UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

# Shortening the Screener and Opioid Assessment for Patients with Pain-Revised (SOAPP-R): A proof-of-principle study for customized computer-based testing

Finkelman, M.D.; Kulich, R.J.; Zacharoff, K.L.; Smits, N.; Magnuson, B.E.; Dong, J.; Butler, S.F.

**DOI**
10.1111/pme.12864

**Publication date**
2015

**Document Version**
Final published version

**Published in**
Pain Medicine

**License**
Article 25fa Dutch Copyright Act (https://www.openaccess.nl/en/policies/open-access-in-dutch-copyright-law-taverne-amendment)

Link to publication

**Citation for published version (APA):**
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*(12), 2344-2356. https://doi.org/10.1111/pme.12864

**General rights**
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)*

Download date:15 Nov 2025

Pain Medicine

Pain Medicine 2015; 16: 2344–2356
Wiley Periodicals, Inc.

# Shortening the Screener and Opioid Assessment for Patients with Pain-Revised (SOAPP-R): A Proof-of-Principle Study for Customized Computer-Based Testing

**Matthew D. Finkelman, PhD,***
**Ronald J. Kulich, PhD,†‡ Kevin L. Zacharoff, MD,§**
**Niels Smits, PhD,¶ Britta E. Magnuson, DMD,****
**Jinghui Dong, PhD,†† and Stephen F. Butler, PhD‡‡**

*Department of Public Health and Community Service, Tufts University School of Dental Medicine, Boston, Massachusetts, USA; †Craniofacial Pain and Headache Division, Department of Diagnostic Sciences, Tufts University School of Dental Medicine, Boston, Massachusetts, USA; ‡Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA; §Medical Affairs, Inc., Newton, Massachusetts, USA; ¶Department of Methods, Faculty of Psychology and Education, VU University, Amsterdam, The Netherlands; **Department of Oral and Maxillofacial Pathology, Oral Medicine, and Craniofacial Pain, Tufts University School of Dental Medicine, Boston, Massachusetts, USA; ††Sackler School of Graduate Biomedical Sciences, Boston, Massachusetts, USA; ‡‡Inflexxion, Inc., Health Analytics Department, Newton, Massachusetts, USA

*Reprint requests to*: Matthew D. Finkelman, PhD, 1 Kneeland Street, Boston, MA 02111, USA. Tel: (617) 636-3449; Fax: (617) 636-6511; E-mail: matthew.finkelman@tufts.edu.

Funding Sources: Research reported in this publication was supported by the National Institute on Drug Abuse of the National Institutes of Health under Award Number R03DA036683. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Data used in this study were collected as part of a NIDA grant (Grant no: R44DA015617, P.I. Butler). Data for this study were provided by Inflexxion, Inc.

Conflict of Interest: KLZ and SFB are employees of Inflexxion, Inc. Inflexxion holds the copyright for the Screener and Opioid Assessment for Patients with Pain-Revised (SOAPP®-R). The authors are indebted to Alexandra Kulich for editing a previous version of this article.

## Abstract

**Background. The Screener and Opioid Assessment for Patients with Pain-Revised (SOAPP-R) is a 24-item self-report instrument that was developed to aid providers in predicting aberrant medication-related behaviors among chronic pain patients. Although the SOAPP-R has garnered widespread use, certain patients may be dissuaded from taking it because of its length. Administrative barriers associated with lengthy questionnaires further limit its utility.**

**Objective. To investigate the extent to which two techniques for computer-based administration (curtailment and stochastic curtailment) reduce the average test length of the SOAPP-R without unduly affecting sensitivity and specificity.**

**Design. Retrospective study.**

**Setting. Pain management centers.**

**Subjects. Four hundred and twenty-eight chronic non-cancer pain patients.**

**Methods. Subjects had taken the full-length SOAPP-R and been classified by the Aberrant Drug Behavior Index (ADBI) as having engaged or not engaged in aberrant medication-related behavior. Curtailment**



Downloaded from http://painmedicine.oxfordjournals.org/ by guest on April 11, 2016

**and stochastic curtailment were applied to the data in post-hoc simulation. Sensitivity and specificity with respect to the ADBI, as well as average test length, were computed for the full-length test, curtailment, and stochastic curtailment.**

**Results. The full-length SOAPP-R exhibited a sensitivity of 0.745 and a specificity of 0.671 for predicting the ADBI. Curtailment reduced the average test length by 26% while exhibiting the same sensitivity and specificity as the full-length test. Stochastic curtailment reduced the average test length by as much as 65% while always exhibiting sensitivity and specificity for the ADBI within 0.035 of those of the full-length test.**

**Conclusions. Curtailment and stochastic curtailment have potential to improve the SOAPP-R's efficiency in computer-based administrations.**

**Key Words. Chronic Pain; Substance Abuse; Opioids; SOAPP-R; Respondent Burden; Risk Stratification**

## Introduction

While chronic opioid therapy has been increasingly sought after by patients with persistent pain, such therapy has seen mixed results with respect to outcome and risk [1,2]. Opioids may have benefits and uses for the treatment of chronic pain [3], yet recent findings indicate a dose-dependent risk for serious harms as well as limited evidence on long-term effectiveness [4]. Moreover, a segment of the patient population can have a tendency to become overly reliant on opioids, exhibit behaviors including misuse and abuse, or follow non-prescribed dosages [5–7]. Patients may also display aberrant behaviors such as diverting drugs or visiting multiple providers for prescriptions [8]. Several articles [9–11] have recommended a "universal precautions" approach when considering long-term opioid therapy for chronic pain patients. Universal precautions assumes that every patient represents some degree of risk. To initiate and modify therapy in a safe and controlled manner, risk assessment strategies are recommended as well as close patient monitoring. A comprehensive evaluation of the chronic pain patient increasingly includes a standardized process for risk assessment for patients who are potential candidates for opioids or for whom opioids for chronic pain have been recommended [1,12,13]. Many modalities, such as urine toxicology, prescription monitoring, self-report measures, and reviewing of risk factors, are available; while no one tool is adequate [6,14], screening questionnaires have been developed to assist the practitioner with this assessment and to help standardize the assessment process. Such questionnaires, however, can be lengthy and complicate adherence, and the evidence to support them has been challenged [4].

The Screener and Opioid Assessment for Patients with Pain (SOAPP) [15–17] is among the most studied of questionnaires for chronic opioid risk. The SOAPP has the limitations of being conceptually derived and dependent on patient report of incriminating behaviors [18]. Thus, Butler et al. developed the Screener and Opioid Assessment for Patients with Pain – Revised (SOAPP-R) to address these limitations [18]. The SOAPP-R is empirically based, easily understood by patients, and less transparent to the patient in terms of how the items are scored than the original SOAPP items [18]. Both the SOAPP and SOAPP-R provide cutoff points that indicate whether the patient is "positive" (i.e., at high risk for aberrant medication-related behaviors) or "negative" (that is, at relatively low risk for such aberrant behaviors). The cutoff is intended to alert a provider about the potential for risk of aberrant medication-related behaviors for a chronic pain patient being considered for long-term opioid therapy and may be useful, along with other medical information, for making prescribing decisions [18,19].

