A Comparison of Short Forms of the Screener and Opioid Assessment for Patients With Pain – Revised (SOAPP-R)

Finkelman, M.D.; Jamison, R.N.; Kulich, R.J.; Butler, S.F.; Smits, N.; Weiner, S.G.

DOI
10.1027/1015-5759/a000519

Publication date
2020

Document Version
Final published version

Published in
European Journal of Psychological Assessment

Multistudy Report

A Comparison of Short Forms of the Screener and Opioid Assessment for Patients With Pain – Revised (SOAPP-R)

Matthew D. Finkelman[1], Robert N. Jamison[2], Ronald J. Kulich[3,4], Stephen F. Butler[5], Niels Smits[6], and Scott G. Weiner[7]

[1]Department of Public Health and Community Service, Tufts University School of Dental Medicine, Boston, MA, USA
[2]Departments of Anesthesiology and Psychiatry, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[3]Craniofacial Pain and Headache Center, Tufts University School of Dental Medicine, Boston, MA, USA
[4]Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, MA, USA
[5]Inflexxion, Inc. Waltham, MA, USA
[6]Department of Methods and Statistics, Research Institute of Child Development and Education, University of Amsterdam, The Netherlands
[7]Department of Emergency Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Abstract: The Screener and Opioid Assessment for Patients with Pain – Revised (SOAPP-R) is a 24-item self-report questionnaire that assesses risk of aberrant medication-related behavior among chronic pain patients. Recently, an 8-item version of the SOAPP-R that weights items differentially was proposed. However, no previous study had compared the 8-item form with other short versions of the SOAPP-R, including a static 12-item short form and computer-based versions customizing the test length to the individual respondent. Moreover, no prior research had investigated combining the 8-item short form with customized computer-based stopping rules to further enhance efficiency. The objectives of this study were to compare the 8-item version with previously recommended short forms of the SOAPP-R, and to develop and evaluate a new version of the SOAPP-R combining the 8-item version with computer-based stopping rules. Versions were compared via sensitivity, specificity, and mean test length using real-data simulation of three datasets. Although results varied across datasets, the 8-item SOAPP-R compared favorably to previously recommended forms. Combining the 8-item form with computer-based stopping rules reduced the mean test length without affecting sensitivity or specificity; thus, the combined approach is recommended. The methodology used to shorten questionnaires via computer-based testing can also be applied to other instruments.

Keywords: Screener and Opioid Assessment for Patients with Pain – Revised (SOAPP-R), opioid misuse, short form, respondent and administrative burden, computer-based testing

It is estimated that 126.1 million adults in the United States experienced pain in the prior three months and that 25.3 million experience chronic pain (Nahin, 2015). Opioids are frequently prescribed to treat pain, but their use is sometimes problematic given the current overdose and abuse epidemic in the United States (Rudd, Seth, David, & Scholl, 2016). Prescribers of opioids are becoming more judicious with their use; while overdose deaths are rising because of illicit use of heroin and fentanyl, there has been a decrease in overdose deaths caused by prescription opioids over the past several years (Hedegaard, Warner, & Minino, 2017). The annual per capita morphine milligram equivalents (MME) prescribed decreased by 18% in 2015 compared with the peak in 2010 (Guy et al., 2017). Still, in 2015 the per capita MME prescribed was 640 mg, which is three times as high as the amount in 1999 (Guy et al., 2017). In 2014, nearly 2 million Americans either abused or were dependent on prescription opioids (Center for Behavioral Health Statistics and Quality, 2016).

Safe use of prescription opioids is guided by several principles as delineated in the CDC Guideline for Prescribing Opioids for Chronic Pain, such as the need to first use non-opioid therapy for the treatment of pain and to use caution when prescribing extended-release and long-acting opioids and avoiding high MME daily doses (Dowell, Haegerich, & Chou, 2016). Other tools clinicians can use

https://doi.org/10.1027/1015-5759/a000519

to ensure safe use of opioids are screening instruments. One commonly used example is the Screener and Opioid Assessment for Patients with Pain – Revised (SOAPP-R; Butler, Fernandez, Benoit, Budman, & Jamison, 2008). The SOAPP-R is a 24-question tool designed to detect high-risk patients who are being evaluated for opioid therapy. The SOAPP-R was prospectively derived and validated, and has been demonstrated to provide excellent discrimination between high and low risk patients (Butler, Budman, Fernandez, Fanciullo, & Jamison, 2009; Passik, Narayana, & Yang, 2014). A high SOAPP-R score ($\geq$ 18 points) also correlates with increased likelihood of drug abuse (Chou et al., 2009) and, in emergency department patients, a high SOAPP-R score is associated with using multiple providers for controlled substance prescriptions (Weiner, Horton, Green, & Butler, 2016). In its original validation study, the SOAPP-R exhibited a sensitivity of 81% and a specificity of 68% for detecting aberrant medication-related behavior (Butler et al., 2008); in its cross-validation study, it had a sensitivity of 80% and a specificity of 52% (Butler et al., 2009).

In general, practitioners considering whether to use a given screener may take into account not only the screener's psychometric properties (e.g., sensitivity and specificity), but also the *respondent and administrative burden* that the screener may produce. Respondent burden refers to the demands that the screener places on the individuals taking it (such as time and effort), while administrative burden refers to the demands placed on those who provide and oversee it (Lohr, 2002). Respondent and administrative burden have been identified by the Scientific Advisory Committee of the Medical Outcomes Trust as one of eight key attributes of health instruments (Lohr, 2002). At 24 items, the SOAPP-R is not unduly onerous for many individuals; however, given that the screener was developed for individuals with chronic pain (Butler et al., 2008, 2009), and respondent burden may be substantially greater for those with physical illness (Carpenter et al., 1998), shortened versions of the instrument may be of use for certain patients. Moreover, given the limited time of providers, and the consequent need for efficiency in health care (Dugdale, Epstein, & Pantilat, 1999), the cumulative time saved by administering a shorter version to multiple patients may amount to a significant decrement in administrative burden.

In light of the value of reducing respondent and administrative burden, a considerable amount of research has been devoted to short forms of the SOAPP-R. Finkelman, Smits, et al. (2017) developed a 12-item form of the screener and provided initial evidence of its potential to achieve sensitivity and specificity similar to those of the full-length SOAPP-R. Another approach that has been studied (Finkelman et al., 2015) is to administer a version of the SOAPP-R by computer, use an internal algorithm to keep track of the respondent's cumulative score in real time as he/she provides answers to the items, and cease testing early if specified by carefully developed *stopping rules*. The stopping rules that will be investigated in the current research are *curtailment* and *stochastic curtailment*, which will be explained thoroughly in the Materials and Methods section. Curtailment and stochastic curtailment can be applied to both the full-length SOAPP-R and the 12-item short form in order to reduce respondent and administrative burden (Finkelman et al., 2018).

