

## Original Research Article

# An Item Bank for Abuse of Prescription Pain Medication from the Patient-Reported Outcomes Measurement Information System (PROMIS<sup>®</sup>)

**Paul A. Pilkonis, PhD,\* Lan Yu, PhD,\*\*† Nathan E. Dodds, BS,\* Kelly L. Johnston, MPH,\* Suzanne M. Lawrence, MS,\* Thomas F. Hilton, PhD,‡ Dennis C. Daley, PhD,\* Ashwin A. Patkar, MD,§,¶ and Dennis McCarty, PhD<sup>§</sup>**

Departments of \*Psychiatry and; †Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; ‡National Institute on Drug Abuse (retired), Indian Harbour Beach, Florida; §Department of Psychiatry, Duke University School of Medicine, Durham, North Carolina; ¶Department of Community and Family Medicine, Duke University School of Medicine, Durham, North Carolina; §School of Public Health, Oregon Health and Science University and Portland State University, Portland, Oregon, USA

Funding sources: PROMIS<sup>®</sup> was funded with cooperative agreements from the National Institutes of Health (NIH) Common Fund Initiative (Northwestern University, PI: David Cella, PhD, U54AR057951, U01AR052177; Northwestern University, PI: Richard C. Gershon, PhD, U54AR057943; American Institutes for Research, PI: Susan (San) D. Keller, PhD, U54AR057926; State University of New York, Stony Brook, PIs: Joan E. Broderick, PhD and Arthur A. Stone, PhD, U01AR057948, U01AR052170; University of Washington, Seattle, PIs: Heidi M. Crane, MD, MPH, Paul K. Crane, MD, MPH, and Donald L. Patrick, PhD, U01AR057954; University of Washington, Seattle, PI: Dagmar Amtmann, PhD, U01AR052171; University of North Carolina, Chapel Hill, PI: Harry A. Guess, MD, PhD (deceased), Darren A. DeWalt, MD, MPH, U01AR052181; Children's Hospital of Philadelphia, PI: Christopher B. Forrest, MD, PhD, U01AR057956; Stanford University, PI: James F. Fries, MD, U01AR052158; Boston University, PIs: Alan Jette, PT, PhD, Stephen M. Haley, PhD

(deceased), and David Scott Tulsy, PhD (University of Michigan, Ann Arbor), U01AR057929; University of California, Los Angeles, PIs: Dinesh Khanna, MD (University of Michigan, Ann Arbor) and Brennan Spiegel, MD, MSHS, U01AR057936; University of Pittsburgh, PI: Paul A. Pilkonis, PhD, U01AR052155; Georgetown University, PIs: Carol M. Moinpour, PhD (Fred Hutchinson Cancer Research Center, Seattle) and Arnold L. Potosky, PhD, U01AR057971; Children's Hospital Medical Center, Cincinnati, PI: Esi M. Morgan DeWitt, MD, MSCE, U01AR057940; University of Maryland, Baltimore, PI: Lisa M. Shulman, MD, U01AR057967; and Duke University, PI: Kevin P. Weinfurt, PhD, U01AR052186). NIH Science Officers on this project have included Deborah Ader, PhD, Vanessa Ameen, MD (deceased), Susan Czajkowski, PhD, Basil Eldadah, MD, PhD, Lawrence Fine, MD, DrPH, Lawrence Fox, MD, PhD, Lynne Haverkos, MD, MPH, Thomas Hilton, PhD, Laura Lee Johnson, PhD, Michael Kozak, PhD, Peter Lyster, PhD, Donald Mattison, MD, Claudia Moy, PhD, Louis Quatrano, PhD, Bryce Reeve, PhD, William Riley, PhD, Peter Scheidt, MD, Ashley Wilder Smith, PhD, MPH, Susana Serrate-Sztejn, MD, William Phillip Tonkins, DrPH, Ellen Werner, PhD, Tisha Wiley, PhD, and James Witter, MD, PhD. An award from the National Institute on Drug Abuse (UG1 DA015815) provided additional support for Dennis McCarty's participation. This article uses data developed under PROMIS. These contents do not necessarily represent an endorsement by the US Government or PROMIS. See [www.nihpromis.org](http://www.nihpromis.org) for additional information on the PROMIS initiative.

*Correspondence to:* Paul A. Pilkonis, PhD, Department of Psychiatry, University of Pittsburgh School of Medicine, Western Psychiatric Institute and Clinic, 3811 O'Hara Street, Pittsburgh, PA 15213,

USA. Tel: 412-246-5833; Fax: 412-246-5840; E-mail: pilkonispa@upmc.edu.

Conflicts of interest: There are no conflicts of interest to report.

## **Abstract**

**Objective.** There is a need to monitor patients receiving prescription opioids to detect possible signs of abuse. To address this need, we developed and calibrated an item bank for severity of abuse of prescription pain medication as part of the Patient-Reported Outcomes Measurement Information System (PROMIS<sup>®</sup>).

**Methods.** Comprehensive literature searches yielded an initial bank of 5,310 items relevant to substance use and abuse, including abuse of prescription pain medication, from over 80 unique instruments. After qualitative item analysis (i.e., focus groups, cognitive interviewing, expert review, and item revision), 25 items for abuse of prescribed pain medication were included in field testing. Items were written in a first-person, past-tense format, with a three-month time frame and five response options reflecting frequency or severity. The calibration sample included 448 respondents, 367 from the general population (ascertained through an internet panel) and 81 from community treatment programs participating in the National Drug Abuse Treatment Clinical Trials Network.

**Results.** A final bank of 22 items was calibrated using the two-parameter graded response model from item response theory. A seven-item static short form was also developed. The test information curve showed that the PROMIS<sup>®</sup> item bank for abuse of prescription pain medication provided substantial information in a broad range of severity.

**Conclusion.** The initial psychometric characteristics of the item bank support its use as a computerized adaptive test or short form, with either version providing a brief, precise, and efficient measure relevant to both clinical and community samples.

