


BMJ Open Development and field testing of primary care screening tools for harms of long-term opioid therapy continuation and tapering to discontinuation: a study protocol

Christine Timko ^{1,2}, Kurt Kroenke,³ Andrea Nevedal,² Mai Chee Lor,² Elizabeth Oliva,⁴ Karen Drexler,^{5,6} Friedhelm Sandbrink,^{5,7} Katherine Hoggatt^{8,9}

To cite: Timko C, Kroenke K, Nevedal A, *et al.* Development and field testing of primary care screening tools for harms of long-term opioid therapy continuation and tapering to discontinuation: a study protocol. *BMJ Open* 2021;**11**:e053524. doi:10.1136/bmjopen-2021-053524

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-053524>).

Received 14 May 2021
Accepted 27 September 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to
Dr Christine Timko;
ctimko@stanford.edu

ABSTRACT

Introduction Despite calls for screening tools to help providers monitor long-term opioid therapy (LTOT) harms, and identify patients likely to experience harms of discontinuation, such screening tools do not yet exist. Current assessment tools are infeasible to use routinely in primary care and focus mainly on behaviours suggestive of opioid use disorder to the exclusion of other potential harms. This paper describes a study protocol to develop two screening tools that comprise one integrated instrument, Screen to Evaluate and Treat (SET). SET1 will indicate if LTOT may be harmful to continue (yes or no), and SET2 will indicate if tapering to discontinue opioids may be harmful to initiate (yes or no). Patients receiving LTOT who screen positive on the SET tools should receive subsequent additional assessment. SET will give providers methods that are feasible to implement routinely to facilitate more intensive and comprehensive monitoring of patients on LTOT and decision-making about discontinuation.

Methods and analysis We will develop the screening tools, SET1 and SET2, concurrently. Tool development will be done in stages: (1) comprehensive literature searches to yield an initial item pool for domains covered by each screening tool; (2) qualitative item analyses using interviews, expert review and cognitive interviewing, with subsequent item revision, to yield draft versions of each tool; and (3) field testing of the draft screening tools to assess internal consistency, test–retest reliability and convergent and discriminant validity.

Ethics and dissemination Ethical approval was obtained from the Institutional Review Boards of Stanford University and the University of California, San Francisco for the VA Palo Alto Health Care System, and the VA San Francisco Healthcare System, respectively. Findings will be disseminated through peer-reviewed manuscripts and presentations at research conferences.

BACKGROUND

Nearly 20% of adults in the USA suffer from chronic non-cancer pain (lasting more than 3 months). Despite known harms, they are

Strengths and limitations of this study

- The primary care tools are intended to be feasible to implement for long-term opioid therapy (LTOT) monitoring and decision-making about discontinuation.
- The tools can be used sequentially or independently and have potential to change practices related to LTOT by helping providers adhere to recommended guidelines.
- Linkage of a national sample of patients' electronic health record data to patient survey data will allow for efficient use of the project's primary data by leveraging the extensive information available from existing records.
- A limitation is the lack of 'gold standard' measures for validity testing or longitudinal data to establish predictive validity of the tools.
- To offset this limitation, we will collect multiple measures of potential consequences of LTOT continuation or tapering, and link patients' survey responses to their diagnoses and healthcare utilisation.

often prescribed long-term opioids.¹ Studies that defined long-term opioid therapy (LTOT) as prescribed opioid use of more than 90 days for chronic pain reported the prevalence of LTOT was 5.4% among all adults in the USA and 79.4% among adults using opioids.² Of approximately 5 million Veterans Health Administration (VA) patients who received an opioid prescription in 2016, about 310 000 (6%) were classified as receiving LTOT.³ Thus high numbers of veterans and other Americans are receiving LTOT despite efforts to decrease opioid prescribing and long-term use.⁴ For example, the Centers for Disease Control and Prevention (CDC) reported that although opioid prescribing has decreased in recent years, it remained three times higher than it was 20 years ago.⁵ Although LTOT benefits some



patients, evidence of its harms is growing.^{2,6} The longer patients take opioids, the greater the potential for developing opioid-related harms.⁷

LTOT harms have been classified using the '5 As', that is, *Analgesia* (opioids do not reduce pain), *Adverse effects* (opioids have side effects, eg, constipation, fatigue, cognitive, breathing and sleep problems, unsteady walking causing falls and injuries, sexual dysfunction, overdose), *Activities of daily living* (opioids do not help to improve daily functioning), *Aberrant behaviours* (non-adherence to opioid treatment plans, eg, self-escalation of dose, frequent requests for medication before refills are due, requests for medication from multiple providers) and *Affect* (eg, depression).⁸ Due to evidence of LTOT harms for many patients, VA/Department of Defence (DoD) and CDC Guidelines recommend frequent monitoring of LTOT patients for the '5 As' to determine pain, adverse opioid effects, function, adherence to opioid treatment plans and aberrant behaviours and co-occurring mental health symptoms and conditions.^{1,9} The main goal of LTOT monitoring is to provide patient-centred care that maximises benefit (reduces pain, improves function and quality of life and offers progress toward individual goals) and minimises harm (reduces side effects and medical and mental health symptoms).

Despite clinical guidelines for care, most LTOT patients in VA and other healthcare settings were found to have not received adequate monitoring.^{6,10-14} In VA, primary care patients on LTOT received, on average, fewer than two of the seven guideline-recommended opioid-monitoring practices (M=1.2, SD=1.5).¹⁰ Monitoring was poor because it was time consuming and did not fit well in the primary care workflow, with pain management often condensed into a 15-to-30 min appointment that also involved other chronic disease management and preventive healthcare needs. Providers cited major barriers to monitoring including inadequate time and resources available; relying on personal impressions of patient risk for aberrant behaviours; and viewing opioid monitoring as a law enforcement rather than safety-enhancing activity.⁶ Together, findings reveal low use of LTOT monitoring among primary care providers and support recommendations for more standardised approaches to LTOT harm reduction. Recommendations for standardised approaches involve using screening tools to identify patients experiencing LTOT-related harms.¹¹ Interdisciplinary teams of clinician—scientists treating LTOT patients similarly reached consensus that a brief instrument, protocolised into routine follow-up, may be the most likely strategy to promote a more active surveillance approach and combat clinical inertia to achieve better patient outcomes.¹⁵

Despite the call for screening tools to help monitor LTOT harms, such screening tools do not yet exist. That is, current monitoring of LTOT patients is inadequate in part because current assessment tools related to prescribed opioids are too long and complex to be feasible to use routinely in primary care. In addition,

current tools tend to focus on only one aspect of monitoring, despite recognition of the '5 As'. Specifically, they focus mainly on aberrant behaviours that may suggest the patient has an opioid use disorder, to the exclusion of the other potential harms of LTOT.