The aforementioned versions of the SOAPP-R assume that the screener's items are given equal weight in the scoring process (in particular, that item scores are summed to produce a total score). On the other hand, Black, McCaffrey, Villapiano, Jamison, and Butler (2018) recently developed an 8-item short form of the instrument in which the items are given different scoring weights based on the results of a logistic regression model. The authors noted that although their form could be administered via either computer or paper-and-pencil, its scoring would necessitate the use of a computer.

Given the body of work that has been conducted to develop and validate SOAPP-R short forms, a recommendation on which version to use in practice would be valuable. However, no prior study has compared the methods in which items are weighted equally with the 8-item SOAPP-R weighting items unequally. Additionally, no prior study has considered combining curtailment or stochastic curtailment with the 8-item version (i.e., applying curtailment or stochastic curtailment to the 8-item form to shorten it further). Such a combined approach would be natural, given that the 8-item version can be administered via computer (and must be scored by computer), and curtailment and stochastic curtailment are methodologies that are facilitated by computer-based testing. The objective of the current study was to fill these gaps by (i) developing stopping procedures that combine curtailment or stochastic curtailment with the 8-item SOAPP-R and (ii) comparing the different short forms of the SOAPP-R with one another as well as with the full-length version. As will be seen, comparison of the different forms involved the assessment of their respective predictive power with regard to manifest (external) measures of aberrant medication-related behavior, as opposed to an approach involving the measurement of one or more latent variables irrespective of predictive power. Hence, the research was not geared toward examining computerized adaptive testing based on item response models involving latent variables, which have been shown to be suboptimal for prediction (Smits, van der Ark, & Conijn, 2018), nor was it geared toward statistics such as reliability or the standard error of measurement. Rather, we have focused on the procedures' relative abilities to predict the manifest measures efficiently. Accordingly, the primary application of the procedures discussed herein is to research and practice settings in which interest lies in predictive measurement rather than the distinct domain of measurement and computerized adaptive testing involving latent variables.

Materials and Methods

A retrospective analysis of three datasets was conducted. The Tufts Health Sciences Institutional Review Board granted non-human participants or exempt status for each of the analyses.

The SOAPP-R and Its Variants

The full-length SOAPP-R is composed of 24 items (Table 1). Each item is scored on a scale from 0 to 4, with a response of "0 = *never*," "1 = *seldom*," "2 = *sometimes*," "3 = *often*," and "4 = *very often*". As an individual's total score on the full-length SOAPP-R is computed by adding his/her item scores, the total score can range from 0 to 96. If the total score is greater than or equal to the cut-off point, a "high risk" classification is obtained; otherwise, a "low risk" classification is obtained. A cut-off point of 18 has been recommended based on previous research (Butler et al., 2008, 2009).

Table 1 shows the items comprising the 12-item short form of the SOAPP-R. This short form was developed via a combination of statistical modeling and content evaluation. Specifically, a least absolute shrinkage and selection operator (lasso) logistic regression model was employed to select items predictive of a measure of aberrant medication-related behaviors, the Aberrant Drug Behavior Index (ADBI). The short form (whose items are summed to produce a total score, which is then compared to a cut-off point) was found to demonstrate adequate sensitivity and specificity as well as adequate content (Finkelman, Smits, et al., 2017). A cut-off point of 9 has been recommended (Finkelman, Jamison, et al., 2017).

As mentioned in the Introduction, another approach to shorten the SOAPP-R is to take advantage of technological advances that allow a respondent's answers to be tracked in real time, and the questionnaire's administration to be adapted accordingly depending on the answers provided. In particular, when the questionnaire is given via computer, it can be stopped judiciously when an algorithm calculating the respondent's cumulative score "on the fly" determines that further items are not needed. For example, suppose that the full-length SOAPP-R is administered alongside a cut-off point of 18. If a given respondent's cumulative score reaches 18 (or above) after the tenth item, then it has become certain that his/her classification will be "high risk," irrespective of his/her answers to the remaining items. Therefore, the screener can be terminated after ten items, and a "high risk" classification can be given immediately, thus reducing the test length from 24 to 10 items for that respondent. Additionally, if a second respondent's cumulative score is 13 after the presentation of 23 items, then his/her total score cannot reach the cut-off point, irrespective of his/her answer to the twenty-fourth item (bearing in mind that only one item remains, and the maximum possible score for that item is 4). Therefore, the screener can be terminated prior to the presentation of the twenty-fourth item, and a "low risk" classification can be given. The above rule for stopping (in which testing is halted once a "high risk" or "low risk" result has become deterministic) is known as curtailment in the statistical and psychometric literature (de Beurs, Fokkema, & O'Connor, 2016). A similar – but more aggressive – stopping rule is known as stochastic curtailment (Fokkema, Smits, Finkelman, Kelderman, & Cuijpers, 2014), in which the respondent's questionnaire is halted once the probability of a "high risk" or "low risk" result reaches or exceeds a certain threshold. For instance, suppose that this threshold has been set at 99%. Suppose further that the respondent's probability of a "high risk" result is between 1% and 99% after each of the first six items, but goes above 99% (or below 1%) after the seventh item. In this case, stochastic curtailment halts the screener after seven items; a "high risk" classification is made if stopping occurred alongside a probability above 99%, and a "low risk" classification is made if stopping occurred alongside a probability below 1%. The probability of a "high risk" classification can be estimated after each stage of testing (i.e., after each item is presented) based on a logistic regression model (see Finkelman, Smits, Kim, and Riley (2012) for details). Both curtailment and stochastic curtailment have been found to substantially reduce the average test length of the full-length SOAPP-R and the 12-item SOAPP-R without unduly compromising sensitivity or specificity (Finkelman et al., 2018).

Finally, the items comprising the 8-item SOAPP-R are shown in Table 1. Like the 12-item form, the 8-item SOAPP-R was developed with both statistical and content characteristics in mind. Specifically, items that were predictive of the ADBI were chosen based on the lasso selection method and the leave-one-out cross-validation (LOOCV) method of the GLMSELECT procedure of SAS 9.4. The eight items that were selected for inclusion can be used in tandem (with the items receiving different weights based on the parameters of the logistic regression model) to estimate a given respondent's probability of misuse. The

Table 1. Items included in the full-length SOAPP-R, the 8-item SOAPP-R, and the 12-item SOAPP-R

Item ("How often...")	Inclusion in 8-Item Short Form	Inclusion in 12-Item Short Form
1. Have mood swings		
2. Felt a need for higher doses of medication to treat your pain		X
3. Felt impatient with your doctors		X
4. Felt that things are just too overwhelming that you can't handle them	X	X
5. Tension in the home	X	X
6. Counted pain pills to see how many are remaining		
7. Been concerned that people will judge you for taking pain medication	X	
8. Feel bored		
9. Taken more pain medication than you were supposed to	X	X
10. Worried about being left alone		
11. Felt a craving for medication		
12. Others expressed concern over your use of medication	X	X
13. Any of your close friends had a problem with alcohol or drugs		
14. Others told you that you had a bad temper		
15. Felt consumed by the need to get pain medication		
16. Run out of pain medication early	X	X
17. Others kept you from getting what you deserve		
18. Had legal problems or been arrested (in your lifetime)		X
19. Attended an AA or NA meeting	X	X
20. Been in an argument that was so out of control that someone got hurt		
21. Been sexually abused	X	X
22. Others suggested that you have a drug or alcohol problem		X
23. Had to borrow pain medications from your family or friends		
24. Been treated for an alcohol or drug problem		X

content of the eight items was judged to be adequate by a set of experts (Black et al., 2018).

