



ELSEVIER

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

Roger Chou,^{*} Gilbert J. Fanciullo,[†] Perry G. Fine,[‡] Christine Miaskowski,[§] Steven D. Passik,^{||} and Russell K. Portenoy[¶]

^{*}*The Oregon Evidence-based Practice Center, Department of Medicine, Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, Oregon.*

[†]*Pain Management Center, Department of Anesthesiology, Dartmouth-Hitchcock Medical Center, Lebanon, New Hampshire.*

[‡]*Pain Research Center, Department of Anesthesiology, University of Utah, Salt Lake City, Utah.*

[§]*Department of Physiological Nursing, School of Nursing, University of California, San Francisco.*

^{||}*Department of Psychiatry and Behavioral Sciences, Memorial Sloan-Kettering Cancer Center, New York, New York; and*

[¶]*Department of Pain Medicine and Palliative Care, Beth Israel Medical Center, New York, New York.*

Abstract: Optimal methods to predict risk of aberrant drug-related behaviors before initiation of opioids for chronic noncancer pain and to identify aberrant behaviors after therapy is initiated are uncertain. We systematically reviewed published literature identified through searches of Ovid MEDLINE and the Cochrane databases through July 2008. Diagnostic test characteristics and accompanying confidence intervals were calculated with data extracted from the studies. Four prospective studies evaluated diagnostic accuracy of risk prediction instruments. Two higher-quality derivation studies found that high scores on the Screener and Opioid Assessment for Patients with Pain (SOAPP) Version 1 and the Revised SOAPP (SOAPP-R) instruments weakly increased the likelihood for future aberrant drug-related behaviors (positive likelihood ratios [PLR], 2.90 [95% CI, 1.91 to 4.39] and 2.50 [95% CI, 1.93 to 3.24], respectively). Low scores on the SOAPP Version 1 moderately decreased the likelihood for aberrant drug-related behaviors (negative likelihood ratio [NLR], 0.13 [95% CI, 0.05 to 0.34]) and low scores on the SOAPP-R weakly decreased the likelihood (NLR, 0.29 [95% CI, 0.18 to 0.46]), but estimates are too imprecise to determine if there is a difference between these instruments. One lower-quality study found that categorization as high risk using the Opioid Risk Tool strongly increased the likelihood for future aberrant drug-related behaviors (PLR, 14.3 [95% CI, 5.35 to 38.4]) and classification as low risk strongly decreased the likelihood (PLR, 0.08 [95% CI, 0.01 to 0.62]). Nine studies evaluated monitoring instruments for identification of aberrant drug-related behaviors in patients on opioid therapy. One higher-quality derivation study found higher scores on the Current Opioid Misuse Measure (COMM) weakly increased the likelihood of current aberrant drug-related behaviors (PLR, 2.77 [95% CI, 2.06 to 3.72]) and lower scores weakly decreased the likelihood (NLR, 0.35 [95% CI, 0.24 to 0.52]). In 8 studies of other monitoring instruments, diagnostic accuracy was poor, results were difficult to interpret due to methodological shortcomings, or standard diagnostic test characteristics were not reported. Definitions for aberrant drug-related behaviors were not standardized across studies and did not account for seriousness of identified behaviors. No reliable evidence exists on accuracy of urine drug screening, pill counts, or prescription drug monitoring programs; or clinical outcomes associated with different assessment or monitoring strategies.

Perspective: Evidence on prediction and identification of aberrant drug-related behaviors is limited. Although several screening instruments may be useful, evidence is sparse and primarily based

This article is based on research conducted at the Oregon Evidence-based Practice Center with funding from the American Pain Society (APS). The authors are solely responsible for the content of this article and the decision to submit for publication. No statement in this article should be construed as an official position of the APS or the American Academy of Pain Medicine. The authors have no known or potential conflicts of interest to declare.

Address reprint requests to Dr Roger Chou, 3181 SW Sam Jackson Park Road, Mail Code B1CC, Portland, OR 97239. E-mail: chour@ohsu.edu
1526-5900/\$36.00

© 2009 by the American Pain Society
doi:10.1016/j.jpain.2008.10.009

on derivation studies, and methodological shortcomings exist in all studies. Research that performs external validation, uses standardized definitions for clinically relevant aberrant drug-related behaviors, and evaluates clinical outcomes associated with different assessment and monitoring strategies is needed.

© 2009 by the American Pain Society

Key words: Analgesics, opioid, pain, risk assessment, drug monitoring, substance-related disorders, drug toxicity, systematic review, aberrant drug-related behaviors.

Editor's Note: The American Pain Society and the American Academy of Pain Medicine present this second of 3 articles in this 3-part report as a guideline for opioid treatment of noncancer pain.

Use of opioids for chronic noncancer pain (CNCP) remains controversial. Data on the long-term effectiveness of opioids for CNCP are sparse, with inconclusive or mixed results.⁵⁰ Although extensive clinical experience suggests that opioids can improve pain and function in some patients,^{21,34} a significant proportion experience no improvement or worsening of symptoms,³ and opioid use is associated with a variety of potentially serious adverse outcomes, including harms related to drug abuse and diversion.^{26,49}

Proper patient selection could mitigate potential risks and enhance potential benefits associated with the prescription of opioids for CNCP.⁵¹ Recent clinical guidelines emphasize the value of risk stratification when contemplating a therapeutic trial of opioids, focusing on assessment of risk for aberrant drug-related behaviors consistent with abuse, addiction, or diversion.^{6,22,31,33,61} Risk stratification may lead to the decision to forego a trial or to offer opioid therapy only with consultative assistance or guide use of various interventions intended to enhance control and monitoring, such as opioid agreements or urine drug screening.

If long-term treatment with an opioid is undertaken for chronic pain, periodic monitoring is essential to optimize benefit and minimize risk during the course of treatment.⁵⁶ Risks and benefits of opioids do not remain static over time due to changes in the severity of the underlying pain condition, development or progression of medical or psychiatric comorbidities, and other factors. Regular monitoring of an array of outcomes is therefore critical to assess the therapeutic response.⁵³ As in performing risk stratification, monitoring for aberrant drug-related behaviors consistent with abuse, addiction, or diversion is considered a core aspect of best practice during opioid therapy.^{6,22,31,33,61} Based on monitoring assessments, treatment may be continued, modified, or possibly discontinued.

Risk stratification and monitoring for aberrant drug-related behaviors may be based on clinical evaluation, the use of formal instruments, or other interventions (such as urine drug screens, pill counts, or prescription drug monitoring programs). Instruments developed to assist clinicians in risk stratification and monitoring generally appear to have strong face, content, and construct validity, but evidence on the accuracy of these instruments for predicting clinical outcomes is limited, and it

is unclear whether the use of these instruments to help guide clinical decision-making improves patient outcomes.⁶³ Uncertainty also exists with regard to optimal monitoring intervals and appropriate use of urine drug screens, pill counts, and prescription drug monitoring programs.^{18,35}

This article reviews current evidence on the accuracy and clinical utility of risk stratification instruments for prediction of future aberrant drug-related behaviors and methods (monitoring instruments, monitoring intervals, urine drug screens, pill counts, and prescription monitoring programs) for identification of aberrant drug-related behaviors during therapy. It is part of a larger evidence review commissioned by the American Pain Society (APS) and the American Academy of Pain Medicine (AAPM) to guide development of recommendations on use of opioids for CNCP.¹²

Materials and Methods

Data Sources and Searches

Searches were conducted (from the inception of each database through July 2008) that combined terms for opioids and chronic pain on Ovid MEDLINE, the Cochrane Central Register of Controlled Trials, and the Cochrane Database of Systematic Reviews (Appendix 1 shows detailed search strategies). Electronic searches were supplemented with reference lists and additional citations suggested by experts.

Evidence Selection

We included the following studies of adults (>18 years old) with CNCP:

- Prospective studies that evaluated the ability of risk stratification instruments to predict aberrant drug-related behaviors in patients prescribed chronic opioid therapy.
- Studies that evaluated the accuracy of monitoring instruments, urine drug screens, prescription drug monitoring, blood level monitoring, and pill counts to identify current aberrant drug-related behaviors in patients on opioid therapy.
- Randomized trials and controlled observational studies that evaluated the effects of risk stratification or monitoring strategies on patient outcomes (pain, function, adverse effects, rates of aberrant drug-related behaviors, mortality).

We excluded non-English language studies, studies published only as conference abstracts, unpublished studies, and studies published only as dissertations.

Data Extraction and Quality Assessment

Two reviewers independently rated the quality of each included study. Discrepancies were resolved by discussion and a consensus process. If data were available from the studies, we used the *diagti* procedure (confidence intervals based on the exact method) in Stata (Stata version 10, StataCorp, College Station, Texas) to calculate sensitivities and specificities and the *cci* procedure (confidence intervals based on the normal approximation) to calculate positive likelihood ratios (PLRs), negative likelihood ratios (NLRs), and diagnostic odds ratios (DORs). If a cell of a 2×2 table had zero events, we added 0.5 to all cells to calculate likelihood and diagnostic odds ratios.

We assessed the quality of studies of risk prediction or diagnostic test accuracy using 9 criteria adapted from methods developed by the United States Preventive Services Task Force²⁵ or evaluated in empiric studies^{36, 67} of sources of variation and bias in studies of diagnostic tests (Appendix 2), including a criterion that assessed whether a study evaluated diagnostic test performance in a population other than the one used to derive the instrument (external validation).^{45,67} We considered studies that met at least five of the nine criteria to be of higher-quality.

