Drug and Alcohol Dependence 178 (2017) 94–100

Contents lists available at ScienceDirect

Drug and Alcohol Dependence

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

Full length article

Cross-validation of short forms of the Screener and Opioid Assessment for Patients with Pain-Revised (SOAPP-R)

Matthew D. Finkelman[a,*], Robert N. Jamison[b], Ronald J. Kulich[c,d], Stephen F. Butler[e], William C. Jackson[f], Niels Smits[g], Scott G. Weiner[h]

[a] Department of Public Health and Community Service, Tufts University School of Dental Medicine, 1 Kneeland St., Boston, MA 02111, USA
[b] Departments of Anesthesiology and Psychiatry, Brigham and Women's Hospital, Harvard Medical School, 75 Francis St., Boston, MA 02115, USA
[c] Craniofacial Pain and Headache Center, Tufts University School of Dental Medicine, 1 Kneeland St., Boston, MA 02111, USA
[d] Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, 15 Parkman St., Boston, MA 02114, USA
[e] Inflexxion, Inc., 890 Winter St., Ste. 235, Waltham, MA 02451, USA
[f] Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, 55 Fruit St. #148, Boston, MA 02114, USA
[g] Department of Methods and Statistics, Research Institute of Child Development and Education, University of Amsterdam, Faculty of Social and Behavioural Sciences, Nieuwe Achtergracht 127, 1018 WS Amsterdam, The Netherland
[h] Department of Emergency Medicine, Brigham and Women's Hospital, 75 Francis Street, NH-226, Boston, MA 02115, USA

ARTICLE INFO

Keywords:
Chronic pain
Opioids
Substance abuse
Risk stratification
Short form
Computer-based testing

ABSTRACT

Background: The Screener and Opioid Assessment for Patients with Pain-Revised (SOAPP-R) is a 24-item assessment designed to assist in the prediction of aberrant drug-related behavior (ADB) among patients with chronic pain. Recent work has created shorter versions of the SOAPP-R, including a static 12-item short form and two computer-based methods (curtailment and stochastic curtailment) that monitor assessments in progress. The purpose of this study was to cross-validate these shorter versions in two new populations.
Methods: This retrospective study used data from patients recruited from a hospital-based pain center (n = 84) and pain patients followed and treated at primary care centers (n = 110). Subjects had been administered the SOAPP-R and assessed for ADB. In real-data simulation, the sensitivity, specificity, and area under the curve (AUC) of each form were calculated, as was the mean test length using curtailment and stochastic curtailment.
Results: Curtailment reduced the number of items administered by 30% to 34% while maintaining sensitivity and specificity identical to those of the full-length SOAPP-R. Stochastic curtailment reduced the number of items administered by 45% to 63% while maintaining sensitivity and specificity within 0.03 of those of the full-length SOAPP-R. The AUC of the 12-item form was equal to that of the 24-item form in both populations.
Conclusions: Curtailment, stochastic curtailment, and the 12-item short form have potential to enhance the efficiency of the SOAPP-R.

1. Introduction

Prescription opioid analgesics are used frequently for patients with pain, and their use has risen at a rapid rate over the past decade, with prescribing levels beginning to stabilize and decrease only recently (Aitken et al., 2016; Kertesz, 2017; Kuehn, 2007). Unfortunately, increased opioid prescribing has also been paralleled by increases in opioid misuse and diversion (Okie, 2010). Recent data show that the rate of opioid abuse has increased, and deaths from opioid overdose have been labeled a national epidemic (Centers for Disease Control Prevention, 2012; Rudd et al., 2016).

To assist providers in determining the risk of aberrant drug-related behavior (ADB) among chronic pain patients, screening questionnaires have been developed. One commonly used questionnaire is the Screener and Opioid Assessment for Patients with Pain—Revised (SOAPP-R), a self-report instrument that classifies respondents as high or low risk for ADB based on a prescribed cutoff (Butler et al., 2008, 2009). The SOAPP-R is a modified version of the original Screener and Opioid Assessment for Patients with Pain (SOAPP) (Butler et al., 2004); the SOAPP-R was empirically derived (as opposed to the SOAPP, which was conceptually derived) and designed to contain more items that are less transparent in their scoring (Butler et al., 2008). The SOAPP-R exhibited sound psychometric characteristics in its validation and cross-validation studies (Butler et al., 2008, 2009).

