Instruments to assess patient-reported safety, efficacy, or misuse of current opioid therapy for chronic pain: A systematic review

William C. Becker a,b,*, Liana Fraenkel a,b, E. Jennifer Edelman b, Stephen R. Holt b, Janis Glover b, Robert D. Kerns a,b, David A. Fiellin b

a VA Connecticut Healthcare System, West Haven, CT, USA
b Yale University School of Medicine, New Haven, CT, USA

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

Abstract

The purpose of this systematic review was to summarize and critically appraise research developing or validating instruments to assess patient-reported safety, efficacy, and/or misuse in ongoing opioid therapy for chronic pain. Our search included the following datasets: OvidSP MEDLINE (1946-August 2012), OvidSP PsycINFO (1967-August 2012), Elsevier Scopus (1947-August 2012), OvidSP HaPI (1985-August 2012), and EBSCO CINAHL (1981-August 2012). Eligible studies were published in English and pertained to adult, nonsurgical/interventional populations. Two authors independently assessed inclusion criteria. Each study was evaluated by 2 authors to assess the sources and content of items, types of psychometric tests, their results, and quality of diagnostic accuracy testing, when applicable. Of 1874 citations found in the initial search, we identified 14 studies meeting our inclusion criteria, describing 9 different instruments. Individual items were derived from surveys of content experts, literature reviews, and adapted non-patient-reported items. Misuse-related items were most prevalent (60/144; 42%), followed by safety (47/144; 33%), with efficacy having the fewest items (17/144; 12%). The studies employed a wide variety of psychometric tests, with most demonstrating statistical significance, but several potential sources of bias and generalizability limitations were identified. Lack of testing in clinical practice limited assessment of feasibility. The dearth of safety and efficacy items and lack of testing in clinical practice demonstrates areas for further research.

1. Introduction

The challenges facing patients and providers in managing ongoing opioid analgesic therapy for chronic pain are complex. Benefit of long-term opioid therapy, for which there are scant data to guide providers [22], must be balanced against myriad potential undesired outcomes including safety concerns, ranging from mild toxicities to overdose and death [41]; inadequate efficacy, which may mean continued patient suffering and unwarranted exposure to toxicities; and misuse of these potent medications. To help patients and providers navigate these challenges and optimize therapy, experts advise a strategy of frequent re-assessment of safety, efficacy, and misuse in patients on opioids to inform treatment decisions [7,38]. To date, however, there is no widely accepted instrument or protocol to facilitate this monitoring strategy.

The strategies for monitoring various aspects of opioid therapy can be divided into those that rely on patient report, for example, asking patients about side effects and therapeutic effects; and those that do not, for example, observing a patient for somnolence, performing urine drug testing, or querying a prescription monitoring database for evidence of multiple prescribers. While the latter strategies are important for high-quality clinical care, in this review we focus solely on instruments that collect patient-reported data because non-patient-reported measures have been recently reviewed elsewhere [30,36]. Additionally, we recognize that patient report is the foundation of monitoring the impact of pain treatment [15] and acknowledge the increasing emphasis on patient-reported outcomes in assessing quality of care [6,12]. As such, the current study was designed to systematically review the psychometric development and testing of patient-reported instruments assessing safety, efficacy, and misuse of opioids and, when possible, to assess the operating characteristics of these instruments compared to a reference standard assessment. This review addresses a void in the literature, as previous reviews have
included only instruments assessing risk of or current misuse [8,82,84,80].

2. Methods

2.1. Identification of studies

We identified studies by searching electronic databases, scanning bibliographies of included studies, contacting leaders in the field, and searching consensus clinical guidelines for potentially relevant instruments missed in the initial search. The search strategy was applied to OvidSP MEDLINE (1946 to August 2012), OvidSP PsycINFO (1967 to August 2012), Elsevier Scopus (1947 to August 2012), OvidSP Health (1985 to August 2012) and EBSCO CINAHL (1981 to August 2012), with the last search occurring on August 22, 2012. The full electronic search strategy for OvidSP MEDLINE is presented in Appendix A (I).

We included studies that developed or validated an instrument designed to assess patient-reported safety, efficacy, or misuse of opioids in ongoing therapy. We excluded studies not published in English; that involved nonhuman subjects; that did not study adults 18 and older; or that were related to perioperative or interventional opioid treatment because such clinical scenarios involve markedly different safety, efficacy, and misuse considerations. Further, we excluded studies that were not related to opioid treatment for chronic pain; that did not study a domain of interest (ie, safety, efficacy, or misuse); or that did not provide data on development or validation. Additionally, we excluded studies of instruments assessing risk of safety, efficacy, or misuse prior to opioid treatment, and instruments employing non-patient-reported items exclusively. We did not exclude any articles based on study design. Two authors (W.B. and E.E.) independently evaluated the abstract of each study and, when necessary, the full text, to determine inclusion; discrepancies were resolved by consensus.