Patients' opioids are being discontinued, with harms

Increasing concern about harms of LTOT relative to LTOT's benefits is leading providers to discontinue patients' use of opioids. However, discontinuation via tapering of opioids also has potential harms. During tapering, patients are at increased risk of *withdrawal symptoms* (eg, generalised pain, chills, cramps, diarrhoea, nausea, vomiting, insomnia, intense cravings), *substance use* (street heroin and other licit and illicit substances), *pain-related distress* and *avoidance coping* associated with *suicidality*, along with *suboptimal use of non-opioid pain management* approaches.¹⁶⁻²⁰ These risks of tapering to discontinuation were highlighted by an international community of stakeholders and key opinion leaders stating their deep concerns about unilateral opioid tapering in LTOT patients.²¹ They described unilateral tapering, even over extended periods, as a large humanitarian issue for the almost 18 million Americans taking LTOT. Countless patients face additional and very serious risks from rapid tapering and related policies mandating dose reductions that are aggressive and unrealistic.²¹ Tapering can precipitate severe opioid withdrawal such that patients destabilise. To escape the resultant suffering, some patients seek relief from illicit and more dangerous opioid sources.^{20,21} Among veterans who had overdosed, transitions to illicit, risky drugs like heroin were driven in part by reduced access to prescription opioids.²² Because tapering is a genuine threat to a large number of vulnerable patients, recent warnings from the Food and Drug Administration and CDC guideline authors as well as empirical studies^{20,21,23-26} call for compassionate systems that include careful selection of patients and patient-centred methods.

In addition to opioid withdrawal symptoms, risks of opioid discontinuation can include use of non-prescribed substances, misuse of prescribed drugs and pain-related distress. Pain-related distress is a persistent pattern of upsetting cognitive and emotional responses (rumination, helplessness and magnification) to current or anticipated pain. It is associated with an array of negative phenomena including patient non-success in prescription opioid tapering.^{27,28} Patients receiving LTOT who have inadequate coping skills may be especially poor candidates for opioid discontinuation.¹⁵ Tapering is daunting to patients receiving LTOT who lack self-efficacy and pain management skills to successfully navigate the process.²⁹ In particular, avoidance (passive) coping (eg, avoid people and activities), more so than approach (active) coping (eg, seek information and social support), predicts responses to pain and opioid management.^{30,31} Thus behavioural interventions to cut down opioid use focus on reducing avoidance coping.²⁸ Among patients with pain and

patients with substance use disorders, there are high rates of suicidal ideation and self-directed violence following opioid discontinuation.^{20 32} Patients receiving LTOT have fears about tapering to discontinuation, such as previous or anticipated lack of effectiveness of non-opioid options, which are associated with their low utilisation.^{20 33} Among veterans using specialty pain services, only 2% had used Complementary and Integrated Health (CIH) services in the previous year.³⁴

CDC and VA/DoD guidelines recommend that providers consider tapering to discontinuation for some patients receiving LTOT. Tapering should be considered when the patient does not experience pain reduction or improvement in function, requests to discontinue opioid therapy, has severe and unmanageable adverse effects of opioids (eg, drowsiness, constipation, cognitive impairment, falls) and does not adhere to the treatment plan and engages in aberrant behaviours (eg, early refills, reporting lost or stolen prescriptions, failing to comply with or having an aberrant urine drug test). Tapering should also be considered when the opioid dosage is a morphine equivalent daily dose of ≥ 90 , and when there are concerns about concomitant substance use disorders, medical comorbidities, use of other medications (eg, benzodiazepines) and mental health conditions that can worsen with opioid therapy (eg, depression). Notably, some of the same patient characteristics that make veterans poor candidates for continued LTOT also make them poor candidates for discontinuing LTOT. These characteristics include substance use and mental health disorders, both of which are highly prevalent among veterans receiving prescribed LTOT.³⁵

Together, findings support recommendations for more standardised approaches to LTOT discontinuation, including using screening tools to identify patients likely to experience harms from discontinuation.^{11 19–21 23–26} Despite the call to use screening tools to help identify patients likely to experience harms of discontinuation, such screening tools do not yet exist.

We plan to develop two screening tools that comprise one integrated instrument, Screen to Evaluate and Treat (SET), to fill these critical gaps in care. SET1 will indicate if LTOT may be harmful to continue (yes or no), and SET2 will indicate if tapering to discontinue opioids may be harmful to initiate (yes or no). (A second stage of this research programme will develop preliminary options for treatment approaches for patients who screen positive on both SET1 and SET2, ie, those for whom there are risks to both continuing LTOT and tapering to discontinuation.) SET1 and SET2 will cover five and six domains, respectively, that were identified as important after extensive literature review and expert input during protocol development (see table 1).

Patients receiving LTOT who screen positive on SET1 should receive an in-depth assessment regarding continued LTOT as described in current guidelines.^{5 36} Patients receiving LTOT who screen positive on SET2 should receive an in-depth assessment as described

Table 1 Domains relevant to SET1 and SET2

	Domain	Salient features
SET1		
1	Pain	Opioids do not reduce pain; pain despite LTOT
2	Opioid side effects	Physical, cognitive, psychological
3	Poor functioning	Opioids do not improve activities of daily living
4	Opioid use disorder	Poor adherence to treatment; aberrant behaviours
5	Depression	Sad mood, hopelessness, feeling worthless
SET2		
1	Opioid withdrawal symptoms	Pain, nausea, insomnia, cravings, etc
2	Substance use	Licit (eg, alcohol) and illicit (eg, street drugs)
3	Pain-related distress	Rumination, helplessness, magnification
4	Coping	Avoidant, passive
5	Suicidality	Ideation, thoughts of death
6	Non-opioid pain management	Non-opioid medications, behavioural interventions, complementary and integrated health treatments

LTOT, long-term opioid therapy; SET, Screen to Evaluate and Treat.

in current guidelines regarding the treatment plan (eg, proceed to initiate tapering to discontinuation, offer medications used for opioid use disorder). The integrated instrument containing two tools will give primary care providers (and other LTOT providers) methods that are feasible to implement routinely to facilitate more intensive and comprehensive monitoring of patients on LTOT and decision-making about discontinuation. The tools can be used sequentially (ie, if SET1 plus further assessment determines that continued LTOT is harmful, then SET2 could be used to help make decisions about tapering to discontinuation) or independently (eg, if a patient requests discontinuation, or a provider has concluded that discontinuation should be considered, SET2 could be used without SET1).

METHODS

We will develop the instrument, SET, containing two screening tools (SET1 and SET2), following Streiner's framework.³⁷ SET1 will be designed to identify patients receiving LTOT (operationally defined as >90 days' prescription of an opioid in the past year) who are being harmed by continued opioid therapy. SET2 will be designed to identify patients who are at risk for adverse

effects from an opioid taper to discontinue their medication. We will develop the screening tools for SET1 and SET2 concurrently. Tool development will be done in stages: (1) comprehensive literature searches to yield an initial item pool for domains covered by each screening tool; (2) qualitative item analyses using interviews, expert review and cognitive interviewing, with subsequent item revision, to yield draft versions of each tool; and (3) field testing of the draft screening tools to assess internal consistency, test–retest reliability and convergent and discriminant validity. This 4-year project is supported for Fiscal Years 2021 through 2024.

Identify preliminary item pools for SET1 and SET2

To select items for SET1 and SET2, we will first conduct a systematic literature search to identify existing instruments that screen for (1) harms of ongoing LTOT; and (2) indications of difficulty with an opioid taper for discontinuation. Using search results, we will follow Patient-Reported Outcomes Measurement Information System methods^{38 39} by using a consensus process to first ‘bin’ items from these instruments into the five domains for SET1 and the six domains for SET2; that is, we will group items according to meaning and specific latent construct. Then, we will ‘winnow’ the binned items, that is, reduce the large item pool to a representative set of items by organising items that were binned into domains into subdomains to help reduce item redundancies, and eliminate items that lack face validity for the domain, are very similar to a better-worded item, or have content that is too narrow. Specifically, two reviewers will independently winnow and bin items; discrepancies between the two reviewers will be resolved through discussion with a third reviewer. We will modify items to adhere to the set of formatting requirements, for example, first-person subject, primarily present tense, simple vocabulary (as concise and simply worded as possible; sixth grade or lower reading level), and not confusing (vague, multi-barreled, outdated, uses slang).