Data Synthesis

We qualitatively synthesized evidence using methods adapted from the US Preventive Services Task Force.²⁵ To assign an overall strength of evidence (good, fair, or poor) to a related body of literature, we considered the number, quality, and size of studies; consistency of results between studies; and directness of evidence. Minimum criteria for fair and good quality ratings are shown in Appendix 3. Consistent results from a number of higher-quality studies across a broad range of populations support a high degree of certainty that the results of the studies are true (the entire body of evidence would be considered "good-quality"). For a "fair-quality" body of evidence, results could be due to true effects or to biases that operated across some or all of the studies. For a "poor-quality" body of evidence, any conclusion is uncertain due to serious methodological shortcomings, sparse data, or markedly inconsistent results.

We classified PLRs >10 and NLRs ≤ 0.1 as "large/strong," PLRs >5 and ≤ 10 and NLRs >0.1 and ≤ 0.2 as "moderate," and PLRs >2 and ≤ 5 and NLRs >0.2 and ≤ 0.5 as "small/weak."³⁰

Results

Results of Literature Search

The literature searches yielded a total of 1,068 potentially relevant citations; of those, 44 were retrieved. After reviewing full-text articles, 4 studies of risk prediction instruments,^{2,7,9,66} 9 studies of monitoring instruments,^{1,4,8,15,27,43,47,64,68} 1 study on accuracy of urine drug screening,¹⁷ and 2 studies on the effect of urine drug screening⁴¹ or adherence monitoring⁴⁰ on clinical outcomes met the pre-specified inclusion criteria. Fifteen potentially relevant studies of risk prediction^{5,16,19,23,24,28,37-}

^{39,42,44,46,57,59,65} and 5 potentially relevant studies of monitoring^{11,14,20,53,54} were excluded based on reasons described in Appendix 4. Studies that evaluated the ability to predict opioid responsiveness were also excluded.^{32,60}

Accuracy of Screening Instruments to Predict Future Aberrant Drug-Related Behaviors

Four prospective studies (658 patients completed follow-up) evaluated the ability of 3 different self-administered instruments to predict aberrant drug-related behaviors (Tables 1 and 2).^{2,7,9,66} The number of risk assessment items in these instruments ranged from 10 to 24; although the specific items varied, they included a personal or family history of drug or alcohol abuse, previous aberrant drug-related behaviors, dysfunctional coping strategies, comorbid psychiatric conditions, cigarette smoking, age, and childhood sexual abuse.⁶³ Three of the 4 studies met our threshold for a higher-quality study,^{2,7,9} but none met all quality criteria. Two studies evaluated diagnostic test performance in the same population used to derive the instrument.^{7,9} It was not clear in any study if outcome assessors were blinded to the results of the screening instrument. In addition, definitions for aberrant drug-related behaviors and abnormal urine toxicology results were not well standardized and did not distinguish relatively mild from more serious behaviors. In one study,⁶⁶ aberrant behaviors were not clearly predefined. Attrition bias is also a concern. In 3 studies, 20% to more than 40% of patients who completed the screening instrument were not assessed for main outcomes.^{2,7,9} In the fourth study, the number of patients lost to follow-up was unclear.⁶⁶ One study only enrolled patients on chronic opioids,⁹ two appeared to enroll patients starting on opioids,^{2,66} and the fourth enrolled a mixed population.⁷ Only one study described baseline severity of pain (average pain 6 on a 0 to 10 scale),⁹ and none attempted to control or adjust for demographic or treatment factors (such as dose or type or opioid prescribed).

Two higher-quality studies evaluated the Screener and Opioid Assessment for Patients with Pain (SOAPP) Version 1 instrument.^{2,7} The first study derived the 14-item, self-administered SOAPP Version 1 (each scored on a 0 to 4 categorical scale, maximum score 56) from 24 original items and evaluated the diagnostic test characteristics of the final instrument in a mixed population of patients on chronic opioids or being considered for therapy (proportion on chronic opioids not reported).⁷ It found a cut-off score of ≥ 7 to be optimal, with a sensitivity of 0.91 (95% CI, 0.78 to 0.98) and specificity of 0.69 (95% CI, 0.54 to 0.81) for identifying aberrant drug-related behaviors after 6 months, based on a questionnaire, staff assessment, and urine toxicology results (PLR, 2.90 [95% CI, 1.91 to 4.39]; NLR, 0.13 [95% CI, 0.05 to 0.34]; and DOR, 21.9 [95% CI, 6.89 to 68.5]).⁷ In a second study, a score ≥ 8 on the previously derived SOAPP Version 1 instrument was associated with a sensitivity and specificity of 0.68 (95% CI, 0.52 to 0.81) and 0.38 (95% CI, 0.29 to 0.49), respectively (PLR, 1.11 [95% CI, 0.86 to 1.43]; NLR, 0.83 [95% CI, 0.50 to 1.36]; and DOR, 1.34 [95% CI, 0.64 to

Table 1. Prospective Studies of Use of Screening Instruments to Predict the Risk of Aberrant Drug-Related Behaviors

<i>AUTHOR, YEAR</i>	<i>NO. OF PATIENTS</i>								
<i>INSTRUMENT EVALUATED</i>	<i>DURATION OF FOLLOW-UP</i>	<i>DEFINITION OF ABERRANT DRUG-RELATED BEHAVIORS</i>	<i>SENSITIVITY</i>	<i>SPECIFICITY</i>	<i>POSITIVE LIKELIHOOD RATIO</i>	<i>NEGATIVE LIKELIHOOD RATIO</i>	<i>DIAGNOSTIC ODDS RATIO</i>	<i>OTHER RESULTS</i>	<i>QUALITY*</i>
Akbik, 2006 ² Screener and Opioid Assessment for Patients with Pain (SOAPP) Version 1 Self-administered, 14 items	N = 397 (155 had urine toxicology results) Duration unclear Patients not on opioids	Urine toxicology screen showing illicit substances and/or unprescribed opioids	0.68 (95% CI, 0.52 to 0.81) for SOAPP Version 1 score \geq 8	0.39 (95% CI, 0.29 to 0.49) for SOAPP Version 1 score \geq 8	1.11 (95% CI, 0.86 to 1.43) for SOAPP Version 1 score \geq 8	0.83 (95% CI, 0.50 to 1.36) for SOAPP Version 1 score \geq 8	1.34 (95% CI, 0.64 to 2.84) for SOAPP Version 1 score \geq 8	SOAPP Version 1 score \geq 8 vs \leq 8 Urine toxicology screen available and abnormal: 30/89 (34%) vs 14/51 (28%), $P < .05$	5/9
Butler, 2004 ⁷ Screener and Opioid Assessment for Patients with Pain (SOAPP) Version 1 Self-administered, 14 items	N = 175 (95 completed 6-month follow-up) 6 months Mixed population	Prescription Drug Use Questionnaire score \geq 11 (of 42) and/or staff assessment of serious drug behavior by 2 or 3 staff members and/or urine toxicology sample with unexpected medications, absence of prescribed medications, and/or illicit substances	0.91 (95% CI, 0.78 to 0.98) for SOAPP Version 1 score \geq 7 0.86 (95% CI, 0.73 to 0.95) for SOAPP Version 1 score \geq 8	0.69 (95% CI, 0.54 to 0.81) for SOAPP Version 1 score \geq 7 0.72 (95% CI, 0.58 to 0.84) for SOAPP Version 1 score \geq 8	2.90 (95% CI, 1.91 to 4.39) for SOAPP Version 1 score \geq 7 3.15 (95% CI, 1.98 to 4.99) for SOAPP Version 1 score \geq 8	0.13 (95% CI, 0.05 to 0.34) for SOAPP Version 1 score \geq 7 0.19 (95% CI, 0.09 to 0.40) for SOAPP Version 1 score \geq 8	21.9 (95% CI, 6.89 to 68.5) for SOAPP Version 1 score \geq 7 16.7 (95% CI, 5.91 to 47.2) for SOAPP Version 1 score \geq 7	Area under receiver operating curve 0.88 (95% CI, 0.81 to 0.95)	5/9

Table 1. Continued

<i>AUTHOR, YEAR</i>	<i>NO. OF PATIENTS</i>								
<i>INSTRUMENT EVALUATED</i>	<i>DURATION OF FOLLOW-UP</i>	<i>DEFINITION OF ABERRANT DRUG-RELATED BEHAVIORS</i>	<i>SENSITIVITY</i>	<i>SPECIFICITY</i>	<i>POSITIVE LIKELIHOOD RATIO</i>	<i>NEGATIVE LIKELIHOOD RATIO</i>	<i>DIAGNOSTIC ODDS RATIO</i>	<i>OTHER RESULTS</i>	<i>QUALITY*</i>
Butler, 2008 ⁹	N = 283 (223 completed 5 month follow-up)	Positive result on the Aberrant Drug Behavior Index: Score on the 42-item Prescription Drug Use Questionnaire of >11, or 2 or more positive results on the 11-item Prescription Opioid Therapy Questionnaire plus an abnormal urine toxicology result (illicit drug or non-prescribed opioid)	0.80 (95% CI, 0.70 to 0.89) for SOAPP-R score ≥ 18	0.68 (95% CI, 0.60 to 0.75) for SOAPP-R score ≥ 18	2.50 (95% CI, 1.93 to 3.24) for SOAPP-R score ≥ 18	0.29 (95% CI, 0.18 to 0.46) for SOAPP-R score ≥ 18	8.71 (95% CI, 4.51 to 16.8)	Area under receiver operating curve: 0.81 (95% CI, 0.75 to 0.87)	6/9
Revised Screener and Opioid Assessment for Patients with Pain (SOAPP-R) Self-administered, 24 items	5 months All patients on opioids								
Webster, 2005 ⁶⁶	N = 185 12 months All patients on opioids	Not defined; 23 different aberrant behaviors reported. Methods for identifying behaviors also not reported.	Not applicable (not dichotomous)	Not applicable (not dichotomous)	High risk (score ≥ 8): 14.3 (95% CI, 5.35 to 38.4) Moderate risk (score 4 to 7): 0.57 (95% CI, 0.44 to 0.74) Low risk (score 0 to 3): 0.08 (95% CI, 0.01 to 0.62)	Not applicable (not dichotomous)	Not applicable (not dichotomous)	Proportion with one or more aberrant behaviors, according to classification using ORT score: Low risk: 6% (1/18) Moderate risk: 28% (35/123) High risk: 91% (40/44)	4/9

* See Table 2 for complete quality criteria scores.