2.2. Data extraction

We developed a data extraction instrument based on models from other systematic reviews and standards in the field for psychometric and diagnostic instruments [9,11,13,14,17,21,34,42]. Three data extractors with clinical and health services research backgrounds piloted the data extraction sheet using 2 randomly chosen studies included in the systematic review. After the piloting phase, the research team met to discuss difficulties with the extraction instrument, to clarify questions prompted by the piloted studies, and to compare extracted data. This process helped refine the study aims, created consensus among the data extractors, and prompted minor modifications to the data extraction instrument. Once the data extraction instrument was finalized, the 2 piloted studies were returned to the pool for re-review. Subsequently, after each pair of randomly assigned raters completed 2 extractions, the first author compared data extracted. Differences were presented to raters to resolve by consensus.

2.3. Extraction variables

The complete list of extraction variables and definitions is available in Appendix A (II). Further detail on some of the extraction variables is provided below.

2.3.1. Source(s) and development of items

We compiled data on how items were identified or, when applicable, the process by which they were modified from other instruments or created de novo. One method of item development, assessment of response processes, defined as reviewing the actions or thought processes of respondents [10], was also considered a test of construct validity (see below).

2.3.2. Categorization of items

We categorized the content area of each item based on whether it directly elicited information about the patient’s own experience taking opioids. Safety-related items covered adverse effects, side effects, and toxicities; efficacy-related items covered benefit of the medication in terms of, for example, pain intensity or functional status; and misuse-related items pertained to using the medication other than how it was prescribed, including co-use of illicit substances and/or alcohol and more severe compulsive use characteristics of addiction. Items assessing content pertaining to issues other than the patient’s current experience taking the medication—for example, a patient’s history of addiction, a patient’s family history of addiction, or a patient’s anger or emotional liability—were categorized as “other.”

2.3.3. Assessment of study and instrument quality

We assessed the quality of the studies and instruments across 5 criteria: 1) categories of psychometric testing performed across all studies of each instrument; 2) results of reliability and validity testing; 3) risk of bias; 4) generalizability to general medical practice settings; and 5) clinical utility. We evaluated the categories of psychometric testing—defined in Appendix A (II)—by comparing testing done on each instrument to a checklist adapted from expert recommendations [10,13,14], including the following 6 categories of tests: 1) test-retest reliability; 2) validity testing based on content; 3) response processes; 4) internal structure (internal consistency, dimensionality); 5) relationship to other variables (responsive, discriminative, criterion, and predictive validity); and 6) diagnostic accuracy. We categorized results of the psychometric testing as “robust” if statistical analyses of all psychometric testing were significant and “equivocal” if they were not.

We assessed risk of bias and generalizability using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool [21,42]. The QUADAS-2 is designed to guide reviewers in evaluating risk of bias (categorized as “low” or “high”), and generalizability limitations (categorized as “no” or “yes”) – with respect to patient selection, conduct, and interpretation of the candidate test; conduct and interpretation of the reference standard; and patient flow and timing. Our frame of reference for generalizability was to general medical settings where the majority of opioids are prescribed. Some studies did not employ tests of diagnostic accuracy or use a reference standard comparison; related QUADAS-2 items were listed as “not applicable” in those instances.

We assessed clinical utility based on whether the instrument had safety, efficacy, and misuse-related items (yes or no); and whether it was demonstrated to be feasible in clinical practice; or, if feasibility was not tested, whether it was likely to be feasible in clinical practice, as judged by the reviewers (“yes” or “no”). Brief instruments that required limited scoring and were easily interpretable were considered feasible. Two “yes” answers equated to high clinical utility; fewer than 2 “yes” answers equated to equivocal clinical utility.

3. Results

We identified 14 studies [1–5,9,18,24,27,29,31,32,35,37] meeting the inclusion criteria describing the development or validation of 9 different instruments (Fig. 1); 4 instruments were tested in more than one study. The kappa statistic for agreement on inclusion vs exclusion was 0.92, indicating high inter-rater agreement.
3.1. Study and instrument characteristics (Table 1)

Detailed descriptions of all studies meeting inclusion criteria are provided in Table 1. Regarding the stated purpose of the instruments, one sought to assess safety, efficacy, and misuse of prescribed opioids: the Pain Assessment and Documentation Tool (PADT) [29]; 3 aimed to assess one aspect of safety – opioid-induced constipation: the Bowel Function Index (BFI) [31,32], the Patient Assessment of Constipation Symptoms (PAC-SYM) [35], and the Bowel Function Diary (BF-Diary) [5]; 4 explicitly targeted misuse: the Current Opioid Misuse Measure (COMM) [3,4,24], the Prescription Drug Use Questionnaire-patient version (PDUQ-p) [1,9], the modified Pain Medication Questionnaire (mPMQ) [27], and the Prescription Opioid Misuse Index (POMI) [18]; and one aimed to assess patients’ perceived difficulties with opioid therapy: the Prescribed Opioid Difficulties Scale (PODS) [2,37].