Interviews to assess preliminary item pools

We will conduct interviews with patients and providers separately in order to (1) ensure the item pools have appropriate coverage of the conceptual areas for each tool, (2) assess the face and content validity of the candidate questions in the item pools (whether items are understandable and acceptable, including the likelihood of patients’ truthful answers), (3) develop items that cover content not already included in the item pools, and (4) identify items with outlying high or low thresholds for endorsement by patients (eg, items that are not optimal because they are likely to be endorsed by almost everyone or only by patients who have experienced severe consequences of LTOT use).

We will conduct about 16 interviews for patients and 16 for providers, for a total sample of approximately 32 participants. The total sample size is based on recent literature reporting tests of purposive sample adequacy and

data saturation; with the planned number of interviews, we will be able to identify $\geq 90\%$ of discoverable themes.^{40 41}

Interviews will be conducted in two US locations, one on the West Coast and one in the Midwest. These sites were selected to provide some regional diversity for the patients and providers. With purposeful criterion sampling, we will use data from the VA Corporate Data Warehouse (CDW), a national level database housing clinical, administrative and financial information, to identify patients at the sites with an indication of LTOT (>90 days of prescribed opioids in the past year). We will select patients with and without an indication of an opioid dose reduction in the past year. To invite patients to participate, we will send an advance notice letter describing the study’s purpose (to develop tools to assist in treatment planning for patients taking opioid pain medication), the interview format, that points raised in the interviews will be kept confidential with no reports identifying patients by name and instructions on how to opt out from further contact. Veterans who do not opt out or decline will be scheduled. Patients will receive US\$20 for interview participation. We will recruit providers from each of three treatment settings: primary care, pain clinics and specialty substance use disorder (SUD) care. To recruit providers willing to participate in the interviews, research team members will email them directly to describe the study and the interview’s structure.

Draft interview guides for the patient and provider interviews will be finalised by information synthesised from the initial literature and opioid scales reviews. For patients, the interviews will focus on potential harms and benefits of their LTOT (to inform SET1), and potential harms and benefits of tapering or discontinuing opioids (to inform SET2). For providers, the interviews will focus on harms and benefits to patients on LTOT (SET1), and of tapering or discontinuing opioids (SET2). Providers’ interviews will also discuss how, given barriers to a new instrument’s implementation in outpatient clinics (eg, time shortages, multiple clinical reminders), the new instrument can best be inserted and used in the clinical workflow. Project personnel (ie, AN, a qualitative methods expert, as the primary analyst, and MCL, also trained in qualitative methods, as the secondary analyst) will review and analyse interview notes and audio recordings taken during the interviews.⁴² Content analysis will be used to analyse the qualitative data and to describe patient or provider preferences with respect to the assessment of LTOT’s harms, and potential tapering-to-discontinuation problems. Specifically, using established methods,⁴³ the primary analyst will take notes during the interview, code them into a matrix immediately after, highlighting areas that need clarification, and write a summary of each interview. The secondary analyst will review the matrix and listen to the audio recording to make additional notes, after which the two analysts will meet to achieve and confirm consensus on identified themes.

Expert panel to review and reduce item pools

An expert panel has been selected to represent different areas of specialisation, including policymakers and providers who may use the screening tools developed. We will use a modified Delphi method for our expert panel. The nominal group-Delphi approach is a well-established, structured method for obtaining expert opinion and consensus.^{44 45} It has been applied to numerous health conditions and interventions, guideline development and treatment outcome ratings, and its strengths include aiding the development of quantitative and qualitative data, providing feedback, and forging agreement.

We will conduct the expert panel in two stages: elicitation and re-elicitation. In the elicitation stage, we will send panel members relevant background literature, the preliminary item pools for SET1 and for SET2 and the interview results. We will ask panellists to rate potential items in the item pool for each tool for importance, suggest other items believed to be important, rate confidence in the questions in the item pools and indicate how, if at all, items need tailoring. Ratings of importance and confidence will be made on 5-point Likert scales that include standards for response options.^{37 46} In the re-elicitation phase, we will review the initial assessments with panellists. The goal of this phase will be to enable panellists to come to agreement on which items have the strongest evidence base for each tool and thus merit further consideration. The project team will provide summaries of the results from the elicitation (including statistics contrasting each individual's response with the group mean, SD and distribution of response choices), and discuss with panellists the item pools' initial ratings pertaining to SET1 and SET2. Items with elicitation and re-elicitation consensus support will be retained, and those lacking support will be eliminated. For each tool, consensus will be defined as having 80% of responses falling within the top two categories on the mean of the two Likert scales, plus confirmation by the panellists' discussions. Following the Delphi process, the project team will compile the items into the two candidate screening tools, SET1 and SET2, for further testing.

Cognitive interviews with patients

We will conduct cognitive interviews on the instrument's SET1 and SET2 with 20 patients^{47 48} to obtain feedback on clarity, understandability and acceptability (eg, avoid value-laden language; determine if patients will answer truthfully) of individual items and the instrument as a whole. Following standard cognitive interviewing methods,^{47 48} we will use post-interview notes to document items that need revision due to bias, lack of comprehension or understanding of the question's intent, poor recall for what is being asked or likely truthfulness. Results will be summarised in a qualitative data matrix.⁴⁹ After five interviews, we will make needed changes and test the revised items in subsequent rounds of five interviews until all interviews are complete.

Field testing with patient surveys

Based on feedback from the expert panel and cognitive interviews, the project team will create final draft versions of SET1 and SET2 for field testing. Because SET1 will be developed for patients receiving LTOT, and SET2 for patients being considered for opioid tapering to discontinuation, we will recruit two independent samples of patients for the field testing of both tools. The first sample (n=about 500) will be patients receiving LTOT (>90 days of prescribed opioids in the past year) with no indication of a recent dose reduction (or a recent taper for discontinuation). The second sample (n=about 500) will include patients on LTOT with an indication of a dose reduction in the past year. To obtain as representative a sample as possible, we will recruit patients for a telephone, rather than in-person, survey; this will allow the sample to be selected without restriction by geography or urban/rural status.

We will identify potentially eligible veteran patients for the project sampling frame using VA electronic health record data from VA CDW. The first sample of 500 patients will include those with an indication of LTOT in CDW data (>90 days prescription for an opioid pain medication in the past year) who do not have an indication of a dose reduction in the past year. The second sample of 500 patients will comprise patients who met criteria for LTOT in the past year and have documentation of a dose reduction in the past year that would be consistent with a taper for discontinuation (indicated by pharmacy documentation of at least two dose reductions of at least 5% in the past year).⁵⁰ To be considered potentially eligible, patients will be ≥ 18 years old, have a valid address and telephone number on file with VA and be able speak and understand English. Using VA CDW diagnosis codes and patient information, we will exclude patients who are in active cancer treatment or enrolled in hospice care, or cognitively impaired.