Curtailment and Stochastic Curtailment of the 8-Item SOAPP-R

As noted above, no prior research has investigated the combination of curtailment or stochastic curtailment with the 8-item SOAPP-R. Such a combination could potentially provide reduced test length (and thereby reduced respondent and administrative burden) without unduly affecting sensitivity and specificity. As Black et al. (2018) have stated that the scoring of the 8-item SOAPP-R necessitates the use of a computer, the combined approach would take full advantage of the benefits of computer-based testing (judicious stopping to enhance efficiency, reduced time required of staff, and automated and error-free scoring; Weiner, Horton, Green, & Butler, 2015).

In order to examine how curtailment and stochastic curtailment could be applied to the 8-item SOAPP-R, it is first necessary to understand the logistic regression model that is used alongside the screener to estimate a given respondent's probability of misuse. Specifically, in this model, the estimated probability of misuse is given by the following equation:

$$\hat{p} = \frac{\exp\left(\hat{\alpha} + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_7 x_7 + \ldots + \hat{\beta}_{21} x_{21}\right)}{1 + \exp\left(\hat{\alpha} + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_7 x_7 + \ldots + \hat{\beta}_{21} x_{21}\right)}. \quad (1)$$

In this equation, $\hat{p}$ represents the respondent's estimated probability of misuse. The x terms represent the respondent's item scores: x_4 denotes the score for item 4 of the full-length SOAPP-R, x_5 denotes the score for item 5, and so forth (note that the 8-item SOAPP-R is comprised of items 4, 5, 7, 9, 12, 16, 19, and 21 of the full-length SOAPP-R). $\hat{\alpha}$ denotes the intercept of the logistic regression model. The $\hat{\beta}$ terms represent the slopes of the model: $\hat{\beta}_4$ denotes the slope for item 4, $\hat{\beta}_5$ denotes the slope for item 5, and so forth. These slopes allow the eight items to be weighted differently in order to optimize the prediction of misuse. The specific numerical $\hat{\alpha}$ value and $\hat{\beta}$ values that were found empirically by Black et al. (2018) to produce the best prediction of misuse are proprietary; therefore, they are not presented here. The ellipsis seen in both the numerator and denominator of Equation 1 indicates that while $\hat{\beta}$ and x terms for items 9, 12, 16, and 19 are not explicitly shown in this equation (due to space considerations), these terms are present in the model. Finally, the exp( ) function of Equation 1 refers to the exponential function.

Once the estimated probability has been obtained for the respondent via the above formula, this probability can be used to produce a "high risk" or "low risk" classification. In particular, a cut-off point p^* is specified along the probability scale; a "high risk" classification is given if the respondent's estimated probability meets or exceeds this cut-off point ($\hat{p} \geq p^*$), and a "low risk" classification is given otherwise ($\hat{p} < p^*$). Black et al. (2018) found empirically that the use of the eight selected items in their logistic regression model, along with a cut-off point of 0.2979 on the probability scale (or 29.79% when written as a percentage), produced adequate sensitivity and specificity.

Equation 1 can be written in the following alternate way that is often more convenient for mathematical purposes:

$$\log\left\{\frac{\hat{p}}{1-\hat{p}}\right\} = \hat{\alpha} + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_7 x_7 + \ldots + \hat{\beta}_{21} x_{21}. \quad (2)$$

Here, log{ } refers to the natural logarithm function. All other terms of Equation 2 have the same meaning as in Equation 1.

Equation 2 is more convenient than Equation 1 in the sense that the right-hand side of Equation 2 is written as a simple summation of terms (without the exponential function). Moreover, the classification decision can also be written in an alternate way; specifically, the criterion that $\hat{p} \geq p^*$ in order for a "high risk" classification to be made is mathematically equivalent to the following rule: give a "high risk" classification if

$$\log\left\{\frac{p^*}{1-p^*}\right\} \leq \hat{\alpha} + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_7 x_7 + \ldots + \hat{\beta}_{21} x_{21}, \quad (3)$$

and give a "low risk" classification if

$$\log\left\{\frac{p^*}{1-p^*}\right\} > \hat{\alpha} + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_7 x_7 + \ldots + \hat{\beta}_{21} x_{21}. \quad (4)$$

The rule presented in Inequalities 3 and 4 always produces the same classification ("high risk" or "low risk") as a rule prescribing that a "high risk" classification be given if and only if $\hat{p} \geq p^*$. Note that if a cut-off of 0.2979 is used, then Inequalities 3 and 4 can be written as follows: give a "high risk" classification if

$$-0.8573 \leq \hat{\alpha} + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_7 x_7 + \ldots + \hat{\beta}_{21} x_{21}, \quad (5)$$

and give a "low risk" classification if

$$-0.8573 > \hat{\alpha} + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_7 x_7 + \ldots + \hat{\beta}_{21} x_{21}, \quad (6)$$

considering that $\log\left\{\frac{0.2979}{1-0.2979}\right\} = -0.8573$.

To explain how curtailment and stochastic curtailment can be applied to the 8-item SOAPP-R, we focus attention on Inequalities 5 and 6. Starting with curtailment, this method stops testing once the classification of the screener in question (here, the 8-item SOAPP-R) is known with certainty. Therefore, if during the administration it becomes certain that Inequality 5 will hold (i.e., that $\hat{\alpha} + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_7 x_7 + \ldots + \hat{\beta}_{21} x_{21}$ will meet or exceed -0.8573 for the respondent being tested), then testing is halted in favor of a "high risk" classification. For example, if the respondent's answers to the first four items of the 8-item SOAPP-R (items 4, 5, 7, and 9 of the SOAPP-R) are so indicative of high risk that $\hat{\alpha} + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_7 x_7 + \ldots + \hat{\beta}_{21} x_{21}$ will necessarily meet or exceed -0.8573 for the respondent (irrespective of his/her answers to the final four items: items 12, 16, 19, and 21), then these final four items are skipped and a "high risk" classification is immediately given. Additionally, if during the administration it becomes certain that Inequality 6 will hold (i.e., that $\hat{\alpha} + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_7 x_7 + \ldots + \hat{\beta}_{21} x_{21}$ will not reach -0.8573 for the respondent being tested), then testing is halted in favor of a "low risk" classification. For example, if the respondent's answers to the first five items of the 8-item SOAPP-R are so indicative of low risk that $\hat{\alpha} + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_7 x_7 + \ldots + \hat{\beta}_{21} x_{21}$ cannot possibly be greater than -0.8573 for the respondent (irrespective of his/her answers to the final three items: items 16, 19, and 21), then these final three items are skipped and a "low risk" classification is immediately given.

