C will not be presented in further detail.

Preliminary evidence of concurrent validity. The correlation between the raw scores of the model 2A EL and AB scales was .615 while the correlation of the SOAPP-R to EL and AB was .635 and .591, respectively. Squaring values and interpreting the resultant coefficients of determination indicate that 40.3 percent of the variance in the EL factor and 34.9 percent of the AB factor are shared with or explained by the variance in the SOAPP-R. It is important to note that the factors of model 2A were not hypothesized to correlate perfectly with the SOAPP-R as the OARS assesses a broader range of psychiatric variables as well as AB and impulsivity. Both measures are designed to assess a person's risk of opioid abuse, but by evaluating somewhat different constructs. For this reason, it is reasonable to expect the variability that is shared between the SOAPP-R and the EL or AB factor to be less than the unshared variability. The percentages reported above provide nascent evidence that the OARS two-factor model measures something similar to the SOAPP-R while also exhibiting distinguishing characteristics that imply the OARS is not identical to the SOAPP-R.

Study 3: factor structure replication

Following Study 2, the authors again consulted with the consensus group to discuss results and review items for content validity after the iterative reduction in item count. The group recommended revisions to the traumatic stress items due to concern about assessing general stress (i.e. in Study 1: "I sometimes have upsetting dreams about events from my past") rather than traumatic stress (i.e. revised in Study 2: "I have nightmares about a past traumatic event"). Items were also added at the request of the consensus group to address suicidal ideation (i.e. "I have been thinking about ending my life") and impulsivity (i.e. "I do things without thinking about the consequences" and "I do not plan activities carefully").

Participants. Data were gathered from a sample of 1821 adults who presented for consultation at one of the 14 outpatient community-based pain clinics in various locations throughout the United States. Patients were randomly divided into two subsamples (S1, $n=911$; S2, $n=910$) stratified by pain clinic. The authors received only

completed records and do not have information about the population that may not have completed the assessment. In this sample, 52.1 percent of subjects were female, 47.4 percent were male, and .5 percent identified as neither male nor female. Patients ranged from 18 to 87 years of age with a mean age of 46.48 years. A total of 48.2 percent of the patients were unemployed, 40.4 percent were smokers, and 12.2 percent reported participating in substance abuse treatment in the past.

Methods. To avoid capitalizing on chance patterns in the data, EFA was conducted on a different sample than the CFA as encouraged by Floyd and Widaman (1995) and Henson and Roberts (2006). An EFA, as described in Study 1 and Study 2, was conducted on sample S1.

A CFA was conducted on sample S2 to test several models that may explain the results of the EFA: model 3A comprised of five first-order factors, models 3B and 3C representing the inter-correlations among the five first-order factors using one and two second-order factors, respectively, and model 3D, a bifactor model including a general factor that directly influences each of the 28 items as well as domain-specific "group factors" that each account for additional common variance shared by a cluster of similar items.

Because models 3A, 3B, and 3C are nested within the bifactor model (3D), the adjusted chi-square difference test was used to formally test each of the three models one at a time against the bifactor solution (Muthen, 2009). Finally, as Cronbach's alpha is not appropriate for estimating the reliability of scores based on a higher-order or bifactor model (Brunner et al., 2012), two extensions of McDonald's (1985, 1999) (ω) omega, $\omega_{\text{hierarchical}}$ (ω_{H}), and ω_{subscale} (ω_{S}), were used (Reise, 2012; Reise et al., 2013; Zinbarg et al., 2006). Correcting for the multidimensionality in the general factor of a bifactor model, ω_{H} indicates the proportion of the total variance due solely to the general factor (Brunner et al., 2012; Gignac, 2015; Reise, 2012). One can use ω_{S} to estimate the internal consistency reliability of each specific factor in a bifactor model independently of all the other factors or subscales (Gignac, 2015; Reise, 2012). Finally, the explained common variance (ECV) ratio was used to quantify the degree of unidimensionality in bifactor data as per Reise (2012) (Figure 1).

Results. Contrary to previous results, the parallel analysis and eigenvalues strongly supported five or six factors. However, the sixth factor comprises solely of redundant test items added to this latest round of data collection, and these items loaded nearly as strongly on one factor in the five-factor model. The resulting factors were generally aligned with the original target constructs (depression, anxiety, traumatic stress, substance use history, and behavioral risk factors).