The initial validation study of the SOAPP-R found that it was an improvement over the original SOAPP and exhibited both strong reliability and validity [18]. In particular, in the initial validation study the coefficient $\alpha$ of the SOAPP-R was 0.88, and the test-retest reliability was also high (intraclass correlation = 0.92). Moreover, the assessment demonstrated predictive validity with respect to an external criterion, the Aberrant Drug Behavior Index (ADBI), which will be described in a later section. In a receiver operating characteristic (ROC) curve analysis with the ADBI as the predictive criterion, the SOAPP-R's area under the curve was 0.81, and the scale demonstrated adequate sensitivity and specificity (0.81 and 0.68, respectively). It has since been cross-validated with a new sample of patients [19], again showing high internal consistency (coefficient $\alpha = 0.86$) and test-retest reliability (intraclass correlation = 0.91). As is anticipated when an assessment is tested in a new population, the SOAPP-R's combination of sensitivity and specificity exhibited shrinkage in cross-validation; nevertheless, its area under the curve of 0.74 was still highly significant and was characterized as having acceptable discrimination by conventional criteria [20]. Both the initial and cross-validation studies concluded that the SOAPP-R is a reliable and valid tool in the prediction of aberrant drug-related behaviors [18,19]. It has been included in both the clinical guidelines of the American Pain Society-American Academy of Pain Medicine Opioids Guidelines Panel [3] and the Canadian guidelines for safe and effective use of opioids [21].

While taking the full 24-item version of the SOAPP-R is a simple task for many respondents, certain individuals may have difficulty completing it, especially taken in the context of multiple other required questionnaires administered in a health care setting. This concern is particularly critical for patients who struggle with reading comprehension and patients with medical ailments, both of whom are known to experience more difficulty with

Downloaded from http://painmedicine.oxfordjournals.org/ by guest on April 11, 2016

questionnaire adherence [22]. Given that the SOAPP-R is specifically designed for persons with chronic pain [18,19]—who typically exhibit physical and mental comorbidities—shorter versions of the SOAPP-R would make the instrument more accessible. The need for shorter versions is also attested to by 1) findings that the response rate [23] and quality of responses [24] can be enhanced by decreasing assessment length, and 2) the Scientific Advisory Committee of the Medical Outcomes Trust's identification of respondent burden as a significant consideration when designing a questionnaire [25].

The development of less time-consuming versions of the SOAPP-R would benefit not only patients, but also providers. Administering screeners in the clinical flow can be challenging, given that current primary care practice guidelines list over 60 different screenings for the primary care setting [26]. The growing recognition of administrative burden and the importance of efficiency in health care delivery [27] necessitate the use of screeners that do not present more items than are necessary.

A short form of the SOAPP containing five items has been introduced [28,29]; however, further efficient assessments to predict the risk of aberrant opioid-related behaviors are needed for two reasons. First, the aforementioned five-item short form is based on the original SOAPP, not the SOAPP-R (only two of its five items appear on the SOAPP-R). A short assessment based on the more rigorously developed SOAPP-R would be beneficial. Second, the previously introduced short form is "static": it gives the exact same set of items to each respondent who takes it. Advances in computerized testing, however, suggest the efficiency of tailored assessments in which the questionnaire is customized at the individual level [30–41]. In computerized *variable-length testing*, the most suitable number of items for a given respondent is determined in real time by monitoring the respondent's answers during the assessment. After each item, a computer program performs internal calculations to decide whether 1) the respondent should be administered another item or 2) the test should be stopped in favor of either a "positive" or a "negative" result for that respondent (as with the full-length SOAPP-R, a "positive" result indicates that the patient is at high risk of future aberrant medication-related behaviors, and a "negative" result suggests lower risk). Two statistical methods for determining when to stop testing are *curtailment* and *stochastic curtailment*. Both of these methods strive to cease testing before the administration of items that cannot, or are unlikely to, influence whether the respondent will ultimately be determined to be at high risk or low risk. To that end, the methods judiciously present fewer items to respondents whose results are clear very quickly, and more items to "borderline" respondents who require further evidence before a "positive" or a "negative" determination can be made. Both curtailment and stochastic curtailment have been shown to lessen the respondent burden of a test while maintaining sensitivity and specificity values comparable to those of the full-length version of the test [30,32,33,35–40]. Within the domain of pain medicine, these methods were recently applied to the Current Opioid Misuse Measure (COMM) and were found to substantially enhance its efficiency of assessment [34]. However, no previous research has investigated their use in the context of the SOAPP-R. The purpose of this study is to fill this gap by examining how curtailment and stochastic curtailment can be applied to the SOAPP-R and quantifying the degree to which they can improve its efficiency. It is noted that the COMM and the SOAPP-R are used for different purposes: the former is designed to assess current aberrant medication-related behaviors involving opioids, whereas the latter is designed to predict such behaviors in the future. Hence, this study seeks to address the current lack of efficient customizable assessment procedures for predicting future aberrant drug-related behaviors.

## Methods

The Institutional Review Board at Tufts Medical Center and Tufts University Health Sciences Campus granted exempt status for this research project.

### *Subjects*

This retrospective study included data from n = 428 subjects who had completed the full (24-item) paper-and-pencil version of the SOAPP-R and had been followed up 5 months later. The purpose of the follow-up was to ascertain whether a given respondent had engaged in aberrant medication-related behavior after taking the SOAPP-R, and thus to evaluate the questionnaire's predictive validity. The assessment used to gauge whether aberrant medication-related behavior had occurred was the Aberrant Drug Behavior Index, which will be described in a later section.

Data came from the original validation study of the SOAPP-R (n = 207) and its cross-validation study (n = 221). The original validation study [18] had recruited patients from pain clinics in three United States states (MA, OH, and PA); all patients had been on a long-term opioid treatment regimen for chronic non-cancer pain. The cross-validation study [19] had recruited patients from pain management centers in five United States states (IN, MA, NH, OH, and PA); all patients had been prescribed opioids for chronic non-cancer pain. The procedures of these studies had been approved by the Human Subjects Committees of the participating centers. All subjects had signed an informed consent form prior to their participation.

### *The SOAPP-R, Curtailment, and Stochastic Curtailment*

Each of the 24 SOAPP-R items asks about the past 30 days and is scored on a 0–4 scale ("Never" = 0, "Seldom" = 1, "Sometimes" = 2, "Often" = 3, "Very Often" = 4). Item scores are summed to produce a total



Downloaded from http://painmedicine.oxfordjournals.org/ by guest on April 11, 2016

**Table 1** Descriptive statistics for all SOAPP-R items (n = 428)

| Item ("In the past 30 days. . .") | Mean (SD) | Median (IQR) |
|---|---|---|
| 1. How often do you have mood swings? | 2.0 (1.0) | 2.0 (2.0) |
| 2. How often have you felt a need for higher doses of medication to treat your pain? | 1.9 (1.1) | 2.0 (2.0) |
| 3. How often have you felt impatient with your doctors? | 1.4 (1.1) | 1.0 (1.0) |
| 4. How often have you felt that things are just too overwhelming that you can't handle them? | 1.5 (1.2) | 1.0 (1.0) |
| 5. How often is there tension in the home? | 1.4 (1.1) | 1.0 (1.0) |
| 6. How often have you counted pain pills to see how many are remaining? | 1.1 (1.1) | 1.0 (2.0) |
| 7. How often have you been concerned that people will judge you for taking pain medication? | 1.2 (1.2) | 1.0 (2.0) |
| 8. How often do you feel bored? | 1.4 (1.1) | 1.0 (1.0) |
| 9. How often have you taken more pain medication than you were supposed to? | 0.8 (0.9) | 1.0 (1.0) |
| 10. How often have you worried about being left alone? | 0.8 (1.1) | 0.0 (1.0) |
| 11. How often have you felt a craving for medication? | 0.7 (1.0) | 0.0 (1.0) |
| 12. How often have others expressed concern over your use of medication? | 0.8 (1.0) | 0.5 (1.0) |
| 13. How often have any of your close friends had a problem with alcohol or drugs? | 0.8 (1.0) | 1.0 (1.0) |
| 14. How often have others told you that you had a bad temper? | 0.7 (1.0) | 0.0 (1.0) |
| 15. How often have you felt consumed by the need to get pain medication? | 0.7 (0.9) | 0.0 (1.0) |
| 16. How often have you run out of pain medication early? | 0.6 (0.9) | 0.0 (1.0) |
| 17. How often have others kept you from getting what you deserve? | 0.6 (0.9) | 0.0 (1.0) |
| 18. How often, in your lifetime, have you had legal problems or been arrested? | 0.3 (0.6) | 0.0 (1.0) |
| 19. How often have you attended an AA or NA meeting? | 0.3 (0.9) | 0.0 (0.0) |
| 20. How often have you been in an argument that was so out of control that someone got hurt? | 0.2 (0.6) | 0.0 (0.0) |
| 21. How often have you been sexually abused? | 0.3 (0.7) | 0.0 (0.0) |
| 22. How often have others suggested that you have a drug or alcohol problem? | 0.3 (0.6) | 0.0 (0.0) |
| 23. How often have you had to borrow pain medications from your family or friends? | 0.2 (0.6) | 0.0 (0.0) |
| 24. How often have you been treated for an alcohol or drug problem? | 0.1 (0.5) | 0.0 (0.0) |
| Total score | 20.4 (11.3) | 18.0 (14.8) |