3.2. Characteristics of patients included in instrument development and validation (Table 1)

The sizes of the development or validation cohorts ranged from 74 to 1144 patients. Some studies recruited samples entirely from primary care, some entirely from pain specialty care, and some from mixed primary care/pain specialty settings. Mean age of the cohorts ranged from 38 to 73 years; however, 4 studies did not publish a mean age, and one study (mean age of 73 years) targeted geriatric patients [27]. The proportion of female participants ranged from 6% to 64%. Most of the study cohorts were comprised of patients with chronic pain on long-term opioid therapy, except for 40/74 of the participants in the POMI’s cohort, in whom it was not clear whether pain existed [18]; pain-related diagnoses and/or assessments were reported in a wide variety of formats, except in 3 studies, where they were not described [18,24,29]. Six studies excluded patients with cancer [1,2,4,31,35,37], 2 studies excluded patients with “serious psychiatric impairment” [4,31], and 4 studies excluded patients with a current Diagnostic and Statistical Manual of Mental Disorders, 4th Revision (DSM-IV)-based substance use disorder diagnosis [9] or other indicator of current substance abuse [5,32,35].

3.3. Characteristics of items within instruments (Table 2)

The sources and development of items varied widely, but surveys of content experts, literature reviews, and adaptation of non-patient-reported items were used most frequently. Patient feedback was used in 2 studies to modify items [2,5]. Misuse-related items were the most prevalent (42%), followed by safety (33%), with efficacy having the fewest items (12%). Most of the items in all the instruments were patient-reported; however, the PADT contained 5/12 efficacy-related items and 17/17 misuse-related items that were not patient-reported [29].

