

Exploring Alternative Approaches to Routine Outpatient Pain Screening

Karl A. Lorenz, MD, MSHS,*†‡ Erin E. Krebs, MD, MPH,§¶** Tanya G. K. Bentley, PhD,†
Cathy D. Sherbourne, PhD,‡ Joy R. Goebel, RN, PhD,†† Lisa Zubkoff, PhD,* Andy B. Lanto, MS,*
and Steven M. Asch, MD, MPH*†‡

*Veterans Administration Greater Los Angeles Healthcare System, Division of General Internal Medicine, Los Angeles, California; †Geffen School of Medicine at UCLA, Los Angeles, California; ‡UCLA Center for Health Policy Research, Los Angeles, California; §Center on Implementing Evidence-Based Practice, Roudebush Veterans Administration Medical Center, Indianapolis, Indiana; ¶Indiana University School of Medicine, Indianapolis, Indiana; **Regenstrief Institute, Inc., Indianapolis, Indiana; ††California State Long Beach Department of Nursing, Long Beach, California, USA

ABSTRACT

Objective. To evaluate potential alternatives to the numeric rating scale (NRS) for routine pain screening.

Design. Cross-sectional.

Setting. Nineteen Veterans Affairs outpatient clinics in Southern California at two hospitals and six community sites.

Patients. Five hundred twenty-eight veterans from primary care, cardiology, and oncology clinics sampled in proportion to the total number of visits made to each clinic during the previous year.

Methods. Veterans were approached following clinic visits to complete researcher-administered surveys about their clinic experience. Using the Brief Pain Inventory (BPI) interference scale of ≥ 5 as a reference standard for important unrelieved pain, we evaluated potential alternative pain screening items and item combinations by analyzing sensitivity and specificity, area under the receiver operating curve (AUC), and likelihood ratios.

Results. Of the veterans, 43.6% had unrelieved pain as measured by the reference standard. Approximately half had painful musculoskeletal diagnoses and one-third had comorbid mental health or substance use disorders. The fifth vital sign detected pain less accurately than did an NRS with a 1-week timeframe and an item assessing pain-related bother over the past week. AUCs were 0.79, 0.86, and 0.86, respectively. A sequential approach combining the pain-related bother and NRS with a 1-week timeframe items had good discriminatory ability.

Conclusions. Alternative single or combined pain screening strategies assessing pain-related bother may improve routine pain detection.

Key Words. Ambulatory Care; Veterans; Pain; Measurement

Reprint requests to: Karl A. Lorenz, MD, MSHS, VA Greater Los Angeles Healthcare System, Division of General Internal Medicine, 11301 Wilshire Blvd, Code 111-G, Los Angeles, CA 90073, USA. Tel: 310-478-3711 x43523; Fax: 310-268-4933; E-mail: Karl.lorenz@med.va.gov

Introduction

The Veterans Affairs (VA) implemented routine pain screening with hopes that it would promote increased awareness and improved management of pain. The VA mandated system-wide screening by designating a single-item, 0–10

numeric rating scale (NRS) for current pain as the “fifth vital sign” in 2003 [1]. In the outpatient setting, all veterans are screened for pain at the same time other vital signs (e.g., blood pressure) are taken. Similarly, California requires routine pain screening and accrediting organizations including the Joint Commission for the Accreditation of Healthcare Organizations emphasize it [2,3]. On the other hand, some have criticized routine pain screening as insufficiently evidence-based or potentially harmful [4,5].

One study found that implementation of routine pain screening using the NRS at an Oregon VA hospital-based primary care clinic did not lead to changes in documented pain management processes [6]. Two subsequent studies, including one at a university-based primary care clinic and a one conducted at VA community and hospital-based primary and specialty clinics (the Helping Veterans Experience Less Pain or HELP-Vets Study), found that routine pain screening using the NRS is only modestly accurate for identifying ambulatory patients with important unrelieved pain [7,8].

On potential reason for the insensitivity of routine pain screening in these studies is that all were conducted in outpatient primary and specialty care settings where chronic persistent or intermittent pain is much more common than acute pain. Routine pain screening that focuses on pain intensity “now” may not be sufficiently sensitive to detect important chronic pain that occurs episodically or varies with activity. In the VA primary care, the vast majority of pain problems are longstanding in nature, so sensitivity for chronic pain is important for any pain screening strategy in this setting [9].

Previous studies suggest another potential reason for the insensitivity of routine pain screening. HELP-Vets found that the NRS performed better when it was asked in a standardized fashion by research personnel than it did as a screening procedure in practice. In half of encounters, staff personalized routine pain screening (e.g., “Mr. Jones, do you hurt?”) instead of using the formal 0–10 scale; this informal practice was associated with pain underestimation [8]. Thus, improving fidelity to the measure or shifting to patient self-administration might improve the accuracy of pain screening. Different pain screening strategies using alternate item combinations also deserve consideration, especially if they prove easier to use, provide clinically relevant information, or translate better to routine practice [10].