IQR = Inter-quartile range.

score for the SOAPP-R. This total score is then compared with a prescribed cutoff point; respondents are considered to have a positive finding of high risk for aberrant behaviors if they meet or exceed the cutoff point, and are considered to be at lower risk (i.e., a negative finding) otherwise. See Table 1 for a list of the SOAPP-R items.

In order for curtailment or stochastic curtailment to be applied operationally to the SOAPP-R, administration of the questionnaire must be conducted by computer so that each respondent's answers can be tracked during his/her assessment. Although the subjects in this study had completed the SOAPP-R via paper-and-pencil, the potential of curtailment and stochastic curtailment could still be assessed via the method of post-hoc *simulation* (see the "Statistical analysis" section below). The remainder of the current subsection is devoted to explaining the logic of curtailment and stochastic curtailment.

When using curtailment, which is sometimes referred to as the countdown method [31], testing proceeds until the respondent's result from the questionnaire (either "positive" or "negative") has been unequivocally determined based on his/her previous answers. Once this point has been reached, the computer program terminates the assessment so that no more items are administered than are necessary. For example, suppose that a cutoff point of $\geq 19$ has been set for the full-length SOAPP-R. Table 2 presents the answers of two hypothetical respondents to this assessment. The table shows each respondent's item scores and cumulative (summed) scores at every stage of the test (i.e., after each sequential item is answered). Respondent #1 is ultimately screened as positive for high risk aberrant behaviors by the full-length test (total score = 61), whereas Respondent #2 is ultimately screened as negative (total score = 10). Note that for Respondent #1, his/her cumulative score after seven items is 19 (having had item scores of 2, 4, 2, 2, 3, 4, and 2). Because negative item scores are not possible for the SOAPP-R, and because Respondent #1's cumulative score has already met the cutoff point after seven items, his/her result has unequivocally been decided at that stage: he/she will necessarily be screened as positive by the full-length test. If curtailment were employed, it would stop the



Downloaded from http://painmedicine.oxfordjournals.org/ by guest on April 11, 2016

**Table 2** Results for two hypothetical respondents (cutoff point of $\geq 19$)

| | Respondent #1 | | | Respondent #2 | | |
|---|---|---|---|---|---|---|
| Item | Item Score | Cumulative Score | Chance of "Positive Result" (%) | Item Score | Cumulative Score | Chance of "Positive Result" (%) |
| 1 | 2 | 2 | 47.0 | 2 | 2 | 47.0 |
| 2 | 4 | 6 | 86.3 | 2 | 4 | 49.1 |
| 3 | 2 | 8 | 87.2 | 0 | 4 | 25.6 |
| 4 | 2 | 10 | 89.1 | 1 | 5 | 20.8 |
| 5 | 3 | 13 | 96.1 | 1 | 6 | 16.8 |
| 6 | 4 | 17 | 99.6 | 1 | 7 | 14.6 |
| 7 | 2 | 19 | 100.0 | 1 | 8 | 13.5 |
| 8 | 3 | 22 | 100.0 | 0 | 8 | 5.9 |
| 9 | 4 | 26 | 100.0 | 0 | 8 | 4.2 |
| 10 | 3 | 29 | 100.0 | 0 | 8 | 3.3 |
| 11 | 1 | 30 | 100.0 | 1 | 9 | 4.6 |
| 12 | 4 | 34 | 100.0 | 0 | 9 | 3.5 |
| 13 | 4 | 38 | 100.0 | 0 | 9 | 1.2 |
| 14 | 3 | 41 | 100.0 | 0 | 9 | 0.9 |
| 15 | 3 | 44 | 100.0 | 0 | 9 | 0.7 |
| 16 | 4 | 48 | 100.0 | 0 | 9 | 0.5 |
| 17 | 1 | 49 | 100.0 | 1 | 10 | 0.6 |
| 18 | 2 | 51 | 100.0 | 0 | 10 | 0.4 |
| 19 | 2 | 53 | 100.0 | 0 | 10 | 0.2 |
| 20 | 2 | 55 | 100.0 | 0 | 10 | $< 0.1$ |
| 21 | 0 | 55 | 100.0 | 0 | 10 | $< 0.1$ |
| 22 | 3 | 58 | 100.0 | 0 | 10 | 0.0 |
| 23 | 3 | 61 | 100.0 | 0 | 10 | 0.0 |
| 24 | 0 | 61 | 100.0 | 0 | 10 | 0.0 |

questionnaire after seven items and screen the respondent as positive, since the final 17 items are not necessary for determining his/her result. For Respondent #2, note that his/her cumulative score after 22 items is 10. Even if this respondent receives the maximum score of four on each of the final two items, his/her score will be 18 and he/she will therefore fall short of the cutoff point of 19. Since the final two items are thus not necessary for determining his/her result, curtailment would stop the questionnaire after 22 items and screen the respondent as negative.

To present the logic of curtailment more formally, let $X*$ represent the cutoff point of the test. Curtailment stops the assessment early, and screens the respondent as positive, if the respondent's cumulative score ever meets or exceeds $X*$ during test administration. Curtailment stops the assessment early, and screens the respondent as negative, if the respondent's "maximum potential score" (i.e., the highest score that the respondent could potentially receive as his/her final cumulative score, given his/her current cumulative score) ever drops below $X*$ during test administration. Mathematically, the latter event occurs if the respondent's current cumulative score, plus four times the number of items remaining in the test, is less than $X*$ (the number four is used because this is the maximum possible score for each SOAPP-R item). If curtailment does not stop the test early at any stage, and therefore the respondent receives all 24 SOAPP-R items, he/she is screened as positive if his/her final cumulative score meets or exceeds $X*$, and is screened as negative otherwise. Theoretical results about the method of curtailment are available in the statistical literature [42,43].

Turning to stochastic curtailment, this method can be motivated by looking again at the two hypothetical respondents in Table 2. For each respondent, there is a column ("Chance of 'positive result' (%)") tracking the probability that the respondent will ultimately be positive on the full-length SOAPP-R (information on how to obtain these probability values is provided later in this subsection). The probability of a positive result is updated after every item answered and is specific to the particular respondent taking the questionnaire. For instance, after four items, Respondent #1 has a cumulative score of 10; using a cutoff point of $\geq 19$, the respondent has a (hypothetical) probability of 89.1% of ultimately being positive on the full-length test. By contrast, Respondent #2 has a cumulative score of only five after four items, and thus has a smaller (hypothetical) probability of 20.8% of ultimately being positive. It can



Downloaded from http://painmedicine.oxfordjournals.org/ by guest on April 11, 2016

be seen from Table 2 that the probability of being positive can be become extreme (i.e., close to 100% or close to 0%) depending on the cumulative score and the stage of the test. When the probability is close to 100%, it may be efficient to stop the assessment and immediately screen the respondent as positive. Conversely, when the probability is close to 0%, it may be efficient to stop the assessment and immediately screen the respondent as negative. In fact, this is exactly the logic of stochastic curtailment: this method halts the assessment once the probability of a positive result becomes sufficiently high or sufficiently low. In the former case, the respondent is screened as positive; in the latter case (which is equivalent to the probability of a negative result becoming sufficiently high), the respondent is screened as negative.