Turning to stochastic curtailment of the 8-item SOAPP-R, this method stops whenever the curtailment procedure described above does. In addition, it stops if either (i) the probability that the 8-item SOAPP-R will produce a "high risk" result goes above a pre-specified threshold (e.g., 95% or 99%), or (ii) the probability that the 8-item SOAPP-R will produce a "low risk" result goes above this threshold. These probabilities are estimated based on logistic regression modeling, using data from individuals who have answered all items on the 8-item SOAPP-R. A separate model is used at each stage of testing (i.e., there is a logistic regression model that is used to estimate the probabilities in question after the first item is administered, a different logistic regression model that is used after the second item is administered, and so forth). The dependent variable in each model is the result of the 8-item SOAPP-R ("high risk" or "low risk"). The independent variable in a given model is the weighted sum of items that have been administered up to that stage of testing. For example, at the third stage of testing (after the third item on the 8-item SOAPP-R has been administered), the independent variable in the logistic regression model is the weighted sum of items that have been administered to that point in the test: $\hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_7 x_7$ (noting that the first three items on the 8-item SOAPP-R are item 4, item 5, and item 7). At the fourth stage of testing, the independent variable is $\hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_7 x_7 + \hat{\beta}_9 x_9$ (noting that the fourth item on the 8-item SOAPP-R is item 9); at the fifth stage of testing, the independent variable is $\hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_7 x_7 + \hat{\beta}_9 x_9 + \hat{\beta}_{12} x_{12}$ (noting that the fifth item on the 8-item SOAPP-R is item 12); and so forth. See Finkelman et al. (2012) for conceptual details about the use of logistic regression to estimate the probabilities that are employed in stochastic curtailment. Note that the logistic regression models described in this section are distinct from the model developed by Black et al. (2018) to estimate a respondent's probability of misuse based on his/her responses to the items on the 8-item SOAPP-R.

Participants

Participant-level information came from three different data sources, which will be referred to as "Dataset 1," "Dataset 2," and "Dataset 3."

Dataset 1

Participant-level data (n = 428) came from the SOAPP-R's initial validation study (Butler et al., 2008) and cross-validation study (Butler et al., 2009). Each participant had been a chronic noncancer pain patient from a pain clinic in Indiana, Massachusetts, New Hampshire, Ohio, or Pennsylvania and had taken the full-length SOAPP-R via paper and pencil. Additionally, each participant had been followed up with five months later to assess aberrant medication-related behavior based on the Aberrant Drug Behavior Index (ADBI), which triangulates information from three sources. These sources are (i) the Prescription Drug Use Questionnaire (PDUQ; Compton, Darakjian, & Miotto, 1998), which is a 42-item self-report instrument with a cut-off of ≥ 11; (ii) the Prescription Opioid Therapy Questionnaire (POTQ; Michna et al., 2004), which is an 11-item physician-report instrument with a cut-off of ≥ 2; and (iii) a urine toxicology screen. The overall result of the ADBI is considered to be positive if the PDUQ is positive, or if both the POTQ and the urine toxicology screen are positive. Further details are provided in previous work (Butler et al., 2008, 2009).

Dataset 2

This dataset included information from chronic noncancer pain patients (n = 84) who had been recruited from a hospital-based pain management center for a study to develop and validate a compliance checklist (the Opioid Compliance Checklist [OCC]; Jamison et al., 2014). Each participant had been prescribed long-term opioid therapy and had taken measures including the full-length SOAPP-R via paper-and-pencil. The external measure of aberrant medication-related behavior was based on four sources: (i) the PDUQ; (ii) the Current Opioid Misuse Measure (COMM), a self-report questionnaire with a cut-off of ≥ 9 (Butler et al., 2007); (iii) the Addiction Behaviors Checklist (ABC), a physician-report assessment with a cut-off of ≥ 2 (Wu et al., 2006); and (iv) a urine toxicology screen. The overall result of the external measure was considered to be positive if the urine toxicology screen was positive, or if at least two of the other three measures were positive. See Jamison et al. (2014) for additional details.

Dataset 3

Participant-level data (n = 110) came from a study (Jamison, Scanlan, Matthews, Jurcik, & Ross, 2016) that had investigated the effect of risk assessment and a structured opioid therapy protocol of compliance checklists and monthly monitoring on the confidence of primary care physicians in the management of chronic noncancer pain patients. Each participant who participated had been prescribed or was eligible for opioid medication and had completed measures including the full-length paper-and-pencil version of the SOAPP-R. After six months, participants were assessed using the COMM; the results of this measure were triangulated with results of the ABC and urine toxicology. Specifically, the external measure of aberrant medication-related behavior was defined to be positive if the urine toxicology results were positive or if the COMM and ABC results were both positive. See Jamison et al. (2016) for further details.

Data Analysis

Each dataset was analyzed using real-data simulation (also called post hoc simulation). This procedure entailed determining how many items would have been administered to each participant in the dataset, if a given stopping procedure had been used. The screening classification ("high risk" or "low risk") that would have resulted from the given stopping procedure was also obtained for each participant. Then, statistics of interest (mean and standard deviation of test length, as well as sensitivity and specificity) were determined for the stopping procedure. This process was repeated for each stopping procedure.

Attention focused on comparing different versions of the 8-item SOAPP-R (the screener in its full-length form, with curtailment, and with stochastic curtailment) with one another and with other versions of the SOAPP-R. Indeed, the other versions of the SOAPP-R examined herein had been investigated previously using the three datasets of this study (Finkelman et al., 2018). However, their performance in comparison to versions of the 8-item SOAPP-R had been an open question.

Given that the 8-item SOAPP-R uses different scoring weights for different items, a natural question was whether the items could be placed in an optimal ordering to maximize the screener's efficiency when utilized alongside curtailment or stochastic curtailment. That is, it was desired to ascertain whether by judiciously ordering the items based on their scoring weights, the mean test lengths of the curtailed and stochastically curtailed versions of the 8-item SOAPP-R could be minimized. Previous research (Finkelman, Kim, He, & Lai, 2013) suggested that when logistic regression is used, greater efficiency is achieved when the items are placed in order of their $\hat{\beta}$ values (from highest to lowest). Therefore, curtailment and stochastic curtailment of the 8-item SOAPP-R were applied in two ways: first, assuming that items would be administered in their standard order (4, 5, 7, 9, 12, 16, 19, and 21), and second, assuming that items would be administered in descending order of their $\hat{\beta}$ values from the logistic regression model predicting aberrant medication-related behavior from the items in the 8-item form.