To recruit eligible patients, data collectors will send an advance notice letter to veterans that describes the study's purpose and includes instructions on how to opt out from further contact (ie, a dedicated voicemail phone line that removes the need to speak with anyone, expressly to minimise burden on patients who do not wish to participate). The letter will state the study's goal of developing tools to assist in treatment planning for patients who are taking opioid pain medication, assure patients that responses will be kept confidential and state that participants will receive US\$20 monetary compensation for completing the survey. Patients who do not opt out from further contact will be scheduled for a telephone interview to administer the draft tools and survey. Interviewers will make up to 12 call attempts to contact eligible patients who have not opted out of further contact. This will include all of an individual's phone numbers (landlines plus cells) and leaving up to three messages (total) on every fourth attempt over several weeks. The number of call attempts is based on the authors' and others' previously-used methods and expert opinion.⁵¹ Specifically, more call attempts increase



recruitment rates and the representativeness of the sample recruited, whereas limiting the sample to participants who are easier to reach excludes those more difficult to schedule, such as those with work, family or other obligations. All messages will include instruction on how to opt out of further contact, and if a patient does opt out, we will cease contact attempts. Once contact is made, the interviewer will describe the study and complete a verbal informed consent process. As part of this process, for the second sample (LTOT with recent dose reduction), we will confirm that the patient has undergone a recent dose reduction.

Interviewers will be trained to administer the draft SET1 and SET2 tools to eligible patients who have provided consent (ie, survey respondents). In addition to the items for the draft SET1 and SET2 tools, the patient survey will include criterion outcomes to allow determination of the optimal cutpoints and operational characteristics of the tools, and covariates to allow assessment of the tools' convergent and discriminant validity. As part of the consent process, Respondents will be asked to consent to have their electronic health record data reviewed and be informed that their personal information will not be shared with anyone outside the research team and will be stored on the VA secure server. In addition, to assess the test-retest reliability of the draft tools, we will recruit about 30 of the respondents to retake the tools 1 week after their initial interview. The interval of 1 week was chosen to be long enough that participants will be unlikely to remember (or be influenced by) their initial responses when providing their second set of responses, but short enough that changes among participants' conditions or symptoms are unlikely to have occurred.^{52 53} Test-retest agreement will be quantified as the intraclass correlation coefficient (ICC) on the total scores of each of the tools (SET1 and SET2) separately.

As noted, as part of the informed consent process, we will ask respondents for permission to link their survey responses to electronic health record data. For respondents who consent, we will obtain data on their demographic characteristics and other clinical and utilisation variables from VA CDW (see [table 2](#)). We will clean all data elements and conduct an initial description using appropriate summary statistics (count, %, mean, SD) to assess overall data quality. As part of this process, we will quantify missing data and impute or weight for missing data as needed.

Patient and public involvement

As part of the preliminary studies for this project, we conducted focus groups composed of patients, caregivers and providers to obtain feedback on items contained in measures that assess the severity of abuse of prescription pain medications. Focus group participants agreed that, generally, measures have too many items with too many response options. They noted three levels of items: would definitely not answer or not answer truthfully because items imply the patient is doing something bad

or wrong; might answer; and would definitely answer. Conclusions were that current measures are not sufficient to meet primary care clinical needs and that new screening tools are needed. To further ensure patients' engagement, in preparing this protocol, project leaders met with veterans and family advisory councils to solicit feedback on the project's significance for patients and the planned approach, and received uniformly enthusiastic agreement to help with accomplishing its aims. We are continuing to meet semiannually with the councils to exchange feedback during the project's execution. This feedback will include findings from the planned interviews and instrument field testing, which require veteran patients' participation and input.

Study measures: patient survey

The patient survey will include items to assess criterion outcomes for screening tool development. Criterion measures for SET1 cover the 5 A's (five domains) and criterion measures for SET2 cover poor discontinuation outcomes (six domains; see Background). The survey will also include covariates for describing the patient survey sample and assessing convergent and discriminant validity. In the absence of gold standards against which to validate the draft screening tools, we will assess the performance of the tools relative to criterion outcomes that cover common, moderate-severe consequences of LTOT and tapering to discontinue.

Criterion outcomes for SET1 will include:

1. **Analgesia** (pain despite LTOT). The total score on the three-item PEG⁵⁴ will be used to assess average pain intensity (P) and interference with enjoyment of life (E) and with general activity (G). The PEG is reliable, with good construct validity for pain-specific measures.
2. **Adverse effects**. We will use the total score from the Side Effects Checklist (SEC), on which patients note whether they experienced any of 12 opioid side effects in the past month. The SEC has high validity and test-retest reliability among primary care patients taking opioids for pain.⁵⁵
3. **Activities** (poor functioning). To assess daily activities functioning, we will use the total summary score indicating role functioning related to physical health on the veterans SF-12, a 12-item self-report questionnaire that has been widely used, disseminated and documented as a reliable and valid measure of the functional effects of illness in VA patients.^{56 57} A lower score indicates poorer functioning.
4. **Aberrant behaviours** (opioid use disorder). We will use the MINI International Neuropsychiatric Interview's (MINI)^{58 59} Diagnostic and Statistical Manual of Mental Disorders (DSM-5)-concordant, 11-item module to assess diagnostic criteria for past-year (current) opioid use disorder. We will consider both the overall MINI score (number of diagnostic criteria endorsed) and a binary classification of opioid use disorder (at least two criteria endorsed).

Table 2 Data used in analysis of SET1 and SET2 tools

Item	Source	Variables
Draft tools		
Screening for LTOT harms	SET1 draft tool	Total score
Screening for taper-to- discontinuation harms	SET2 draft tool	Total score
SET1: Criterion outcomes		
Pain	Patient survey: PEG	Total score
Side effects	Patient survey: SEC	Total score
Side effects	CDW: Falls or accidents ³⁵	Past year falls and vehicle accidents based on documented ICD-10 codes
Poor functioning	Patient survey: Veterans SF-12, physical health	Total score
Opioid use disorder	Patient survey: MINI	Total score; binary outcome
Opioid use disorder diagnosis	CDW: Outpatient and inpatient encounter data	Past-year diagnosis based on documented ICD-10 codes
Depression	Patient survey: PHQ-9	Total score
SET2: Criterion outcomes		
Opioid withdrawal symptoms	Patient survey: SOWS	Total score
Substance use	Patient survey: BAM, substance use	Total score
Pain catastrophising	Patient survey: PCS-4	Total score
Poor coping	Patient survey: CRI, Avoidance Coping	Total score
Suicidality	Patient survey: PHQ-9 item	Item score
Low non-opioid pain management	Patient survey: CHAPS	Total score
Low non-opioid pain management	CDW: CIH, Chiropractic, OT, Pain clinic, PT use	CPT codes and ICD-10 procedure codes
Patient survey covariates		
Demographics	Patient survey: Demographics	Age, sex, race/ethnicity, education, marital status, use of VA
Social desirability bias	Patient survey: M-C	Total score
VA CDW covariates		
Demographics	CDW patient table	Service-connection
Medical or psychiatric diagnoses	Outpatient and inpatient encounter data	ICD-10 diagnosis codes for pain and pain-related conditions, insomnia, mental health conditions (depression, PTSD, anxiety)
Encounter location	Outpatient and inpatient encounter data	Clinic stop codes/bed sections for SUD specialty care, admissions to ED for overdose or withdrawal syndrome
Procedure codes	Outpatient and inpatient encounter data	CPT codes for mental healthcare procedures related to SUD
Alcohol screening test (AUDIT-C)	CDW Mental Health Domain	Alcohol misuse screening scores
Other mental health/SUD screening tests	CDW Mental Health Domain	Mental health conditions (depression, PTSD) and SUD (full BAM, AUDIT)
SUD pharmacotherapy	Prescription records from DSS, PBM files	Prescriptions for MOUD of opioid withdrawal and dependence

AUDIT-C, Alcohol Use Disorders Identification Test-Concise; BAM, Brief Addiction Monitor; CDW, Corporate Data Warehouse; CHAPS, Complementary Health Approaches for Pain Survey; CIH, Complementary and Integrated Health; CPT, Current Procedural Terminology; CRI, Coping Responses Inventory; DSS, Decision Support System; ED, emergency department; ICD-10, International Classification of Diseases, 10th Revision; LTOT, long-term opioid therapy; M-C, Marlowe-Crowne; MINI, Mini-International Neuropsychiatric Interview; MOUD, medications for opioid use disorder; PBM, Pharmacy Benefits Management; PCS-4, Pain Catastrophising Scale-4 items; PEG, pain, enjoyment, general activity; PHQ-9, Patient Health Questionnaire-9 items; PT, physical therapy; PTSD, post-traumatic stress disorder; SEC, Side Effects Checklist; SET, Screen to Evaluate and Treat; SF-12, Short Form-12 items; SOWS, Subjective Opioid Withdrawal Scale; SUD, substance use disorder.