Stochastic curtailment stops more aggressively than curtailment: it stops whenever curtailment does, and stops earlier than curtailment in some instances. Therefore, stochastic curtailment makes greater reductions in respondent burden than curtailment does. However, these reductions in respondent burden may come at a price: unlike curtailment, which always gives the same result (positive or negative) as the full-length test, the result of stochastic curtailment does not necessarily match that of the full-length test. Hence, the sensitivity and specificity of stochastic curtailment might be lower than those of the full-length test. We note that stochastic curtailment was originally proposed for the stopping of clinical trials prior to their scheduled end [44] and was suggested for questionnaire usage in the context of personality assessment [31].

A natural question to ask when using stochastic curtailment is how high (or low) the probability of a positive result must be in order for the test to be terminated. Previous work [31,32] suggested stopping the assessment if the probability of a positive result becomes greater than or equal to 95% (determining the respondent is at high risk) or less than or equal to 5% (determining the respondent is at low risk). Based on this rule, Respondent #1 and Respondent #2 of Table 2 would receive only five items and nine items, respectively. A more liberal rule [33] is to stop when the probability of a positive result becomes greater than or equal to 90% or less than or equal to 10% (which would result in five items for Respondent #1 and eight items for Respondent #2). A more conservative rule [33] is to stop when the probability in question becomes greater than or equal to 99% or less than or equal to 1% (which would result in six items for Respondent #1 and 14 items for Respondent #2). Under all of these rules, Respondent #1 would be screened as positive and Respondent #2 would be screened as negative, matching the results of the full-length test.

A second natural question regards how to determine, at any stage of the test, the probability of the respondent ultimately being screened as positive by the full-length assessment. In other words, a statistical method to determine the numbers in the "Chance of 'positive result' (%)" column of Table 2 is needed. In previous studies [32–34], these numbers were obtained by conducting predictive modeling on training data (i.e., pilot data that are specifically taken to estimate the probabilities in question, prior to stochastic curtailment being used in practice). Finkelman et al. [33] compared two predictive modeling approaches (nonparametric estimation and logistic regression) and found logistic regression to be more effective in reducing respondent burden. We therefore focus attention on the latter procedure herein. In this procedure, a separate logistic regression model is estimated at each stage of the questionnaire; the independent variable in the logistic regression is the cumulative score at the given stage, and the dependent variable is the screening result of the full-length test (positive or negative). See Finkelman et al. [33] for further details.

Because it would be computationally inefficient to conduct logistic regression analyses during a respondent's assessment, all necessary calculations are performed ahead of time (before stochastic curtailment is used operationally for any respondent). That is, upon estimating all probabilities via logistic regression of the pilot data, the set of cumulative scores for which early stopping should occur is written as a simple list of decision rules for each stage of testing [33]. These decision rules are then checked for their internal consistency from stage to stage. For example, it would be undesirable to utilize a set of rules whereby respondents with a cumulative score of 3 at the sixth stage are stopped for a negative result, but respondents with a cumulative score of 3 at the seventh stage continue testing. Such a scenario would be internally inconsistent, considering that a cumulative score of 3 after seven items is at least as indicative of a negative result as a cumulative score of 3 after six items. If the initial decision rules produced by logistic regression contain an internal inconsistency, a simple adjustment of the rules is made so that they exhibit coherence from stage to stage [34]. In the above example, the rules would be updated so that either a cumulative score of 3 after seven items would result in early stopping, or a cumulative score of 3 after six items would not result in early stopping. The latter adjustment is generally favored in order to take a conservative approach [34]. Once finalized, the decision rules are implemented in practice using a computer program that delivers the questionnaire (and stops it when appropriate) without undue computational burden.

### *The Aberrant Drug Behavior Index (ADBI)*

In order to evaluate the predictive validity of the full-length SOAPP-R, curtailment, and stochastic curtailment, an external measure of aberrant medication-related behavior was needed. Such an external measure was provided by the ADBI, which was administered to respondents at follow-up. Specifics about this index have been provided in previous articles [18,19]. Briefly, the ADBI consists of three separate assessments: the

Downloaded from http://painmedicine.oxfordjournals.org/ by guest on April 11, 2016

Prescription Drug Use Questionnaire (PDUQ), the Prescription Opioid Therapy Questionnaire (POTQ), and a urine toxicology screen. The PDUQ is a 42-item self-report questionnaire that uses an interview format [45]. Based on published guidelines for assessing addiction in patients with chronic pain [46], the PDUQ includes items on evaluation of the pain condition, opioid use patterns, patient psychiatric history, and patient history of substance abuse, as well as family history and social/family factors [45]. Each item contributing to the total score counts an affirmative answer as one point, with the exception of one item (which asks about having explored or tried nonpharmacological pain management techniques) that is scored negatively. A cutoff point of $\geq 11$ for the total score was used previously [5,19] based on the results of Compton et al. [45], and was also employed in this study. The POTQ is a physician-reported instrument consisting of 11 dichotomously scored items, including questions related to multiple unsanctioned dose escalations, early refills with the absence of acute changes in the medical condition, episodes of lost or stolen prescriptions, frequent unscheduled visits to the clinic or emergency room, excessive phone calls, obtaining opioids from supplemental sources, and inflexibility about treatment options [7]. A cutoff point of $\geq 2$ for the total score was used based on previous studies [18,19]. Finally, the urine toxicology screen was defined to be positive for patients with evidence of having taken 1) an illicit substance, such as cocaine, or 2) an additional opioid medication that had not been prescribed [5,18,19]. The overall ADBI result was then considered to be positive if either the PDUQ was positive or both the POTQ and urine toxicology screen were positive [18,19].

*Statistical Analysis*

We conducted a retrospective analysis of the aforementioned $n = 428$ subjects who had previously been assessed via both the SOAPP-R and the ADBI. The goal of the retrospective analysis was to compare curtailment and stochastic curtailment with the full-length SOAPP-R in terms of testing efficiency. To accomplish this goal, a post-hoc simulation was conducted: a computer program was written to find the screening result (positive or negative) and test length that would have been observed for each subject, if computer-based testing had been used and curtailment (or stochastic curtailment) had been employed to determine when to stop testing. The results were then compared with those of the full-length SOAPP-R. Such post-hoc simulation is an established technique for evaluating the efficiency of questionnaire delivery methods [34,47,48].

All methods under study (the full-length SOAPP-R, curtailment, and stochastic curtailment) were evaluated in terms of their screening properties (sensitivity and specificity with respect to the ADBI) and their respondent burden (average and standard deviation of test length). Curtailment and stochastic curtailment were also assessed based on their sensitivity and specificity with respect to the full-length SOAPP-R, as well as the percentage of subjects for whom early stopping (i.e., stopping prior to the final item) occurred. Note that by definition, the full-length SOAPP-R stops early 0% of the time; therefore, it necessarily has an average test length of 24 items with a standard deviation of 0 items.