As described in the Materials and Methods section, stochastic curtailment requires that logistic regression modeling be conducted at each stage of testing to obtain the estimated probability of a "high risk" result (such modeling is distinct from the logistic regression model alluded to in the previous paragraph, which predicts aberrant medication-related behavior from items in the 8-item form). The former logistic regression modeling required by stochastic curtailment was performed using Dataset 1 because this dataset was based on the SOAPP-R's validation study (Butler et al., 2008) and cross-validation study (Butler et al., 2009), and because it had the greatest sample size. The results were used to derive the stopping rules of stochastic curtailment, that is, the specific rules determining when this procedure stops early in favor of a "high risk" or "low risk" result. Two versions of stopping curtailment were examined: a conservative version that stopped early only if the probability of a "high risk" or "low risk" result reached or exceeded a threshold of 99%, and a more liberal version that stopped early using a lower threshold of 95%. These two versions will be referred to as SC-99 and SC-95, respectively. Results were obtained using R (Version 3.3.1) software (R Core Team, 2013).

Each version of the SOAPP-R was implemented alongside its standard cut-off point. In particular, a cut-off point of $\geq$ 18 was utilized for the full-length SOAPP-R, and a cut-off point of $\geq$ 9 was utilized for the 12-item short form. For the 8-item SOAPP-R, a cut-off point of 0.2979 on the probability scale was used. Each of these cut-off points had been recommended in previous empirical research based on their combination of sensitivity and specificity. The cut-off point of $\geq$ 18 for the full-length SOAPP-R had been suggested by Butler et al. (2008, 2009); the cut-off point of $\geq$ 9 for the 12-item short form had been suggested by Finkelman, Jamison, et al. (2017); and the cut-off point of 0.2979 on the probability scale for the 8-item SOAPP-R had been suggested by Black et al. (2018).

Results

Among participants with information on age, the $M \pm SD$ age in years was 51.4 ± 13.0 for Dataset 1 (n = 425), 49.9 ± 8.8 for Dataset 2 (n = 84), and 53.4 ± 9.5 for Dataset 3 (n = 109). Among participants with information on gender, 183 of 426 (43.0%) in Dataset 1 were male, as opposed to 45 of 84 (53.6%) in Dataset 2 and 41 of 110 (37.3%) in Dataset 3. One hundred forty-five of the 428 participants in Dataset 1 (33.9%) exhibited aberrant medication-related behavior according to the external criterion used in that dataset; the analogous results were 43 of 84 (51.2%) for Dataset 2 and 40 of 110 (36.4%) for Dataset 3.

Table 2 displays screening characteristics (sensitivity and specificity), as well as mean and standard deviation of test length for each SOAPP-R version, separately for each dataset. Results of procedures studied previously with these datasets are presented along with results of the 8-item SOAPP-R, its curtailed version alongside each item ordering method, and its stochastically curtailed version alongside each item ordering method. Comparisons between procedures are discussed below.

In Dataset 1, the sensitivity and specificity of the full-length SOAPP-R were 0.79 and 0.59, respectively; the analogous values for the 12-item short form were 0.80 and 0.59, respectively. The 8-item SOAPP-R exhibited lower sensitivity (0.72) and greater specificity (0.67) than both the full-length SOAPP-R and the 12-item short form. Applying curtailment, SC-99, and SC-95 to the 8-item SOAPP-R (using the standard item ordering) resulted in mean test lengths of 6.8, 6.8, and 6.5 items, respectively, along with an increase in specificity of 0.01 for SC-95 compared to the 8-item SOAPP-R without any early stopping. When ordering items by their logistic regression model coefficients, mean test lengths of the stopping rules were 6.2, 5.9, and 4.7 items, respectively, with an analogous decrease in sensitivity of 0.01 and an increase in specificity of 0.02 for SC-95. The aforementioned mean test length of 4.7 items was the smallest mean test length of any version of the SOAPP-R for Dataset 1.

Turning to Dataset 2, the respective sensitivity and specificity were 0.67 and 0.59 for the full-length SOAPP-R; 0.67 and 0.56 for the 12-item short form; and 0.70 and 0.54 for the 8-item SOAPP-R. When applying the stopping rules to

Table 2. Results of each version of the SOAPP-R, by dataset (n = 428 for Dataset 1, n = 84 for Dataset 2, n = 110 for Dataset 3)

Dataset	Max. Possible Number of Items	Stopping Rule	Sensitivity	Specificity	*M* ± *SD* test length
1	24	Full-length	0.79	0.59	24.0 ± 0.0
		Curtailment	0.79	0.59	16.9 ± 6.5
		SC-99	0.79	0.59	13.6 ± 6.4
		SC-95	0.78	0.58	10.4 ± 6.6
	12	Full-length	0.80	0.59	12.0 ± 0.0
		Curtailment	0.80	0.59	8.4 ± 3.5
		SC-99	0.80	0.59	7.5 ± 3.2
		SC-95	0.78	0.59	5.9 ± 3.3
	8 (standard ordering)	Full-length	0.72	0.67	8.0 ± 0.0
		Curtailment	0.72	0.67	6.8 ± 1.6
		SC-99	0.72	0.67	6.8 ± 1.6
		SC-95	0.72	0.68	6.5 ± 1.7
	8 (ordering by item coefficient)	Full-length	0.72	0.67	8.0 ± 0.0
		Curtailment	0.72	0.67	6.2 ± 1.9
		SC-99	0.72	0.67	5.9 ± 1.9
		SC-95	0.71	0.69	4.7 ± 2.0
2	24	Full-length	0.67	0.59	24.0 ± 0.0
		Curtailment	0.67	0.59	16.8 ± 6.3
		SC-99	0.67	0.59	13.1 ± 5.8
		SC-95	0.65	0.56	9.6 ± 5.6
	12	Full-length	0.67	0.56	12.0 ± 0.0
		Curtailment	0.67	0.56	8.4 ± 3.3
		SC-99	0.67	0.56	7.6 ± 3.2
		SC-95	0.63	0.56	6.2 ± 3.4
	8 (standard ordering)	Full-length	0.70	0.54	8.0 ± 0.0
		Curtailment	0.70	0.54	6.6 ± 1.6
		SC-99	0.70	0.54	6.6 ± 1.6
		SC-95	0.67	0.56	6.3 ± 1.7
	8 (ordering by item coefficient)	Full-length	0.70	0.54	8.0 ± 0.0
		Curtailment	0.70	0.54	6.0 ± 2.2
		SC-99	0.70	0.54	5.8 ± 2.4
		SC-95	0.70	0.56	5.3 ± 2.4
3	24	Full-length	0.68	0.44	24.0 ± 0.0
		Curtailment	0.68	0.44	15.9 ± 6.6
		SC-99	0.68	0.44	12.8 ± 6.2
		SC-95	0.68	0.47	8.8 ± 6.5
	12	Full-length	0.68	0.46	12.0 ± 0.0
		Curtailment	0.68	0.46	7.8 ± 3.6
		SC-99	0.68	0.46	7.0 ± 3.1
		SC-95	0.68	0.47	4.9 ± 3.1
	8 (standard ordering)	Full-length	0.73	0.53	8.0 ± 0.0
		Curtailment	0.73	0.53	6.7 ± 1.7
		SC-99	0.73	0.53	6.7 ± 1.7
		SC-95	0.70	0.53	6.3 ± 1.9
	8 (ordering by item coefficient)	Full-length	0.73	0.53	8.0 ± 0.0
		Curtailment	0.73	0.53	6.2 ± 1.9
		SC-99	0.73	0.53	6.0 ± 2.2
		SC-95	0.70	0.57	5.1 ± 2.1