**Key Words.** Prescription Pain Medication; Opioid Use; Substance Use; Item Response Theory; Measurement; PROMIS

## **Introduction**

The Patient-Reported Outcomes Measurement Information System (PROMIS<sup>®</sup>) is a National Institutes of Health (NIH) roadmap initiative designed to improve the assessment of self-reported outcomes using state-of-the-art

psychometric methods (for detailed information, see [www.nihpromis.org](http://www.nihpromis.org)). PROMIS is the most ambitious attempt to date to apply models from item response theory (IRT) to health-related assessment, including physical, mental, and social health. The PROMIS methodology involves iterative steps of comprehensive literature searches; item pooling; development of a conceptual framework describing the main themes in the item pool; qualitative assessment of items using expert review, focus groups, and cognitive interviewing; and quantitative evaluation of items using techniques from both classical test theory (CTT) and IRT [1–4]. Currently, there are more than 50 PROMIS item banks measuring constructs relevant to all diseases and health states—e.g., physical functioning, pain, fatigue, sleep disturbance, emotional distress (depression, anxiety, and anger), alcohol and substance use, and social participation—providing a comprehensive profile of health status [1,2,5–10].

We report here on the development and calibration of an item bank assessing severity of abuse of prescription pain medication. There are existing measures both to screen for and to monitor abuse of opioid medications [11–14], but they suffer from certain limitations. For example, some of the current instruments are proprietary, e.g., Screener and Opioid Assessment for Patients with Pain [12] and Current Opioid Misuse Measure [13], although it should be noted that these two measures are freely available. Another significant limitation is that “legacy” measures have been developed using CTT. Therefore, these measures must be administered in a static way; when the same items are asked repeatedly (whether or not they are relevant to a particular respondent or phase of treatment), the burden to respondents is greater and the risk of invalid responding increases. IRT-calibrated item banks underlie the use of computerized adaptive testing (CAT) [15–17], in which the presentation of items is tailored individually to respondents and their levels of the latent construct being measured. The result is an efficient procedure for reducing both the total number of items administered and measurement error following the administration of each successive item [18–21]. Simulation studies indicate that CAT using as few as five polytomous items can achieve excellent precision and that scores derived from CAT correlate strongly with the conventional total score from a measure [22–25]. The goal of PROMIS is to create item banks that provide a comprehensive profile of health status, that are psychometrically sound, and that are publicly available on the internet [26].

Drugs of abuse are too various to allow the development of separate item banks for individual substances. Therefore, in our previous work to create generic item banks for substance use [8], we took a broad approach and created items that referred to “drugs” in general. This approach is consistent with the clinical epidemiology of many secondary and tertiary care facilities for addiction medicine where patients appearing for treatment often abuse multiple substances [27]. However, the abuse of one specific class of drugs—prescribed opioid pain medications—and the morbidity and mortality associated with such abuse have become

pronounced public health problems over the past 20 years: “Rates of unintentional overdose on prescription opioids increased almost fourfold from 2000 to 2010, accounting for more than half of all overdose deaths and exceeding overdose deaths attributed to all other illicit drug categories combined” [28]. The ready availability of prescription opioid drugs has also facilitated the phenomenon of “transferring” addictions from such drugs to heroin (and other illicit substances), which, in some circumstances, may be less expensive and easier to obtain [29–31]. However, using heroin rather than prescription opioids puts the user at increased risk of not only overdose but also infectious diseases and other medical problems.

Brady et al. [28] discussed the need for “universal risk evaluation” prior to starting opioid therapy and for continued monitoring of patients during treatment to detect possible signs of misuse. The PROMIS mandate has been to develop measures of severity that can be used as a common metric across all chronic diseases and health states. As such, the measures provide an index of severity for multiple health-related constructs, and they are not intended primarily for screening or diagnostic purposes. Therefore, our intent was not to create a measure of “risk” but rather to develop a measure of severity of abuse that would be relevant to monitoring both clinical and community samples using prescription pain medications. Consistent with this aim, we developed a pool of items specific to abuse of opioid pain medications and administered these items to respondents who acknowledged having 1) a legitimate prescription for such medications and 2) potential indicators of misuse during the previous three months.

## Methods

### Development of Item Pool

#### Comprehensive Literature Searches

The Pittsburgh PROMIS research site developed a methodology for performing comprehensive literature searches to ensure content validity and broad coverage of the general domain of substance use and the specific area of abuse of prescription pain medication. We performed searches in the MEDLINE, PsycINFO, and Health and Psychosocial Instruments (HaPI) databases. Details of the methodology are reported in Klem et al. [32], and all search algorithms are available upon request. The searches generated 3,388 abstracts that could be linked to more than 80 unique measures of substance use. Cited reference searches were run on the primary reference for each measure to determine its acceptance and use by the scientific community. Copies of the measures were gathered from both electronic and print sources, and the measures were reviewed at the item level. The initial item pool for substance use and abuse, including abuse of prescription pain medication, contained 5,310 items.

### Focus Groups

To ensure comprehensive coverage of the conceptual area, we conducted focus groups and performed thematic analyses of the topics discussed [33,34]. Members of two groups (N = 15) were recruited from outpatient substance use treatment programs. The average age was 41 years (SD = 9 years). Females (53%) and members of minority groups were well represented (race = African American 53%; ethnicity = Hispanic 6%). Two-thirds of the participants (67%) had no formal education beyond high school.

Using semistructured scripts, facilitators prompted participants to discuss their experiences with use of prescription pain medication and the characteristics of problematic use. Research staff reviewed process notes from the groups and audio recordings, paying special attention to positive and negative appraisals (consequences of use, intended vs actual effects of use) and contexts and triggers of substance-related experiences. The goal was to enrich our item pool with content not represented on traditional questionnaires. We were also eager to identify potential items with lower thresholds for endorsement, unlike the usual high-threshold items on many questionnaires that are most relevant at the severe end of the continuum of substance use.

### Qualitative Item Review

A key step in editing the item bank was qualitative review of the items done by members of the research team (see [35] for a description of the qualitative procedures used by the PROMIS network). This process involved elimination of redundant items, items that were too narrow (often because of being substance-specific), items that were confusing or vague, and items that were poorly written (e.g., double-barreled items). Our goal was to create a pool of 20–30 items for field testing. With this goal in mind, we reduced the item pool to 32 prescription pain medication items that underwent cognitive interviewing.

### Standardization of Items

Items were written in a first-person, past-tense format with a three-month time frame (e.g., “in the past 3 months, I ran out of my prescription pain medication early”), and five response options reflecting either frequency (never, rarely, sometimes, often, almost always) or severity (not at all, a little bit, somewhat, quite a bit, very much). This standardization of items was consistent with the usual efforts to promote internal consistency across PROMIS measures [5,35]. In addition, a review of intellectual property issues was completed for all items [36,37]. The large majority of items were generic—that is, they were similar to several extant items but not identifiable with any one in particular.