* Corresponding author.
E-mail addresses: matthew.finkelman@tufts.edu (M.D. Finkelman), rjamison@bwh.harvard.edu (R.N. Jamison), rkulich@mgh.harvard.edu (R.J. Kulich), sfbutler@inflexxion.com (S.F. Butler), wcjackson@mgh.harvard.edu (W.C. Jackson), n.smits@uva.nl (N. Smits), sweiner@bwh.harvard.edu (S.G. Weiner).

http://dx.doi.org/10.1016/j.drugalcdep.2017.04.016
Received 16 January 2017; Received in revised form 17 April 2017; Accepted 26 April 2017
Available online 13 June 2017
0376-8716/

At 24 items, the SOAPP-R's length is manageable for many patients; nevertheless, the introduction of shorter versions may save time and improve utilization rates (Finkelman et al., 2015; Finkelman et al., 2017b). Indeed, the importance of test length is highlighted by the Scientific Advisory Committee of the Medical Outcomes Trust's categorization of respondent and administrative burden as a key attribute of a health questionnaire (Aaronson et al., 2002). While an abbreviated form of the original SOAPP has been developed (Koyyalagunta et al., 2013), shorter versions of the empirically derived SOAPP-R were desired, leading to two recent studies that suggested different approaches to reducing the length of the latter screener. The simpler of the two approaches is to use a static short form containing a subset of the SOAPP-R items; a retrospective study found that such a static short form consisting of 12 items exhibited sensitivity and specificity comparable to those of the full-length SOAPP-R (Finkelman et al., 2017b. The more complex approach is to administer the SOAPP-R via computer, track an individual's responses as he/she proceeds through the test, and stop the assessment if a computer algorithm determines that further items are unnecessary. For instance, if a respondent's item scores are high enough that he/she reaches the cutoff during testing—or low enough that it has become impossible for him/her to reach the cutoff—the test can be terminated and the appropriate classification can be made. This type of stopping rule, which has been well-studied in the psychometric literature, is referred to as *curtailment* or the *countdown method* (Ben-Porath et al., 1989; Butcher et al., 1985; Finkelman et al., 2015; Forbey et al., 2012). A variation on the above approach is to terminate testing if the SOAPP-R's classification ("high risk" or "low risk") has been determined with certainty from a respondent's previous answers, *or* if the classification has been determined up to a specified level of probability. This variation is referred to as *stochastic curtailment* (Finkelman et al., 2015). A retrospective study found that curtailment and stochastic curtailment produce considerable reductions in average test length while maintaining sensitivity and specificity similar to those of the full-length SOAPP-R (Finkelman et al., 2015).

While previous studies suggested the potential for shorter versions of the SOAPP-R (Finkelman et al., 2015; Finkelman et al., 2017b), their conclusions are limited by the fact that the results of each were based on a single dataset (which was common to both studies). Moreover, the research on the 12-item static short form indicated unstable results with respect to specific cutoffs, and recommended that the preliminary cutoff for this short form ($\geq$10 points) be validated in further study (Finkelman et al., 2017b). Both of the previous studies on short versions of the SOAPP-R emphasized that cross-validation should be conducted in other populations (Finkelman et al., 2015; Finkelman et al., 2017b). The objective of this study was to compare the static short form, curtailment, stochastic curtailment, and the full-length SOAPP-R in two different populations.

2. Material and methods

This retrospective study examined the performance of the full-length SOAPP-R and its short versions using two separate datasets. The Tufts Health Sciences Institutional Review Board granted exempt status or non-human subjects research status for the analysis of each dataset.

