Drug and Alcohol Dependence 130 (2013) 167–177

Contents lists available at SciVerse ScienceDirect

Drug and Alcohol Dependence

journal homepage: www.elsevier.com/locate/drugalcdep

Item banks for alcohol use from the Patient-Reported Outcomes Measurement Information System (PROMIS®): Use, consequences, and expectancies

Paul A. Pilkonis[a,*], Lan Yu[a], Jason Colditz[a], Nathan Dodds[a], Kelly L. Johnston[a], Catherine Maihoefer[a], Angela M. Stover[a], Dennis C. Daley[a], Dennis McCarty[b]

[a] Department of Psychiatry, University of Pittsburgh Medical Center, Pittsburgh, PA 15213, United States
[b] Department of Public Health and Preventive Medicine, Oregon Health and Science University, Portland, OR 97239, United States

ARTICLE INFO

Article history:
Received 20 February 2012
Received in revised form 1 November 2012
Accepted 1 November 2012
Available online 1 December 2012

Keywords:
Alcohol use
Alcohol consequences
Alcohol expectancies
Item response theory
Measurement

ABSTRACT

Background: We report on the development and calibration of item banks for alcohol use, negative and positive consequences of alcohol use, and negative and positive expectancies regarding drinking as part of the Patient-Reported Outcomes Measurement Information System (PROMIS®).
Methods: Comprehensive literature searches yielded an initial bank of more than 5000 items from over 200 instruments. After qualitative item analysis (including focus groups and cognitive interviewing), 141 items were included in field testing. Items for alcohol use and consequences were written in a first-person, past-tense format with a 30-day time frame and 5 response options reflecting frequency. Items for expectancies were written in a third-person, present-tense format with no time frame specified and 5 response options reflecting intensity. The calibration sample included 1407 respondents, 1000 from the general population (ascertained through an internet panel) and 407 from community treatment programs participating in the National Institute on Drug Abuse (NIDA) Clinical Trials Network (CTN).
Results: Final banks of 37, 31, 20, 11, and 9 items (108 total items) were calibrated for alcohol use, negative consequences, positive consequences, negative expectancies, and positive expectancies, respectively, using item response theory (IRT). Seven-item static short forms were also developed from each item bank.
Conclusions: Test information curves showed that the PROMIS item banks provided substantial information in a broad range of severity, making them suitable for treatment, observational, and epidemiological research.

1. Introduction

The Patient-Reported Outcomes Measurement Information System (PROMIS®) is an NIH Roadmap initiative designed to improve self-reported outcomes using state-of-the-art psychometric methods (for detailed information, see http://www.nihpromis.org/). PROMIS has developed and calibrated item banks assessing physical, mental, and social health, consistent with the World Health Organization's tripartite framework (Cella et al., 2007). For example, there are item banks assessing physical functioning, pain, fatigue, sleep disturbance, emotional distress (depression, anxiety, and anger), and social participation, providing a comprehensive profile of health status (Buysse et al., 2010; Cella et al., 2007, 2010; Fries et al., 2009; Pilkonis et al., 2011; Revicki et al., 2009). PROMIS is the most ambitious attempt to date to apply models from item response theory (IRT) to health-related assessment. The PROMIS methodology involves iterative steps of comprehensive literature searches; item pooling; development of conceptual frameworks; 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 (Cella et al., 2007, 2010; Hilton, 2011; Reeve et al., 2007). We report here on the development and calibration of five item banks capturing prominent aspects of alcohol use (consumption, craving, efforts at control, internal and external triggers for drinking), negative and positive consequences of alcohol use, and negative and positive expectancies regarding drinking.

There is informative previous work using IRT models for the scaling and calibration of criteria for substance abuse and dependence, including those for alcohol (Krueger et al., 2004; Langenbucher et al., 2004; Martin et al., 2006; Saha et al., 2006). These criteria are sufficiently unidimensional for calibration with IRT models, but they are high-threshold items most appropriate for clinical samples. The use of such items results in "tests" with narrow

* Corresponding author at: Western Psychiatric Institute and Clinic, 3811 O'Hara Street, Pittsburgh, PA 15213, United States. Tel.: +1 412 246 5833; fax: +1 412 246 5840.
E-mail address: pilkonispa@upmc.edu (P.A. Pilkonis).

0376-8716/$ – see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.drugalcdep.2012.11.002

bandwidth relevant only at the severe end of the continuum of alcohol use and consequences. From a psychometric perspective, our goal was to identify items that were more normally distributed and less positively skewed in a sample that included both members of the general population who used alcohol as well as alcohol abusers. Such items will provide more information across a broader range of the continuum of alcohol use, and for this reason, they will constitute more sensitive measures of treatment outcome and result in a single metric that could be used across treatment, observational, and epidemiological settings. Thus, we were interested in alcohol use not only at the level of clinical disorders but also at lower levels of consumption, where alcohol use may still be an important health-related behavior (or risk factor) relevant to a wide range of medical and psychiatric conditions.