Before the above results could be obtained, it was necessary to "train" each method on the data. That is, it was necessary to perform initial calculations on the data so that each method was properly defined. For example, in order to find the sensitivity and specificity of the full-length SOAPP-R, it was first required that the cutoff point for this screener be determined. This determination was made via the Youden J index [49]: all possible cutoff points were examined and the one maximizing the quantity sensitivity + specificity − 1 was selected. The cutoff point that was chosen for the full-length SOAPP-R was then applied to curtailment and stochastic curtailment as well (i.e., this cutoff point was also used in curtailment and stochastic curtailment when a subject's assessment was not stopped early). The training process for stochastic curtailment involved the additional step of fitting logistic regression models, as described previously.

Two different analyses were performed. In the first analysis, the statistical models were trained on the full dataset ($n = 428$), and the methods under study were then evaluated on this same dataset. This approach has the advantage of using all data in model training. However, it is prone to the so-called "capitalization on chance" problem, in which the model performs more favorably in the study dataset than would subsequently be observed in practice [50]. Therefore, a second analysis was also undertaken in which 10-fold cross-validation was used. In 10-fold cross-validation, the dataset is randomly divided into 10 subsets of equal (or approximately equal) size. Nine of the subsets are pooled together, and the resulting "pooled" dataset is used for model training (including both cutoff point determination and logistic regression analysis, in this study); the tenth subset is then used to evaluate the performance of each method. By thus separating the data used for training from the data used for evaluation, the capitalization on chance problem is avoided [50]. The process is repeated 10 times, with each subset taking a turn as the evaluation dataset, and then results are aggregated across the 10 iterations. Sensitivities, specificities, and average test lengths from the cross-validation were compared with those obtained when training and evaluating each method on the full dataset.

Three versions of stochastic curtailment were examined. The most conservative version stopped when the probability of a positive result became greater than or equal to 99%, or less than or equal to 1%. The most liberal version replaced these thresholds with the numbers 90% and 10%, while a moderate version used the numbers 95% and 5%. These three versions will be referred to as $SC_{1,99}$, $SC_{10,90}$, and $SC_{5,95}$, respectively.



Downloaded from http://painmedicine.oxfordjournals.org/ by guest on April 11, 2016

**Table 3** Stopping rules of curtailment and stochastic curtailment

| | Curtailment | | $SC_{1,99}$ | | $SC_{5,95}$ | | $SC_{10,90}$ | |
|---|---|---|---|---|---|---|---|---|
| Stage of Testing | Stop: Negative Result | Stop: Positive Result | Stop: Negative Result | Stop: Positive Result | Stop: Negative Result | Stop: Positive Result | Stop: Negative Result | Stop: Positive Result |
| 1 | NA | NA | NA | NA | NA | NA | NA | NA |
| 2 | NA | NA | NA | NA | CS = 0 | CS = 8 | CS ≤ 1 | CS ≥ 7 |
| 3 | NA | NA | NA | CS = 12 | CS ≤ 1 | CS ≥ 10 | CS ≤ 2 | CS ≥ 9 |
| 4 | NA | NA | CS = 0 | CS ≥ 14 | CS ≤ 2 | CS ≥ 12 | CS ≤ 3 | CS ≥ 11 |
| 5 | NA | CS ≥ 19 | CS ≤ 1 | CS ≥ 16 | CS ≤ 4 | CS ≥ 13 | CS ≤ 5 | CS ≥ 12 |
| 6 | NA | CS ≥ 19 | CS ≤ 3 | CS ≥ 16 | CS ≤ 5 | CS ≥ 14 | CS ≤ 6 | CS ≥ 13 |
| 7 | NA | CS ≥ 19 | CS ≤ 4 | CS ≥ 18 | CS ≤ 6 | CS ≥ 15 | CS ≤ 7 | CS ≥ 14 |
| 8 | NA | CS ≥ 19 | CS ≤ 5 | CS ≥ 19 | CS ≤ 7 | CS ≥ 17 | CS ≤ 8 | CS ≥ 16 |
| 9 | NA | CS ≥ 19 | CS ≤ 5 | CS ≥ 19 | CS ≤ 8 | CS ≥ 18 | CS ≤ 9 | CS ≥ 17 |
| 10 | NA | CS ≥ 19 | CS ≤ 6 | CS ≥ 19 | CS ≤ 8 | CS ≥ 19 | CS ≤ 9 | CS ≥ 18 |
| 11 | NA | CS ≥ 19 | CS ≤ 6 | CS ≥ 19 | CS ≤ 9 | CS ≥ 19 | CS ≤ 10 | CS ≥ 18 |
| 12 | NA | CS ≥ 19 | CS ≤ 6 | CS ≥ 19 | CS ≤ 9 | CS ≥ 19 | CS ≤ 11 | CS ≥ 19 |
| 13 | NA | CS ≥ 19 | CS ≤ 8 | CS ≥ 19 | CS ≤ 11 | CS ≥ 19 | CS ≤ 12 | CS ≥ 19 |
| 14 | NA | CS ≥ 19 | CS ≤ 9 | CS ≥ 19 | CS ≤ 11 | CS ≥ 19 | CS ≤ 13 | CS ≥ 19 |
| 15 | NA | CS ≥ 19 | CS ≤ 9 | CS ≥ 19 | CS ≤ 12 | CS ≥ 19 | CS ≤ 13 | CS ≥ 19 |
| 16 | NA | CS ≥ 19 | CS ≤ 10 | CS ≥ 19 | CS ≤ 12 | CS ≥ 19 | CS ≤ 14 | CS ≥ 19 |
| 17 | NA | CS ≥ 19 | CS ≤ 10 | CS ≥ 19 | CS ≤ 13 | CS ≥ 19 | CS ≤ 14 | CS ≥ 19 |
| 18 | NA | CS ≥ 19 | CS ≤ 11 | CS ≥ 19 | CS ≤ 14 | CS ≥ 19 | CS ≤ 15 | CS ≥ 19 |
| 19 | NA | CS ≥ 19 | CS ≤ 12 | CS ≥ 19 | CS ≤ 14 | CS ≥ 19 | CS ≤ 16 | CS ≥ 19 |
| 20 | CS ≤ 2 | CS ≥ 19 | CS ≤ 13 | CS ≥ 19 | CS ≤ 15 | CS ≥ 19 | CS ≤ 16 | CS ≥ 19 |
| 21 | CS ≤ 6 | CS ≥ 19 | CS ≤ 14 | CS ≥ 19 | CS ≤ 16 | CS ≥ 19 | CS ≤ 17 | CS ≥ 19 |
| 22 | CS ≤ 10 | CS ≥ 19 | CS ≤ 15 | CS ≥ 19 | CS ≤ 17 | CS ≥ 19 | CS ≤ 17 | CS ≥ 19 |
| 23 | CS ≤ 14 | CS ≥ 19 | CS ≤ 15 | CS ≥ 19 | CS ≤ 18 | CS ≥ 19 | CS ≤ 18 | CS ≥ 19 |
| 24 | CS ≤ 18 | CS ≥ 19 | CS ≤ 18 | CS ≥ 19 | CS ≤ 18 | CS ≥ 19 | CS ≤ 18 | CS ≥ 19 |

Results are based on the complete dataset (n = 428).

NA = not applicable (no early stopping can occur); CS = cumulative score.

A computer program written in R (Version 2.13.1) was used to carry out the analysis. In addition to providing information on the screening properties and respondent burden of each method, the program calculated descriptive statistics on each item. Specifically, the mean, median, standard deviation, and inter-quartile range of each item were computed.

## Results

Of the 425 subjects with valid age information, the mean (SD) age was 51.4 (13.0) years. Of the 426 subjects with valid gender information, 243 were female (57.0%). The result of the ADBI was negative for 283 of the 428 subjects in the dataset (66.1%). Among these 428 subjects, the mean (SD) total score for the full-length SOAPP-R was 20.4 (11.3).