Table 2. Quality* of Prospective Studies of the Use of Screening Instruments to Predict the Risk of Aberrant Drug-Related Behaviors

AUTHOR/YEAR	EVALUATES POPULATION OTHER THAN THE ONE USED TO DERIVE THE INSTRUMENT	DESCRIBES SEVERITY OF SYMPTOMS, OPIOID DOSE/DURATION, AND UNDERLYING CONDITIONS	ADEQUATE DESCRIPTION OF SCREENING INSTRUMENT	ADEQUATE DESCRIPTION OF METHOD FOR IDENTIFYING ABERRANT DRUG-RELATED BEHAVIORS	APPROPRIATE CRITERIA INCLUDED IN SCREENING INSTRUMENT	ADEQUATE DESCRIPTION OF SCREENING INSTRUMENT	APPROPRIATE CRITERIA USED TO IDENTIFY ABERRANT DRUG-RELATED BEHAVIORS	ABERRANT DRUG-RELATED BEHAVIORS ASSESSED IN ALL ENROLLEES	BLINDED ASSESSMENT OF ABERRANT DRUG-RELATED BEHAVIORS	SCORE (MAX 9)
Akbik, 2006 ²	Yes	No	Yes	Yes	Yes	Yes	No	No	Don't know	5/9
Butler, 2004 ⁷	No	No	Yes	Yes	Yes	Yes	Yes	No	Don't know	5/9
Butler, 2008 ⁹	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Don't know	6/9
Webster, 2005 ⁶⁶	Yes	No	Yes	No	Yes	No	Don't know	Don't know	Don't know	4/9

* See Appendix 2 for description of quality criteria

2.84)].² However, these results are difficult to interpret because aberrant drug-related behaviors were identified solely on the basis of urine drug screen results, urine drug screens were not obtained in most patients, and duration of follow-up was unclear.

A third study derived the 24-item, self-administered revised SOAPP (SOAPP-R) from 97 original items and evaluated the diagnostic test characteristics of the final instrument in patients already prescribed chronic opioid therapy (average duration, 6 years).⁹ The SOAPP-R was designed in part to include less transparent items on drug abuse compared with the SOAPP Version 1, to potentially reduce the likelihood of overt patient deception. At a cut-off score of ≥ 18 (each item scored from 0 to 4, maximum score 96), sensitivity was 0.80 (95% CI, 0.70 to 0.89) and specificity was 0.68 (95% CI, 0.60 to 0.75) for identification of any aberrant drug-related behavior based on results of 2 questionnaires and a urine drug screen (PLR, 2.50 [95% CI 1.93 to 3.24]; NLR, 0.29 [95% CI, 0.18 to 0.46]; and DOR, 8.71 [95% CI, 4.51 to 16.8]). The area under the receiver operating curve (0.81; 95% CI, 0.75 to 0.87) was similar to results for the SOAPP Version 1 (0.88; 95% CI, 0.81 to 0.95),⁷ but may not be directly comparable due to use of different criteria to define aberrant drug-related behaviors and differences in the proportion of patients on chronic opioid therapy at enrollment.

A fourth, lower-quality study evaluated the self-administered Opioid Risk Tool (ORT), which consists of 10 items (maximum score, 26).⁶⁶ Items in this instrument were chosen and weighted before evaluation of diagnostic test characteristics, and cut-off scores for different risk categories appeared to be selected on an a priori basis. Aberrant drug-related behaviors were identified in 6% (1/18) of patients categorized as low risk (score, 0 to 3), compared with 28% (35/123) of patients categorized as moderate risk (score, 4 to 7) and 91% (41/44) of those categorized as high risk (score ≥ 8) after 12 months. A high-risk score strongly increased the likelihood of subsequent aberrant drug-related behaviors (PLR, 14.3 [95% CI, 5.35 to 38.4]), a moderate risk score had little effect (PLR, 0.57 [95% CI, 0.44 to 0.74]), and a low risk score strongly decreased the likelihood (PLR, 0.08 [95% CI, 0.01 to 0.62]). An important shortcoming of this study is that it did not use standardized methods (eg, questionnaires or urine drug screening) to identify aberrant drug-related behaviors, and aberrant behaviors were not clearly predefined.

No study evaluated the utility of formal risk stratification instruments compared with informal clinical assessments alone, or compared one screening instrument with another.

Accuracy of Screening Instruments to Identify Current Aberrant Drug-Related Behaviors in Patients Prescribed Opioids

We identified 9 studies (N = 1,530) that evaluated accuracy of screening instruments to identify aberrant drug-related behaviors in patients prescribed long-term opioid therapy for CNCP (Tables 3 and 4).^{1,4,8,15,27,43,47,64,68} Although 5 studies met our threshold for higher quality,^{1,8,15,47,64} all studies had methodological shortcomings.

No study described whether investigators assessing the reference standard for aberrant drug-related behaviors were blinded to results of the screening instrument. In addition, methods for identifying aberrant drug-related behaviors varied across studies and did not distinguish well between new and preexisting aberrant drug-related behaviors (particularly substance abuse or illicit drug use) or between less and more serious behaviors. In 2 studies, methods for identifying drug-related behaviors were not well described.^{4,43} Five studies incorporated urine toxicology results with illicit drugs or unprescribed opioids into definitions of aberrant drug-related behaviors.^{4,8,43,47,64} All of the studies evaluated different screening instruments, with the exception of 2 studies that assessed the Pain Medication Questionnaire.^{1,27} Of the 8 instruments evaluated, 2 were self-administered,^{1,8} 4 interviewer-administered,^{15,47,64,68} and in 2 the method of administration was unclear.^{4,43} The instruments varied in complexity, with the number of assessment items ranging from 3⁴⁷ to 42.¹⁵ One screening instrument focused on history of alcohol or substance abuse⁴⁷ and one focused on psychosocial factors.⁶⁴ The others assessed multiple domains including coping strategies, pain medication behaviors, abuse of substances other than prescribed opioids, and/or psychosocial factors.^{1,4,8,15,27,43,47,64,68} One instrument⁶⁴ was based on a subset of psychiatric items included in another screening instrument (the Prescription Drug Use Questionnaire¹⁵). Only one study reported pain scores (average, 6 on a 0 to 10 scale).⁸ No study reported doses of opioids prescribed and none adjusted or controlled for demographic and intervention variables.

One higher-quality study derived the 17-item, self-administered Current Opioid Misuse Measure (COMM) from 40 original items and evaluated the diagnostic test characteristics of the final instrument.⁸ It found an area under-the-receiver operating curve of 0.81 (95% CI, 0.74 to 0.86). Based on an optimal cut-off score of ≥ 10 (of a maximum possible score 68), the sensitivity and specificity were 0.74 (95% CI, 0.63 to 0.84) and 0.73 (95% CI, 0.65 to 0.80), respectively, with a PLR of 2.77 (95% CI, 2.06 to 3.72), NLR of 0.35 (95% CI, 0.24 to 0.52), and DOR of 7.90 (95% CI, 4.25 to 14.7).

A second, lower-quality study found the interviewer-administered Addiction Behavior Checklist (ABC, 20 items) associated with a sensitivity of 0.88 and specificity of 0.86 (PLR, 6.29; NLR, 0.14) at the optimal cut-off score of ≥ 3 of 20 (confidence intervals not calculable).⁶⁸ Items included in the ABC were selected before evaluation in the study. The interpretation of this study is challenging, however, because the presence of aberrant drug-related behaviors was defined by the response of the treating pain physician to a single question of uncertain reliability or validity: "Do you think patient is using medications appropriately?"

In 4 other studies, the screening instrument showed poor diagnostic accuracy^{47,64} or results are difficult to interpret due to serious methodological shortcomings.^{4,43} One higher-quality study found that positive responses to at least 2 of 3 preselected questions had only modest sensitivity and specificity for various behaviors associated with opioid misuse or abuse, resulting in small or trivial likelihood ratios (Table 3).⁴⁷ Another higher-quality

study found that the presence of psychiatric comorbidity (defined as 2 or more positive responses on the 5 psychiatric items of the previously developed Prescription Drug Use Questionnaire) was associated with a sensitivity of 0.74 (95% CI, 0.63 to 0.82) and a specificity of 0.57 (95% CI, 0.49 to 0.65) for positive findings on the Drug Misuse Index (which combines results from the SOAPP, COMM, other risk assessment instruments, and urine toxicology results).⁶⁴ The PLR was 1.72 (95% CI, 1.37 to 2.17) and the NLR was 0.46 (95% CI, 0.31 to 0.67). One study found a 6-item instrument associated with small positive and negative likelihood ratios for aberrant drug-related behaviors,⁴ and another found a 4-item instrument associated with a large PLR and small NLR (Table 3).⁴³ However, both of these studies used a retrospective case-control design, were rated lower-quality, and derived and validated the instrument in the same population.

In 3 studies, higher scores on various screening instruments generally correlated with presence of variably defined aberrant drug-related behaviors, but sensitivity, specificity, and other standard measures of diagnostic accuracy were not reported and could not be calculated (Table 3).^{1,15,27} No study evaluated the utility of formal monitoring instruments compared with informal clinical assessments alone, or compared one screening instrument to another.

Effectiveness of Risk Assessment and Monitoring for Improving Clinical Outcomes or Reducing Risk of Aberrant Drug Behaviors

We identified no studies meeting the prespecified inclusion criteria.