We therefore set out to characterize alternative routine screening strategies to the fifth vital sign for identifying important unrelieved pain. As a reference standard, we compared each strategy with the interference scale of the Brief Pain Inventory (BPI), a reliable and valid measure for evaluating pain.

Methods

HELP-Vets set out to characterize routine pain screening among a random outpatient visit-based sample of patients from March 2006 to June 2007 at two hospitals and six affiliated community sites in three large counties (Los Angeles, Ventura, and Orange). Of 19 participating clinics, five offer oncology and cardiology services and 14 offer primary care. HELP-Vets had two components, only one of which is addressed in the current analysis. We first determined the total number of visits at each of the study sites during the year before the study. We then targeted each clinic with the goal of obtaining a total number of surveys in each clinic that reflected the proportion of total visits at each clinic and that would also allow us to reach our overall sample size goal. We supplemented that with a convenience sample of cardiology outpatients to evaluate pain in cardiac conditions ($N = 122$). This analysis focuses on the 528 patients in our proportional visit-based sample.

Research assistants approached veterans leaving participating clinics after their provider visits. To be eligible, veterans had to have vital signs taken and be examined by a consenting treatment provider that day, pass a brief cognitive screening test [11], and speak English. In order to sample adequate subjects with painful health conditions, yet include healthy subjects, we selected all veterans who self-reported their health as fair or poor and selected every other veteran among those who self-reported their health as excellent, very good, or good. Research assistants interviewed eligible veterans in person, immediately after they provided consent. We abstracted chart review data from the VA electronic health record.

Of the 6138 patients approached in clinic waiting rooms, 862 refused screening and 4337 were ineligible, 2265 of whom had not had vital signs taken when they were approached, 942 of whom were not in the clinic for a clinician visit, 310 of whom were not visiting participating clinics, 171 of whom exhibited behavioral problems (e.g., aggression toward an interviewer) or

failed the cognitive screening test, 103 of whom were visiting nonparticipating providers, 61 of whom had previously participated, and 49 of whom had hearing impairment. Among the remaining 939 eligible patients, 650 (69.2%) completed the interview. The 528 from the proportional visit-based sample were included in this analysis. Further details of HELP-Vets methods are available [8].

Measures

Measures were derived from established surveys or developed based on theoretical constructs and empiric qualitative preparatory work. A final survey was piloted for clarity and accuracy, including item completion rates, skip patterns, and informal cognitive testing to ensure item comprehension [8].

No “gold standard” for pain exists, so we chose to focus on the ability of pain screening measures to detect functionally-limiting pain, which we refer to as “important unrelieved pain” throughout this manuscript. Other forms of pain (e.g., postoperative pain, mild pain that doesn’t interfere with function) may also be clinically important, but they are not the focus of this study conducted in the chronic outpatient care setting.

BPI—Reference Standard

We chose the BPI interference score as our reference for important unrelieved pain because it is a generic measure of pain-related function that has been validated in numerous pain conditions [12–15]. The BPI interference score includes seven items (range 0–10) on which patients rate how much pain interfered with general activity, mood, walking ability (or ability to transport in a wheelchair or scooter), normal work activities (both outside the house and housework), interpersonal relations, sleep, and enjoyment of life during the past week. The BPI interference score is the average of these seven items. Consistent with previous studies, we used a score of ≥ 5 on the BPI interference scale as the cutoff for moderate or greater pain interference [16,17].

Alternative Pain Screening Approaches

Current NRS and Average NRS

Research assistants asked patients to rate their current pain, using language identical to the standard NRS fifth vital sign screening tool: “On a scale of 0 to 10, where 0 means no pain and 10 equals the worst possible pain, what is your

current pain level?” VA policy regards pain intensity of 4 or greater as warranting additional evaluation [18]. We also tested a 0–10 NRS for “average pain during the last week,” which is an item from the BPI intensity scale [16].

Pain Bother

Bother is a patient-centered, widely used general construct to evaluate pain and other symptoms’ effects on health-related quality of life. We asked the patients, “How much did overall pain distress or bother you during the past week?” using a response set of not at all, a little bit, somewhat, quite a bit, and very much [19].

Pain Importance

We developed an item to directly elicit clinical importance from the patient’s perspective, using an approach from similar pain and symptoms studies [7]. We queried patients about sites of pain, and if they reported any pain, asked for each pain location, “In your opinion, how important was it for your doctor or treatment provider to address that pain today?” using a response set of not at all important, somewhat important, very important, and extremely important.