Cognitive Interviews

Seven participants from outpatient substance use treatment programs were recruited for cognitive interviews. All 32 items were reviewed by the seven individuals. An effort was made to include adequate numbers of women and members of underrepresented minority groups (71% and 57%, respectively). Fifty-seven percent reported no formal education beyond high school, and participants had an average reading level of the 11th grade as assessed by the Wide Range Achievement Test (WRAT-4) [38,39]. An interviewer met with participants and asked each to “think aloud” while responding to items. Participants were then prompted for feedback on the language and clarity of items and the relevance of the content. Adaptations arising from cognitive interview feedback resulted in the clarification of ambiguities (e.g., “I misplaced my prescription pain medication and needed to ask for more” became “I told my healthcare provider that I lost my pain medication and I needed more” to distinguish healthcare providers from other potential sources). Of the 32 items reviewed by participants, 81% (26 items) were acceptable as presented, 13% (four items) were rewritten, and 6% (two items) were deleted. All of the re-written items required a revision to the item stem (content) vs the time frame or response options. The cognitive interviews and continued refinement of the items produced a pool of 25 items for field testing.

Sampling

For calibration purposes, the goal was to collect a sample across the full spectrum of severity of abuse, including both members of the community and identified patients in recognized clinical settings. Therefore, we administered the banks to an internet (YouGov) sample of 367 participants from the general population. YouGov is a national, web-based polling firm in Palo Alto, CA. We also included a clinical sample of 81 patients in treatment for substance use disorders in three different regions: Addiction Medicine Services at the University of Pittsburgh Medical Center; the CODA treatment programs in Portland, Oregon; and the Southlight and Coastal Horizons treatment programs in North Carolina. These sites are members of the National Drug Abuse Treatment Clinical Trials Network [40,41]. Both the internet and clinical samples were composed of participants who reported possession of a prescription for pain medication and potential misuse of such medication within the past three months. Potential misuse was assessed by asking whether the medication had been used differently than how it was prescribed (including a different dosage, a different frequency of use, or use combined with other drugs) or whether it was used for the feeling or effect it caused (e.g., “to get high”). The study was approved by the institutional review boards of all the settings involved, and written consent was obtained from all participants. Demographic characteristics of the YouGov sample and the clinical sample (combined across all sites) are summarized in Table 1.

Table 1 Demographic characteristics

Characteristic	Clinical sample (N = 81) %	Community sample (N = 367) %
<b>Sex</b>		
Male	45.7	44.7
<b>Ethnicity</b>		
Hispanic	6.2	15.0
<b>Race</b>		
American Indian/Alaska Native	2.5	0.0
Asian	1.2	0.5
Black/African American	11.1	14.4
White	76.5	76.8
Other/multiracial	8.6	8.2
<b>Education</b>		
High school diploma or less	49.4	43.3
Further educational attainment	50.6	56.7
Mean age (SD), years	38.6 (11.7)	49.3 (16.0)

Members of the clinical sample were younger ( $P < .001$ ) than participants from the community, and there were also fewer Hispanic participants in the clinical sample ( $P < .04$ ).

Measures

The prescription pain medication item bank used for field testing contained 25 items reflecting motivations to use (two items), intended effects of use (four items), drug-seeking behaviors (12 items), and patterns of consumption (seven items). Participants also completed two legacy measures relevant to abuse of prescription pain medication specifically and other substances more generally: the Pain Medication Questionnaire (PMQ; 11), and the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST v. 3.0) [42]. In addition, they reported the ages at which they first used alcohol and drugs, and they provided a medical history that asked about lifetime diagnoses and functional impairment in nine areas (e.g., cardiovascular problems, cancer, mental health).

Data Analysis

General Strategy

We made no prediction about whether the 25 items tapping abuse of prescription pain medication would reflect a single underlying dimension. Therefore, our primary goal was to identify the most robust latent constructs within the item pool and to document sufficient unidimensionality for one or more of them to allow us to proceed with IRT analyses. First, we inspected frequency distributions of individual items for sparse cells.



We then investigated dimensionality by dividing the sample randomly into two subsamples, one for exploratory factor analysis (unweighted least squares EFA,  $N=232$ ) and the second for confirmatory factor analysis (CFA,  $N=216$ ). Both EFA and CFA were conducted using Mplus 6.1 with promax rotation [43]. In the CFA, the items were treated as categorical variables, and the robust weighted least squares (WLSMV) estimator was used. Scree plots, eigenvalues, and factor loadings were examined. We focused on the ratio of eigenvalues in EFA and the relative proportion of variance accounted for by the factors extracted. We also paid most attention to the size of factor loadings in both EFA and CFA and the information values for individual items from the IRT models.

### Item Response Theory Analysis

The most commonly used IRT model for polytomous items (i.e., items with three or more ordinal response categories) is the two-parameter graded response model (GRM) [44]. The GRM has a slope parameter and  $N - 1$  threshold parameters for each item, where  $N$  is the number of response categories (five in the present analyses). The slope parameter measures item discrimination, i.e., how well the item differentiates between higher vs lower levels of severity (or  $\theta$  in IRT terms). Useful items have large slope parameters. Threshold parameters measure item difficulty, i.e., the ease vs difficulty of endorsing different response options for an item. For example, the first threshold parameter for an item tells us where along the  $\theta$  scale of severity a respondent is more likely to endorse a response of “rarely” rather than “never.” Items were calibrated using IRTPRO v. 2.1 [45].

Differential item functioning (DIF) occurs when characteristics such as age, gender, or ethnicity, which may seem extraneous to the assessment of the construct under consideration, actually do affect the measurement of the construct. An item functions differentially if the item is more (or less) discriminating or more (or less) difficult to endorse in one group (e.g., men) compared with a reference group (e.g., women) when the different subgroups have been matched on the latent variable under investigation. We conducted DIF analyses for both uniform (difficulty) and nonuniform (discrimination) DIF on the basis of age, gender, and education (high school education or less vs further educational attainment). We focused on these variables because the relevant comparison groups were adequately represented; our sample included 45% men vs 55% women and 44% respondents with a high school education or less vs 56% with further education. We also used a median split on age (younger than 42 years vs 42 years or older) to compare younger and older respondents. Other potential comparison groups (e.g., white vs nonwhite respondents, Hispanic vs non-Hispanic respondents) were less equally divided. In addition, we conducted DIF

analyses on the basis of severity of overall substance use (using a median split on the total substance involvement scores computed from the ASSIST). This analysis assessed the performance of items based on differential exposure to drugs and documented their consistency across the full spectrum of severity of substance use. Two different DIF procedures were employed—the IRT likelihood ratio method [46] and an ordinal logistic regression procedure [47]—and items were considered for removal if they showed significant DIF ( $P < .01$ ) by both methods [48].