Table 1 shows information for all 24 items of the full-length SOAPP-R. The items with the highest means were "have mood swings" (mean = 2.0) and "felt a need for higher doses of medication" (mean = 1.9). The items with the lowest means were "been treated for an alcohol or drug problem" (mean = 0.1), "had to borrow pain medications from your family or friends" (mean = 0.2), and "been in an argument that was so out of control that someone got hurt" (mean = 0.2). All medians and inter-quartile ranges were between 0 and 2.

Using the complete dataset (n = 428) and the Youden J index, a cutoff point of ≥ 19 was obtained for the full-length SOAPP-R. Based on this cutoff point, the full-length SOAPP-R screened as positive 108 of the 145 subjects that were identified as positive by the ADBI (sensitivity = 0.745). The full-length SOAPP-R screened as negative 190 of the 283 subjects that were identified as negative by the ADBI (specificity = 0.671).

Table 3 provides the stopping rules for curtailment and each version of stochastic curtailment ($SC_{1,99}$, $SC_{5,95}$, and $SC_{10,90}$). This table is written as a list of decision rules: at each stage of testing, the cumulative scores for which early stopping occurs are provided. For instance, after stage 20 of testing (i.e., after 20 items have been administered), curtailment stops to screen the



Downloaded from http://painmedicine.oxfordjournals.org/ by guest on April 11, 2016

**Table 4** Sensitivity, specificity, and respondent burden of each method

| | Predicting the Full-Length SOAPP-R | | Predicting the ADBI | | Average Test Length | SD of Test Length | % Test Lengths <24 |
|---|---|---|---|---|---|---|---|
| | Sensitivity | Specificity | Sensitivity | Specificity | | | |
| Full-length SOAPP-R | 1 | 1 | 0.745 | 0.671 | 24.0 | 0.0 | 0.0 |
| Curtailed | 1 | 1 | 0.745 | 0.671 | 17.7 | 6.3 | 80.6 |
| $SC_{1,99}$ | 1 | 1 | 0.745 | 0.671 | 14.1 | 6.4 | 86.4 |
| $SC_{5,95}$ | 0.980 | 0.996 | 0.724 | 0.671 | 10.8 | 6.6 | 100.0 |
| $SC_{10,90}$ | 0.935 | 0.960 | 0.710 | 0.668 | 8.3 | 6.1 | 100.0 |

Model training and evaluation were both performed on the complete dataset (n = 428).

respondent as negative if his/her cumulative score (CS) is $\leq 2$; it stops to screen the respondent as positive if his/her cumulative score is $\geq 19$. The analogous rules are $\leq 13$ and $\geq 19$ for $SC_{1,99}$; $\leq 15$ and $\geq 19$ for $SC_{5,95}$; and $\leq 16$ and $\geq 19$ for $SC_{10,90}$. One adjustment was made for the purpose of internal consistency: a "CS $\leq 15$" rule was used for $SC_{1,99}$ at stage 22, rather than an initial "CS $\leq 16$" rule obtained from logistic regression, in order to be consistent with the "CS $\leq 15$" rule at stage 23. Note that the stopping rules presented in Table 3 were derived from the full dataset (n = 428); they take advantage of all available data and therefore are most suitable for practical usage. The stopping rules resulting from cross-validation are not presented for the purpose of parsimony; in all cases, they were similar to the rules derived from the full dataset.

Table 4 presents results for the analysis in which both model training and evaluation were performed on the full dataset. As is always the case, curtailment was perfectly concordant with the full-length screener (sensitivity and specificity of 1 for predicting the full-length SOAPP-R). Therefore, for predicting the ADBI, curtailment exhibited the same sensitivity (0.745) and specificity (0.671) as the full-length screener. Additionally, curtailment lessened the respondent burden of the SOAPP-R: it reduced the average test length from 24 to 17.7 items, with early stopping in 80.6% of tests. $SC_{1,99}$ further enhanced the efficiency of the assessment: it was perfectly concordant with the full-length SOAPP-R while administering an average of 14.1 items and stopping early in 86.4% of tests. $SC_{5,95}$ and $SC_{10,90}$ were more aggressive in stopping and therefore did not always match the screening result of the full-length SOAPP-R. The sensitivity and specificity of $SC_{5,95}$ for predicting the full-length SOAPP-R were 0.980 and 0.996, respectively; the corresponding values for $SC_{10,90}$ were 0.935 and 0.960. For predicting the ADBI, $SC_{5,95}$ had the same specificity as the full-length SOAPP-R and a sensitivity 0.021 lower; $SC_{10,90}$ had specificity and sensitivity 0.003 and 0.035 lower, respectively, than the full-length SOAPP-R. Both of these methods lessened respondent burden by at least 55% compared with the full-length assessment: the average test lengths for $SC_{5,95}$ and $SC_{10,90}$ were 10.8 and 8.3, respectively. Each method stopped the test early for 100% of respondents.

Table 5 presents results of the 10-fold cross-validation. All 10 iterations resulted in a cutoff point of $\geq 19$ based on the Youden J index (results not shown). Both the full-length screener and curtailment exhibited the same properties in cross-validation as had been observed when model training and evaluation were performed on the full dataset (i.e., their Table 5 values are identical to their Table 4 values). All stochastic curtailment methods exhibited cross-validation sensitivities and specificities within 0.015 of their Table 4 values. $SC_{1,99}$ was no longer perfectly concordant with the full-length screener:

**Table 5** Sensitivity, specificity, and respondent burden of each method

| | Predicting the full-length SOAPP-R | | Predicting the ADBI | | Average test length | SD of test length | % Test lengths <24 |
|---|---|---|---|---|---|---|---|
| | Sensitivity | Specificity | Sensitivity | Specificity | | | |
| Full-length SOAPP-R | 1 | 1 | 0.745 | 0.671 | 24.0 | 0.0 | 0.0 |
| Curtailed | 1 | 1 | 0.745 | 0.671 | 17.7 | 6.3 | 80.6 |
| $SC_{1,99}$ | 0.985 | 1 | 0.731 | 0.675 | 14.1 | 6.4 | 89.0 |
| $SC_{5,95}$ | 0.975 | 0.991 | 0.717 | 0.668 | 10.8 | 6.6 | 96.5 |
| $SC_{10,90}$ | 0.940 | 0.956 | 0.710 | 0.661 | 8.4 | 6.1 | 100.0 |

Ten-fold cross-validation was performed (n = 428).



Downloaded from http://painmedicine.oxfordjournals.org/ by guest on April 11, 2016

the former's sensitivity and specificity for predicting the latter were 0.985 and 1, respectively, in cross-validation. Compared with the full-length screener, $SC_{1,99}$ exhibited slightly lower sensitivity (0.731 vs 0.745)—but slightly higher specificity (0.675 vs 0.671)—for predicting the ADBI. Regarding respondent burden, all stochastic curtailment methods exhibited average test lengths (and standard deviations) within 0.1 of their Table 4 values; all percentages of early stopping were within 3.5% of their Table 4 values.

## Discussion

Screening is typically required when the burden of illness is high, as it is when considering chronic opioid therapy. The burden of testing must be commensurate with the benefit; tests should be inexpensive, accurate, and brief [51]. With the advent of required electronic health records, most future assessment instruments will necessarily be woven into the patient's medical record. Hence, close attention must be paid to a model that can lend itself to integrating cost-effective screening into the record [12].

A benefit of computerized instruments is that they can be customized at the level of the individual respondent and therefore can garner enhanced measurement efficiency [52–57]. Such customized assessment was previously studied for the COMM [34], but not for the SOAPP-R. Since these two screeners have distinct purposes (the former is designed to assess current aberrant medication-related behavior, whereas the latter is designed to predict it in the future), the development of a customized SOAPP-R is important for the efficient prediction of aberrant behavior. Efficiency is especially critical for the SOAPP-R because this screener is typically taken by patients with chronic pain, and individuals who are physically ill are known to be particularly sensitive to the effects of respondent burden [22]. The importance of keeping questionnaires brief may be further heightened when respondents are assessed for multiple health problems in a single visit; additionally, reducing the length of a questionnaire may be valuable as a means to alleviate the potential emotional stress associated with taking it [58].