2. Methods

2.1. Development of item pool

2.1.1. 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 alcohol domain. We performed searches in the MEDLINE, PsycINFO, and Health and Psychosocial Instruments (HaPI) databases. Details of the methodology are reported in Klem et al. (2009), and all search algorithms are available upon request. The searches generated 785 abstracts that could be linked to more than 200 unique measures of substance use. Cited reference searches were run on the primary reference for each measure in order 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.

2.1.2. Conceptual organization of items. The initial alcohol item pool contained 5241 items. We organized the items into conceptually meaningful categories using a hierarchical approach informed by previous empirical work (e.g., factor analyses) and clinical formulation. Previous work had divided alcohol use items into subdomains and factors relevant both to the DSM-IV categorization of alcohol use disorders, e.g., alcohol consumption, impairment associated with drinking (Green et al., 2011; Krueger et al., 2004; Muthén, 2006; Saha et al., 2007), and to broader themes surrounding the use of alcohol, e.g., precipitants to alcohol use, alcohol motives, expectancies regarding alcohol use (Jones et al., 2001; Pabst et al., 2009).

Our hierarchical structure for alcohol use included eight subdomains; consumption, craving and efforts to control drinking, triggers (internal and external), negative consequences, positive consequences, negative expectancies, positive expectancies, and general attitudes about alcohol. We also created 105 distinct facets within the subdomains. For example, within the consumption subdomain, we included facets for frequency, quantity, and patterns of alcohol use (e.g., binge versus consistent drinking).

2.1.3. Focus groups. To ensure comprehensive coverage of the conceptual area, we conducted focus groups and performed thematic analyses of the topics discussed (see Castel et al., 2008; Kelly et al., 2011). Members of four groups were recruited from outpatient substance use treatment programs. Two additional groups were comprised of social drinkers, community participants who reported drinking at least one alcoholic beverage in the past 30 days but who had no lifetime history of substance use disorders and no risk factors for current alcohol problems according to the National Institute on Alcohol Abuse and Alcoholism (NIAAA) screening guidelines (2003). Participants (total $n = 65$) were between the ages of 25 and 64 (mean = 45, SD = 10). They were predominantly female (68%) and members of minority groups (race = African American 58%, Caucasian 37%, Other 5%; ethnicity = Hispanic 5%). A majority had an annual household income of less than $20,000 (66%) and no formal education beyond high school (51%).

Using semi-structured scripts, facilitators prompted participants to discuss their experiences with alcohol and characteristics of problematic drinking. Research staff reviewed process notes from the groups and audio recordings, paying special attention to positive and negative appraisals (consequences of alcohol use, general expectations regarding alcohol) and contexts of drinking-related experiences. The goal was to enrich our item pool with content not represented on traditional questionnaires. For this purpose, we paid particular attention to accounts that suggested lower threshold items (e.g., did embarrassing things when drinking, rudeness, drinking routinely at the end of a busy day).

2.1.4. 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 DeWalt et al., 2007, 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 by virtue of being disease-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 about 250 items for field testing, with approximately 150 items for the alcohol bank and an additional 100 items devoted to demographic characteristics, health status, medical history, history of substance use, and "legacy" measures of alcohol use and abuse (to investigate convergent validity with the new item bank). With this goal in mind, we reduced the item pool to 147 items, covering 103 of the original 105 facets.

2.1.5. Standardization of items. Items for alcohol use and consequences were written in a first-person, past-tense format with a 30-day time frame and 5 response options reflecting frequency (e.g., In the past 30 days, I lied about my drinking: *never*, *rarely*, *sometimes*, *often*, *almost always*). Expectancies and general attitudes about alcohol use, however, represent enduring beliefs, and as such, these items used a third-person, present-tense format, no time frame, and an intensity scale (e.g., Drinking puts people in a bad mood: *not at all*, *a little bit*, *somewhat*, *quite a bit*, *very much*). A small number of consumption items used a scale of actual number of drinks (e.g., drinks in a typical week, largest number of drinks in a single day). This standardization of items was consistent with our usual efforts to promote internal consistency across PROMIS measures (DeWalt et al., 2007; Pilkonis et al., 2011). In addition, a review of intellectual property issues was completed for all items (Berzon et al., 1994; Revicki and Schwartz, 2009). The large majority of items were generic, that is, they were similar to several extant items but not identifiable with any one in particular.

2.1.6. Cognitive interviews. Twenty-eight participants were recruited for cognitive interviews, and items were reviewed by at least 9 individuals with a variety of characteristics: at least 3 female, 4 minority, 3 social drinkers, 5 less than high school graduate reading level, and 2 less than 9th grade reading level as assessed by the Wide Range Achievement Test (WRAT-4; Wilkinson, 1993; Wilkinson and Robertson, 2006). An interviewer met with participants and asked each to "think aloud" while responding to items, then prompted for feedback on the language and clarity of items and the relevance of the content. Adaptations arising from cognitive interview feedback included the removal of modifiers that increased the threshold of items (e.g., "I had a *strong* urge to continue drinking"), clarifying ambiguities (e.g., "Drinking eases *physical* pain," to differentiate this from emotional pain),