the 8-item SOAPP-R along with the standard item ordering, mean test lengths were 6.6 for curtailment, 6.6 for SC-99, and 6.3 for SC-95; the latter exhibited a decrease in sensitivity of 0.03 and an increase in specificity of 0.02 compared to the 8-item SOAPP-R without any early stopping. When applying these stopping rules along with item ordering by logistic regression coefficient, mean test lengths were 6.0 for curtailment, 5.8 for SC-99, and 5.3 for SC-95; the latter exhibited an analogous increase of 0.02 in specificity and no change in sensitivity. The aforementioned 5.3 items represented the smallest mean test length of any version of the SOAPP-R for Dataset 2.

Finally, in Dataset 3, the respective sensitivity and specificity were 0.68 and 0.44 for the full-length SOAPP-R; 0.68 and 0.46 for the 12-item short form; and 0.73 and 0.53 for the 8-item SOAPP-R. Using the standard item ordering and applying the stopping rules to the 8-item SOAPP-R, mean test lengths were 6.7 for curtailment, 6.7 for SC-99, and 6.3 for SC-95, with the latter resulting in a 0.03 decrease in sensitivity as compared to the 8-item SOAPP-R without any early stopping. When ordering items by their logistic regression coefficients, mean test lengths were 6.2 for curtailment, 6.0 for SC-99, and 5.1 for SC-95, with the latter resulting in an analogous 0.03 decrease in sensitivity and a 0.04 increase in specificity. The mean test length of 5.1 items was the second-smallest mean test length of any version of the SOAPP-R for Dataset 3, following the 4.9 items exhibited by SC-95 of the 12-item short form.

Discussion

Previous research has been devoted to developing short forms of the SOAPP-R in order to reduce respondent and administrative burden (Black et al., 2018; Finkelman et al., 2018). However, prior to the current study, no research had compared the 8-item SOAPP-R to other short versions of the SOAPP-R, such as the 12-item short form and the curtailed and stochastically curtailed versions of the full-length SOAPP-R and 12-item form. Furthermore, no prior study had investigated the use of curtailment and stochastic curtailment in combination with the 8-item SOAPP-R. The purpose of the current research was to fill these gaps by developing curtailed and stochastically curtailed versions of the 8-item SOAPP-R, then comparing all of the different short versions to one another using three separate datasets.

Results varied across the three datasets examined. When using previously recommended cut-off points for all forms, the 8-item SOAPP-R exhibited lower sensitivity and higher specificity than the full-length SOAPP-R and the 12-item short form in Dataset 1. In Dataset 2, however, this trend reversed, albeit with smaller differences between the 8-item SOAPP-R and the other forms. Finally, in Dataset 3, the 8-item SOAPP-R's sensitivity and specificity were both higher than the corresponding values of the other forms. The finding that a short version of the instrument can exhibit greater sensitivity and specificity than its longer versions is reflective of the fact that certain items on the SOAPP-R (which exhibits multidimensionality; see Butler et al., 2008) are more predictive of external criteria for aberrant medication-related behavior than others (Black et al., 2018; Finkelman, Smits, et al., 2017).

Turning to the comparison of the 8-item SOAPP-R with its curtailed and stochastically curtailed versions, both curtailment and SC-99 always achieved the same sensitivity and specificity as the version of the screener without any early stopping. These two stopping rules produced small but non-negligible benefits in efficiency, with SC-99 providing greater item savings. SC-95 produced a larger reduction in mean test length, but also affected the sensitivity and specificity of the screener (sometimes resulting in higher values than the 8-item SOAPP-R without any early stopping, and sometimes resulting in lower values). Given SC-95's inconsistent effect on the questionnaire's screening characteristics, a more conservative procedure such as curtailment or SC-99 may be preferred.

In all cases, ordering the items by their logistic regression coefficients provided a greater reduction in mean test length than administering the items in their standard order. For example, SC-99 of the 8-item SOAPP-R exhibited mean test lengths between 6.6 and 6.8 items when the standard ordering was used, as opposed to a range of 5.8–6.0 items when placing the items in descending order of their coefficients. It should be noted that before any item ordering is used in operational screening, it should be scrutinized by content experts to ensure that the items "hang together" in a sensible manner. Indeed, an item may be interpreted in different ways depending on which items came before it in a questionnaire; such "context effects" (Ortner, 2008) should be considered along with efficiency when selecting an item ordering.

Ultimately, which version of the SOAPP-R to use operationally may depend on the technological constraints of the individual or practice administering it. If computer-based testing is not available, then the full-length screener or its 12-item short form is preferred. Indeed, while the 8-item SOAPP-R could be administered to patients via paper-and-pencil, its scoring requires a computer (Black et al., 2018); any procedure involving curtailment or stochastic curtailment requires a computer for both administration and scoring. If computer-based testing (which has been found to be feasible for the SOAPP-R; Weiner et al., 2015) is available, then SC-99 of the 8-item SOAPP-R may be recommended, given this procedure's strong balance of screening characteristics and mean test length observed herein. When items were ordered by their logistic regression coefficients, SC-99 always reduced the mean test length of the 8-item SOAPP-R by at least 2.0 items (25% of the screener's total test length) without affecting sensitivity or specificity in any dataset. Even a modest improvement in test efficiency arising from the use of SC-99 may be welcomed in some contexts. At the same time, the small absolute difference in mean test length between the 8-item SOAPP-R itself and SC-99 of this form may result in a choice by some practitioners to prefer the former, as it is simpler to employ and generally results in only a slight increase in the number of items administered compared to the latter. Similarly, while the 8-item form is substantially shorter than the 12-item form when considered on a percentage basis (being one-third shorter), the absolute difference of four items might not be considered substantial enough to be a primary factor in form selection in some settings. Moreover, respondent burden may not always be an issue when administering the longer version of a questionnaire. Indeed, participants might view a questionnaire as a signal of interest in important elements of their lives, and lengthy questionnaires have been used successfully especially in settings in which the participants are interested in the content at hand (Sprangers & Schwartz, 2017). It is for these reasons that we emphasize that the shorter versions of the SOAPP-R are not intended to replace the longer versions for all participants and scenarios. Nevertheless, they may be a useful "tool in the toolbox" in contexts in which an abbreviated form is desired. Additionally, we note that the four-item difference is not the only distinction between the 8-item and 12-item forms. In particular, the 8-item SOAPP-R weights the items differentially in its scoring procedure in an effort to enhance its screening characteristics, whereas the 12-item SOAPP-R gives equal weight to each item. The 8-item form may thus offer an advantage in sensitivity and specificity in some settings, as was observed in Dataset 3. All of these considerations factor into our recommendation to use SC-99 of the 8-item SOAPP-R when this procedure can be implemented via computer-based testing and respondent and/or administrative burden is an important factor, with the caveat that computer-based testing is not always available and that practitioners' decisions may also be influenced by other context-specific factors.