The goal of this research was to develop a family of methods that can shorten the SOAPP-R while maintaining adequate concordance with the full-length screener's result (positive or negative). The most liberal of these methods, $SC_{10,90}$, reduced the average test length by 65% while matching the result of the full-length screener in 94.9% of cases (whether performing model training on the entire dataset or using cross-validation). The more conservative $SC_{5,95}$ reduced the average test length by 55% while matching the full-length screener's result in over 98% of cases (again, whether performing model training on the entire dataset or using cross-validation). For $SC_{1,99}$, the reduction in average test length was 41%; this method matched the full-length screener's result in 100% of cases when training on the full dataset, and in over 99% of cases in cross-validation. Finally, the most conservative method was curtailment, which reduced the average test length by 26%. Because curtailment's screening result always matches that of the full-length SOAPP-R, the concordance between the two is guaranteed to be 100% in any dataset.

Which variable-length procedure to use in practice depends on the desired balance between lessening the average test length and maintaining the sensitivity and specificity of the assessment. The most liberal procedure under study, $SC_{10,90}$, achieved the greatest reduction in respondent burden; however, $SC_{5,95}$ exhibited an average test length within 2.5 items of that of $SC_{10,90}$ while garnering considerably greater concordance with the full-length test. In order to best preserve the screening properties of the SOAPP-R, the more conservative short versions ($SC_{5,95}$, $SC_{1,99}$, and curtailment) may be recommended.

One limitation of the study was its retrospective nature: each method's performance was assessed based on the results of a post-hoc simulation. It is possible that the results obtained in a prospective study, with the SOAPP-R administered via computer, would differ from those obtained retrospectively. Additionally, while the curtailment stopping rules of Table 3 are suitable for operational usage in any population for which a $\geq 19$ cutoff point is appropriate, the stopping rules for $SC_{1,99}$, $SC_{5,95}$, and $SC_{10,90}$ are population-specific and hence should be validated prior to their use in a given population. Finally, because the two populations studied herein were drawn from similar regions of the country, results may not be generalizable to the United States pain population or to populations from other regions.

While adjunctive measures like the SOAPP-R may improve our ability to identify high-risk patients, an instrument of any length has inherent limitations. The results of the SOAPP-R are intended as a complement to information from other sources, such as history and physical examination, psychiatric/substance abuse history, clinical interview, review of prior medical records, and laboratory findings [18,19]. The material from these other sources would be included in clinical documentation, allowing any information from items omitted in the shortened SOAPP-R to be incorporated into the medical record. Results of the SOAPP-R, whether in shortened or full-length form, are not intended as a replacement for clinical judgment.

This research represents a first step toward utilizing variable-length testing techniques in conjunction with the SOAPP-R. Given the considerable improvements in average test length achieved by curtailment and stochastic curtailment in post-hoc simulation, the next step is to develop a functional computer-based version of each method. Future studies will then prospectively evaluate the comparability between the paper-and-pencil form of the SOAPP-R and all of its computerized

Downloaded from http://painmedicine.oxfordjournals.org/ by guest on April 11, 2016

versions (including a computerized full-length SOAPP-R as well as curtailment and stochastic curtailment). Such work will promote the efficient prediction of aberrant drug-related behavior among chronic pain patients.

Use of prescription opioid analgesics for chronic pain remains controversial. For example, Deyo et al. [59] have noted that despite the proliferation of guidelines calling for increased screening for risk, overall prescription rates and adverse events associated with opioid use (i.e., misuse, abuse, and overdose) have not decreased. We concur with these authors in endorsing selective prescription of opioids, use of lower doses when possible, use of prescription drug monitoring programs (PDMPs), avoidance of co-prescription with other neurologic depressants, and consideration of the use of abuse deterrent reformulations that make the tablets and capsules more difficult to snort, smoke, or inject. We also acknowledge that screening alone is insufficient in determining a risk profile. Systematic risk screening may help to standardize risk evaluation, and in combination with the efforts described above, plus a detailed clinical interview, appropriate monitoring of urine drug testing (UDT) and treatment agreements, it seems possible to potentially reduce inappropriate prescribing and opioid-related adverse events. Indeed, recently published post-marketing surveillance data [60] suggest that the large increases in the rates of opioid diversion and abuse observed from 2002 to 2010 have flattened or decreased from 2011 through 2013. This might suggest that a variety of interventions, perhaps including greater levels of systematic screening, may be having an impact on the prescription opioid problem. Judicious screening efforts provide the physician with an opportunity to address the inevitable risks.

## Conclusions

Curtailment and stochastic curtailment have the potential to substantially reduce the respondent and administrative burden of the SOAPP-R, without unduly affecting its screening properties, in computer-based administrations of the questionnaire.

## References


1 Atluri S, Sudarshan G, Manchikanti L. Assessment of the trends in medical use and misuse of opioid analgesics from 2004 to 2011. Pain Physician 2014; 17:E119–28.

2 Fine P, Webster L, Argoff C. American Academy of Pain Medicine response to PROP petition to the FDA that seeks to limit pain medications for legitimate noncancer pain sufferers. Pain Med 2012;13: 1259–64.

3 Chou R, Fanciullo, GJ, Fine PG, et al. Clinical guidelines for the use of chronic opioid therapy in chronic noncancer pain. J Pain 2009;10:113–30.

4 Chou R, Turner JA, Devine EB, et al. The effectiveness and risks of long-term opioid therapy for chronic pain: A systematic review for a National Institutes of Health Pathways to Prevention Workshop. Ann Intern Med 2015;162:276–86.

5 Butler SF, Budman, SH, Fernandez KC, et al. Development and validation of the Current Opioid Misuse Measure. Pain 2007;130:144–56.

6 Cheatle MD. Psychological dependence and prescription opioid misuse and abuse. Pain Med 2014; 15:541–3.

7 Michna E, Ross EL, Hynes WL, et al. Predicting aberrant drug behavior in patients treated for chronic pain: Importance of abuse history. J Pain Symptom Manage 2004;28:250–8.

8 Peirce GL, Smith MJ, Abate MA, Halverson J. Doctor and pharmacy shopping for controlled substances. Med Care 2012;50:494–500.

9 Gourlay DL, Heit HA. Universal precautions revisited: Managing the inherited pain patient. Pain Med 2009;10:S115–23.

10 Gourlay DL, Heit HA, Almahrezi A. Universal precautions in pain medicine: A rational approach to the treatment of chronic pain. Pain Med 2005;6:107–12.

11 Webster L, St Marie B, McCarberg B, et al. Current status and evolving role of abuse-deterrent opioids in managing patients with chronic pain. J Opioid Manag 2011;7:235–45.

12 Cahana A, Dansie EJ, Theodore BR, Wilson HD, Turk DC. Redesigning delivery of opioids to optimize pain management, improve outcomes, and contain costs. Pain Med 2013;14:36–42.

13 Jamison RN, Serraillier J, Michna E. Assessment and treatment of abuse risk in opioid prescribing for chronic pain. Pain Res Treat 2011;2011: 941808

14 Sehgal N, Manchikanti L, Smith HS. Prescription opioid abuse in chronic pain: A review of opioid abuse predictors and strategies to curb opioid abuse. Pain Physician 2012;15:ES67–92.

15 Butler SF, Budman SH, Fernandez K, Jamison RN. Validation of a screener and opioid assessment measure for patients with chronic pain. Pain 2004; 112:65–75.