Accuracy of Urine Drug Screening to Detect Illicit Drug Use or the Presence or Absence of Prescribed and Nonprescribed Opioids

Data on the accuracy of urine drug screening compared with a reference standard are extremely limited. One retrospective study (N = 226) found that analyses of urine drug samples (performed with gas chromatography-mass spectrometry) were associated with sensitivities of 86% for cannabinoids, 76% for benzodiazepines, and 88% for opioid use compared with patient self-report during psychiatric examination.¹⁷ However, interpretation of these results is challenging because it is not clear if the investigators who obtained the patients' self-reports were blinded to the results of urine drug screening, or when illicit drug or opioid use last occurred relative to timing of urine sampling.

Effectiveness of Urine Drug Screening or Adherence Monitoring to Reduce Aberrant Drug-Related Behaviors

One observational study of 500 consecutive patients receiving opioids for CNCP reported marijuana in 11% of samples, cocaine in 5%, and methamphetamines or amphetamines in 2% in a setting in which all patients agreed to random urine drug screening.⁴¹ Compared with an earlier cohort in the same setting, the prevalence

Table 3. Studies on Accuracy of Screening Instruments to Identify Aberrant Drug-Related Behaviors in Patients Prescribed Opioids

<i>AUTHOR, YEAR</i>											
<i>INSTRUMENT EVALUATED</i>	<i>NO. OF PATIENTS</i>	<i>DEFINITION OF ABERRANT DRUG-RELATED BEHAVIORS</i>			<i>SENSITIVITY</i>	<i>SPECIFICITY</i>	<i>POSITIVE LIKELIHOOD RATIO</i>	<i>NEGATIVE LIKELIHOOD RATIO</i>	<i>DIAGNOSTIC ODDS RATIO</i>	<i>OTHER RESULTS</i>	<i>QUALITY*</i>
Adams, 2004 ¹	111 patients on opioids	Physician Risk Assessment tool used to identify opioid misuse; based on a set of six dimensions, each rated on a 5-point Likert scale	Not calculable	Not calculable	Not calculable	Not calculable	Not calculable	Not calculable	Not calculable	Known opioid misuse (n = 12) versus no known history of opioid misuse (matched sample)	6/9
Pain Medication Questionnaire (PMQ)	Cross-sectional									(matched sample)	
Self-administered, 26 items										Mean PMQ score: 33.9 vs 25.5 (P = .045 based on 1-sided t test)	
Atluri, 2004 ⁴	107 cases, 103 controls	Inappropriate opioid use included inappropriate urine drug screen (not defined), intentional 'doctor shopping', alteration of opioid prescription to obtain more opioids, criminal activity involving prescription opioids (89% inappropriate urine drug screen)	0.77 (95% CI, 0.68 to 0.84), for score ≥ 4	0.84 (95% CI, 0.76 to 0.91) for score ≥ 4	4.93 (95% CI, 3.11 to 7.83) for score ≥ 4	0.28 (95% CI, 0.19 to 0.39) for score ≥ 4	17.8 (95% CI, 8.93 to 35.6) for score ≥ 4		Risk of inappropriate opioid use	Score ≥ 4 (of 6) positive items (high risk) vs score < 4 (low risk): OR, 16.6 (95% CI, 8.3 to 33)	2/9
Method of administration unclear, 6 items	Case-control										
Butler, 2007 ⁸	227	Aberrant Drug Behavior Index positive if Patient Drug Use Questionnaire score > 11 or urine toxicology screen	0.77 (95% CI, 0.66 to 0.86) for COMM score ≥ 9	0.66 (95% CI, 0.58 to 0.73) for COMM score ≥ 9	2.25 (95% CI, 1.74 to 2.90) for COMM score ≥ 9	0.35 (95% CI, 0.23 to 0.5) for COMM score ≥ 9	6.41 (95% CI, 3.44 to 11.9) for COMM score ≥ 9		Area under receiver operating curve: 0.81 (95% CI, 0.74 to 0.86)		5/9
Current Opioid Misuse Measure (COMM)	Cross-sectional (for assessing diagnostic accuracy)										
Self-administered, 17 items			0.74 (95% CI, 0.63 to 0.84) for COMM score ≥ 10	0.73 (95% CI, 0.65 to 0.80) for COMM score ≥ 10	2.77 (95% CI, 2.06 to 3.72) for COMM score ≥ 10	0.35 (95% CI, 0.24 to 0.52) for COMM score ≥ 10	7.90 (95% CI, 4.25 to 14.7) for COMM score ≥ 10				

Table 3. Continued

<i>AUTHOR, YEAR</i>		<i>INSTRUMENT EVALUATED</i>		<i>NO. OF PATIENTS</i>		<i>DEFINITION OF ABERRANT DRUG-RELATED BEHAVIORS</i>		<i>SENSITIVITY</i>	<i>SPECIFICITY</i>	<i>POSITIVE LIKELIHOOD RATIO</i>	<i>NEGATIVE LIKELIHOOD RATIO</i>	<i>DIAGNOSTIC ODDS RATIO</i>	<i>OTHER RESULTS</i>	<i>QUALITY*</i>
Compton, 1998 ¹⁵	52	Prescription Drug Use Questionnaire (PDUQ)	Cross-sectional	positive (presence of illicit drug or non-prescribed opioid) and Prescription Opioid Therapy Questionnaire score ≥ 3	Not calculable	Not calculable	Not calculable	Not calculable	Not calculable	Not calculable	Not calculable	Score (range for number of positive items) on 40-item Prescription Drug Use Questionnaire ($P < .0005$ on ANOVA)	7/9	
Interviewer-administered, 40 items		American Society of Addiction Medicine criteria for substance abuse and substance dependence as evaluated by a single addiction medicine specialist										Nonaddicted: 6 to 15 Substance-abusing: 11 to 25 Substance-dependent: 15 to 28		
Holmes, 2006 ²⁷	271	Pain Medication Questionnaire (PMQ)	Prospective cohort	Individuals with a known history of substance abuse (alcohol, prescription drugs, illicit drugs) based on self-admission, referring physician report, or initial psychologist evaluation; Physician Risk Assessment score; requests for early prescription refills	Not calculable	Not calculable	Not calculable	Not calculable	Not calculable	Not calculable	Not calculable	Known history of substance abuse (n = 68) versus no known history of substance abuse (n = 68)	3/9	
Self-administered, 26 items												Pain Medication Questionnaire score (mean): 28.8 vs 23.9 ($P = .01$) High vs low Pain Medication Questionnaire score		

Table 3. Continued

<i>AUTHOR, YEAR</i>											
<i>INSTRUMENT EVALUATED</i>	<i>NO. OF PATIENTS</i>	<i>DEFINITION OF ABERRANT DRUG-RELATED BEHAVIORS</i>			<i>SENSITIVITY</i>	<i>SPECIFICITY</i>	<i>POSITIVE LIKELIHOOD RATIO</i>	<i>NEGATIVE LIKELIHOOD RATIO</i>	<i>DIAGNOSTIC ODDS RATIO</i>	<i>OTHER RESULTS</i>	<i>QUALITY*</i>
Manchikanti, 2004 ⁴³	150	Controlled substance abuse defined as: Misuse of controlled substances in a clinical setting, including obtaining controlled substances from other physicians or other identifiable sources, dose escalations with inappropriate use, and/or violation of controlled substance agreement	0.49 (95% CI, 0.37 to 0.60) for score ≥ 2	1.00 (95% CI, 0.95 to 1.0) for score ≥ 2	69.2 (95% CI, 4.33 to 1106) for score ≥ 2	0.52 (95% CI, 0.42 to 0.64) for score ≥ 2	134 (95% CI, 8.04 to 2241) for score ≥ 2	Request for early refills: 61.5% vs 33.3% ($P = .02$); OR, 3.2 (95% CI, 1.21 to 8.44)	3/9		
Based on Atluri et al ⁴	Case-control	Illicit drug abuse not defined						Total score 0 or 1 out of 8 items: 100% vs 94% vs 20% vs 23% (P values $>.05$ for all comparisons)			
Method of administration unclear, 4 items								Total score ≥ 2 out of 8: 0% vs 6% vs 80% vs 77% ($P < .05$ for 6% vs 0% and for 80% or 77% vs 0% or 6%)			
Michna, 2004 ⁴⁷	145	A: unanticipated positive results in urine toxicology tests	2-3 positive responses	2-3 positive responses	2-3 positive responses	2-3 positive responses	2-3 positive responses	High risk (2-3 positive responses) vs low risk (0-1 positive responses)	7/9		