3.4. Instrument reliability and validity testing and results (Table 2)

The studies employed a range of tests within the 6 psychometric testing categories. Reliability testing was performed on 4 instruments, with the COMM (intraclass correlation = 0.87), the PDUQ-p (r = 0.67, P < 0.001), the BFI [32], and the BF-Diary [5] demonstrating good test-re-test reliability [24,27]. Content validity assessment, performed in 5 instruments [2,4,5,9,29], employed experts ranging from office staff to physician subject matter experts. Validity based on response processes was tested in the PODS [2] and the BF-Diary [5], in which items were modified based on
<table>
<thead>
<tr>
<th>Instrument</th>
<th>Study/ies</th>
<th>Purpose</th>
<th>Participants and demographics</th>
<th>Exclusion criteria</th>
<th>Administration during testing</th>
<th>Description of pain-related disorders</th>
<th>Description of opioid therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prescribed Opioids Difficulties Scale (PODS)</td>
<td>Banta-Green et al. (2010) [2]</td>
<td>Develop instrument to assess difficulties with chronic opioid therapy as perceived from patients’ perspectives, based on 2 expected domains: psychosocial problems attributed to opioids and concerns about control of opioid medications</td>
<td>1144 patients on chronic opioids enrolled in Group Health for at least 1 year; age range 21–80 y; % female = 61; % white = 84; % employed = 39</td>
<td>Cancer diagnosis</td>
<td>Telephone survey, trained interviewer; time unpublished</td>
<td>Most bothersome pain problem in past 3 months: back pain 29.6%; widespread/multiple 25.4%; leg 19.6%. Reporting 2+ pain problems in the past 6 months: 98.0%</td>
<td>Average daily dose in year prior (mg/d): 0–49 (33.8%); 50–99 (33.7%); ≥ 100 (32.5%). Most frequently used opioids in the past 3 mo: long-acting morphine (24.7%); hydrocodone (20.2%); oxycodone (16.9%); long-acting oxycodone (16.1%); methadone (10.0%)</td>
</tr>
<tr>
<td></td>
<td>Sullivan et al. (2010) [37]</td>
<td>Assess psychosocial difficulties patients attribute to use of opioid medications using the PODS</td>
<td>1144 patients on chronic opioids enrolled in Group Health for at least 1 year; age range 21–80 y; % female = 63; % white = 83; % employed = 45</td>
<td>Cancer diagnosis</td>
<td>Telephone survey, trained interviewer; time unpublished</td>
<td>Most bothersome pain problem in past 6 months: back pain 28.9%; widespread/multiple 20.0; leg 24.5. Reporting 2+ pains in last 6 months: 98.3%</td>
<td>Average daily dose in year prior (mg/d): 0–49 (74.2%); 50–99 (15.0%); ≥ 100 (10.8%). Most frequently used opioids in past 3 months: hydrocodone combination (34.9%); oxycodone (21.6%); long-acting morphine (16.7%); long-acting oxycodone (8.5%); codeine combination (5.2%)</td>
</tr>
<tr>
<td>Pain Assessment and Documentation Tool (PADT)</td>
<td>Passik et al. (2004) [29]</td>
<td>Develop a charting tool focused on outcomes in pain treatment</td>
<td>366 patients on opioids from both primary care and pain specialty; mean age not reported; % female = 64; % white = 84; % employed = 29</td>
<td>None published</td>
<td>Clinician administered; alpha version (59 items) took 10-20 minutes to complete</td>
<td>Not published</td>
<td>Not published</td>
</tr>
<tr>
<td>Current Opioid Misuse Measure (COMM)</td>
<td>Butler et al. (2007) [4]</td>
<td>Develop an assessment tool to monitor misuse of opioids among patients on chronic opioid therapy</td>
<td>227 patients on opioids from 3 pain specialty practices; mean age = 41 y; % female = 62; % white = 83; % employed = not published</td>
<td>Cancer diagnosis</td>
<td>Patient self-administered; time unpublished</td>
<td>Pain intensity 0–10, past 24 hours: worst 7.3 (SD 2.1); least 4.5 (SD 2.2); average 5.9 (SD 1.8); current 5.8 (SD 2.2); Pain interference (0–10); general activity 6.5 (SD 2.6); mood 5.1 (SD 2.7); walking 6.1 (SD 3.0); normal work 7.2 (SD 2.6); relations with others 4.0 (SD 3.2); sleep 6.3 (SD 3.1); enjoyment of life 6.2 (SD 3.0)</td>
<td>Years taking opioids: mean 5.7 (SD 9.2; range 5 months-66 years)</td>
</tr>
<tr>
<td></td>
<td>Butler et al. (2010) [3]</td>
<td>Validate the COMM in a new population</td>
<td>226 patients on opioids from 3 pain specialty practices; mean age = 52 y; % female = 48; % white = 83; % employed = not published</td>
<td>None published</td>
<td>Patient self-administered; time unpublished</td>
<td>Pain intensity 0–10, past 24 hours: worst 7.0 (SD 2.3); least 4.5 (SD 2.4); average 5.9 (SD 1.9); current 5.5 (SD 2.3); Pain interference (0–10); general activity 5.9 (SD 2.6); mood 4.6 (SD 2.9); walking 5.9 (SD 3.1); normal work 6.7 (SD 2.9); relations with others 3.8 (SD 3.1); sleep 5.9 (SD 3.0); enjoyment of life 6.0 (SD 3.1)</td>
<td>Years taking opioids: mean 5.4 (SD 5.8; range 1 mo-38 years)</td>
</tr>
<tr>
<td></td>
<td>Meltzer et al. (2011) [24]</td>
<td>Validate the COMM in a primary care sample; Test COMM against measure of prescription drug use</td>
<td>238 patients on opioids from the primary care clinics of an urban, academic medical center;</td>
<td>None published</td>
<td>Designed to be patient self-administered but may have been interviewer</td>
<td>Chronic pain, defined as 3 months or greater based on chart review</td>
<td>15% of the subjects received the equivalent of 20 tablets of 5 mg oxycodone in &lt;2 fills; 12.6% received 21–60 tablets in &lt;3</td>
</tr>
<tr>
<td>Instrument/Questionnaire</td>
<td>Study Details</td>
<td>Sample Description</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prescription Drug Use Questionnaire – patient version (PDUQ-p)</td>
<td>Compton et al. (2008) [9] Adapt the PDUQ to be patient-administered; assess predictive validity of PDUQ-p</td>
<td>Patients on opioids in pain specialty setting; mean age = 53 y; % female = 6; race/ethnicity not reported; % employed = 19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Banta-Green et al. (2009) [1]</td>