Unacceptable Pain and Inadequate Relief

We created two exploratory items from constructs identified in previous qualitative research, which identified pain tolerability and minimizing medication use as important pain management endpoints for patients [20]. Using cards and illustrations, we asked patients to rate their average pain intensity on a scale of 0–10 in the last week, then we asked medication users to rate the average decile percentage of relief they experienced on a scale of 0–100. We then asked patients to describe the level of pain intensity and degree of medication relief at which they would “consider treatment successful,” and created difference scores of current intensity minus desired intensity and current relief minus goal relief as measures of Unacceptable Pain (i.e., current > goal intensity) and Inadequate Relief (i.e., goal relief > current relief).

Other Measures

In addition to demographics (age, sex, self-reported race/ethnicity), we used a validated six-item screen to determine the presence of mild cognitive impairment [11]. We assessed self-reported health using a single item (excellent to poor) from the Short Form-12 [21]. We evaluated the presence of mental health conditions using the Patient-Health Questionnaire-2 for depression,

the Generalized Anxiety Disorders-2 for anxiety, and the Post-Traumatic Stress Disorders-2 for PTSD [22–24]. Patients were classified as having probable depression, anxiety, or PTSD if they scored 3–6 on the PHQ-2 or GAD-2 or 4–8 on the PTSD-2. We determined previous diagnoses of alcoholism or other substance abuse and other comorbid conditions from diagnostic codes gathered from chart review.

Analysis

We used BPI-interference ≥ 5 as the reference standard for important unrelieved pain for all analyses. We fit receiver operator characteristic (ROC) curves for each of the pain screening measures given the full range of possible values and calculated the area under the curve (AUC), a measure of overall accuracy (range 0.5 for a worthless test to 1.0 for a perfect test). We used chi-square tests to compare AUCs [25]. We also calculated sensitivity, specificity, and accuracy (the proportion correctly classified) for specific cutoff values for each of the measures. Using this information, we chose cutoff values for the measures, preferentially weighting sensitivity.

We determined multilevel likelihood ratios (LR) for the alternate single and combined pain screening strategies using methods described by Peirce and Cornell [26]. The LR is the probability of a test result among a population with a condition of interest divided by the probability of that test result among a population without the condition. A test result associated with an LR >1.0 increases the probability of a condition, whereas one with an LR <1.0 decreases the probability; if the LR is 1.0, that test result does not change the probability of the condition of interest [27,28].

To better understand the clinical usefulness of the alternative strategies, we also evaluated the preferred single item approaches among veterans in three clinically relevant subgroups, those with: 1) fair-poor self-reported health; 2) an alcohol or other substance abuse diagnosis in the medical record; and 3) self-reported misuse of prescription medications, alcohol, or street drugs to manage pain. We fit the ROC curves for each subgroup and used chi-square tests to compare AUCs for the three strategies.

Results

The mean age of participants was 62 years and 95% were male (Table 1). They were ethnically

Table 1 Characteristics of participants (N = 528)

Characteristic	Mean or %
Age in years (mean)	62
Sex, male (%)	95
Race/ethnicity (%)	
White	52
African American	25
Hispanic	16
Other	7
Mild cognitive impairment* (%)	18
Health status* (%)	
Excellent, very good, or good	48
Fair or poor	52
Comorbid health conditions* (%)	
Cancer	12
Cardiovascular conditions	28
Musculoskeletal conditions	45
Mental health conditions* (%)	
Depression	33
Post-traumatic stress disorder	31
Other anxiety disorder	29
Diagnoses of substance abuse (%)	
Alcoholism	29
Other substance abuse	20

* Measures included the following: cognitive status: Callahan 6-item screener [11]; health status: SF-1 [22]; comorbid conditions: International Statistical Classification of Diseases-9 diagnoses in the chart problem list; depression: PHQ-2 [22]; post-traumatic stress disorder: 4 item screener [24]; other anxiety disorder: GAD-2 [23].

diverse with 48% nonwhite. Approximately half rated their overall health as fair or poor. Musculoskeletal conditions were present in nearly half, and approximately one-third of the cohort suffered from one or more of comorbid depression, anxiety, and substance use disorders. The percentage of participants with important unrelieved pain according to our reference standard (BPI interference ≥ 5) was 43.6%.

Single-Measure Pain Screening Strategies

Figure 1 shows ROC curves for the three best performing alternate pain screening strategies, Current NRS, Average NRS, and Pain Bother, compared with BPI interference ≥ 5 as the reference standard. The AUC for the Current NRS (i.e., the fifth vital sign) was 0.78 (95% confidence interval [CI] = 0.75, 0.82), indicating fair accuracy. The AUCs for both Average NRS and Pain Bother were 0.86 (95% CI = 0.83, 0.89); both had significantly better accuracy than the current routine screening approach ($P < 0.001$ for both compared with Current NRS).