Models 3A: five first-order factors. Model 3A (comprises five first-order factors) allowed the correlations between the

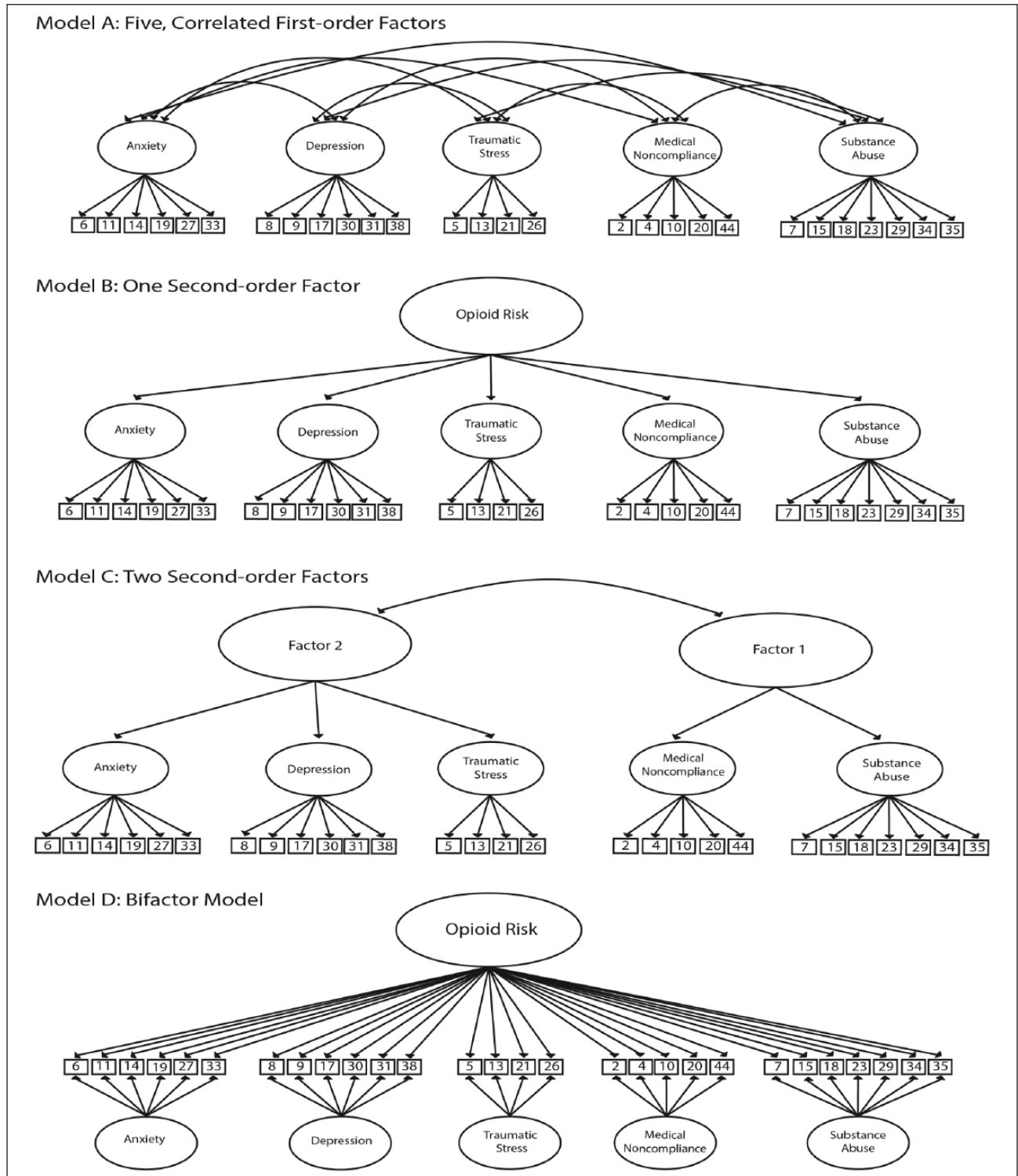


Figure 1. Path diagrams for models 3A, 3B, 3C, and 3D.

five factors to be freely estimated, but it did not allow for any correlated residuals. Good relative model-data fit (CFI=.959 and TLI=.954) and reasonably good absolute model-data fit (RMSEA=.070 and weighted root mean square residual WRMR=1.830) were found (see Table 2 for summary).