## Results

### *Alcohol, Drug, and Tobacco Use in the Two Samples*

Table 2 summarizes the lifetime use of substances reported on the ASSIST. The most frequently used drugs (lifetime) in the clinical sample were opioids (for nonmedical purposes, 91%), cannabis (85%), and cocaine (82%), and in the community sample, cannabis (45%), sedatives (34%), and amphetamines (23%). During the past three months, the clinical sample reported using opioids (for nonmedical purposes, 83%), sedatives (56%), and cannabis (52%) most frequently; in the community sample, the comparable percentages were sedatives (19%), cannabis (17%), and opioids (for nonmedical purposes, 12%). Alcohol use in the past three months was common in both samples: 69% of the clinical sample and 52% of the community sample. Regarding tobacco use, 86% of the clinical sample were current smokers vs 31% for the community sample. The samples differed in median age of first substance use (clinical sample = 14 years, community sample = 17 years) and the percentage using prior to age 15 years (clinical sample = 61%, community sample = 26%), a commonly cited risk factor [49–51]. The lifetime prevalence of injecting any drug was 44% in the clinical sample and 18% in the community sample.

### *Pain and Medical History in the Two Samples*

Table 3 summarizes the presence of chronic pain and the levels of acute pain reported in the two samples. Reports of chronic pain were prevalent in both the clinical and community samples, with substantial percentages of acute pain in the “moderate” to “very severe” range during the past seven days. Table 3 also summarizes the history of selected medical conditions reported in the two samples. About 40% of both samples reported “poor” or “fair” general health, with an average of 1.6 physical health conditions (median = 1). Large percentages of both samples reported lifetime “problems” in mental health (clinical = 82%, community = 47%).

### *Frequency Distributions of Items*

Among the initial item pool of 25 items, there were no items with any response categories having less than 1% response, but there were five items having at least one response category with less than 3% response.

**Table 2** Substance use reported on the ASSIST

Substance	Ever used		Used in past 3 mo	
	Clinical sample (N = 81) %	Community sample (N = 367) %	Clinical sample (N = 81) %	Community sample (N = 367) %
Opioids	91.3	20.3	82.7	11.7
Cannabis	85.0	44.9	51.9	16.9
Cocaine	81.5	21.3	29.6	2.7
Sedatives	74.1	34.4	55.6	19.1
Amphetamines	65.4	22.8	33.3	5.4
Hallucinogens	48.1	19.1	3.7	2.5
Inhalants	13.6	6.3	1.2	2.2
Tobacco	92.6	63.4	86.4	31.1
Alcohol	92.6	74.0	69.1	52.0

The use of opioids reported on the ASSIST refers to nonmedical use. All members of the current samples reported having a legitimate prescription for opioid pain medications *and* potential indicators of misuse during the past three months.

**Table 3** Pain ratings and medical status

	Clinical sample (N = 81)	Community sample (N = 367)
<i>Pain ratings</i>		
Chronic pain (longer than 6 mo)	73%	70%
“Moderate” to “very severe” pain		
Pain at its worst <sup>a</sup> (past 7 d)	80%	76%
Average pain <sup>a</sup> (past 7 d)	78%	79%
Pain right now <sup>a</sup>	57%	66%
<i>Medical status</i>		
“Poor” or “fair” general health <sup>b</sup>	42%	37%
Physical health problems (lifetime)		
Sleep problems	57%	43%
Liver problems	33%	15%
Heart problems	25%	50%
Sexually transmitted disease	24%	9%
Diabetes	12%	23%
Cancer	4%	14%
Stroke	0%	6%
HIV/AIDS	0%	3%
Mental health problems (lifetime)	82%	47%

<sup>a</sup> The response options for these items were “had no pain,” “mild,” “moderate,” “severe,” or “very severe.” The percentages reflect endorsements at the level of “moderate” or higher.

<sup>b</sup> The response options for this item were “poor,” “fair,” “good,” “very good,” or “excellent.” The percentages reflect endorsements at the level of “poor” or “fair.”

However, the sparse cells for all five items had at least seven respondents, ranging from seven to 13. Therefore, we retained all five response categories for all items for further analyses.

*Factor Analyses*

**Exploratory Factor Analyses**

The initial EFA of the 25 items (N = 232) yielded three factors with eigenvalues greater than one. However, the first factor (with an eigenvalue of 16.4) dominated, with a ratio of the first to second eigenvalue of 9.3. Inspection of one- through three-factor solutions suggested that the one-factor solution was the most interpretable. Based on these results, however, we chose to delete one item with a small factor loading (.09; “saved my unused prescription pain medication just in case I needed it later”) and one item with content not exclusive to the use of pain medication (“used street drugs because they treated my pain better than my prescription pain medication”). We repeated the EFA with the remaining 23 items, and this analysis produced a single factor with all item loadings greater than .63.

**Confirmatory Factor Analysis**

We performed a single-factor CFA on the reduced item pool, using the second half of the sample (N = 216). With the smaller pool of 23 items, we found sufficient evidence of unidimensionality to allow us to proceed with IRT calibrations. The factor loading for one item was .47, with all other factor loadings greater than .60. The conventional fit indices were generally strong, although the RMSEA was modest: CFI = .930, TLI = .923, and RMSEA = .118. Nonetheless, the results were adequate to document that responses to these items were largely a function of a single underlying latent construct.