5. **Affect.** We will use the Patient Health Questionnaire (PHQ-9)^{60 61} to assess depression among LTOT patients. The PHQ-9 closely mirrors DSM criteria; it includes an item on suicidality, which will be used for SET2. It has excellent internal consistency and test-retest reliability.

Criterion outcomes for SET2 will include the following:

1. For patients' ratings of opioid withdrawal symptoms with lower doses, we will use the 16-item Subjective Opioid Withdrawal Scale (SOWS). The American Society of Addiction Medicine recommends the SOWS due to its strong psychometric properties.⁶²

2. We will use the Brief Addiction Monitor (BAM; required in VA for measurement-based SUD care) to assess non-prescribed opioid and non-opioid substance use, to which patients may turn when opioid doses are reduced.⁶³
3. For pain-related distress, we will use the Pain Catastrophising Scale's (PCS-4) total score. The PCS-4 has four items examining three components (rumination, magnification and helplessness) and good psychometric properties (reliability, validity, predictive validity).⁶⁴
4. Avoidance coping will be assessed with the total Avoidance Coping scale score from the Coping Responses Inventory. This scale has six items, with higher scores indicating poorer coping (eg, isolated from social support). It has been found psychometrically sound in numerous studies of medical patients.⁶⁵
5. As noted, suicidality will be assessed with the last item from the PHQ-9. (For any participant endorsing this item, the validated P4 screener will be used to grade suicidal risk, and for anyone assessed at intermediate to high risk, a safety plan will be implemented.)
6. To measure low use of non-opioid pain management strategies, we will use the Complementary Health Approaches for Pain Survey (CHAPS), which assesses use of 12 non-opioid strategies for pain management (eg, yoga, meditation; items will be added for non-opioid medications, chiropractic care and behavioural interventions, eg, cognitive-behavioural therapy, mindfulness). The CHAPS performs well with patients in pain management and rheumatology practices.⁶⁶

Social desirability bias will be assessed with the 13-item Marlowe-Crowne Social Desirability Scale–Short Form (M-C).^{67 68} It will be used to test items' and SET1 and SET2's associations with patients' desires to answer questions in a socially desirable way, that is, less willing to answer questions truthfully and less accurate self-representations. We will determine if higher scores (more social desirability) are strongly associated with SET1 and SET2 item responses; if so, those items will be candidates for removal.

To create the final analytical data set, VA CDW data (table 2) will be securely extracted and analysed. After creating a linked analytical data set containing both patient survey and CDW data, we will clean the data, check the data for logical inconsistencies, reconcile any problems and construct study variables. We will describe distributions of all variables using univariate and multiple-variable statistics. We will assess the potential for response bias by quantifying differences in demographic characteristics, prescription history and electronic health record diagnoses between patients who are survey responders or non-responders. If systematic differences are detected, we will compute inverse-probability of response weights that will be incorporated into subsequent analyses to adjust for non-response. The merged data set of survey and CDW data for study participants will be stripped of direct identifiers and stored securely.

Final item selection

The goal of data analysis will be to perform final item selection for the draft SET1 and SET2 tools, to determine the optimal cutpoint on each tool for detecting patients for whom LTOT is harmful (SET1) and for detecting patients who may experience difficulty with an opioid taper for discontinuation (SET2). Each of the tools, SET1 and SET2, will be developed using each of the patient samples we recruit: a general group of patients receiving LTOT (Sample 1) and a group of patients receiving LTOT with a recent dose reduction (Sample 2). We will develop both tools using data from both samples to ensure the items selected will perform adequately for patients with fewer or greater potential harms and consequences of LTOT.

Empirical item selection

The recommended method to ensure brevity of both tools is the combination of the qualitative content and statistical psychometric analyses proposed.^{69 70} There is no consensus definition of what constitutes a brief measure, but several streams of evidence and practical considerations suggest 8–10 items as a reasonable upper boundary.^{69–71} To select the final items for each tool, we will evaluate each of the items on each of the tools in terms of (1) concurrent validity with the specified criterion measures for the tool (table 1); (2) the contribution of each item to the internal consistency of the tool as reflected by Cronbach's alpha; and (3) the contribution of each item to the test–retest reliability of the tool. Following the process used to develop other screening tools,⁷² we will retain items for each tool that have: a correlation with the criterion outcome measure for that tool of 0.20 or higher, a higher correlation with the criterion measures than with the M-C (so the correlation with the criterion is greater than the correlation with the measure of social desirability), a Cronbach's alpha of at least 0.70 (all tool items together) and a test–retest intra-class correlation of at least 0.50 (full tool). The project team will use these empirical standards for each of the items on each of the draft screening tools to select final versions of SET1 and SET2.

SET1 and SET2 cutpoint selection

To select cutpoints for the screening tools, we will use the criterion outcomes specific to each tool (table 2). We will conduct area under the curve (AUC) analyses to select the optimal cutpoint for each criterion outcome for each tool. We will select outcome-specific cutpoints to optimise both sensitivity and specificity by selecting thresholds that yield sufficiently high AUC. We will then choose the tool cutpoint that captures all criterion outcomes; most likely, this will be the cutpoint for the outcome that has the lowest optimal cutpoint. This method will work best if there is no major outlier, for example, cutpoints across the five criterion outcomes are 7, 8, 8, 9, 10. If there is a major outlier as a low cutpoint (eg, outcome-specific cutpoints are 3, 7, 8, 9, 11), we may instead choose the

tool cutpoint that captures the majority of outcomes (eg, pick 7 rather than 3 as the cutpoint because 3 would yield too many false positives for the other four outcomes). A third case that could occur is that we select the median cutpoint, should the individual cutpoints show a wider range (eg, 3, 7, 10, 13, 15); here, 10 might be the tool cutpoint in order to balance false positives and negatives.