The factors are all positively correlated with the correlations ranging in magnitude from .523 to .846 (Table 4).

Models 3B and 3C: five first-order factors loading on one (3B) or two (3C) second-order factors. The results summarized

Table 4. Between-factor correlation coefficients for model 3A.

	F1	F2	F3	F4	F5
F1	1.000				
F2	.756	1.000			
F3	.737	.523	1.000		
F4	.716	.539	.825	1.000	
F5	.813	.609	.764	.846	1.000

F1: medical non-compliance; F2: substance use history; F3: traumatic stress; F4: anxiety; F5: depression.

in Table 2 indicate reasonably good relative fit for model 3B (CFI=.958 and TLI=.954) and 3C (CFI=.950 and TLI=.945). Similarly, reasonably good absolute model-data fit was found for 3B (RMSEA=.071 and WRMR=1.894) and 3C (RMSEA=.077 and WRMR=2.056).

Model 3D: bifactor model with a general factor and five domain-specific factors. The value of ω_H is .897 indicating that approximately 89 percent of the variance in the total scores computed from the 28 items is due to the general factor. The values of ω_S range from .128 to .462, but four of the five specific factors have a value less than .230. These findings indicate that the proportion of the variance contributed by each of the specific factors themselves is relatively small. The sum of the squared standardized loadings on the general factor is 12.421, and the sum of the squared standardized loadings on the five domain-specific factors is 17.151; the ECV ratio is 12.421/(12.421 + 5.213), or .704. Estimated factor loadings obtained from the bifactor model for the general factor and each of the domain-specific factors are reported in Table 5. All of the unstandardized factor loadings were statistically significant.

Preliminary suggestions for scoring. Given that the bifactor model fits the data better than the other tested models, and the fact that the general factor tends to be dominant relative to the specific factors, a single, composite score should be reported for each examinee, comprised of the unit-weighted sum of responses to all 28 items. If the response anchors are coded 0–3, the range of possible scores will be 0–84. The score obtained by an individual respondent can be interpreted as a measure of his or her risk of opioid abuse; the higher the score, the greater the risk. However, it is important to note that until further study compares these 28 items to other measures (both behavioral, as in the SOAPP-R, and biological, as in urine drug tests), this assumption will not be confirmed. After controlling for the general factor, the loadings on the specific factors are generally small. Consequently, the omega values for the specific factors are generally low as shown in Table 6. Therefore, the items do not provide sufficient information to strongly support the reporting of sub-scores for each specific factor. Future data collection and analysis will refine scoring algorithms.

Discussion

Effective and predictive risk stratification screening is vital for anyone being considered for opioid pain management given the significant abuse liability and related negative consequences, up to and including opioid-related overdose and death. Although there are validated measures currently available, which do provide useful information and demonstrate predictive validity (e.g. Butler et al., 2009), they do not deliver a comprehensive assessment of multiple biopsychosocial risk factors of opioid abuse. The OARS was developed in response to the need for careful and comprehensive assessment for opioid risk stratification and was created based on constructs shown in the literature to be most relevant to substance abuse broadly, and opioid abuse specifically.

The models

An iterative process has arrived at a measure consisting of 28 items. Since the 28 items were intended to assess different facets of a single over-arching construct (the constructs related to risk of opioid abuse initially derived), these moderate-to-high correlations in 3A are not surprising. However, the size of these between-factor correlations indicates the possibility that a hierarchical model would account for these relationships among the factors and better fit the data.

Next, we explored models that provided a different way to explain the inter-correlations among the five first-order factors, with one (3B) or two (3C) second-order factors. The second-order factor is assumed to directly influence two or more first-order factors, but have no influence on the items specifically. It is assumed that the influence of the second-order factors on the various items is mediated through the first-order factors.

Initial development procedures of the OARS demonstrate preliminary, yet promising internal validity of a bifactor model (3D). As above, the value of ω_H was just below 90 percent indicating that vast majority of variance in the total scores computed from the 28 items is due to the general factor. Reise (2012) indicates that “Generally speaking, the higher the ECV, the ‘stronger’ the general factor relative to the group (i.e. specific) factors and thus, the more confidence a researcher has in applying a unidimensional measurement model to multidimensional data” (p. 687).