A notable limitation of the study is that it was performed retrospectively. In particular, the analysis of each dataset utilized responses from participants who had taken the full-length SOAPP-R. Results of each short form were obtained post hoc by identifying the items of that short form and analyzing them as if they had been presented consecutively. Such post hoc results might not be representative of respondents' answers to items on a short form when those items are administered prospectively in a single unit (Ortner, 2008). A second limitation is that the external criteria used to determine aberrant medication-related behavior differed among datasets. Moreover, none of the criteria could identify aberrant behavior with certainty. However, Dataset 1's criterion was the ADBI, which is a well-studied measure that was used in both the SOAPP-R's initial validation study (Butler et al., 2008) and its cross-validation study (Butler et al., 2009). The criteria of Datasets 2 and 3 were similar to the ADBI in that they triangulated information from multiple sources. Third, the sample sizes of Datasets 2 and 3 were lower than that of Dataset 1. Therefore, the results using these datasets exhibited less precision than those using Dataset 1. However, it is noteworthy that in all three datasets, SC-99 of the 8-item SOAPP-R consistently had the same sensitivity and specificity as the version of this screener without any early stopping, while also exhibiting similar mean test lengths across datasets. Lastly, we note that only in rare instances did a procedure have a sensitivity of 0.80 (and in no instance did a procedure have a specificity of 0.80 or above) in the datasets examined herein, indicating that none of them exhibited particularly strong discriminative ability.

As there are many competing interests for the time of both patients and providers, any reduction of burden without compromising predictive characteristics will be advantageous. Furthermore, the utility of the methodologies described herein may grow as the development of computer-based tools continues and patients become more accepting of interacting with computers. Still, an important next step will be to apply these new techniques prospectively in new patient populations and a variety of settings to confirm their usability and validity.

In sum, the combination of SC-99 and the 8-item SOAPP-R has potential to enhance efficiency when screening for aberrant medication-related behavior via computer. Further prospective study should be conducted. The methodology described herein may also be applied to other questionnaires in an effort to lessen respondent and administrative burden without reducing sensitivity and specificity.

References

Black, R. A., McCaffrey, S. A., Villapiano, A. J., Jamison, R. N., & Butler, S. F. (2018). Development and validation of an eight-item brief form of the SOAPP-R (SOAPP-8). *Pain Medicine, 19*, 1982–1987. https://doi.org/10.1093/pm/pnx194

Butler, S. F., Budman, S. H., Fernandez, K. C., Fanciullo, G. J., & Jamison, R. N. (2009). Cross-validation of a screener to predict opioid misuse in chronic pain patients (SOAPP-R). *Journal of Addiction Medicine, 3*, 66–73. https://doi.org/10.1097/ADM.0b013e31818e41da

Butler, S. F., Budman, S. H., Fernandez, K. C., Houle, B., Benoit, C., Katz, N., & Jamison, R. N. (2007). Development and validation of the Current Opioid Misuse Measure. *Pain, 130*, 144–156. https://doi.org/10.1016/j.pain.2007.01.014

Butler, S. F., Fernandez, K., Benoit, C., Budman, S. H., & Jamison, R. N. (2008). Validation of the revised Screener and Opioid Assessment for Patients with Pain (SOAPP-R). *Journal of Pain, 9*, 360–372. https://doi.org/10.1016/j.jpain.2007.11.014

Carpenter, J. S., Andrykowski, M. A., Wilson, J., Hall, L. A., Rayens, M. K., Sachs, B., & Cunningham, L. L. (1998). Psychometrics for two short forms of the Center for Epidemiologic Studies-Depression Scale. *Issues in Mental Health Nursing, 19*, 481–494. https://doi.org/10.1080/016128498248917

Center for Behavioral Health Statistics and Quality. (2016). *Key substance use and mental health indicators in the United States: Results from the 2015 National Survey on Drug Use and Health* (HHS Publication No. SMA 16-4984, NSDUH Series H-51). Retrieved from http://www.samhsa.gov/data/

Chou, R., Fanciullo, G. J., Fine, P. G., Miaskowski, C., Passik, S. D., & Portenoy, R. K. (2009). 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. https://doi.org/10.1016/j.jpain.2008.10.009

Compton, P., Darakjian, J., & Miotto, K. (1998). Screening for addiction in patients with chronic pain and "problematic" substance use: Evaluation of a pilot assessment tool. *Journal of Pain and Symptom Management, 16*, 355–363. https://doi.org/10.1016/S0885-3924(98)00110-9

de Beurs, D. P., Fokkema, M., & O'Connor, R. C. (2016). Optimizing the assessment of suicidal behavior: The application of curtailment techniques. *Journal of Affective Disorders, 196*, 218–224. https://doi.org/10.1016/j.jad.2016.02.033

Dowell, D., Haegerich, T. M., & Chou, R. (2016). CDC Guideline for Prescribing Opioids for Chronic Pain – United States, 2016. *Morbidity and Mortality Weekly Report (MMWR) Recommendations and Reports, 65*, 1–49. https://doi.org/10.15585/mmwr.rr6501e1

Dugdale, D. C., Epstein, R., & Pantilat, S. Z. (1999). Time and the patient-physician relationship. *Journal of General Internal Medicine, 14*(Suppl. 1), S34–S40. https://doi.org/10.1046/j.1525-1497.1999.00263.x

Finkelman, M. D., Jamison, R. N., Kulich, R. J., Butler, S. F., Jackson, W. C., Smits, N., & Weiner, S. G. (2017). Cross-validation of short forms of the Screener and Opioid Assessment for Patients with Pain-Revised (SOAPP-R). *Drug and Alcohol Dependence, 178*, 94–100. https://doi.org/10.1016/j.drugalcdep.2017.04.016