Table 3. Continued

<i>AUTHOR, YEAR</i>		<i>INSTRUMENT EVALUATED</i>		<i>NO. OF PATIENTS</i>		<i>DEFINITION OF ABERRANT DRUG-RELATED BEHAVIORS</i>		<i>SENSITIVITY</i>	<i>SPECIFICITY</i>	<i>POSITIVE LIKELIHOOD RATIO</i>	<i>NEGATIVE LIKELIHOOD RATIO</i>	<i>DIAGNOSTIC ODDS RATIO</i>	<i>OTHER RESULTS</i>	<i>QUALITY*</i>
<i>METHOD OF ADMINISTRATION</i>	<i>TYPE OF STUDY</i>													
Abuse questions Items (3 questions) Interviewer- administered, 3 items	Cross-sectional	B: episodes of lost or stolen prescription												
		C: multiple unsanctioned escalations in dose		A: 0.53 (95% CI, 0.35 to 0.71)	A: 0.75 (95% CI, 0.66 to 0.83)	A: 2.14 (95% CI, 1.36 to 3.39)	A: 0.62 (95% CI, 0.42 to 0.92)	A: 3.44 (95% CI, 1.54 to 7.71)	A: 38% vs 15%, P < .05					
		D: frequent unscheduled pain center or emergency room visits		B: 0.47 (95% CI, 0.29 to 0.65)	B: 0.74 (95% CI, 0.64 to 0.81)	B: 1.77 (95% CI, 1.09 to 2.85)	B: 0.72 (95% CI, 0.51 to 1.02)	B: 2.44 (95% CI, 1.10 to 5.44)	B: 33% vs 17%, P < .05					
		E: concern expressed by a significant other about the patient's use of opioids		C: 0.40 (95% CI, 0.25 to 0.58)	C: 0.72 (95% CI, 0.63 to 0.80)	C: 1.46 (95% CI, 0.89 to 2.39)	C: 0.82 (95% CI, 0.62 to 1.10)	C: 1.77 (95% CI, 0.82 to 3.84)	C: 33% vs 22%, P > .05					
		F: excessive phone calls		D: 0.40 (95% CI, 0.19 to 0.64)	D: 0.70 (95% CI, 0.62 to 0.78)	D: 1.35 (95% CI, 0.74 to 2.46)	D: 0.85 (95% CI, 0.58 to 1.24)	D: 1.59 (95% CI, 0.61 to 4.11)	D: 18% vs 12%, P > 0.05					
				E: 0.44 (95% CI, 0.22 to 0.69)	E: 0.71 (95% CI, 0.62 to 0.79)	E: 1.53 (95% CI, 0.85 to 2.73)	E: 0.78 (95% CI, 0.51 to 1.20)	E: 1.95 (95% CI, 0.73 to 5.19)	E: 18% vs 10%, P > .05					
Wasan, 2007 ⁶⁴ Psychiatric items from the Prescription Drug Use Questionnaire (PDUQ) Interviewer- administered, 5 items	Prospective cohort	Drug Misuse Index: Misuse or abuse defined as positive scores on the self-reported Screener and Opioid Assessment for Pain Patients and the Current Medication Misuse Measure; or positive scores on the urine toxicology screen (presence of illicit		F: 0.36 (95% CI, 0.11 to 0.69)	F: 0.69 (95% CI, 0.61 to 0.77)	F: 1.19 (95% CI, 0.52 to 2.70)	F: 0.92 (95% CI, 0.58 to 1.45)	F: 1.30 (95% CI, 0.38 to 4.41)	F: 9% vs 7%, P > .05					
				0.74 (95% CI, 0.63 to 0.83)	0.57 (95% CI, 0.48 to 0.66)	1.72 (95% CI, 1.37 to 2.17)	0.46 (95% CI, 0.31 to 0.67)	3.77 (95% CI, 2.11 to 6.72)	High psychiatric comorbidity (≥ 2 positive items of 5 psychiatric items on the PDUQ) vs low psychiatric comorbidity (<2 positive items)					
											Drug Misuse Index positive: 52% vs 22% (P < .001)			

Table 3. Continued

AUTHOR, YEAR	INSTRUMENT EVALUATED	NO. OF PATIENTS	DEFINITION OF ABERRANT DRUG-RELATED BEHAVIORS	SENSITIVITY	SPECIFICITY	POSITIVE LIKELIHOOD RATIO	NEGATIVE LIKELIHOOD RATIO	DIAGNOSTIC ODDS RATIO	OTHER RESULTS	QUALITY*
Wu, 2006 ⁶⁸	Addiction Behaviors Checklist (ABC)	136	Interviewer's global clinical judgment (yes or no to "Do you think patient is using medications appropriately?")	0.88 for ABC score ≥ 3 (confidence intervals not calculable)	0.86 for ABC score ≥ 3 (confidence intervals not calculable)	Not calculable	Not calculable	Not calculable	None	4/9

* See Table 4 for complete quality criteria scores.

of marijuana in urine was lower (11% vs 18%, *P* value not reported), but the prevalence of other illicit drugs was similar. A second study that appeared to be conducted in the same patient cohort found that institution of adherence monitoring (signed controlled substance agreement, periodic monitoring, periodic drug testing, pill counts, and education when necessary) was associated with a rate of controlled substance abuse of 9% (defined as receiving controlled substances from any place or source other than the prescribing physician), compared with 18% in an earlier cohort.⁴⁰ Results of both of these studies are difficult to interpret because they used historical controls, did not report statistical significance of differences in rates of aberrant behaviors, did not describe monitoring protocols well, and did not describe how the monitoring protocols (and other factors) differed compared with the historical cohort. We identified no other studies that met the prespecified inclusion criteria.

Accuracy or Effectiveness of Pill Counts, Limited Prescriptions, Monitoring Blood Levels, Prescription Drug Monitoring to Reduce Aberrant Drug-Related Behaviors

We identified no studies that met the prespecified inclusion criteria.

Effectiveness of Monitoring at Different Intervals on Clinical Outcomes

We identified no studies that met the prespecified inclusion criteria.

Effectiveness of Outcomes Assessment Tools on Clinical Outcomes

The Pain Assessment and Documentation Tool (PADT) was developed to assist clinicians in the evaluation and documentation of outcomes related to use of opioids in 4 key domains (analgesia, activities of daily living, adverse events, and aberrant drug-related behaviors).^{52,53} However, no study has evaluated the effect that using the PADT or any other outcomes assessment tool has on clinical outcomes.

Discussion

Based on the findings from this systematic review of the literature, only limited evidence exists to determine optimal methods for prediction and identification of aberrant drug-related behaviors in patients with CNCP who are being considered for or are being prescribed chronic opioid therapy.

There is fair-to-poor evidence from 2 derivation studies that high scores on the SOAPP Version 1⁷ and the SOAPP-R⁹ instruments weakly increase the likelihood for any future aberrant drug-related behavior (PLRs, 2.90 [95% CI, 1.91 to 4.39] and 2.50 [95% CI, 1.93 to 3.24], respectively). Low scores on the SOAPP Version 1 moderately decrease the likelihood of aberrant drug-related behaviors (NLR, 0.13 [95% CI, 0.05 to 0.34]),⁷ and low scores on the SOAPP-R weakly decrease the likelihood (NLR, 0.29

Table 4. Quality* of Studies on Accuracy of Screening Instruments to Identify Aberrant Drug-Related Behaviors in Patients Prescribed Opioids

AUTHOR/YEAR	EVALUATES POPULATION OTHER THAN THE ONE USED TO DERIVE THE INSTRUMENT	DESCRIBES SEVERITY OF SYMPTOMS, OPIOID DOSE/DURATION, AND UNDERLYING CONDITIONS	ADEQUATE DESCRIPTION OF SCREENING INSTRUMENT	ADEQUATE DESCRIPTION OF METHOD FOR IDENTIFYING ABERRANT DRUG-RELATED BEHAVIORS	APPROPRIATE CRITERIA INCLUDED IN SCREENING INSTRUMENT	APPROPRIATE CRITERIA USED TO IDENTIFY ABERRANT DRUG-RELATED BEHAVIORS	ABERRANT drug-related behaviors assessed in all enrollees	BLINDED ASSESSMENT OF ABERRANT DRUG-RELATED BEHAVIORS	SCORE (MAX 9)
Adams, 2004 ¹	No	Yes	Yes	Yes	Yes	Yes	Yes	Do not know	6/9
Atluri, 2004 ⁴	No	No	Yes	No	Yes	Do not know	Do not know	Do not know	2/9
Butler, 2007 ⁸	No	Yes	Yes	Yes	Yes	Yes	Do not know	Do not know	5/9
Compton, 1998 ¹⁵	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Do not know	7/9
Holmes, 2006 ²⁷	Yes	Yes	Yes	No	Yes	No	Do not know	Do not know	4/9
Manchikanti, 2004 ⁴³	No	Yes	No	No	Yes	Do not know	Yes	Do not know	3/9
Michna, 2004 ⁴⁷	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Do not know	7/9
Wasan, 2007 ⁶⁴	Yes	Yes	Yes	Yes	Yes	Yes	No	Do not know	6/9
Wu, 2006 ⁶⁸	No	Yes	Yes	Yes	Yes	No	Do not know	Do not know	4/9

* See Appendix 2 for description of quality criteria.

[95% CI, 0.18 to 0.46]).⁹ Because the confidence intervals overlap, it is uncertain that the revised version improves diagnostic accuracy. Another study found the SOAPP Version 1 to be poorly predictive, but it is difficult to interpret due to methodological shortcomings.² Categorization of patients as high or low risk using the ORT instrument strongly affects the likelihood of future aberrant drug-related behaviors (PLR, 14.3 [95% CI, 5.35 to 38.4] and 0.08 [95% CI, 0.01 to 0.62], respectively).⁶⁶ However, evidence on the ORT is limited to one lower-quality study and requires verification.

Limited evidence also exists to guide decisions regarding optimal monitoring strategies. There is fair-to-poor evidence from one derivation study that scores on the COMM weakly predict absence or presence of any current aberrant drug-related behavior (PLR, 2.77 [95% CI, 2.06 to 3.72]) and NLR, 0.35 [95% CI, 0.24 to 0.52]).⁸ Studies of other monitoring instruments either did not report diagnostic accuracy,^{1,15,27} found diagnostic accuracy to be poor,^{47,64} or are difficult to interpret due to important methodological shortcomings.^{4,43,68} For example, although a study of the ABC instrument appeared to show superior test performance compared with the COMM, it used as its reference standard for aberrant drug-related behaviors a subjective question of uncertain validity and reliability (“Do you think patient is using medications appropriately?”).⁶⁸ In addition, if this single question were truly a valid reference standard for aberrant drug-related behaviors, a more complex screening instrument would not be necessary.