Higher CFI/TLI and lower RMSEA/WRMR for the bifactor relative to the other models support the inference that the bifactor model is a better fit for the data than the other three models. These results (presented in Table 7) indicate that the bifactor solution produces a statistically significant decrement in the chi-square measure of misfit compared to each of the other three models. In other words, each of the three rival models fits significantly worse than the bifactor model.

Table 5. Unstandardized and standardized factor loadings for the bifactor model (3D).

Item	General factor	Domain-specific factors				
		Anxiety	Depression	Traumatic stress	Medical non-compliance	Substance use history
6	1.303 (.714) ^a	.797 (.437)				
11	1.348 (.725)	.800 (.430)				
14	1.168 (.734)	.407 (.256)				
19	1.049 (.663)	.634 (.401)				
27	1.589 (.775)	.821 (.401)				
33	1.250 (.746)	.498 (.297)				
8	1.113 (.712)		.454 (.290)			
9	1.390 (.745)		.742 (.397)			
17	1.867 (.854)		.546 (.249)			
30	.978 (.694)		.180 (.128)			
31	1.404 (.783)		.493 (.275)			
38	1.912 (.773)		1.208 (.489)			
5	1.104 (.648)			.830 (.487)		
13	1.253 (.703)			.780 (.438)		
21	1.274 (.724)			.688 (.391)		
26	.698 (.555)			.310 (.246)		
2	.675 (.474)				.755 (.530)	
4	.765 (.449)				1.146 (.673)	
10	1.127 (.718)				.442 (.281)	
20	1.231 (.768)				.234 (.146)	
44	.880 (.655)				.175 (.130)	
7	.677 (.423)					1.051 (.657)
15	.988 (.478)					1.513 (.733)
18	1.082 (.689)					.546 (.348)
23	1.062 (.495)					1.577 (.734)
29	.999 (.596)					.899 (.537)
34	.752 (.542)					.596 (.430)
35	.696 (.532)					.473 (.362)

^aCompletely standardized factor loadings are shown in parentheses; those not enclosed in parentheses are unstandardized.

Table 6. Model-based reliability estimates.

Error type	Construct	Omega-hierarchical	Omega-scale
General factor	Risk of opioid abuse	.897	
Specific factors	Anxiety		.191
	Depression		.128
	Traumatic stress		.223
	Medical non-compliance		.211
	Substance use history		.462

Comparison of models in Study 2 and Study 3

There are notable similarities and differences in the models described in Study 2 and Study 3. The selected model in Study 2 specified emotional and behavioral items on two separate first-order factors, and although a bifactor model with five specific factors was identified in Study 3, the results are not as divergent as it may seem. Model 3B was comprised of five first-order factors overlaying two second-order factors (one emotional and the other behavioral). A potential

explanation of this is that the larger sample size and slightly modified item content in Study 3 may have provided higher “resolution” data, allowing each construct to be more easily differentiated from the shared variance, causing a shift from two first-order factors to five in both EFA and CFA. Yet despite that increased granularity, the higher level functioning of these constructs has remained relatively stable throughout all three studies presented above with EB and AB continuing to emerge as strong conceptual frameworks upon which to base an understanding of risk of opioid abuse.

Table 7. Adjusted chi-square difference tests comparing three models to the bifactor model.

Model	No. of freely estimated parameters	Adjusted chi-square difference	Degrees of freedom	Probability F5
3A	122	166.381	18	.0000
3B	118	224.239	22	.0000
3C	117	433.725	23	.0000

3A: correlated factors model (five first-order factors with no second-order factors); 3B: two second-order model (five first-order factors with two second-order factors); 3C: one second-order factor model (five first-order factors with one second-order factor).

Many of the factors that contribute to opioid risk overlap with one another outside of the context of this specific instrument (i.e. psychiatric conditions such as depression and PTSD are highly comorbid with substance use disorders), so inter-correlated factors and high cross-loadings are not surprising. Construct-relevant multidimensionality (Morin et al., 2015) may explain this result as well as the similarities observed between models 2A and 3B. The risk factors for opioid abuse do not occur in a vacuum and are not unitary constructs, but are a collection of highly correlated, highly comorbid, and often mutually exacerbating symptoms that contribute to opioid risk (Morin et al., 2015).