2.1. Versions of the SOAPP-R

2.1.1. Full-length form

The 24 items comprising this form are listed in Table 1. Each item can be answered "Never," "Seldom," "Sometimes," "Often," or "Very Often;" the scores for these answer choices are 0, 1, 2, 3, and 4, respectively. The total score on the SOAPP-R is the sum of the item scores; a higher total score indicates greater risk of ADB. The validation and cross-validation studies of the screener recommended a cutoff of $\geq$18 (Butler et al., 2008, 2009).

2.1.2. Curtailment

A curtailment rule is conducted on a fixed order of items. It stops testing once the screener's result ("high risk" or "low risk") has become determined from the respondent's previous answers. Applying this rule to the SOAPP-R with a $\geq$18 cutoff, a computerized version of the questionnaire would stop presenting items (in favor of a "high risk" result) if the respondent's cumulative score reached or exceeded 18. It would also stop presenting items (in favor of a "low risk" result) if the respondent's total score could not reach 18 even if the respondent answered "Very Often" to all remaining items.

Curtailment is a *variable-length testing* method: it produces different test lengths for different respondents. The number of items that a given respondent receives is dependent on his/her answers. In particular, a respondent whose screening result is determined quickly will receive a shorter test length than a respondent whose screening result is not determined until a large number of items have been presented. The maximum number of items that curtailment can administer is equal to the number of items on the full-length screener (24 items for the SOAPP-R). The minimum number of items that curtailment can administer depends on the cutoff; Section 3.3 will present the minimum possible number of items for the particular cutoff used in this study.

2.1.3. Stochastic curtailment

Like curtailment, stochastic curtailment is conducted on a fixed order of items. In stochastic curtailment of the SOAPP-R, early stopping occurs not only if the screener's result has become determined from previous answers, but also if the probability of obtaining one of the results ("high risk" or "low risk") has become adequately high. For the SOAPP-R, previous research (Finkelman et al., 2015) recommended setting the stopping threshold at 99% or 95% (i.e., terminating the screener if the probability of one of the results becomes at least 99%, or if it becomes at least 95%). The use of stochastic curtailment with the former threshold will be referred to as SC-99; its use with the latter threshold will be referred to as SC-95. At each stage of testing, the probability of a "high risk" result, based on the respondent's previous answers, is estimated based on a logistic regression model; see Finkelman et al. (2012) for details. As will be explained in Section 3.3, the set of scores that result in early stopping via stochastic curtailment, at each stage of testing, can be written as a simple look-up table. Finkelman et al. (2015) presented such look-up tables for curtailment, SC-99, and SC-95, but their tables are only applicable when a $\geq$19 cutoff is used. Therefore, in the current study, the data from Finkelman et al. (2015) were re-analyzed to produce look-up tables using the standard $\geq$18 SOAPP-R cutoff.

In sum, stochastic curtailment is a variable-length testing method that is less conservative than curtailment. As in curtailment, the number of items presented to a respondent by stochastic curtailment is based on the respondent's pattern of answers. The maximum number of items that stochastic curtailment can administer, when used in conjunction with the SOAPP-R, is 24; the minimum number of items is dependent on the cutoff. See Section 3.3 for the minimum possible number of items that SC-99 and SC-95 can administer when applied to the SOAPP-R with the cutoff used in this study.

2.1.4. Static short form

The development of the static short form of the SOAPP-R (i.e., the selection of items for this form) was based on both (i) statistical criteria and (ii) a scrutiny of content by an external set of pain practitioners (Finkelman et al., 2017b). The statistical component utilized data from 428 individuals who had taken the full-length SOAPP-R, and had also been classified as "negative" or "positive" for ADB by the Aberrant Drug Behavior Index (ADBI), as part of the screener's original validation study or cross-validation study (Butler et al., 2008, 2009). Using these data, candidate short forms of different lengths were developed and evaluated. In particular, for every test length of fewer than 24 items (i.e., for each test length between one item and 23 items), a candidate