<td>Assess psychometric properties of the PDUQ-p in different setting; examine factor structure of PDUQ-p and relationship to DSM-IV opioid abuse/dependence diagnoses</td>
<td>Patients on chronic opioids, enrolled in Group Health for at least 3 years; mean age = 55 y; % female = 62; % white = 89; % employed = 40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modified Pain Medication Questionnaire (mPMQ)</td>
<td>Park et al. (2011) [27] Adapt the Pain Medication Questionnaire for geriatric population</td>
<td>Patients on opioids for at least a month from both primary care and pain specialty; mean age = 73 y; % female = 29; % white = 51; % employed not reported</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instruments assessing one content area (safety or misuse)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prescription Opioid Misuse Index (POMI)</td>
<td>Knisely et al. (2008) [18] Assess a brief interview focused on prescription use behaviors</td>
<td>Patients on opioids; 40 from addiction treatment, 34 from pain specialty; mean age = 38 y; % female = 45; % white = 95; % employed not reported</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bowel Function Index (BFI)</td>
<td>Rentz et al. (2009) [32] Evaluate the psychometric characteristics of the BFI in patients with opioid-induced constipation</td>
<td>Patients on opioids; referral source not reported; mean age = 57 y; % female = 61; % white not reported; % employed not reported</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rentz et al. (2011) [31]</td>
<td>Assess the psychometric properties of the BFI with respect to patient-reported outcomes in opioid-induced constipation</td>
<td>Patients on opioids; referral source not reported; mean age = 64 y; % female = 66; % white = 99; % employed not reported</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slappendel et al. (2006) [35]</td>
<td>Evaluate the reliability, validity and responsiveness of PAC-SYM in assessing</td>
<td>Patients on opioids; referral source not reported; mean age = 54 y; % employed = not published</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instruments assessing one content area (safety or misuse)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prescription Opioid Misuse Index (POMI)</td>
<td>Knisely et al. (2008) [18] Assess a brief interview focused on prescription use behaviors</td>
<td>Patients on opioids; 40 from addiction treatment, 34 from pain specialty; mean age = 38 y; % female = 45; % white = 95; % employed not reported</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bowel Function Index (BFI)</td>
<td>Rentz et al. (2009) [32] Evaluate the psychometric characteristics of the BFI in patients with opioid-induced constipation</td>
<td>Patients on opioids; referral source not reported; mean age = 57 y; % female = 61; % white not reported; % employed not reported</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rentz et al. (2011) [31]</td>
<td>Assess the psychometric properties of the BFI with respect to patient-reported outcomes in opioid-induced constipation</td>
<td>Patients on opioids; referral source not reported; mean age = 64 y; % female = 66; % white = 99; % employed not reported</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slappendel et al. (2006) [35]</td>
<td>Evaluate the reliability, validity and responsiveness of PAC-SYM in assessing</td>
<td>Patients on opioids; referral source not reported; mean age = 54 y; % employed = not published</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued on next page)
patient feedback. Validity testing based on internal structure was used on all the instruments, with internal consistency ranging from Cronbach’s $\alpha = 0.56$ [1] to 0.96 [4], and dimensionality tested in a variety of methods of factor analysis. Validity testing based on relationship to other variables was performed in 3 instruments, with criterion validity established for the PODS’ relationship to the Patient Health Questionnaire, a self-reported measure of depressive symptoms ($r = 0.317$, $P < 0.001$) [37]; the COMM’s relationship to the Aberrant Drug Behavior Index (Cohen’s $D = 1.25$) [4]; the PDUQ-p’s relationship to the original Prescription Drug Use Questionnaire ($r = 0.64$, $P < 0.001$) [9]; and the BFI’s relationship to the PAC-SYM [31]. However, criterion validity was not established for the PDUQ-p’s relationship to the Composite International Diagnostic Interview (CIDI) for lifetime opioid abuse and dependence, as factor analysis of the CIDI compared to the PDUQ-p revealed low concordance [1]. Discriminative validity was established across a range of scores on the BFI compared to in-depth clinical interview [31,32]; scores on the PAC-SYM compared to a clinical constipation score [35]; and correlation between items on the BF-Diary compared to constipation status [5]. Predictive validity was tested and established in the PDUQ-p as baseline PDUQ-p score was significantly associated with odds of medication agreement violation-related discharge from the practice [9]. However, predictive validity was not established for the PAC-SYM, as its global score’s correlation with dropout or treatment switches due to constipation was low [35]. Responsive validity was tested in the BFI, where it demonstrated significantly improved scores with increasing doses of the opioid antagonist naloxone [32]. Tests of diagnostic accuracy were performed on 3 instruments. In 2 separate studies, the COMM demonstrated good accuracy compared to the Aberrant Drug Behavior Index (area under the curve = 0.81, $P < 0.05$) [3] and DSM-IV prescription drug use disorder (area under the curve = 0.84; 95% confidence interval 0.76–0.91) [24]. The PDUQ-p demonstrated modest accuracy compared to chart reviews ascertaining medication agreement violation-related discharge from practice (sensitivity = 51.4%, specificity 59.8%) [4]; the COMM was accurate when compared to the Aberrant Drug Behavior Index (area under the curve = 0.89, $P < 0.001$) [18].