Limitations

This work has some limitations that must be addressed. Although the final sample size is large, the majority of patients used for this validation study presented for care at outpatient community pain clinics across Utah, Idaho, and Nevada. It will be valuable to seek a more diverse sample in terms of pain populations as well as geographical location to improve generalizability of results. Additionally, this study only evaluates an adult population. Given the prevalence rates of opioid abuse among teens, it will be important to evaluate the OARS' effectiveness at determining risk in a younger population.

Unfortunately, information about the race, marital status, and current use of opioid analgesics for pain were not available to the authors for Study 3. Additionally, it is not known how many subjects did not complete the OARS as a part of Study 3, because the investigators were only provided with data from subjects who completed the full assessment and there may be unique characteristics about the population who did not complete the measure. Finally, the investigators were not able to obtain additional measures in Study 3, as in Study 2, so it is neither possible to evaluate convergent validity from the data in Study 3, nor to evaluate the generalizability of the preliminary evidence of validity in Study 2 to the results of Study 3. However, preliminary convergent validity with the SOAPP-R was established using an earlier iteration of the OARS. These determinations of criterion and predictive validity will need to be firmly established in future research.

To further provide support of the instruments validity, it will be helpful to compare patient responses on similar

constructs across measures and to engage in a more controlled study, perhaps including the results of a structured interview and other assessment further exploring personal and familial psychiatric and substance abuse history, comorbid medical conditions, history of ABs including legal involvement, coping skills for managing both physical and emotional pain and distress, and so on.

Ongoing and future research endeavors will elicit evidence of convergent and predictive validity, and will include analysis of specificity and sensitivity when comparing OARS results to biological factors like urine drug tests, to reports from the Department of Professional Licensing (DOPL) Controlled Substance Database, and to other measures including but not limited to the SOAPP-R. Future research should continue to evaluate the external validity of this instrument, comparing it to other relevant data types (e.g. self- and clinician-administered mood and symptom scales) and other measures of abuse risk level. Additionally, the OARS will be evaluated for use with other pain populations (e.g. acute pain, geriatric populations, etc.). Given the many reasons people may seek prescription opioids and may be motivated to exaggerate or otherwise misreport their symptoms and experiences, it will be valuable to develop a brief malingering, desirability, and consistency-in-reporting add-on component to help identify potential drug-seeking individuals and non-illicit inadequate reporting by patients. Finally, it will be important to examine the utility of the OARS across treatment as a safety and compliance monitoring measure. It may be useful to adapt the measure to be more appropriate as an effective measure for monitoring.

Appropriate and effective risk screening can help to lower the rates of opioid abuse, overdose, and death, reduce incidence of inaccurate or missed diagnoses, and doctor shopping, and decrease costs for patients, providers, and insurers by providing additional information concerning patient's backgrounds and potential risk factors (Chou et al., 2009a, 2009b). The OARS was developed as a comprehensive measure to assess a variety of elements shown by the literature to be predictive of increased risk of abuse, and meets a demand for effective risk stratification in the face of the significant public health crisis that is prescription opioid. The OARS is a promising tool in opioid risk management that could be used to improve patient outcomes and reduce risk of opioid abuse.