The operating characteristics of each tool will be summarised using six values: sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio and negative likelihood ratio. Because the tools' goal is to identify patients with opioid-related or discontinuation-related harms, we consider the tools' sensitivity (probability of a patient who meets the criterion testing positive on the tool) to be of primary importance. Therefore, the project team will review the results of the receiving operating characteristic (ROC) curve analysis and choose thresholds that balance the tools' empirical performance with the team's assessment of the tools' clinical appropriateness. The decision about which cut-off scores to use should depend on both empirical analytical results and on experts' qualitative judgements about what is best for patients.⁷²

Convergent and discriminant validity

To assess convergent validity, we will estimate the correlation between the total SET1 or SET2 score and the covariates listed in table 2. We will also assess the association between screen-positive status on SET1 or SET2 with the same covariates. We will compute point and interval estimates of the prevalence ratio as the ratio of the prevalence of an index characteristic or condition among the patients screening positive to the prevalence among patients screening negative. Because there is no generally agreed on threshold for a 'clinically meaningful' prevalence ratio, we will focus on reporting point and interval estimates of the associations between screen-positive status and other covariates. To assess discriminant validity, we will compute correlations between the total SET1 or SET2 score and the Brief Sensation Seeking Scale (BSSS) score.⁷³ Sensation seeking has been demonstrated to be unrelated to opioid aberrant behaviours among chronic pain patients.⁷⁴ The four-item BSSS will be given on the patient survey. It has been shown to be internally consistent, related as expected to other indices, and perform similarly to longer measures of sensation seeking.⁷³ Discriminant validity will be demonstrated by (a) low correlations between SET1 or SET2 and BSSS scores, and (b) lower SET1-BSSS and SET2-BSSS correlations than correlations calculated to assess convergent validity.

Sample size/power calculations

The sample size required for the patient survey was determined to provide sufficient precision or power to assess internal consistency of the candidate items (a key item selection criterion for SET1 and SET2), estimate the area under the ROC curve (AUC, used to determine optimal cutpoints for a positive screen on each tool), contrast

the prevalence of patient characteristics or conditions for patients who do versus do not screen positive on the tools (an item selection criterion) and assess test-retest reliability.

We first estimated the size of Sample 1 (LTOT patients) and Sample 2 (LTOT patients with a past-year dose reduction) needed to examine internal consistency (KR-20 or Cronbach's alpha) with a given degree of precision, as reflected by the width of the corresponding 95% CI estimate.³⁷ The sample size required is inversely related to the scale's number of items; thus, we computed a conservative estimate for the required sample size that would be adequate even for ultra-brief tools (<5 items). We could estimate an alpha (or KR-20) of 0.90 (high internal consistency) with a 95% CI of width 0.15 with ~35 patients, a width of 0.10 with ~65 patients and a width of 0.05 with ~250 patients, per sample. We could estimate an alpha of 0.70 (a common lower bound for desired internal consistency) with a 95% CI of width 0.30 with ~65 patients, a width of 0.20 with ~130 patients and a width of 0.10 with ~500 patients, per sample. Based on this assessment, we considered 500 patients per sample to be the candidate sample size.

We next assessed whether the candidate sample size would adequately test whether the tools are able to discriminate patients with or without LTOT harms, and patients likely to be harmed or not by tapering to discontinuation. A sample size of about 425 or larger would yield $\geq 90\%$ power to test the hypothesis that the AUC is ≤ 0.50 with an alpha-level of 0.05 over a range of hypothesised AUC values (poor (0.60) to near perfect (0.95) accuracy) and case:non-case ratios (from 0.25 or 1 non-case for every 4 cases, to 4.0 or 4 non-cases for every case).⁷⁵ We also assessed the sample size necessary for estimating the AUC with a given precision, summarised by the width of the 95% CI estimate, assuming a balanced allocation. With about 480 patients we could estimate an AUC of 0.60 with a CI width of .10. Thus, 500 patients per sample would be more than adequate.

Finally, we assessed whether 500 patients per sample would provide adequate power for comparing the prevalence (using the prevalence ratio) of characteristics or conditions between patients who do versus do not screen positive. Assuming 500 patients per sample and a screen-positive:screen-negative ratio of 0.1, we would have $\geq 80\%$ power to detect a prevalence ratio of ≥ 3 for conditions with a baseline prevalence of 5%, a prevalence ratio of ≥ 2.2 for conditions with a baseline prevalence of 10% and a prevalence ratio of 1.4 for conditions with a baseline prevalence of 40%. With a higher screen-positive:screen-negative ratio of 1, we would have $\geq 80\%$ power to detect a prevalence ratio of ≥ 2 for conditions with a baseline prevalence of 5%, a prevalence ratio of ≥ 1.6 for conditions with a baseline prevalence of 10% and a prevalence ratio of 1.4 for conditions with a baseline prevalence of 20%. Thus, we chose a final sample size of 500 patients per sample.

From the patient survey respondents, we will recruit 30 to assess test-retest reliability of each tool. The sample



size will allow estimating an ICC of 0.90, SE of 0.05, or an ICC of 0.75, SE of 0.10.

DISCUSSION

This project will develop tools for the management of patients receiving LTOT and patients for whom an opioid taper for discontinuation is considered. The tools can be implemented as one integrated instrument for primary care-based screening to indicate the need for, and initiation of, more thorough monitoring of LTOT harms and benefits for the patient (SET1) and the need for more intensive patient-centred decision-making before initiating discontinuation (SET2). The instrument comprised of these tools will thus be a source of clinical knowledge that contributes critical and novel information to the patient's comprehensive treatment plan. The project uses qualitative and quantitative methods for tool development with input from patients, providers and leadership within the VA and other institutions. The linkage of VA CDW data to patient survey data will allow for efficient use of the project's primary data by leveraging the extensive data available from electronic health records.

The most important limitation of the proposed project is the lack of 'gold standard' measures for validity testing or longitudinal data to establish the predictive validity of the tools for the presence of LTOT's harms or harms of tapering to discontinuation. To offset this limitation, we will collect multiple measures of potential consequences of LTOT or tapering to discontinue, and link survey responses to rich VA CDW data on patients' diagnoses and healthcare utilisation. As with any health survey, the quality of data collected will depend on the response rate and our ability to adjust for any systematic non-response using available data. Although the survey sampling frame will be national in scope, the sample size will limit the ability to explore associations between SET1 or SET2 screening status and characteristics or conditions within patient subgroups. Finally, the patient survey will necessarily include sensitive questions about opioid use and related experiences; although the interviewers will be thoroughly trained (eg, be non-judgmental, assure respondents that information will be kept confidential), patients may nonetheless have intentional or inadvertent errors in reporting.

In conclusion, this protocol outlines the design of multiple steps in developing integrated screening tools for harms of opioid continuation and tapering to discontinuation. These tools are intended for use in primary care, pain and substance use disorder clinics. Potential next steps will be to examine the extent to which use of the screening tools is associated with improved patient outcomes (eg, fewer overdoses), healthcare system utilisation (eg, fewer emergency department visits and hospitalisations for opioid-related acute crises) and costs. Another future avenue

will be to partner on quality improvement projects to implement the tools in health systems' primary care and pain clinics. For example, mixed methods formative evaluations may be needed to identify contextual factors that will determine implementation strategies for the SET1 and SET2 tools. Addressing the needs of patients receiving LTOT will likely require new clinical practice guidelines that recognise patients who are at risk of LTOT harms even without meeting diagnostic criteria for opioid use disorder (ie, screen positive on SET1), yet would also be at risk of harms if tapering to discontinuation were initiated (also screen positive on SET2).

Ethics and dissemination

Ethical approval was obtained from the Institutional Review Board of Stanford University for the VA Palo Alto Health Care System, and the University of California, San Francisco for the VA San Francisco Healthcare System. Findings will be disseminated through peer-reviewed manuscripts and presentations at research conferences. In addition, we will disseminate findings to veteran patients and family members during regularly scheduled meetings with local Veterans' Councils and annual VA Research Week events at which researchers meet directly with veterans and caregivers to exchange needs and ideas, and using local VA social media feeds.