**Table 4** Calibrated items: Abuse of prescription pain medication

Item stem	Slope (discrimination)	Location thresholds			
		Level 1	Level 2	Level 3	Level 4
I abused prescription pain medication*	3.66	0.03	0.48	1.08	1.62
I ran out of my prescription pain medication early*	3.46	-0.17	0.29	1.02	1.48
I got prescription pain medication from someone other than my healthcare provider*	3.41	0.31	0.64	1.26	2.04
I used more of my prescribed pain medication than I was supposed to*	3.08	-0.51	0.04	0.94	1.52
My prescription pain medication was gone too soon	3.00	-0.39	0.18	0.88	1.43
I used pain medication against my healthcare provider's advice	2.90	0.28	0.80	1.42	2.07
I experienced cravings for pain medication*	2.86	0.00	0.38	1.03	1.74
I borrowed prescription pain medication from someone	2.82	0.16	0.62	1.52	2.26
When my prescription for pain medication ran out, I felt anxious*	2.76	-0.30	0.32	0.89	1.47
I used someone else's prescription pain medication	2.75	0.11	0.57	1.37	2.05
I used more pain medication before the effects wore off*	2.54	-0.49	0.12	0.96	1.79
I hid my use of prescribed pain medication from others	2.52	0.21	0.54	1.17	1.68
I wanted more prescription pain medication to relieve my pain	2.41	-0.83	-0.23	0.74	1.53
I needed more prescription pain medication to relieve my pain	2.37	-0.94	-0.32	0.69	1.51
I felt better with a higher dose of pain medication than prescribed	2.24	-0.78	-0.05	0.68	1.50
I went to the emergency room to get additional pain medication	2.18	0.58	1.07	1.76	2.55
My prescription pain medication was less effective than it used to be	2.03	-0.73	0.13	0.92	1.60
Other people obtained pain medication for me from their own healthcare providers	2.00	0.76	1.13	2.06	2.91
I told my healthcare provider that I lost my pain medication and I needed more	1.81	0.93	1.37	2.05	2.82
I kept a hidden supply of pain medication	1.69	0.36	0.88	1.77	2.61
I got the same prescription pain medication from more than one healthcare provider	1.43	0.94	1.41	2.24	3.44
I counted the hours to know when I could take my next dose of pain medication	1.26	-0.65	0.19	1.24	2.28

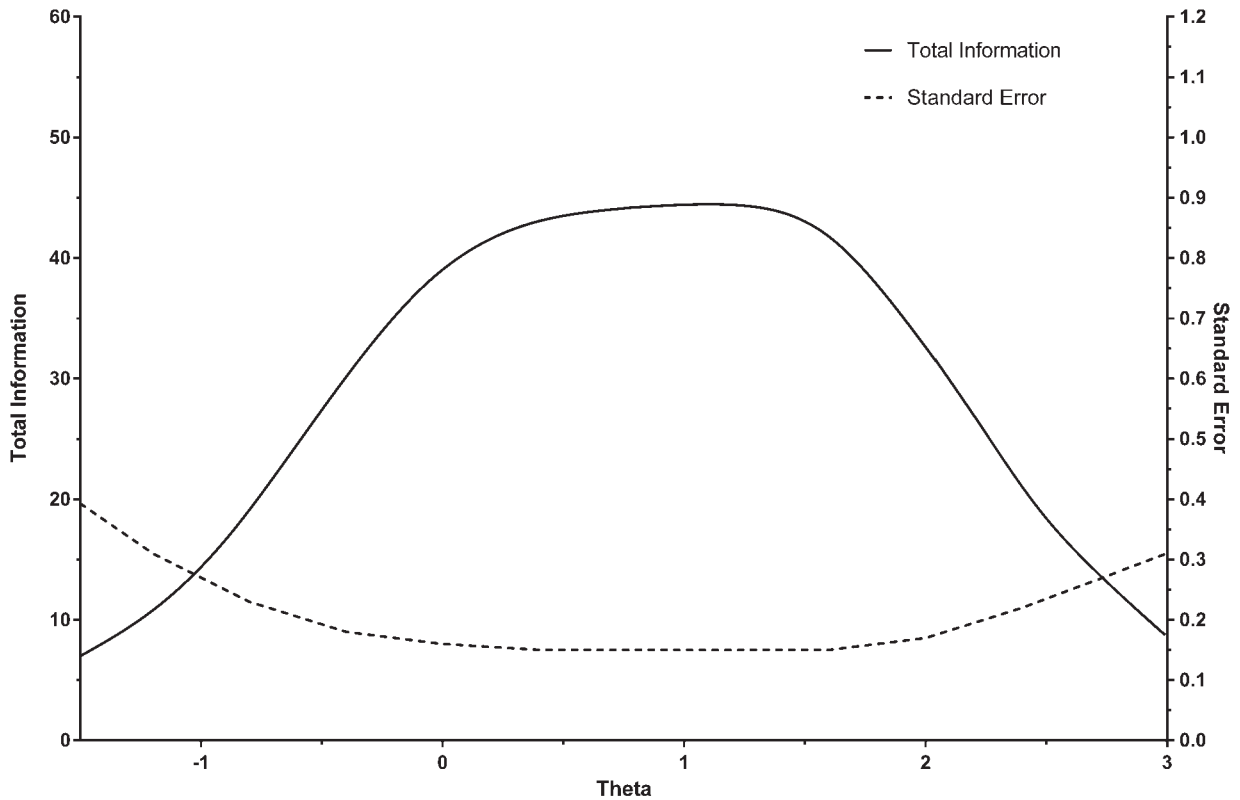
Items are rank-ordered on the basis of their slope (discrimination) parameters. Items included in the short form are marked with an asterisk (\*).

### IRT Calibrations

The remaining 23 items from the prescription pain medication pool were calibrated using the two-parameter GRM. Discrimination parameters ranged from 1.26 to 3.66. One item ("used additional medications to help my prescription pain medication work better") showed local dependence (residual correlations) with other items and was removed. We inspected the individual item information functions for the remaining 22 items. All items contributed meaningful information and were retained for the final bank. DIF analyses showed that the IRTPRO

and logistic regression methods did not jointly identify any items displaying DIF on the basis of age, gender, education, or more vs less severe substance use; thus, no items were eliminated for this reason. The Flesch-Kincaid readability test documented that the grade level for the final 22 items was grade 8.1.

Table 4 summarizes the items in the final bank, together with their IRT parameters. Figure 1 displays the test information curve (and the plot of the corresponding standard error), with  $\theta$  (severity) represented in its usual standardized form, i.e., with a mean of 0 and SD of 1.



**Figure 1** Test information curve for the item bank for abuse of prescription pain medication.

A standard error of .30 corresponds approximately to a CTT reliability of .90. At this threshold, the effective range of measurement for the item bank is about -1 to +3 SDs. This range is broader than that typically found for measures of substance abuse, which often assess high levels of severity.

*Selection of Items for Short Forms*

For some applications where computerized adaptive testing (CAT) is not feasible, static short forms may be a useful alternative. To develop short forms, we rank ordered all 22 items on four criteria: discrimination parameter, the percentage of times the item would have been selected in a simulated CAT based on the observed data from our calibration sample, expected information under the standard normal distribution with a mean of 0 and SD of 1, and expected information under a normal distribution with a larger SD, i.e., a mean of 0 and SD of 1.5 (23). The CAT simulations were performed using the Firestar program [24]. For the CAT simulations, we set the minimum number of items to be administered to eight and the maximum number of items to be administered to be the full bank. We selected seven items for the short form based on the convergence of the four psychometric criteria, the content of candidate items, and location parameters (i.e., we tried to include some items with lower thresholds to

increase the precision of the short form closer to the floor). The internal consistency of the short form was excellent, with an alpha coefficient of .87. The correlation between the theta score derived from the short form and the corresponding full item bank was .95. In Table 4, asterisks identify the items selected for the short form.