Acknowledgements

P.H.-B. and L.A. equally contributed to this work.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Alford DP and Livingston EH (2013) Misuse of opioid medication. *The Journal of the American Medical Association* 309: 2055.
- Ballantyne JC and Mao J (2003) Opioid therapy for chronic pain. *New England Journal of Medicine* 349: 1943–1953.
- Bohnert AB, Valenstein M, Bair MJ, et al. (2011) Association between opioid prescribing patterns and opioid overdose-related deaths. *The Journal of the American Medical Association* 305: 1315–1321.
- Bresler DE (1979) *Free Yourself from Pain*. New York: Simon & Schuster.
- Browne MW, Cudeck R, Bollen KA, et al. (1993) Alternative ways of assessing model fit. *Sage Focus Editions* 154: 136.
- Brown TA (2006) *Confirmatory Factor Analysis for Applied Research*. New York: Guilford Press.
- Brown TA (2015) *Confirmatory Factor Analysis for Applied Research*. New York: Guilford Press.
- Brunner M, Nagy G and Wilhelm O (2012) A tutorial on hierarchically structured constructs. *Journal of Personality* 80: 796–846.
- Butler SF, Budman SH, Fernandez KC, et al. (2009) Cross-validation of a Screener to Predict Opioid Misuse in Chronic Patients (SOAPP-R). *Journal of Addiction Medicine* 3: 66–73.
- Butler SF, Fernandez K, Benoit C, et al. (2008) Validation of the revised Screener and Opioid Assessment for Patients with Pain (SOAPP-R). *Journal of Pain* 9: 360–372.
- Centers for Disease Control and Prevention (CDC) (2011) Vital signs: Overdoses of prescription opioid pain relievers—United States, 1999–2008. *Morbidity and Mortality Weekly Report* 60(43): 1487–1492.
- Centers for Disease Control and Prevention (CDC) (2012a) CDC grand rounds: Prescription drug overdoses—A U.S. epidemic. *Morbidity and Mortality Weekly Report* 61(1): 10–13.
- Centers for Disease Control and Prevention (CDC) (2012b) Prescription drug abuse and overdose: Public health perspective. *CDC's Primary Care and Public Health Initiative*, 24 October. Available at: <http://www.cdc.gov/primarycare/materials/opoidabuse/>
- Chen LH, Hedegaard H and Warner M (2014) Drug-poisoning deaths involving opioid analgesics: United States, 1999–2011. *NCHS Data Brief*: 1–8. Available at: <http://www.cdc.gov/nchs/data/databriefs/db190.pdf>
- Chou R, Ballantyne JC, Fanciullo GJ, et al. (2009a) Research gaps on use of opioids for chronic noncancer pain: Findings from a review of the evidence for an American pain society and American academy of pain medicine clinical practice guideline. *Journal of Pain* 10: 147–159.
- Chou R, Fanciullo GJ, Fine PG, et al. (2009b) Clinical guidelines for the use of chronic opioid therapy in chronic noncancer pain. *Journal of Pain* 10: 113–130.
- Chou R, Fanciullo GJ, Fine PG, et al. (2009c) Opioids for chronic noncancer pain: Prediction and identification of aberrant drug-related behaviors: A review of the evidence for an American pain society and American academy of pain medicine clinical practice guideline. *Journal of Pain* 10: 131–146.
- Department of Veterans Affairs and Department of Defense (VA/DoD) (2010) VA/DoD clinical practice guideline for management of opioid therapy for chronic pain. *Department of Veterans Affairs Department of Defense*, May. Available at: http://www.va.gov/painmanagement/docs/cpg_opioidtherapy_fulltext.pdf
- Edlund MJ, Steffick D, Hudson T, et al. (2007) Risk factors for clinically recognized opioid abuse and dependence among veterans using opioids for chronic non-cancer pain. *Pain* 129: 355–362.
- Edwards MC, Wirth RJ, Houts CR, et al. (2012) Categorical data in the structural equation modeling framework. In: Hoyle RH (ed.) *Handbook of Structural Equation Modeling*. New York: Guilford Press, pp. 195–208.
- Floyd FJ and Widaman KF (1995) Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment* 7: 286–299.
- Fornell CL and Larcker DF (1981) Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research* 18: 39–50.
- Gignac GE (2015) On the inappropriateness of using items to calculate total scale score reliability via coefficient alpha for multidimensional scales. *European Journal of Psychological Assessment* 30(2): 130–139.
- Graziotti PJ and Goucke CR (1997) The use of oral opioids in patients with chronic non-cancer pain. Management strategies. *Medical Journal of Australia* 167: 30–34.
- Hair JFJ, Black WC, Babin BJ, et al. (2006) *Multivariate Data Analysis*. Upper Saddle River, NJ: Pearson Prentice-Hall.
- Henson RK and Roberts JK (2006) Use of exploratory factor analysis in published research common errors and some comment on improved practice. *Educational and Psychological Measurement* 66: 393–416.
- Institute of Medicine (IOM) (2011) *Report from the Committee on Advancing Pain Research, Care, and Education: Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education and Research*. Washington, DC: The National Academies Press.
- Jovey RD, Ennis J, Gardner-Nix J, et al. (2003) Use of opioid analgesics for the treatment of chronic noncancer pain—A consensus statement and guidelines from the Canadian pain society, 2002. *Pain Research Management* 8 (Suppl. A): 3A–28A.
- Kalso E, Allan L, DelleMijn PL, et al. (2003) Recommendations for using opioids in chronic non-cancer pain. *European Journal of Pain* 7: 381–386.
- MacCallum RC, Wegener DT, Uchino BN, et al. (1993) The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin* 114: 185–199.