Author affiliations

¹Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, California, USA

²Department of Veterans Affairs, Menlo Park, California, USA

³Indiana University Center for Health Services and Outcomes Research, Indianapolis, Indiana, USA

⁴Program Evaluation and Resource Center, Department of Veterans Affairs, Menlo Park, California, USA

⁵Veterans Health Administration, Washington, DC, USA

⁶Emory University School of Medicine, Atlanta, Georgia, USA

⁷George Washington University, Washington, DC, USA

⁸Department of Medicine, University of California, San Francisco, California, USA

⁹Department of Veterans Affairs, San Francisco, California, USA

Contributors CT, KK, EO, AN and KH were major contributors to the study design and conceptualisation. KD and FS originated and are guiding the collaborative project. MCL is making substantial contributions to the acquisition and analysis of data for the work. CT drafted the initial manuscript which was read and revised for important intellectual content by all authors, who also approved the final manuscript. All authors agree to be accountable for all aspects of the work.

Funding This research was supported by the US Department of Veterans Affairs (VA), Health Services Research and Development (HSR&D) Service (IIR 18-253 to CT and KH; RCS 00-001 to CT).

Competing interests None declared.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sets meeting standards of the US Department of Veterans Affairs, Veterans Health Administration (VA), for disclosure to the public will be made available post publication. Guidance on request and distribution processes will be provided by VA Office of Research and Development.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Christine Timko <http://orcid.org/0000-0002-6271-7141>

REFERENCES

- Karmali RN, Bush C, Raman SR, *et al*. Long-term opioid therapy definitions and predictors: a systematic review. *Pharmacoepidemiol Drug Saf* 2020;29:252–69.
- Mojtabai R. National trends in long-term use of prescription opioids. *Pharmacoepidemiol Drug Saf* 2018;27:526–34.
- Hadlandsmayth K, Mosher H, Vander Weg MW, *et al*. Decline in prescription opioids attributable to decreases in long-term use: a retrospective study in the veterans health administration 2010–2016. *J Gen Intern Med* 2018;33:818–24.
- Stokes A, Berry KM, Hempstead K, *et al*. Trends in prescription analgesic use among adults with musculoskeletal conditions in the United States, 1999–2016. *JAMA Netw Open* 2019;2:e1917228.
- Dowell D, Haegerich TM, Chou R. *Cdc guideline for prescribing opioids for chronic pain*. Atlanta, GA: CDC, 2016.
- Krebs EE, Bergman AA, Coffing JM, *et al*. Barriers to guideline-concordant opioid management in primary care—a qualitative study. *J Pain* 2014;15:1148–55.
- Lembke A, Humphreys K, Newmark J. Weighing the risks and benefits of chronic opioid therapy. *Am Fam Physician* 2016;93:982–90.
- Manchikanti L, Abdi S, Atluri S, *et al*. American Society of Interventional Pain Physicians (ASIPP) guidelines for responsible opioid prescribing in chronic non-cancer pain: Part I—evidence assessment. *Pain Physician* 2012;15:S1–65.
- Centers for Disease Control and Prevention. *Quality improvement and care coordination: implementing the CDC guideline for prescribing opioids for chronic pain*. Atlanta, GA: National Center for Injury Prevention and Control, Division of Unintentional Injury Prevention, 2018.
- Krebs EE, Ramsey DC, Milosheff JM, *et al*. Primary care monitoring of long-term opioid therapy among veterans with chronic pain. *Pain Med* 2011;12:740–6.
- Starrels JL, Becker WC, Weiner MG, *et al*. Low use of opioid risk reduction strategies in primary care even for high risk patients with chronic pain. *J Gen Intern Med* 2011;26:958–64.
- Morasco BJ, Duckart JP, Dobscha SK. Adherence to clinical guidelines for opioid therapy for chronic pain in patients with substance use disorder. *J Gen Intern Med* 2011;26:965–71.
- Black E, Khor KE, Demirkol A. Responsible prescribing of opioids for chronic non-cancer pain: a scoping review. *Pharmacy* 2020;8:150.
- Cross AJ, Buchbinder R, Bourne A, *et al*. Barriers and enablers to monitoring and deprescribing opioid analgesics for chronic non-cancer pain: protocol for a qualitative evidence synthesis using the theoretical domains framework. *BMJ Open* 2019;9:e034039.
- Becker WC, Fiellin DA, Black AC, *et al*. Initial development of patient-reported instrument assessing harm, efficacy, and misuse of long-term opioid therapy. *J Rehabil Res Dev* 2016;53:127–36.10.1682/JRRD.2014.11.0285
- Manhapra A, Arias AJ, Ballantyne JC. The conundrum of opioid tapering in long-term opioid therapy for chronic pain: a commentary. *Subst Abuse* 2018;39:152–61.
- Berna C, Kulich RJ, Rathmell JP. Tapering long-term opioid therapy in chronic noncancer pain: evidence and recommendations for everyday practice. *Mayo Clin Proc* 2015;90:828–42.
- Petrosky E, Harpaz R, Fowler KA, *et al*. Chronic pain among suicide decedents, 2003 to 2014: findings from the National violent death reporting system. *Ann Intern Med* 2018;169:448–55.
- Henry SG, Paterniti DA, Feng B, *et al*. Patients' experience with opioid tapering: a conceptual model with recommendations for clinicians. *J Pain* 2019;20:181–91.
- Kroenke K, Alford DP, Argoff C, *et al*. Challenges with implementing the centers for disease control and prevention opioid guideline: a consensus panel report. *Pain Med* 2019;20:724–35.
- Darnall BD, Juurlink D, Kerns RD, *et al*. International stakeholder community of pain experts and leaders call for an urgent action on forced opioid tapering. *Pain Med* 2019;20:429–33.
- Bennett AS, Elliott L, Golub A. Opioid-involved overdose among male Afghanistan/Iraq-era US military veterans. *Subst Use Misuse* 2017;52:1701–11.
- Dowell D, Haegerich T, Chou R. No shortcuts to safer opioid prescribing. *N Engl J Med Overseas Ed* 2019;380:2285–7.
- Watch HR. *Not allowed to be compassionate: chronic pain, the overdose crisis, and unintended harms in the US*. New York, NY: Human Rights Watch, 2018.
- Ashton-James CE, Chemke-Dreyfus A, Costa D, *et al*. Time for change: an experimental investigation of chronic pain patients' emotional and attitudinal responses to simulated opioid-tapering advice. *Pain* 2019;160:1586–93.
- FDA. *Fda identifies harm reported from sudden discontinuation of opioid pain medicines and requires label changes to guide prescribers on gradual, individualized tapering*. Silver Spring, MD: FDA Drug Safety Communication, 2019.
- Darnall BD, Colloca L. Optimizing placebo and minimizing nocebo to reduce pain, catastrophizing, and opioid use. *Int Rev Neurobiol* 2018;139:129–57.
- Azam MA, Weinrib AZ, Montbriand J. Acceptance and commitment therapy to manage pain and opioid use after major surgery. *Can J Pain* 2017;1:37–49.
- Sullivan MD, Turner JA, DiLodovico C, *et al*. Prescription opioid taper support for outpatients with chronic pain: a randomized controlled trial. *J Pain* 2017;18:308–18.
- Vorster AC, Walker SP, Esterhuysen KGF. Coping responses as predictors of psychosocial functioning amongst individuals suffering from chronic pain. *Southern African Journal of Anaesthesia and Analgesia* 2009;15:25–30.
- Jensen MP, Turner JA, Romano JM. Changes after multidisciplinary pain treatment in patient pain beliefs and coping are associated with concurrent changes in patient functioning. *Pain* 2007;131:38–47.10.1016/j.pain.2006.12.007
- Demidenko MI, Dobscha SK, Morasco BJ, *et al*. Suicidal ideation and suicidal self-directed violence following clinician-initiated prescription opioid discontinuation among long-term opioid users. *Gen Hosp Psychiatry* 2017;47:29–35.
- Frank JW, Levy C, Matlock DD, *et al*. Patients' perspectives on tapering of chronic opioid therapy: a qualitative study. *Pain Med* 2016;17:1838–47.
- Ashrafioun L, Allen KD, Pigeon WR. Utilization of complementary and integrative health services and opioid therapy by patients receiving veterans health administration pain care. *Complement Ther Med* 2018;39:8–13.
- Beehler GP, Rodrigues AE, Mercurio-Riley D, *et al*. Primary care utilization among veterans with chronic musculoskeletal pain: a retrospective chart review. *Pain Med* 2013;14:1021–31.
- VA/DoD. *VA/DoD clinical practice guideline for opioid therapy for chronic pain*. Washington DC: VA, 2017.
- Streiner DL. *Health measurement scales: a practical guide to their development and use*. New York, NY: Oxford, 2015.
- DeWalt DA, Rothrock N, Yount S, *et al*. Evaluation of item candidates: the promis qualitative item review. *Med Care* 2007;45:S12e21.
- Reeve BB, Hays RD, Bjorner JB, *et al*. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the patient-reported outcomes measurement information system (PROMIS). *Med Care* 2007;45:S22–31.
- Guest G, Namey E, McKenna K. How many focus groups are enough? *Field Methods* 2017;9:3–22.
- Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. *Qual Health Res* 2016;26:1753–60.
- Morgan DL, Groups F, Interaction S. *The sage handbook of interview research*. 2 edn. Thousand Oaks, CA: Sage, 2012: 161–76.
- Nevedal AL, Reardon CM, Opra Widerquist MA, *et al*. Rapid versus traditional qualitative analysis using the consolidated framework for implementation research (CFIR). *Implementation Sci* 2021;16:67.
- Allen J, Dyas J, Jones M. Building consensus in health care: a guide to using the nominal group technique. *Br J Community Nurs* 2004;9:110–4.
- Linstone H, Turoff M. *The Delphi method. reading*. MA: Addison-Wesley, 1975.
- McDowell I. *Measuring health: a guide to rating scales and questionnaires*. NY: Oxford, 2014.
- Beatty PC, Willis GB. Research synthesis: the practice of cognitive interviewing. *Public Opin Q* 2007;71:287–311.
- Willis GB, Artino AR. What do our Respondents think we're asking? using cognitive interviewing to improve medical education surveys. *J Grad Med Educ* 2013;5:353–6.
- Averill JB. Matrix analysis as a complementary analytic strategy in qualitative inquiry. *Qual Health Res* 2002;12:855–66.