*Preliminary Validity Evidence*

To provide preliminary results regarding convergent validity, we examined the relationship between the  $\theta$  score from the new abuse of prescription pain medication item bank and the total score from the PMQ—the correlation was .73. We also computed ASSIST substance involvement (SI) scores and examined their correlations with the  $\theta$  score from the new item bank. The correlation between the total ASSIST SI score and the item bank was .44. As expected, the SI score for opioids (which focuses on nonmedical use of opioids) showed the strongest relationship with the abuse of prescription pain medication item bank:  $r = .59$ . The next largest correlations appeared with the SI scores for tobacco ( $r = .45$ ) and sedatives ( $r = .33$ ). Among current smokers, the mean  $\theta$  score was .46 vs  $-.32$  for nonsmokers, a difference of .78 SD units ( $P < .001$ ). Among current sedative users, the mean  $\theta$  score was .60 vs  $-.21$  for nonusers, a difference of .81 SD units ( $P < .001$ ). Finally, the correlation between the prescription pain medication item bank and the PROMIS



alcohol use short form was .44 (although the correlation with the ASSIST alcohol SI score was smaller, at .23). Given the sample size, all correlations were significant with  $P < .001$  (two-tailed).

We examined the impact of age at first use of substances on scores from the abuse of prescription pain medication item bank. We compared the high-risk group of those using substances prior to age 15 years (34% of the sample) with the remainder of the sample. As expected, those with earlier onset had higher scores on the item bank. The mean  $\theta$  score was .42 for early onset vs .00 for later onset, a difference of .42 SD units ( $P < .001$ ). Lifetime history of a mental health “problem” was also associated with greater abuse. Among respondents who reported having a “problem” with their mental health, the mean  $\theta$  score was .18 vs -.20 for those denying such problems, a difference of .38 SD units ( $P < .001$ ).

By contrast, correlations between the abuse of prescription pain medication item bank and perceptions of global health and pain (both acute and chronic) were small, with absolute values in a range from .05 to .18. In addition, there was no relationship between medical burden and scores on the new item bank. A median split based on our survey of eight physical health conditions revealed no difference between the 44% of the sample with two or more lifetime medical conditions (mean  $\theta$  score = -.02) and the 56% of the sample with zero or one medical conditions (mean  $\theta$  score = .03).

## Discussion

Our mixed (qualitative and quantitative) PROMIS methodology produced a 22-item bank (and a seven-item short form) for the assessment of severity of abuse of prescription pain medication. Psychometric analyses documented the unidimensionality of the 22 items, making them suitable for calibration with the IRT graded response model. Nonetheless, the content of the items is somewhat varied, including items related to excessive consumption, craving, and efforts to procure larger amounts of medication. There are trade-offs between bandwidth (item banks that have good content validity and capture a somewhat varied pool of clinical indicators) and fidelity (item banks that are more limited in content). We tried to achieve an appropriate balance by ensuring that the item bank was suitable for unidimensional scaling without unduly narrowing the construct. There is a difference between studying the dimensionality of a correlation matrix vs determining the degree to which “scores” are influenced by a single common factor, and even multidimensional data can result in scores that still reflect essentially one common influence [52]. We believe that our PROMIS item bank strikes a reasonable compromise in this regard.

Screening to identify high-risk patients prior to starting opioid therapy and monitoring of patients during treatment to detect possible signs of misuse are now considered to be important components of care [28]. A

systematic review of research on the use of opioids for chronic noncancer pain found limited work on tools to screen and monitor patients for risk of opioid misuse and recommended development of more effective instruments [53,54]. Our development of an item bank for abuse of prescription pain medication is responsive to the need for improved monitoring of patients receiving opioid pain medication. For repeated administration, a bank of items calibrated using IRT models has several advantages. Such items can be administered as CATs in which different items may be selected at different times, which is helpful for reducing the “practice effects” and tedium that can occur with repeated use of identical items in fixed-form tests. Also, the PROMIS experience with CATs suggests that only four to six items are necessary to generate precise estimates of the construct being measured. Brevity is an advantage for repeated assessment.

Our initial results on the lack of relationships between the new item bank and physical health suggest that pain and medical burden *per se* are not linked to abuse of prescription opioids. Thus, it may be more fruitful to look elsewhere for general risk factors. In this context, use of any other drug (including tobacco and alcohol) was associated with higher scores on the new item bank, but relatively stronger associations appeared with tobacco and sedative use. These results with the ASSIST SI scores are consistent with recent findings from the National Surveys on Drug Use and Health [55], which documented that nicotine dependence and sedative use disorders are risk factors for problematic prescription opioid use. Han et al. [55] also cited depression as a risk factor, and in the present analyses problems in mental health were associated with higher scores on the abuse of prescription pain medication item bank.

Limitations in the current work include a constraint on the implementation of the new item bank and the need for further studies of its validity. The bank includes a screening question that asks about the respondent’s having a legitimate prescription for opioid pain medication, and it proceeds only if the answer is “yes.” Many users, of course, obtain such substances without a prescription. For such respondents, the generic PROMIS substance use item banks (for severity of illicit drug use and positive appeal of use) would be appropriate [8], but these item banks do not yield specific information about procuring and consuming prescription pain medications. The two other PROMIS item banks for substance use have a broader reach, and they fulfill the assessment needs of investigators and clinicians examining substance use and abuse in a more general way. The new item bank for abuse of prescription pain medications fills a narrower niche—one relevant to respondents who have a prescription for such drugs and who should be monitored for how they procure and consume such drugs. On a related methodological note, it should be acknowledged that, for the development of the current item bank, we did not validate the accuracy

of the screening questions for misuse beyond the participant's self-report. Our goal, however, was not to characterize the nature or diversity of misuse but rather to accrue a sample from both the clinic and the community that was willing to endorse some departures from the use of pain medication as prescribed.