- McDonald RP (1985) *Factor Analysis and Related Methods*. Hillsdale, NJ: Lawrence Erlbaum.
- McDonald RP (1999) *Test Theory: A Unified Treatment*. Mahwah, NJ: Lawrence Erlbaum.
- Michna E, Jamison RN, Pham L-D, et al. (2007) Urine toxicology screening among chronic pain patients on opioid therapy: Frequency and predictability of abnormal findings. *Clinical Journal of Pain* 23: 173–179.
- Moore TM, Jones T, Browder JH, et al. (2009) A comparison of common screening methods for predicting aberrant drug-related behavior among patients receiving opioids for chronic pain management. *Pain Medicine* 10: 1426–1433.
- Morin AJ, Arens AK, Tran A and Caci H (2015) Exploring sources of construct-relevant multidimensionality in psychiatric measurement: A tutorial and illustration using the Composite Scale of Morningness. *Journal of Methods in Psychiatry Research* PMID: 26265387.
- Muthen LKM (2009) *Mplus User's Guide*. Los Angeles, CA: Muthén & Muthén.
- Owen GT, Burton AW, Schade CM, et al. (2012) Urine drug testing: Current recommendations and best practices. *Pain Physician* 15: ES119–133.
- Passik SD, Kirsh KL, McDonald MV, et al. (2000) A pilot survey of aberrant drug-taking attitudes and behaviors in samples of cancer and AIDS patients. *Journal of Pain and Symptom Management* 19: 274–286.
- Reise SP (2012) The rediscovery of bifactor measurement models. *Multivariate Behavioral Research* 47: 667–696.
- Reise SP, Bonifay WE and Haviland MG (2013) Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment* 95: 129–140.
- Richardson LP, Russo JE, Katon W, et al. (2012) Mental health disorders and chronic opioid use among adolescents and young adults with chronic pain. *Journal of Adolescent Health* 50: 553–558.
- Salovey P (1992) Reporting chronic pain episodes on health surveys. In: Seiber WJ (ed.) *Vital Health Statistics* (6th edn). Hyattsville, MD: National Center for Health Statistics.
- Seal KH, Shi Y, Cohen G, et al. (2012) Association of mental health disorders with prescription opioids and high-risk opioid use in US veterans of Iraq and Afghanistan. *The Journal of the American Medical Association* 307: 940–947.
- Sehgal N, Manchikanti L and Smith HS (2012) Prescription opioid abuse in chronic pain: A review of opioid abuse predictors and strategies to curb opioid abuse. *Pain Physician* 15: ES67–92.
- Society BP (2010) *Opioids for persistent pain: Good practice*. London: The British Pain Society.
- Sullivan MD, Edlund MJ, Zhang L, et al. (2006) Association between mental health disorders, problem drug use, and regular prescription opioid use. *Archives of Internal Medicine* 166: 2087–2093.
- U.S. Food and Drug Administration (FDA) (2012) Extended-release (ER) and long-acting (LA) opioid analgesics risk evaluation and mitigation strategy (REMS). U.S. Food and Drug Administration, 9 July. Available at: <http://www.fda.gov/downloads/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/UCM311290.pdf>
- Wasan AD, Wootton J and Jamison RN (2005) Dealing with difficult patients in your pain practice. *Regional Anesthesia and Pain Medicine* 30: 184–192.
- Webster LR and Webster RM (2005) Predicting aberrant behaviors in opioid-treated patients: Preliminary validation of the Opioid Risk Tool. *Pain Medicine* 6: 432–442.
- White AG, Birnbaum HG, Mareva MN, et al. (2005) Direct costs of opioid abuse in an insured population in the United States. *Journal of Managed Care Pharmacy* 11: 469–479.
- Zinbarg RE, Yovel I, Revelle W, et al. (2006) Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for ω_h . *Applied Psychological Measurement* 30: 121–144.