### 3.5. QUADAS-informed quality assessment (Table 3)

With respect to risk of bias in patient selection, all except 5 of the studies were low risk, as they employed a consecutive or random sample selection procedure and appropriate exclusion criteria. In the study of the POMI [18], it was not clear how patients were selected; thus, risk of bias was considered high. Patient selection in the study of the PADT [29] and one study of the BFI [31] were considered to be at high risk of bias because a convenience sample of patients was used. Patient selection in one study of the PDUQ-p was considered to be at high risk of bias because patients with current substance use disorders were excluded [9]. Patient selection in the other study of the BFI [32] and the PAC-SYM [35] were considered to be at high risk of bias because participants were recruited from pharmaceutical company-sponsored clinical trials, which typically contain healthier individuals. With regard to the conduct or interpretation of the candidate test, all but one study was rated low risk for bias. Specifically, the PADT contained some items that were nonspecific prompts (eg, “increased dose without authorization”), requiring providers to use their own syntax, thus introducing variability and bias in conduct [29]. Because 9 of the studies [1,2,5,27,29,31,32,35,37] did not use a reference standard instrument to which to compare the candidate instrument, risk of bias in the conduct or interpretation of the reference standard could not be assessed. Five studies that did em-
<table>
<thead>
<tr>
<th>Instrument</th>
<th>Studies; total # items</th>
<th>Source(s) and development of items</th>
<th>Item domain (number of items/percentage patient-reportable)</th>
<th>Reliability and validity metrics and results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Instruments assessing safety, efficacy and misuse</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Prescribed Opioids
Difficulties Scale (PODS) | Banta-Green et al. [2010] [2]; 16 | ■ Qualitative input from respondents re: items from Banta-Green 2009  
■ Coverage of items reviewed by co-authors  
■ Literature review | 8/100% 1/100% 7/100% n/a                                                                 | Content:  
2 content experts/co-authors reviewed coverage of items  
Response processes:  
First author reviewed participants’ comments about items to inform modifications  
Internal consistency (2 subscales identified):  
Problems subscale: Cronbach alpha > 0.8; concerns subscale 0.70; total scale 0.87 (derivation) and 0.82 (derivation, recoded) and 0.87 (validation) and 0.81 (validation, recoded)  
Dimensionality:  
Multiple exploratory factor analyses on derivation and validation split samples  
Criteron:  
PODS scores were compared to pain intensity, pain interference, depression symptoms (PHQ-8) and 2 chart-review measures of substance abuse → weak correlation with pain interference; strong correlation with PHQ-8  
Content (59 items):  
6 experts reviewed items in alpha version → eliminated 18 items and added 1  
Internal consistency (1 section):  
Cronbach α = 0.86  
Dimensionality (1 section):  
Exploratory factor analysis → all items loaded onto 1 factor  
Test-retest (40 items):  
55 participants took the alpha-version twice; intraclass correlation = 0.87 (no significance testing)  
Content (40 items):  
22 experts reviewed the alpha-version; every item scored 2.4 or higher on 1–5 importance scale  
Internal consistency:  
Cronbach α = 0.96  
Dimensionality (40 items):  
6 clusters identified using concept mapping  
Criterion validity:  
Comparing COMM scores to criterion measure (ADBI) →Cohen’s D = 1.25  
Diagnostic accuracy (17 items):  
Receiver-operator curves demonstrating true positive/false positive rates for each score on the Current Opioid Misuse Measure compared to Aberrant Drug Behavior Index (ADBI) → area under the curve = .81  
Internal consistency: Cronbach α = 0.83  
Diagnostic accuracy: Receiver-operator curve demonstrating true positive/false positive rates for each score on the COMM compared to Aberrant Drug Behavior Index (ADBI) → area under the curve = 0.79 (P < 0.001)  
Diagnostic accuracy: Receiver-operator curve demonstrating true positive/false positive rates for each score on the Current Opioid Misuse Measure compared to CIDI interview-ascertained prescription drug use disorder → area under the curve = 0.84 (95% confidence interval, 0.76, 0.91) |
| Pain Assessment and Documentation Tool | Sullivan et al. [2010] [37]; 16 Passik et al. [2004] [29]; 42 | ■ Literature review  
■ Survey of experts  
■ Author contribution | 12/100% 12/42% 17/0% 1/0% (changes to treatment plan) |                                                                                                                                                   |                                                                                                                                                                                                                                                                       |
| Current Opioid Misuse Measure (COMM) | Butler et al. [2007] [4]; 17 | ■ Interdisciplinary group of content experts identified candidate items  
■ Second expert panel sorted candidate items by content  
■ Concept mapping software used to develop self-reported item pool.  
■ Third panel voted on importance and quality of wording of items | 3/100% 0/0 11/100% 3/100% (anger, emotional lability) |                                                                                                                                                   |                                                                                                                                                                                                                                                                       |
<p>|                          | Butler et al. [2010] [3]; 17 | COMM [see Butler et al. [2007] [4]] | 3/100% 0/0 11/100% 3/100% (anger, emotional lability) |                                                                                                                                                   |                                                                                                                                                                                                                                                                       |
|                          | Meltzer et al. [2010] [24]; 16 | COMM with 1 item different | 3/100% 0/0 10/100% 3/100% (anger, emotional lability) |                                                                                                                                                   |                                                                                                                                                                                                                                                                       |</p>
<table>
<thead>
<tr>
<th>Instrument Studies; total # items</th>
<th>Source(s) and development of items</th>
<th>Item domain (number of items/percentage patient-reportable)</th>
<th>Reliability and validity metrics and results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prescription Drug Use Questionnaire - patient version (PDUQ-p)</strong></td>
<td>Compton et al. (2008) [9]; 31</td>
<td>PDUQ, with redundant or low-performing items removed</td>
<td>0/0 2/100% 15/100% 14/100% (other treatments tried, family history of addiction, anger towards doctors, personal history of addiction, family members assist in care)</td>
</tr>
<tr>
<td><strong>Banta-Green et al. (2009) [1]; 25</strong></td>
<td>PDUQ-p, modified for phone interviewing purposes and based on factor analysis</td>
<td>0/0 2/100% 15/100% 8/100% (other treatments tried, family history of addiction, anger towards doctors, personal history of addiction, family members assist in care)</td>
<td></td>
</tr>
<tr>
<td><strong>Pain Medication Questionnaire, modified</strong></td>
<td>Park et al. (2010) [27]; 7</td>
<td>Pain Medication Questionnaire</td>
<td>0/100% 2/100% 4/100% 1/100% (find it helpful to call doctor to discuss pain)</td>
</tr>
<tr>
<td><strong>Prescription Opioid Misuse Index</strong></td>
<td>Kniseley et al. (2008) [18]; 6</td>
<td>Frequently used questions in investigators' clinical practice</td>
<td>0/0 0/0 6/100% n/a</td>
</tr>
<tr>
<td><strong>Bowel Function Index (BFI)</strong></td>
<td>Rentz et al. (2009) [32]; 3</td>
<td>Established criteria of known assessment tools for opioid-induced constipation</td>
<td>3/3 0/0 0/0 n/a</td>
</tr>
<tr>
<td><strong>Rentz et al. (2011) [31]; 3</strong></td>
<td>BFI (see Rentz et al. (2009) [32])</td>
<td>3/3 0/0 0/0 n/a</td>
<td></td>
</tr>
</tbody>
</table>