In addition, data from the calibration sample provided only preliminary evidence of validity. Validity studies can take many forms, but consistent with past PROMIS precedents [56], it would be valuable to conduct 1) psychometric studies comparing the operating characteristics of the PROMIS item bank for abuse of prescription pain medication with other commonly used measures in this area and 2) longitudinal studies of change in circumstances where one would (or would not) expect variability over time. Given that validation is an evolving process, it remains important to test the operating characteristics of the new item bank in a variety of contexts—medical, psychiatric, and epidemiological—and across the lifespan, including adolescent, young adult, and geriatric samples. Such work is consistent with the general mandate for PROMIS: to provide a single metric for assessing symptoms and health-related quality of life across all chronic diseases and along the full spectrum of severity (characteristic of community as well as clinical samples). Although PROMIS measures are intended to assess severity (rather than diagnosis), it is common for investigators and clinicians to ask about their relevance for screening and diagnostic purposes and about thresholds on the severity metric that should motivate clinical concern or intervention. Thus, further investigation of the sensitivity of the new item bank in predicting subsequent misuse, abuse, and related events would also be valuable.

In summary, the development of a new item bank for abuse of prescription pain medication adds to the existing body of PROMIS measures and is responsive to the need for additional instruments to assess the risk associated with prescribing opioid pain medication and to monitor potential misuse during treatment. The initial psychometric characteristics of the item bank support its use as a CAT or short form, with either version providing a brief, precise, and efficient measure relevant to both clinical and community samples. Further studies of the validity of the item bank are now appropriate to develop a better understanding of its measurement properties.

**Acknowledgments**

We would like to thank our collaborators at the recruitment sites: Center for Psychiatric and Chemical Dependency Services, Pittsburgh, PA (Dorothy J. Sandstrom, MS; Janis McDonald; Trey Ghee, MSW); CODA, Inc., Portland, OR (Katharina Wiest, PhD; Rosalie Gordon, BA; Kasie Cloud, MSW), Coastal Horizons Center, Inc., Wilmington, NC (Kenny S. G. House, BA, LCAS, CCS; Michael Younkins, BS, BBA), and Southlight Healthcare, Raleigh, NC (Alyssa H. Kalata, PhD, Scott Luetgenau, BSW, Reynolds (Tad) Clodfelter, Jr., PsyD).

**References**

- 1 Cella D, Gershon R, Lai JS, Choi S. The future of outcomes measurement: Item banking, tailored short forms, and computerized adaptive assessment. *Qual Life Res* 2007;16(Suppl 1):133–44.
- 2 Cella D, Riley W, Stone A, et al. The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item bank: 2005–2008. *J Clin Epidemiol* 2010;63:1179–94.
- 3 Reeve BB, Hays RD, Bjorner JB, et al. Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the Patient-Reported Outcome Measurement Information System (PROMIS). *Med Care* 2007;45:S22–31.
- 4 Hilton TF. The promise of PROMIS® for addiction. *Drug Alcohol Depend* 2011;119:229–34.
- 5 Pilkonis PA, Choi SW, Reise SP, et al. Item banks for measuring emotional distress from the patient-reported outcomes measurement information system: Depression, anxiety, anger. *Assessment* 2011; 18(3):263–83.
- 6 Revicki D, Chen W, Harnam N, et al. Development and psychometric analysis of the PROMIS pain behavior item bank. *Pain* 2009;146(1–2):158–69.
- 7 Fries JF, Cella D, Rose M, Krishnan E, Bruce B. Progress in assessing physical function in arthritis: PROMIS short forms and computerized adaptive testing. *J Rheumatol* 2009;36(9):2061–6.
- 8 Pilkonis P, Yu L, Dodds NE, et al. Item banks for substance use from the Patient-Reported Outcomes Measurement Information System (PROMIS): Severity of use and positive appeal of use. *Drug Alcohol Depend* 2015;156:184–192.
- 9 Pilkonis PA, Yu L, Colditz J, et al. Item banks for alcohol use from the Patient-Reported Outcomes Measurement Information System (PROMIS): Use, consequences, and expectancies. *Drug Alcohol Depend* 2013;130:167–77.
- 10 Buysse D, Yu L, Moul DE, et al. Development and validation of patient-reported outcome measures for sleep disturbance and sleep-related impairments. *Sleep* 2010;33(6):781–92.
- 11 Adams LL, Gatchel RJ, Robinson RC, et al. Development of a self-report screening instrument for assessing potential opioid medication misuse in chronic pain patients. *J Pain Symptom Manage* 2004;27(5):440–59.