Downloaded from http://painmedicine.oxfordjournals.org/ by guest on April 11, 2016

16 Jamison RN, Link CL, Marceau LD. Do pain patients at high risk for substance misuse experience more pain? A longitudinal outcomes study. Pain Med 2009;10:1084–94.

17 Moore TM, Jones T, Browder JH, Daffron S, Passik SD. A comparison of common screening methods for predicting aberrant drug-related behavior among patients receiving opioids for chronic pain management. Pain Med 2009;10:1426–33.

18 Butler SF, Fernandez K, Benoit C, Budman SH, Jamison RN. Validation of the revised Screener and Opioid Assessment for Patients with Pain (SOAPP-R). J Pain 2008;9:360–72.

19 Butler SF, Budman SH, Fernandez KC, Fanciullo GJ, Jamison RN. Cross-validation of a screener to predict opioid misuse in chronic pain patients (SOAPP-R). J Addict Med 2009;3:66–73.

20 Hosmer DW, Lemeshow S. Applied Logistic Regression, 2nd edition. New York: John Wiley & Sons; 2000.

21 National Opioid Use Guideline Group (NOUGG). Canadian Guideline for Safe and Effective Use of Opioids for Chronic Non-Cancer Pain. 2010. Available at: http://nationalpaincentre.mcmaster.ca/opioid/ (accessed December 2014).

22 Carpenter JS, Andrykowski MA, Wilson J, et al. Psychometrics for two short forms of the Center for Epidemiologic Studies-Depression Scale. Issues Ment Health Nurs 1998;19:481–94.

23 Adams LLM, Gale D. Solving the quandary between questionnaire length and response rate in educational research. Res High Educ 1982;17:231–40.

24 Herzog AR, Bachman JG. Effects of questionnaire length on response quality. Public Opin Q 1981;45:549–59.

25 Aaronson N, Alonso, J, Burnam A, et al. Assessing health status and quality-of-life instruments: Attributes and review criteria. Qual Life Res 2002;11:193–205.

26 Esherick JS, Clark DS, Slater ED. CURRENT practice guidelines in primary care 2014. Available at: http://accessmedicine.mhmedical.com/book.aspx?bookid=925 (accessed December 2014).

27 Dugdale DC, Epstein R, Pantilat SZ. Time and the patient-physician relationship. J Gen Intern Med 1999;14(suppl 1):S34–40.

28 Akbik H, Butler SF, Budman SH, et al. Validation and clinical application of the Screener and Opioid Assessment for Patients with Pain (SOAPP). J Pain Symptom Manage 2006;32:287–93.

29 Koyyalagunta D, Bruera E, Aigner C, et al. Risk stratification of opioid misuse among patients with cancer pain using the SOAPP-SF. Pain Med 2013;14:667–75.

30 Ben-Porath YS, Slutske WS, Butcher JN. A real-data simulation of computerized adaptive administration of the MMPI. Psychol Assess 1989;1:18–22.

31 Butcher JN, Keller LS, Bacon SF. Current developments and future directions in computerized personality assessment. J Consult Clin Psychol 1985;53:803–15.

32 Finkelman MD, He Y, Kim W, Lai AM. Stochastic curtailment of health questionnaires: A method to reduce respondent burden. Stat Med 2011;30:1989–2004.

33 Finkelman MD, Smits N, Kim W, Riley B. Curtailment and stochastic curtailment to shorten the CES-D. Appl Psychol Meas 2012;36:632–58.

34 Finkelman MD, Kulich RJ, Zoukhri D, Smits N, Butler SF. Shortening the Current Opioid Misuse Measure via computer-based testing: A retrospective proof-of-concept study. BMC Med Res Methodol 2013;13:126

35 Fokkema M, Smits N, Finkelman MD, Kelderman H, Cuijpers P. Curtailment: A method to reduce the length of self-report questionnaires while maintaining diagnostic accuracy. Psychiatry Res 2014;215:477–82.

36 Forbey JD, Handel RW, Ben-Porath YS. A real data simulation of computerized adaptive administration of the MMPI-A. Comput Hum Behav 2000;16:83–96.

37 Forbey JD, Ben-Porath YS. Computerized adaptive personality testing: A review and illustration with the MMPI-2 computerized adaptive version. Psychol Assess 2007;19:14–24.

38 Forbey JD, Ben-Porath YS, Arbisi PA. The MMPI–2 computerized adaptive version (MMPI–2–CA) in a Veterans Administration medical outpatient facility. Psychol Assess 2012;24:628–39.

39 Roper BL, Ben-Porath YS, Butcher JN. Comparability of computerized adaptive and conventional testing with the MMPI-2. J Pers Assess 1991;57:278–90.

40 Roper BL, Ben-Porath YS, Butcher JN. Comparability and validity of computerized adaptive testing with the MMPI-2. J Pers Assess 1995;65:358–71.

Downloaded from http://painmedicine.oxfordjournals.org/ by guest on April 11, 2016

41 Thompson NA. A practitioner's guide for variable-length computerized classification testing. Prac Assess Res Eval 2007;12. Available at: http://pareonline.net/pdf/v12n1.pdf (accessed December 2014).

42 Eisenberg B, Ghosh BK. Curtailed and uniformly most powerful sequential tests. Ann Stat 1980;8: 1123–31.

43 Eisenberg B, Simons G. On weak admissibility of tests. Ann Stat 1978;6:319–32.

44 Lan KKG, Simon R, Halperin M. Stochastically curtailed tests in long–term clinical trials. Commun Stat Seq Anal 1982;1:207–19.

45 Compton P, Darakjian J, Miotto K. Screening for addiction in patients with chronic pain and "problematic" substance use: Evaluation of a pilot assessment tool. J Pain Symptom Manage 1998;16:355–63.

46 Miotto K, Compton P, Ling W, Conolly M. Diagnosing addictive disease in chronic pain patients. Psychosomatics 1996;37:223–35.

47 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.

48 Smits N, Zitman FG, Cuijpers P, den Hollander-Gijsman ME, Carlier IV. A proof of principle for using adaptive testing in Routine Outcome Monitoring: The efficiency of the Mood and Anxiety Symptoms Questionnaire—Anhedonic Depression CAT. BMC Med Res Methodol 2012;12:4.

49 Youden WJ. Index for rating diagnostic tests. Cancer 1950;3:32–5.

50 Hastie T, Tibshirani R, Friedman JH. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edition. New York: Springer; 2009.

51 Dans LF, Silvestre MA, Dans AL. Trade-off between benefit and harm is crucial in health screening recommendations. Part I: General principles. J Clin Epidemiol 2011;64:231–9.

52 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):133–41.

53 Cella D, Yount S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH Roadmap cooperative group during its first two years. Med Care 2007;45:S3–11.

54 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;36:2061–6.

55 Fries JF, Witter J, Rose M, et al. Item response theory, computerized adaptive testing, and PROMIS: Assessment of physical function. J Rheumatol 2014; 41:153–8.

56 Hung M, Franklin JD, Hon SD, et al. Time for a paradigm shift with computerized adaptive testing of general physical function outcomes measurements. Foot Ankle Int 2014;35:1–7.

57 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 Outcomes Measurement Information System (PROMIS). Med Care 2007;45:S22–31.

58 Kohout FJ, Berkman LF, Evans DA, Cornoni-Huntley J. Two shorter forms of the CES-D (Center for Epidemiological Studies Depression) depression symptoms index. J Aging Health 1993;5:179–93.

59 Deyo RA, Von Korff M, Duhrkoop D. Opioids for low back pain. BMJ 2015;350:g6380.

60 Dart RC, Surratt HL, Cicero TJ, et al. Trends in opioid analgesic abuse and mortality in the United States. N Engl J Med 2015;372:241–8.

Downloaded from http://painmedicine.oxfordjournals.org/ by guest on April 11, 2016