Several aspects of studies reviewed made it difficult to interpret results. First, all studies had methodological shortcomings, decreasing confidence in their results. For example, higher-quality studies of the SOAPP Version 1,⁷ SOAPP-R,⁹ and COMM⁸ derived and validated the instruments in the same population. Estimates of diagnostic accuracy from such derivation studies can be inflated using this methodology because the most predictive items in the derivation population are retrospectively selected to be included in the instrument and tested for validity in the same population.⁴⁵ The same items may not be equally predictive when applied prospectively to other populations. Similarly, threshold values for classifying results of the screening instrument as positive or negative are selected on a post hoc basis to maximize sensitivity and specificity in a derivation population, but may not perform as well when applied prospectively to other populations. Verification of diagnostic test performance of previously derived instruments in other populations and settings using prespecified thresholds is needed. Second, use of poorly standardized criteria to define aberrant drug-related behaviors is problematic, as it makes comparisons of results across studies difficult. In addition, because the methods used to define aberrant drug-related behaviors did not distinguish relatively less serious from more serious behaviors or identify the reasons for such behaviors, the clinical importance of their identification is unclear. Third, most studies were performed in pain clinic settings, and results may not be directly applicable to primary care or other settings. Furthermore, both of the higher-quality prospective

studies of risk stratification instruments included patients already prescribed opioids, which may limit their applicability to patients being considered for (but not yet prescribed) opioids.^{7,9} Finally, although self-administered instruments may be more efficient for clinicians, no evidence exists to compare the uptake, reliability, or accuracy of self-administered versus interviewer-administered or clinician-completed instruments.

Even if multiple higher-quality studies were to show that a risk prediction or monitoring instrument is highly accurate for predicting or identifying aberrant drug-related behaviors, it does not necessarily mean that it will improve clinical outcomes. The effects of using such an instrument depend not only on its diagnostic accuracy, but also on the seriousness of the behaviors identified; how well the behaviors correlate with actual drug abuse, addiction, or diversion; how applying the instrument influences clinical decision-making; and how those clinical decisions affect patient outcomes.⁵⁸ Studies showing that use of a risk prediction or monitoring instrument alters clinician behavior and improves patient outcomes would provide strong evidence to support its use. At this time, no such studies are available.

We identified no reliable data on the accuracy of urine drug screening, pill counts, or prescription drug monitoring programs to identify aberrant drug-related behaviors, or on effects of using such interventions on patient outcomes. No study evaluated effects of different monitoring intervals on patient outcomes, or on effects of different methods to assess and document outcomes.

Our systematic review has some potential limitations. We excluded non-English language studies, as well as unpublished studies and studies published only as abstracts. However, language restrictions do not necessarily lead to biased findings,⁴⁸ and we are not aware of non-English language or unpublished studies likely to change any of our main conclusions. In addition, the quality of unpublished studies is often difficult to assess due to incomplete reporting, and results can change between initial presentation and final journal publication.⁶² We also limited the scope of this article to risk prediction and monitoring as they pertain to aberrant drug-related behaviors. Evidence on other important components of a comprehensive benefit-to-harm evaluation such as assessing likelihood of therapeutic benefits, adverse effects, or opioid responsiveness (analgesia or symptom relief achievable with tolerable adverse effects) is reviewed elsewhere.¹³

A strength of our review is that we calculated unreported sensitivities, specificities, and likelihood ratios (as well as corresponding confidence intervals) when data were available to do so. This provides quantitative information with which to compare diagnostic test characteristics across studies, shows precision of the estimates, and facili-

tates evaluations of how the application of the instruments might influence clinical decision-making. For example, in a population with a pre-test prevalence for aberrant drug-related behaviors after starting opioid therapy of 3%,⁵⁵ the post-test probability after a high score on either the SOAPP Version 1 or the SOAPP-R would be 7% to 8% using likelihood ratio estimates.^{7,9} Low scores on either SOAPP instrument would decrease the post-test probability to below 1%. With the ORT instrument, categorization as high risk would increase the post-test probability to 31%, and categorization as low risk would decrease the post-test probability to 0.2%.⁶⁶ The clinical utility of risk prediction and monitoring instruments depends on whether shifts from pre- to post-test probabilities would cross thresholds likely to alter clinical decision-making, and will vary depending on the population.²⁹ In a higher-risk population with a 20% pre-test probability for aberrant drug-related behaviors, the post-test probability after a high score on either SOAPP instrument would be around 40%, and after a low score 3% to 7%.

Use of opioids for CNCP is steadily increasing.¹⁰ Clinicians are in need of high-quality evidence to help guide decisions regarding patient selection and monitoring during opioid therapy. Available evidence on prediction and identification of aberrant drug-related behaviors is limited by sparse data, presence of methodological shortcomings, and absence of evidence on effects of different assessment and monitoring methods on clinical outcomes. Future research should avoid the methodological shortcomings of previously published studies, use standardized definitions for clinically relevant aberrant drug-related behaviors, externally validate previously derived instruments, and evaluate how using these instruments affects patient outcomes. Other important research needs are to evaluate effects of different monitoring intervals on patient outcomes and to evaluate the accuracy and effectiveness of urine drug screens, pill counts, and prescription monitoring programs.

Acknowledgments

The authors thank Laurie Hoyt Huffman for reviewing literature and data abstraction; Rongwei Fu for performing statistical analyses; and Jayne Schablaske, Michelle Pappas, and Tracy Dana for administrative support with this manuscript.

Supplementary Data

Supplementary data accompanying this article is available online at www.jpain.org, www.sciencedirect.com, and at [doi:10.1016/j.jpain.2008.10.009](https://doi.org/10.1016/j.jpain.2008.10.009). The supplementary data include Appendices 1–4.

References

1. Adams LL, Gatchel RJ, Robinson RC, Polatin P, Gajraj N, Deschner M, Noe C: Development of a self-report screening instrument for assessing potential opioid medication misuse in chronic pain patients. *J Pain Symptom Manage* 27: 440-459, 2004

2. Akbik H, Butler SF, Budman SH, Fernandez K, Katz NP, Jamison RN: Validation and clinical application of the screener and opioid assessment for patients with pain (SOAPP). *J Pain Symptom Manage* 32:287-293, 2006

3. Allan L, Richarz U, Simpson K, Slappendel R: Transdermal fentanyl versus sustained release oral morphine in strong-

- opioid naive patients with chronic low back pain. *Spine* 30: 2484-2490, 2005
4. Atluri SL, Sudarshan G: Development of a screening tool to detect the risk of inappropriate prescription opioid use in patients with chronic pain. *Pain Physician* 7:333-338, 2004
 5. Belgrade MJ, Schamber CD, Lindgren BR: The DIRE score: Predicting outcomes of opioid prescribing for chronic pain. *J Pain* 7:671-681, 2006
 6. British Pain Society: Recommendations for the appropriate use of opioids for persistent non-cancer pain (the first revised edition). London: The British Pain Society, March 2005
 7. Butler SF, Budman SH, Fernandez K, Jamison RN: Validation of a screener and opioid assessment measure for patients with chronic pain. *Pain* 112:65-75, 2004
 8. Butler SF, Budman SH, Fernandez KC, Houle B, Benoit C, Katz N, Jamison RN: Development and validation of the current opioid misuse measure. *Pain* 130:144-156, 2007
 9. 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* 9: 360-372, 2008
 10. Caudill-Slosberg MA, Schwartz LM, Woloshin S: Office visits and analgesic prescriptions for musculoskeletal pain in US: 1980 vs 2000. *Pain* 109:514-519, 2004
 11. Chabal C, Erjavec MK, Jacobson L, Mariano A, Chaney E: Clinical criteria, incidence, and predictors. *Clin J Pain* 13: 150-155, 1997
 12. Chou C, Fanciullo G, Fine P, Adler J, Ballantyne J, Davies P, Donovan M, Fishbain D, Foley K, Fudin J, Gilson A, Kelter A, Mauskop A, O'Connor P, Passik S, Pasternak G, Portenoy R, Rich B, Roberts R, Saper J, Todd K, Miaskowski C, for the American Pain Society-American Academy of Pain Medicine Opioids Guidelines Panel. Clinical guidelines for the use of chronic opioid therapy in chronic noncancer pain. *J Pain* 10:113-130, 2009
 13. Chou R, Huffman L. The use of opioids for chronic non-cancer pain: Evidence review. Glenview, IL: American Pain Society, 2009
 14. Coombs R: A new screening instrument for identifying potential opioid abusers in the management of chronic non-malignant pain with general medical practice. *Pain Res Manag* 1:155-162, 1996
 15. 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* 16:355-363, 1998
 16. Edlund MJ, Steffick D, Hudson T, Harris KM, Sullivan M: Risk factors for clinically recognized opioid abuse and dependence among veterans using opioids for chronic non-cancer pain. *Pain* 129:355-362, 2007
 17. Fishbain DA, Cutler RB, Rosomoff HL, Rosomoff RS: Validity of self-reported drug use in chronic pain patients. *Clin J Pain* 15:184-191, 1999
 18. Fishman SM, Wilsey B, Yang J, Reisfield GM, Bandman TB, Borsook D: Adherence monitoring and drug surveillance in chronic opioid therapy. *J Pain Sympt Manage* 20:293-307, 2000
 19. Fleming MF, Balousek SL, Klessig CL, Mundt MP, Brown DD: Substance use disorders in a primary care sample receiving daily opioid therapy. *J Pain* 8:573-582, 2007
 20. Friedman R, Li V, Mehrotra D: Treating pain patients at risk: Evaluation of a screening tool in opioid-treated pain patients with and without addiction. *Pain Med* 4:182-185, 2003
 21. Furlan AD, Sandoval JA, Mailis-Gagnon A, Tunks E: Opioids for chronic noncancer pain: A meta-analysis of effectiveness and side effects. *CMAJ* 174:1589-1594, 2006
 22. Graziotti P, Goucke R: The use of oral opioids in patients with chronic nonmalignant pain: Management strategies. *Med J Austr* 167:30-34, 1997
 23. Gustorff B: Intravenous opioid testing in patients with chronic non-cancer pain. *Eur J Pain* 9:123-125, 2005
 24. Hariharan J, Lamb G, Neuner J: Long-term opioid contract use for chronic pain management in primary care practice: A five year experience. *J Gen Intern Med* 22:485-490, 2007
 25. Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, Atkins D: Current methods of the third US Preventive Services Task Force. *Am J Prev Med* 20:21-35, 2001
 26. Hojsted J, Sjogren P: Addiction to opioids in chronic pain patients: A literature review. *Eur J Pain* 11:490-518, 2007
 27. Holmes CP, Gatchel RJ, Adams LL, Stowell AW, Hatten A, Noe C, Lou L: An opioid screening instrument: Long-term evaluation of the utility of the pain medication questionnaire. *Pain Pract* 6:74-88, 2006
 28. Ives TJ, Chelminski PR, Hammitt-Stabler CA, Malone RM, Perhac JS, Potisek NM, Shilliday BB, DeWalt DA, Pignone MP: Predictors of opioid misuse in patients with chronic pain: A prospective cohort study. *BMC Health Serv Res* 6:46, 2006
 29. Jaeschke R: Users' guide to the medical literature, III: How to use an article about diagnostic test, B: What are the results and will they help me in caring for my patients? *JAMA* 271:703-707, 1994
 30. Jaeschke R, Guyatt GH, Sackett DL: Users' guides to the medical literature, III: How to use an article about a diagnostic test, B: What are the results and will they help me in caring for my patients? *JAMA* 271:703-707, 1994
 31. Jovey RD, Ennis J, Gardner-Nix J, Goldman B, Hays H, Lynch M, Moulin D: Use of opioid analgesics for the treatment of chronic noncancer pain: A consensus statement and guidelines from the Canadian Pain Society, 2002. *Pain Res Manag* 8:3A-28A, 2003
 32. Kalman S, Osterberg A, Sorensen J, Boivie J, Bertler A: Morphine responsiveness in a group of well-defined multiple sclerosis patients: A study with iv morphine. *Eur J Pain* 6:69-80, 2002
 33. Kalso E, Allan L, Dellemijn PL, Faura CC, Ilias WK, Jensen TS, Perrot S, Plaghki LH, Zenz M: European Federation of Chapters of the International Association for the Study of Pain. Recommendations for using opioids in chronic non-cancer pain. *Eur J Pain* 7:381-386, 2003
 34. Kalso E, Edwards JE, Moore RA, McQuay HJ: Opioids in chronic non-cancer pain: Systematic review of efficacy and safety. *Pain* 112:372-380, 2004
 35. Katz N, Fanciullo GJ: Role of urine toxicology testing in the management of chronic opioid therapy. *Clin J Pain* 18: 576-582, 2002
 36. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, Bossuyt PM: Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 282: 1061-1066, 1999