- 12 Butler SF, Budman SH, Fernandez KC, Jamison RN. Validation of a screener and opioid assessment measure for patients with chronic pain. *Pain* 2004; 112(1-2):65–75.
- 13 Butler SF, Budman SH, Fernandez KC, et al. Development and validation of the current opioid misuse measure. *Pain* 2007;130(1–2):144–56.
- 14 Webster LR. Predicting aberrant behaviors in opioid-treated patients: Preliminary validation of the opioid risk tool. *Pain Med* 2005;6(6):432–42.
- 15 Kingsbury GG, Weiss DJ. An alternate-forms reliability and concurrent validity comparison of Bayesian adaptive and conventional ability tests Minneapolis. Research Report 80-5. University of Minnesota, Department of Psychology, Psychometrics Methods Program, Computerized Adaptive Testing Laboratory; 1980.
- 16 Kingsbury GG, Weiss DJ. A comparison of IRT-based adaptive mastery testing and a sequential mastery testing procedure. In: Weiss DJ, ed. *New Horizons in Testing: Latent Trait Theory and Computerized Adaptive Testing*. New York: Academic Press; 1983:257–83.
- 17 McBride JR, Martin JR. Reliability and validity of adaptive ability tests in a military setting. In: Weiss DJ, ed. *New Horizons in Testing: Latent Trait Theory and Computerized Adaptive Testing*. New York: Academic Press; 1983:223–36.
- 18 Lord FM. *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1980.
- 19 Weiss DJ. Computerized adaptive testing for effective and efficient measurement in counseling and education. *Meas Eval Couns Dev* 2004;37:70–84.
- 20 Weiss DJ. Adaptive testing by computer. *J Consult Clin Psychol* 1985;53:774–89.
- 21 Gibbons RD, Weiss DJ, Kupfer DJ, et al. Using computerized adaptive testing to reduce the burden of mental health assessment. *Psychiatr Serv* 2008; 59:361–8.
- 22 Bjorner JB, Chang CH, Thissen D, Reeve BB. Developing tailored instruments: Item banking and computerized adaptive assessment. *Qual Life Res* 2007;16:95–108.
- 23 Choi SW, Reise SP, Pilkonis PA, Hays RD, Cella D. Efficiency of static and computer adaptive short forms compared to full-length measures of depressive symptoms. *Qual Life Res* 2010;19:125–36.
- 24 Choi S. Firestar: Computerized adaptive testing simulation program for polytomous item response theory models. *Appl Psychol Meas* 2009;33(8):644–5.
- 25 Gardner W, Shear K, Kelleher KJ, et al. Computerized adaptive measurement of depression: A simulation study. *BMC Psychiatry* 2004;4:13.
- 26 Revicki DA, Sloan J. Practical and philosophical issues surrounding a national item bank: If we build it will they come? *Qual Life Res* 2007;16:167–74.
- 27 Stinson FS, Grant BF, Dawson DA, et al. Comorbidity between DSM-IV alcohol and specific drug use disorders in the United States: Results from the national epidemiologic survey on alcohol and related conditions. *Drug Alcohol Depend* 2005; 80(1):105–16.
- 28 Brady KT, McCauley JL, Back SE. Prescription opioid misuse, abuse, and treatment in the United States: An update. *Am J Psychiatry* 2016;173 (1):18–26.
- 29 Cicero TJ, Ellis MS, Harney J. Shifting patterns of prescription opioid and heroin abuse in the United States. *N Engl J Med* 2015;373(18):1789–90.
- 30 Cicero TJ, Ellis MS, Surratt HL, Kurtz SP. The changing face of heroin use in the United States: A retrospective analysis of the past 50 years. *JAMA Psychiatry* 2014;71(7):821–6.
- 31 Mars SG, Bourgois P, Karaninos G, Montero F, Ciccarone D. “Every ‘never’ I ever said came true”: Transitions from opioid pills to heroin injecting. *Int J Drug Policy* 2014;25(2):257–66.
- 32 Klem ML, Saghabi E, Abromitis R, et al. Building PROMIS item banks: Librarians as co-investigators. *Qual Life Res* 2009;18(7):881–8.
- 33 Castel LD, Williams KA, Bosworth HB, et al. Content validity in the PROMIS social health domain: A qualitative analysis of focus group data. *Qual Life Res* 2008;17(5):737–49.
- 34 Kelly MA, Morse JQ, Stover A, et al. Describing depression: Congruence between patient experiences and clinical assessments. *Br J Clin Psychol* 2011;50 (1):46–66.
- 35 DeWalt DA, Rothrock N, Yount S, Stone AA, on behalf of the PROMIS Cooperative Group. Evaluation of item candidates: The PROMIS qualitative item review. *Med Care* 2007;45:S12–21.
- 36 Berzon R, Patrick D, Guyatt G, Conley JM. Intellectual property considerations in the

- development and use of HRQL measures for clinical trial research. *Qual Life Res* 1994;3(4):273–7.
- 37 Revicki D, Schwartz CE. Intellectual property rights and good research practice. *Qual Life Res* 2009;18:1279–80.
- 38 Wilkinson GS. *The Wide Range Achievement Test: Manual*, 3rd edition. Wilmington, DE: Wide Range; 1993.
- 39 Wilkinson GS, Robertson GJ. *WRAT4: Wide Range Achievement Test Professional Manual*. Lutz, FL: Psychological Assessment Resources; 2006.
- 40 McCarty D, Fuller B, Kaskutas LA, et al. Treatment programs in the national drug abuse treatment clinical trials network. *Drug Alcohol Depend* 2008;92:200–7.
- 41 Tai B, Straus MM, Liu D, et al. The first decade of the national drug abuse treatment clinical trials network: Bridging the gap between research practice to improve drug abuse treatment. *J Subst Abuse Treat* 2010;38(Suppl 1):S4–13.
- 42 WHO ASSIST Working Group. *The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): Development, reliability and feasibility*. *Addiction* 2002;97(9):1183–94.
- 43 Muthén LK, Muthén B. *Mplus User's Guide*, 4th edition. Los Angeles, CA: Muthén & Muthén; 2006.
- 44 Samejima F. *Estimation of Latent Ability Using a Response Pattern of Graded Scores* (Psychometric Monograph No. 17). Richmond, VA: Psychometric Society; 1969.
- 45 Cai L, Thissen D, du Toit SHC. *IRTPRO: Flexible, Multidimensional, Multiple Categorical IRT Modeling* [Computer Software]. Lincolnwood, IL: Scientific Software International; 2011.
- 46 Thissen D, Steinberg L, Wainer H. Detection of differential item functioning using the parameters of item response models. In: Holland PW, Wainer H, eds. *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum; 1993:67–113.
- 47 Zumbo BD. *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-type (Ordinal) Item Scores*. Ottawa, ON: Directorate of Human Resources Research and Education, Department of National Defense; 1999.
- 48 Teresi JA, Ocepek-Welikson K, Kleinman M, et al. Analysis of differential item functioning in the depression item bank from the Patient Reported Outcome Measurement Information System (PROMIS): An item response theory approach. *Psychol Sci Q* 2009;51(2):148–80.
- 49 Swendsen J, Conway KP, Degenhardt L, et al. Socio-demographic risk factors for alcohol and drug dependence: The 10-year follow-up of the national comorbidity survey. *Addiction* 2009;104(8):1346–55.
- 50 Grant BF, Dawson DA. Age of onset of drug use and its association with DSM-IV drug abuse and dependence: Results from the national longitudinal alcohol epidemiologic survey. *J Subst Abuse* 1998;10(2):163–73.
- 51 King KM, Chassin L. A prospective study of the effects of age of initiation of alcohol and drug use on young adult substance dependence. *J Stud Alcohol Drugs* 2007;68(2):256–65.
- 52 Reise SP, Moore TM, Haviland MG. Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *J Pers Asses* 2010;92(6):544–59.
- 53 Chou R, Ballantyne JC, Fanciullo GJ, Fine PG, Miaskowski C. 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. *J Pain* 2009;10(2):147–59.
- 54 Chou R, Fanciullo GJ, Fine PG, et al. 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. *J Pain* 2009;10(2):131–46.
- 55 Han B, Compton WM, Jones CM, Cai R. Nonmedical prescription opioid use and use disorders among adults aged 18 through 64 years in the United States, 2003–2013. *JAMA* 2015;314(14):1468–78.
- 56 Pilkonis PA, Yu L, Dodds NE, et al. Validation of the depression item bank from the Patient-Reported Outcomes Measurement Information System (PROMIS) in a three-month observational study. *J Psychiat Res* 2014;56:112–9.

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.