Test-retest: Baseline to 4 months ($r = 0.67, P < 0.001$), 8 months ($r = 0.61, P < 0.001$) and 12 months ($r = 0.40, P = 0.001$)

Criterion validity: Comparison of PDUQ-p to PDUQ scores ($r = 0.64, P < 0.001$ at 4 months)

Predictive validity: Baseline PDUQ-p score was significantly associated with odds of medication agreement violation-related discharge from the practice

Diagnostic accuracy: Using a cutoff score of 10, sensitivity of PDUQ-p for medication agreement violation-related discharge from practice = 51.4%, specificity 59.8%; for opioid specific problem-related discharge: sensitivity = 66.7%, specificity = 59.7%

Content (25 items): Study team and external experts removed 6 items because content was poor fit with studied clinical population or endorsement levels were too high or low.

Internal consistency (25 items): Cronbach $\alpha = 0.56$

Dimensionality (19 items): Factor analysis of 19 items from PDUQ revealed 3 factors and 4 items that did not load on these 3.

Criterion (15 items): Factor analysis comparing CIDI DSM-IV abuse/dependence criteria and 15 items from PDUQ revealed low concordance

Internal consistency: Cronbach $\alpha = 0.73$

Dimensionality: Exploratory factor analysis revealed a 2 factor solution with 7 items (reduced from 26) having the highest factor loading belonging to either of the 2 factors

Internal consistency: Cronbach $\alpha = 0.85$; principal component analysis $\rightarrow$ eliminated 2 items

Discriminative: Receiver-operator curve demonstrating true positive/false positive rates for each score on the Prescription Opioid Misuse Index compared to DSM-IV checklist-ascertained opiate addiction $\rightarrow$ area under the curve = .89 ($P < .001$)

Test-retest: In a sub-sample, no significant difference in BFI score at 2 time points where stability was expected

Internal consistency: Cronbach $\alpha > 0.7$

Discriminative: Each item significantly correlated with degrees of constipation as established by diary entries

Responsive: BFI scores improved in a statistically significant trend with increasing doses of naloxone

Internal consistency: Cronbach $\alpha$ for total BFI score = 0.86; inter-item and item-total correlations were all statistically significant

Discriminative: BFI item and total scores significantly discriminated between 3 constipation severity categories based on clinical interview

Criterion: Correlations between the BFI and Patient Assessment of Constipation Symptoms (PAC-SYM) were in the low-moderate to high range and all were significant ($r = 0.26$ to $0.66; P < 0.01$ to $0.0001$)
Table 2 (continued)

Internal consistency:

- 12/12: 0/0: 0/0: n/a: Cronbach's non-opioid-taking population (2006) [35];
- 0.57-0.71: Discriminative: PAC-SYM scores correlated with presence/absence of constipation based on
- One-way ANOVA test of differences: F = 114; P < 0.0001.

PAC-SYM validated in

- Slappendel previously validated in non-opioid-taking population [36; 3]:
- 12/12: 0/0: 0/0: 1/1: (list of laxatives used)
- Test-retest: Intraclass correlation > 0.71 for all items except stool consistency (0.52)
- In patients with reduced response to review items.
- Predictive: In patients with reduced response to review items.
- Discriminative: Factor analysis revealed a 3-factor solution.
- Dimensionality:
- Survey of experts to review items
- Camilleri et al. (2011) [5];
- Survey of experts to review items
- Rome criteria for functional constipation
- Intraclass correlation > 0.71 for all items except stool consistency (0.52)
- Content:
- Response process:
- Patient input

Table 4 contains a summary of quality assessments of the instruments.

4. Discussion

To monitor patients and make informed, patient-centered clinical decisions, providers need well-validated and feasible instruments for measuring patient-reported safety, efficacy, and misuse of opioids. Our systematic review of the published literature identified 14 studies developing or validating 9 instruments targeting patient-reported safety, efficacy, and misuse. A shortcoming of all the instruments was that none had been tested in clinical practice, suggesting a need for further instrument development and validation. As such, important questions remain unanswered, specifically: what is the feasibility and acceptability of use of these instruments for patients and providers? Do these instruments accurately and consistently identify safety, efficacy, and misuse in clinical practice? And, perhaps the most important question that can only be answered through clinical trials: Does the use of these instruments improve clinical outcomes [33]? The few studies using tests of diagnostic accuracy limits utility but may also be explained by, in some cases, the absence of a suitable reference standard to which to compare the instrument or, in others, by the fact that the instrument was designed more for documentation than diagnostic purposes.