37. Mahowald ML, Singh JA, Majeski P: Opioid use by patients in an orthopedics spine clinic. *Arthritis Rheum* 52:312-321, 2005
38. Manchikanti L, Cash KA, Damron KS, Manchukonda R, Pampati V, McManus CD: Controlled substance abuse and illicit drug use in chronic pain patients: An evaluation of multiple variables. *Pain Physician* 9:215-225, 2006
39. Manchikanti L, Giordano J, Boswell MV, Fellows B, Manchukonda R, Pampati V: Psychological factors as predictors of opioid abuse and illicit drug use in chronic pain patients. *J Opioid Manage* 3:89-100, 2007
40. Manchikanti L, Manchukonda R, Damron KS, Brandon D, McManus CD, Cash K: Does adherence monitoring reduce controlled substance abuse in chronic pain patients? *Pain Physician* 9:57-60, 2006
41. Manchikanti L, Manchukonda R, Pampati V, Damron KS, Brandon DE, Cash KA, McManus CD: Does random urine drug testing reduce illicit drug use in chronic pain patients receiving opioids? *Pain Physician* 9:123-129, 2006
42. Manchikanti L, Pampati V, Damron KS, Beyer CD, Barnhill RC: Prevalence of illicit drug use in patients without controlled substance abuse in interventional pain management. *Pain Physician* 6:173-178, 2003
43. Manchikanti L, Pampati V, Damron KS, McManus CD: Evaluation of variables in illicit drug use: Does a controlled substance abuse screening tool identify illicit drug use? *Pain Physician* 7:71-75, 2004
44. Maruta T, Swanson DW, Finlayson RE: Drug abuse and dependency in patients with chronic pain. *Mayo Clin Proc* 54:241-244, 1979
45. McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS: Users' guides to the medical literature XXII: How to use articles about clinical decision rules. *JAMA* 284:79-84, 2000
46. Michna E, Jamison RN, Pham L-D, Ross EL, Janfaza D, Nedeljkovic SS, Narang S, Palombi D, Wasan AD: Urine toxicology screening among chronic pain patients on opioid therapy: frequency and predictability of abnormal findings. *Clin J Pain* 23:173-179, 2007
47. Michna E, Ross EL, Hynes WL, Nedeljkovic SS, Soumekh S, Janfaza D, Palombi D, Jamison RN: Predicting aberrant drug behavior in patients treated for chronic pain: importance of abuse history. *J Pain Sympt Manage* 28:250-258, 2004
48. Moher D, Pham B, Klassen TP, Schulz KF, Berlin JA, Jadad AR, Liberati A: What contributions do languages other than English make on the results of meta-analyses? *J Clin Epidemiol* 53:964-972, 2000
49. Moore RA, McQuay HJ: Prevalence of opioid adverse events in chronic non-malignant pain: Systematic review of randomised trials of oral opioids. *Arthritis Res Ther* 7:R1046-R1051, 2005
50. Noble M, Tregear SJ, Treadwell JR, Schoelles K: Long-term opioid therapy for chronic noncancer pain: A systematic review and meta-analysis of efficacy and safety. *J Pain Sympt Manage* 35:214-228, 2008
51. Passik SD, Kirsh KL: The need to identify predictors of aberrant drug-related behavior and addiction in patients being treated with opioids for pain. *Pain Med* 4:186-189, 2003
52. Passik SD, Kirsh KL, Whitcomb L, Portenoy RK, Katz NP, Kleinman L, Dodd SL, Schein JR: A new tool to assess and document pain outcomes in chronic pain patients receiving opioid therapy. *Clin Ther* 26:552-561, 2004
53. Passik SD, Kirsh KL, Whitcomb L, Schein JR, Kaplan MA, Dodd SL, Kleinman L, Katz NP, Portenoy RK: Monitoring outcomes during long-term opioid therapy for noncancer pain: Results with the Pain Assessment and Documentation Tool. *J Opioid Manag* 1:257-266, 2005
54. Phillips JE, Bogema S, Fu P, Furmaga W, Wu AHB, Zic V, Hammett-Stabler C: Signify ER Drug Screen Test evaluation: Comparison to triage drug of abuse panel plus tricyclic antidepressants. *Clin Chim Acta* 328:31-38, 2003
55. Portenoy RK, Farrar JT, Backonja M-M, Cleeland CS, Yang K, Friedman M, Colucci SV, Richards P: Long-term use of controlled-release oxycodone for noncancer pain: Results of a 3-year registry study. *Clin J Pain* 23:287-299, 2007
56. Portenoy RK: Opioid therapy for chronic nonmalignant pain: A review of the critical issues. *J Pain Symptom Manage* 11:203-217, 1996
57. Reid MC, Engles-Horton LL, Weber MB, Kerns RD, Rogers EL, O'Connor PG: Use of opioid medications for chronic noncancer pain syndromes in primary care. *J Gen Intern Med* 17:173-179, 2002
58. Reilly BM, Evans AT: Translating clinical research into clinical practice: Impact of using prediction rules to make decisions. *Ann Intern Med* 144:201-209, 2006
59. Schieffer BM, Pham Q, Labus J, Baria A, Van Vort W, Davis P, Davis F, Naliboff BD: Pain medication beliefs and medication misuse in chronic pain. *J Pain* 6:620-629, 2005
60. Sorensen J, Kalman S, Tropp H, Bengtsson M: Can a pharmacological pain analysis be used in the assessment of chronic low back pain? *Eur Spine J* 5:236-242, 1996
61. Management of Opioid Therapy for Chronic Pain Working Group: VA/DoD Clinical Practical Guidelines for the management of opioid therapy for chronic pain. Contract No: V101(93)P-1633(version 1.0), 2003
62. Toma M, McAlister FA, Bialy L, Adams DB, Vandermeer B, Armstrong PW: Transition from meeting abstract to full-length journal article for randomized controlled trials. *JAMA* 295:1281-1287, 2006
63. Turk DC, Swanson KS, Gatchel RJ: Predicting opioid misuse by chronic pain patients: A systematic review and literature synthesis. *Clin J Pain* 24:497-508, 2008
64. Wasan AD, Butler SF, Budman SH, Benoit C, Fernandez K, Jamison RN: Psychiatric history and psychologic adjustment as risk factors for aberrant drug-related behavior among patients with chronic pain. *Clin J Pain* 23:307-315, 2007
65. Wasan AD, Davar G, Jamison R: The association between negative affect and opioid analgesia in patients with discogenic low back pain. *Pain* 117:450-461, 2005
66. Webster LR, Webster RM: Predicting aberrant behaviors in opioid-treated patients: Preliminary validation of the Opioid Risk Tool. *Pain Med* 6:432-442, 2005
67. Whiting P, Rutjes A, Reitsma J, Glas A, Bossuyt P, Kleijnen J: Sources of variation and bias in studies of diagnostic accuracy: A systematic review. *Ann Intern Med* 140:189-202, 2004
68. Wu SM, Compton P, Bolus R, Schieffer B, Pham Q, Baria A, Van Vort W, Davis F, Shekelle P, Naliboff BD: The addiction behaviors checklist: Validation of a new clinician-based measure of inappropriate opioid use in chronic pain. *J Pain Sympt Manage* 32:342-351, 2006

Appendix 1. Search Strategies

COCHRANE DATABASES

Cochrane Database of Systematic Reviews (CDSR), Through 3rd Quarter 2008

1. opioid\$.mp.
2. narcotic\$.mp.
3. (alfentanil or α -prodine or β -casomorphins or buprenorphine or carfentanil or codeine or deltorphin or dextromethorphan or dezocine or dihydrocodeine or dihydromorphine or enkephalin\$ or ethylketocyclazocine or ethylmorphine or etorphine or fentanyl or heroin or hydrocodone or hydromorphone or ketobemidone or levorphanol or lofentanil or meperidine or meptazinol or methadone or methadyl acetate or morphine or nalbuphine or opium or oxycodone or oxymorphone or pentazocine or phenazocine or phenoperidine or piritramide or promedol or propoxyphene or remifentanil or sufentanil or tilidine or tramadol).mp.
4. or/1-3.
5. (((intract\$ or chronic\$ or severe\$ or unbearable\$) adj3 pain\$) or agony or agoniz\$).mp.
6. 4 and 5.
7. (back or spin\$).mp. [mp = title, short title, abstract, full text, keywords, caption text].
8. 6 and 7.
9. from 8 keep 1-66.
10. from 8 keep 1-66.
11. from 8 keep 1-66 (66).