- 50 VA. *Opioid decision tool*. Washington DC: VA, 2016.
- 51 Donovan DM, Hatch-Maillette MA, Phares MM, *et al*. Lessons learned for follow-up phone booster counseling calls with substance abusing emergency department patients. *J Subst Abuse Treat* 2015;50:67–75.
- 52 Wallace L, Wexler R, Miser W, *et al*. Development and validation of the patient opioid education measure. *J Pain Res* 2013;2013:663–81.
- 53 Hasin D, Shmulewitz D, Stohl M, *et al*. Test-Retest reliability of DSM-5 substance disorder measures as assessed with the PRISM-5, a clinician-administered diagnostic interview. *Drug Alcohol Depend* 2020;216:108294.
- 54 Krebs EE, Lorenz KA, Bair MJ, *et al*. Development and initial validation of the peg, a three-item scale assessing pain intensity and interference. *J Gen Intern Med* 2009;24:733–8.
- 55 Jamison RN, Dorado K, Mei A, *et al*. Influence of opioid-related side effects on disability, mood, and opioid misuse risk among patients with chronic pain in primary care. *Pain Rep* 2017;2:e589.
- 56 Perlin J, Kazis KM, Skinner X. *Health status and outcomes of veterans*. Washington DC: OQP, 2000.
- 57 Kazis LE, Miller DR, Skinner KM, *et al*. Patient-reported measures of health: the veterans health study. *J Ambul Care Manage* 2004;27:70–83.
- 58 Lecrubier Y, Sheehan DV, Weiller E, *et al*. The mini international neuropsychiatric interview (mini). a short diagnostic structured interview: reliability and validity according to the CIDI. *Eur. psychiatr.* 1997;12:224–31.
- 59 Sheehan DV, Lecrubier Y, Harnett Sheehan K, *et al*. The validity of the mini international neuropsychiatric interview (mini) according to the SCID-P and its reliability. *European Psychiatry* 1997;12:232–41.
- 60 Kroenke K, Spitzer RL, Williams JBW. The PHQ-9. *J Gen Intern Med* 2001;16:606–13.
- 61 Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. *Psychiatr Ann* 2002;32:509–15.
- 62 Handelsman L, Cochrane KJ, Aronson MJ, *et al*. Two new rating scales for opiate withdrawal. *Am J Drug Alcohol Abuse* 1987;13:293–308.
- 63 Cacciola JS, Alterman AI, DePhillippis D, *et al*. Development and initial evaluation of the brief addiction monitor (bam). *J Subst Abuse Treat* 2013;44:256–63.
- 64 AGJ B, Becker SJE, Bruijnzeel H. Creation of the abbreviated measures of the pain catastrophizing scale and the short health anxiety inventory: the PCS-4 and SHAI-5. *J Musculoskelet Pain* 2014;22:145–51.
- 65 Moos RH. Coping responses inventory. In: *Psychological assessment resources*, 2004.
- 66 Feinberg T, Jones DL, Lilly C, *et al*. The complementary health approaches for pain survey (CHAPS): validity testing and characteristics of a rural population with pain. *PLoS One* 2018;13:e0196390.
- 67 Crowne DP, Marlowe D. A new scale of social desirability independent of psychopathology. *J Consult Psychol* 1960;24:349–54.
- 68 Reynolds WM. Development of reliable and valid short forms of the marlowe-crowne social desirability scale. *J Clin Psychol* 1982;38:119–25.
- 69 Kroenke K, Monahan PO, Kean J. Pragmatic characteristics of patient-reported outcome measures are important for use in clinical practice. *J Clin Epidemiol* 2015;68:1085–92.
- 70 Goetz C, Coste J, Lemetayer F, *et al*. Item reduction based on rigorous methodological guidelines is necessary to maintain validity when shortening composite measurement scales. *J Clin Epidemiol* 2013;66:710–8.
- 71 Mitchell AJ, Coyne JC. Do ultra-short screening instruments accurately detect depression in primary care? a pooled analysis and meta-analysis of 22 studies. *Br J Gen Pract* 2007;57:144–51.
- 72 Butler SF, Budman SH, Fernandez KC, *et al*. Development and validation of the current opioid misuse measure. *Pain* 2007;130:144–56.
- 73 Stephenson MT, Hoyle RH, Palmgreen P, *et al*. Brief measures of sensation seeking for screening and large-scale surveys. *Drug Alcohol Depend* 2003;72:279–86.
- 74 Marino EN, Rosen KD, Gutierrez A, *et al*. Impulsivity but not sensation seeking is associated with opioid analgesic misuse risk in patients with chronic pain. *Addict Behav* 2013;38:2154–7.
- 75 Obuchowski NA, Lieber ML, Wians FH. ROC curves in clinical chemistry: uses, misuses, and possible solutions. *Clin Chem* 2004;50:1118–25.