With regard to feasibility, the main limitation was the length and respondent burden of the available instruments. As most opioid therapy is prescribed in general medical settings, monitoring must be brief to account for the reality of competing demands [19]. Several of the reviewed instruments were developed and validated in referral-based settings or for research purposes where time constraints may be less of a concern, which may explain in part why brevity was not a central focus. Shorter versions of the PODS and COMM could be studied to see if accuracy is preserved with reduced respondent burden. Furthermore, the PODS was interviewer-administered in its studies; demonstration of its feasibility as a patient-administered instrument would enhance its clin-
Another limitation related to feasibility that should be addressed in future instrument development is designing assessments that directly guide decision-making. Busy clinicians need to have a clear sense of the next steps informed by the assessment. Our review highlighted a systematic weakness of the instruments with respect to testing patient comprehension of items. If an assessment is designed for patient report, then establishment of patient comprehension of items—by testing the items via patient administration and then using response processes to guide modification—is critical to ensure construct validity [10]. One study used response processes [2], mentioned briefly, and one study used this technique in an assessment designed primarily for research [5]. More thorough response processes in the pilot testing of items would add to the rigor of future instrument development. We identified significant gaps in the content covered by available instruments as another shortcoming. In terms of safety-related limitations, of a published list of 7 clinically relevant side effects of opioids (sedation, nausea/vomiting, delirium, myoclonus, pruritus, respiratory depression, constipation) [23], only 5 were assessed in any of the instruments. Secondary data analyses to determine frequency and scope of side effects, perhaps using novel methods such as natural language processing, and qualitative work to further determine what symptoms patients find distressing, would likely lead to better accuracy. Beyond coverage of side effects, quantification of their severity and impact on the patient would further inform clinical decision-making.

Efficacy-related items were the fewest in number across instruments, possibly reflecting a belief that lack of efficacy is either (a) not as important as identifying safety or misuse issues, or (b) something patients will discuss without prompting, or (c) already incorporated through the use of standardized measures like the numerical pain rating scale. However, lack of efficacy exposes patients to risk needlessly. Furthermore, data suggest that patients do not reliably discuss lack of efficacy with providers and may simply hoard medications or not fill prescriptions [20], each of which carries its own implications for low quality of care. Consensus is
building around the use of specific, mutually agreed-upon functional goals between patient and provider as benchmarks for efficacy [7,25], but how to incorporate such goals into routine screening has not been studied. Brief measures studied in pain care broadly, such as the 3-item PEG [19] or global assessment of benefit [39], may prove well suited for opioid-specific use.

Nearly half of the items developed and validated in these instruments were related to opioid misuse. While this content area is important to both patient and public health, the value of patient-report may be compromised if patients perceive that opioid therapy may be discontinued if they provide truthful information about misuse [16]. Objective determinations such as urine drug testing, querying prescription monitoring databases, and documenting emergency department visits and early refill requests through electronic health records have the potential to augment patient-reported items in these instruments. Furthermore, if the sole target of assessment is misuse, as was the case with the COMM (though we identified 3 items related to safety), the time spent focusing on this may come at the expense of equally important assessments of safety and efficacy. An aspect of misuse that can be captured only through patient-reported items is the patient’s perception of their use as potentially unhealthy or addictive, which was the intent of 3 items of the PODS [2]. A combination of objective misuse measures and the patient’s perception of unhealthy or addictive use holds promise in improving accuracy and efficiency of assessment.

This systematic review has limitations. First, our search may have missed qualifying published instruments, especially if not published in English. To address this concern we queried experts, a large number of electronic databases, and searched the bibliographies of relevant literature. Second, we were not able to evaluate the utility of these instruments in assessing outcomes aside from safety, efficacy, or misuse of opioids or their use of non-patient-reported items.

Despite these limitations, this work adds to the existing literature by describing the breadth of work in the field and identifying opportunities for research to address important clinical needs. We found 14 studies describing the development and/or validation of 9 instruments designed to assess patient-reported safety, efficacy, and misuse of opioids during treatment of chronic pain. The dearth of safety and efficacy-related items and lack of testing in clinical practice demonstrates the need for further research.

Conflict of interest statement

The authors report no conflict of interest related to the study.

Acknowledgements

Dr. Becker was supported by a Veterans Health Administration Health Services Research & Development Career Development Award (08-276); Dr. Fraenkel was supported by National Institute of Arthritis and Musculoskeletal and Skin Diseases grant K24 AR060231-01; Dr. Kerns was supported by VA Health Services Research and Development Research Enhancement Award Program (REA 08-266); Dr. Fiellin was supported by National Institute on Drug Abuse grant R01-DA020576-01A1 and National Institute on Alcohol Abuse and Alcoholism grant U01-AA020795-01.

The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.pain.2013.02.031.

References