Cochrane Central Register of Controlled Trials (CCRCT), Through 3rd Quarter 2008

Basic search strategy

1. exp Narcotics/
2. exp Analgesics, Opioid/
3. narcotic\$.mp.
4. opioid\$.mp.
5. or/1-4).
6. (((intract\$ or chronic\$ or severe\$ or unbearable\$) adj3 pain\$) or agony or agoniz\$).mp.
7. 5 and 6 (921).

Specific Searches (Each Search Combined the Basic Search Strategy With the Additional Steps Shown)

Studies on Risk Prediction and Monitoring

8. exp "Sensitivity and Specificity"/
9. Prognosis/
10. exp risk/
11. "outcome and process assessment (health care)"/ or "outcome assessment (health care)"/ or "process assessment (health care)"/
12. diagnostic accuracy.mp.
13. receiver operating characteristic.mp. or ROC Curve/
14. 7 and (or/8-13).
15. from 14 keep 1-32 (32).

Studies on Abuse

8. exp Patient Compliance/
9. exp Health Services Misuse/
10. exp "drug and narcotic control"/
11. or/8-10.
12. (abuse\$ or abusing or misus\$ or diversion\$ or divert\$).mp.
13. exp Substance-Related Disorders/
14. 7 and (or/8-13).
15. from 14 keep 1-25 (25).

Studies on Pill Counts and Prescription Drug Monitoring

8. ((medication\$ or opioid\$ or pain\$) adj7 (contract\$ or agree\$)).mp.
9. exp Drug Monitoring/
10. (adher\$ adj5 monitor\$).mp.
11. ((pill or pills or tablet\$ or dose or doses or prescript\$) adj7 (limit\$ or count\$ or ration\$ or monitor\$)).mp.
12. 7 and (or/8-11).
13. from 12 keep 1-23 (23).

Studies on Urine Drug Screening

8. exp Substance Abuse Detection/ (211).
9. (urine adj7 (screen\$ or test\$ or detect\$)).mp. (998).
10. 8 or 9 (1154).
11. 7 and 10 (1).
12. from 11 keep 1 (1).

Search Strategies

Appendix 1. Continued

COCHRANE DATABASES

Ovid MEDLINE**Ovid MEDLINE, 1950 to July Week 3 2008 (Includes Systematic Reviews and Primary Studies)****Basic Search Strategy**

1. exp Narcotics/
2. exp Analgesics, Opioid/
3. narcotic\$.mp.
4. opioid\$.mp.
5. or/1-4.
6. (((intract\$ or chronic\$ or severe\$ or unbearable\$) adj3 pain\$) or agony or agoniz\$).mp.
7. 5 and 6 (5532).

Specific Searches (Each Search Combined the Basic Search Strategy With the Additional Steps Shown)**Studies on Risk Prediction and Monitoring**

8. exp "Sensitivity and Specificity"/
9. Prognosis/
10. exp risk/
11. "outcome and process assessment (health care)"/ or "outcome assessment (health care)"/ or "process assessment (health care)"/
12. diagnostic accuracy.mp.
13. receiver operating characteristic.mp. or ROC Curve/
14. 7 and (or/8-13).
15. from 14 keep 1-298 (298).

Studies on Abuse

8. exp Patient Compliance/
9. exp Health Services Misuse/
10. exp "drug and narcotic control"/
11. or/8-10.
12. 7 and 11.
13. (abuse\$ or abusing or misus\$ or diversion\$ or divert\$).mp.
14. 7 and 13.
15. exp Substance-Related Disorders/
16. 7 and 15.
17. 12 or 14 or 16.
18. from 17 keep 1-696 (696).

Studies on Risk Prediction and Monitoring

8. exp "Sensitivity and Specificity"/
9. Prognosis/
10. exp risk/
11. "outcome and process assessment (health care)"/ or "outcome assessment (health care)"/ or "process assessment (health care)"/
12. diagnostic accuracy.mp.
13. receiver operating characteristic.mp. or ROC Curve/
14. 7 and (or/8-13).
15. from 14 keep 1-298 (298).

Studies on Abuse

8. exp Patient Compliance/
9. exp Health Services Misuse/
10. exp "drug and narcotic control"/
11. or/8-10.
12. 7 and 11.
13. (abuse\$ or abusing or misus\$ or diversion\$ or divert\$).mp.
14. 7 and 13.
15. exp Substance-Related Disorders/
16. 7 and 15.
17. 12 or 14 or 16.
18. from 17 keep 1-696 (696).

Appendix 2. Criteria for Grading Quality of Studies Reporting Diagnostic Accuracy of Risk Stratification and Monitoring Instruments

1. Does the study evaluate diagnostic test performance in a population other than the one used to derive the instrument?
 2. Does the study evaluate a consecutive clinical series of patients or a random subset?
 3. Does the study adequately describe symptom severity, underlying condition, and duration and doses of opioids (if prescribed)?
 4. Does the study adequately describe the instrument evaluated?
 5. Does the study include appropriate criteria in the instrument (must include prior history of addiction or substance abuse and at least one other psychosocial item)?
 6. Does the study adequately describe the method used to identify aberrant drug-related behaviors?
 7. Does the study use appropriate criterion to identify aberrant drug-related behaviors (uses either a validated questionnaire or urine drug screen plus other corroborating data [such as a questionnaire, prescription drug monitoring program, pill counts, family interview, etc]).
 8. Does the study evaluate outcomes or the reference standard in all patients enrolled (up to 10% loss considered acceptable)?
 9. Does the study evaluate outcomes blinded results of the screening instrument?
-

References: Harris et al,²⁵ Lijmer et al,³⁶ Whiting et al⁶⁷ and McGinn et al.⁴⁵

Appendix 3. Criteria for Grading the Overall Strength of a Body of Evidence

<i>GRADE</i>	<i>DEFINITION</i>
Good	Evidence includes consistent results from well-designed, well-conducted studies in representative populations that directly assess effects on health outcomes (at least 2 consistent, higher-quality trials).
Fair	Evidence is sufficient to determine effects on health outcomes, but the strength of the evidence is limited by the number, quality, size, or consistency of included studies; generalizability to routine practice; or indirect nature of the evidence on health outcomes (at least 1 higher-quality trial of sufficient sample size; 2 or more higher-quality trials with some inconsistency; at least 2 consistent, lower-quality trials, or multiple consistent observational studies with no significant methodological flaws).
Poor	Evidence is insufficient to assess effects on health outcomes because of limited number or power of studies, large and unexplained inconsistency between higher-quality trials, important flaws in trial design or conduct, gaps in the chain of evidence, or lack of information on important health outcomes.

Adapted from methods developed by the US Preventive Services Task Force.²⁵

Appendix 4. Excluded Studies With Reasons for Exclusion

<i>STUDY</i>	<i>REASON FOR EXCLUSION</i>
Risk stratification instruments	
Belgrade, 2006 ⁵ (DIRE)	Retrospective and did not evaluate diagnostic accuracy for identifying aberrant drug-related behaviors
Edlund, 2007 ¹⁶	Retrospective and did not assess a risk prediction instrument
Fleming, 2007 ¹⁹	Retrospective and did not assess a risk prediction instrument
Gustorff, 2005 ²³	Did not assess diagnostic accuracy
Hariharan, 2007 ²⁴	Retrospective and did not assess a risk prediction instrument
Ives, 2006 ²⁸	Did not assess a risk prediction instrument
Mahowald, 2005 ³⁷	Retrospective and did not assess a risk prediction instrument
Manchikanti, 2007 ³⁹	Cross-sectional and did not assess a risk prediction instrument
Manchikanti, 2006 ³⁸	Cross-sectional and did not assess a risk prediction instrument
Manchikanti, 2003 ⁴²	Retrospective and did not assess a risk prediction instrument
Maruta, 1979 ⁴⁴	Retrospective and did not assess a risk prediction instrument
Michna, 2007 ⁴⁶	Retrospective and did not assess a risk prediction instrument
Reid, 2002 ⁵⁷	Retrospective and did not assess a risk prediction instrument
Schieffer, 2005 ⁵⁹	Retrospective and did not assess a risk prediction instrument
Wasan, 2005 ⁶⁵	Did not assess predictive value for aberrant drug-related behaviors
Monitoring instruments	
Chabal, 1997 ¹¹	Did not evaluate diagnostic accuracy for identifying aberrant drug-related behaviors
Coombs, 1996 ¹⁴ (SISAP)	Did not evaluate patients with chronic noncancer pain
Friedman, 2003 ²⁰ (STAR)	Did not evaluate diagnostic accuracy for identifying aberrant drug-related behaviors
Passik, 2005 ⁵³ (PADT)	Did not evaluate diagnostic accuracy for identifying aberrant drug-related behaviors
Urine drug screens	
Phillips, 2003 ⁵⁴	No clinical data provided