Finkelman, M. D., Jamison, R. N., Magnuson, B., Kulich, R. J., Butler, S. F., Smits, N., & Weiner, S. G. (2018). Computer-based testing and the 12-item Screener and Opioid Assessment for Patients with Pain-Revised (SOAPP-R): A combined approach to improving efficiency. *Journal of Applied Biobehavioral Research*. Advance online publication. https://doi.org/10.1111/jabr.12145

Finkelman, M. D., Kim, W., He, Y., & Lai, A. M. (2013). Item ordering in stochastically curtailed health questionnaires with an observable outcome. *Journal of Computerized Adaptive Testing, 1*, 38–66. https://doi.org/10.7333/1304-0103038

Finkelman, M. D., Kulich, R. J., Zacharoff, K. L., Smits, N., Magnuson, B. E., Dong, J., & Butler, S. F. (2015). Shortening the Screener and Opioid Assessment for Patients with Pain-Revised (SOAPP-R): A proof-of-principle study for customized computer-based testing. *Pain Medicine, 16*, 2344–2356. https://doi.org/10.1111/pme.12864

Finkelman, M. D., Smits, N., Kim, W., & Riley, B. (2012). Curtailment and stochastic curtailment to shorten the CES-D. *Applied Psychological Measurement, 36*, 632–658. https://doi.org/10.1177/0146621612451647

Finkelman, M. D., Smits, N., Kulich, R. J., Zacharoff, K. L., Magnuson, B. E., Chang, H., & Butler, S. F. (2017). Development of short-form versions of the Screener and Opioid Assessment for Patients with Pain-Revised (SOAPP-R): A proof-of-principle study. *Pain Medicine, 18*, 1292–1302. https://doi.org/10.1093/pm/pnw210

Fokkema, M., Smits, N., Finkelman, M. D., Kelderman, H., & Cuijpers, P. (2014). Curtailment: A method to reduce the length of self-report questionnaires while maintaining diagnostic accuracy. *Psychiatry Research, 215*, 477–482. https://doi.org/10.1016/j.psychres.2013.11.003

Guy, G. P. Jr., Zhang, K., Bohm, M. K., Losby, J., Lewis, B., Young, R., . . . Dowell, D. (2017). Vital Signs: Changes in opioid prescribing in the United States, 2006–2015. *Morbidity and Mortality Weekly Report, 66*, 697–704. https://doi.org/10.15585/mmwr.mm6626a4

Hedegaard, H., Warner, M., & Minino, A. M. (2017). Drug overdose deaths in the United States, 1999–2015. *NCHS Data Brief, 273*, 1–8.

Jamison, R. N., Martel, M. O., Edwards, R. R., Qian, J., Sheehan, K. A., & Ross, E. L. (2014). Validation of a brief Opioid Compliance Checklist for patients with chronic pain. *Journal of Pain, 15*, 1092–1101. https://doi.org/10.1016/j.jpain.2014.07.007

Jamison, R. N., Scanlan, E., Matthews, M. L., Jurcik, D. C., & Ross, E. L. (2016). Attitudes of primary care practitioners in managing chronic pain patients prescribed opioids for pain: A prospective longitudinal controlled trial. *Pain Medicine, 17*, 99–113. https://doi.org/10.1111/pme.12871

Lohr, K. N. (2002). Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research, 11*, 193–205. https://doi.org/10.1023/A:1015291021312

Michna, E., Ross, E. L., Hynes, W. L., Nedeljkovic, S. S., Soumekh, S., Janfaza, D., . . . Jamison, R. N. (2004). Predicting aberrant drug behavior in patients treated for chronic pain: Importance of abuse history. *Journal of Pain and Symptom Management, 28*, 250–258. https://doi.org/10.1016/j.jpainsymman.2004.04.007

Nahin, R. L. (2015). Estimates of pain prevalence and severity in adults: United States, 2012. *Journal of Pain, 16*, 769–780. https://doi.org/10.1016/j.jpain.2015.05.002

Ortner, T. M. (2008). Effects of changed item order: A cautionary note to practitioners on jumping to computerized adaptive testing for personality assessment. *International Journal of Selection and Assessment, 16*, 249–257. https://doi.org/10.1111/j.1468-2389.2008.00431.x

Passik, S. D., Narayana, A., & Yang, R. (2014). Aberrant drug-related behavior observed during a 12-week open-label extension period of a study involving patients taking chronic opioid therapy for persistent pain and fentanyl buccal tablet or traditional short-acting opioid for breakthrough pain. *Pain Medicine, 15*, 1365–1372. https://doi.org/10.1111/pme.12431

R Core Team. (2013). R: A language and environment for statistical computing. Austria: Vienna. [Computer software manual]. Retrieved from https://www.R-project.org/

Rudd, R. A., Seth, P., David, F., & Scholl, L. (2016). Increases in drug and opioid-involved overdose deaths – United States, 2010–2015. *Morbidity and Mortality Weekly Report, 65*, 1445–1452. https://doi.org/10.15585/mmwr.mm655051e1

Smits, N., van der Ark, L. A., & Conijn, J. M. (2018). Measurement versus prediction in the construction of patient-reported outcome questionnaires: Can we have our cake and eat it? *Quality of Life Research, 27*, 1673–1682. https://doi.org/10.1007/s11136-017-1720-4

Sprangers, M. A. G., & Schwartz, C. E. (2017). Toward mindfulness in quality-of-life research: Perspectives on how to avoid rigor becoming rigidity. *Quality of Life Research, 26*, 1387–1392. https://doi.org/10.1007/s11136-016-1492-2

Weiner, S. G., Horton, L. C., Green, T. C., & Butler, S. F. (2015). Feasibility of tablet computer screening for opioid abuse in the emergency department. *Western Journal of Emergency Medicine, 16*, 18–23. https://doi.org/10.5811/westjem.2014.11.23316

Weiner, S. G., Horton, L. C., Green, T. C., & Butler, S. F. (2016). A comparison of an opioid abuse screening tool and prescription drug monitoring data in the emergency department. *Drug and Alcohol Dependence, 159*, 152–157. https://doi.org/10.1016/j.drugalcdep. 2015.12.007

Wu, S. M., Compton, P., Bolus, R., Schieffer, B., Pham, Q., Baria, A., . . . Naliboff, B. D. (2006). The addiction behaviors checklist: Validation of a new clinician-based measure of inappropriate opioid use in chronic pain. *Journal of Pain and Symptom Management, 32*, 342–351. https://doi.org/10.1016/j.jpainsymman.2006.05.010

History

Received July 24, 2018
Revision received October 25, 2018
Accepted November 4, 2018
Published online May 10, 2019
EJPA Section/Category Miscellaneous/Other

Acknowledgments

Data for this study were provided by Inflexxion, Inc.

Matthew D. Finkelman

Department of Public Health and Community Service
Tufts University School of Dental Medicine
1 Kneeland St.
Boston, MA 02111
USA
matthew.finkelman@tufts.edu