Test characteristics for three best performing single item screening strategies are shown in Table 2. We present information on cutoff values of 3, 4, and 5 for the two NRS screeners. Average NRS was more sensitive and slightly less specific

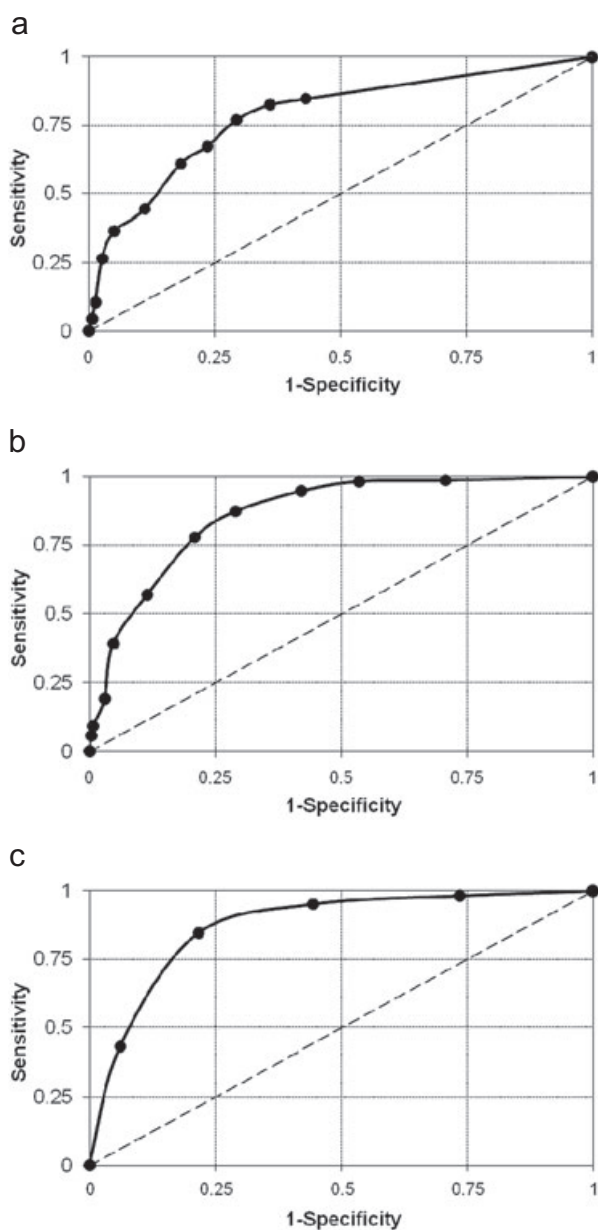


Figure 1 Receiver operator characteristic (ROC) curves for alternate pain screening strategies compared with Brief Pain Inventory interference ≥ 5 as the reference standard. (a) Current pain rating (Current numeric rating scale [NRS]). (b): Average pain over the last week rating (Average NRS). (c) Overall Pain Bother over the last week (Pain Bother).

than Current NRS at each threshold value. A cutoff of “somewhat” for Pain Bother was highly sensitive with marginal specificity. Using a cutoff of “quite a bit” substantially improved specificity at the expense of sensitivity.

The other single item screeners did not perform as well. Pain Importance did not discrimi-

nate well between participants with important unrelieved pain and those with less severe pain (those with no pain were not asked the Pain Importance question). Pain Importance had a low AUC at 0.66 and LRs were near 1.0 for both “somewhat” and “very” important cutoffs.

Unacceptable Pain was highly sensitive (93%) and moderately specific (45%). Sixteen of 513 veterans with complete data for the measure indicated that their current pain was lower than goal pain. Veterans who reported scores in the unexpected direction did not clearly differ from others in terms of relationship with age or cognition (data not shown).

Inadequate Relief from medications had poor sensitivity (68%) and specificity (45%) and produced frequent unexpected results. Approximately 20% (75/376) of participants with complete data for this measure reported that they were achieving excessive relief from medications, which may indicate that they did not understand the questions.

Multistage Screening Strategies

To assess whether a two-stage strategy might improve the accuracy of pain screening, we explored combination screening strategies using the three best single screeners: Current NRS, Average NRS, and Pain Bother (Table 3). We calculated test statistics for a sequential strategies starting with one of the two most sensitive single item screeners, Pain Bother with a cutoff of “somewhat.” Average NRS with a cutoff of 3 was equally sensitive, but Pain Bother has the advantage of addressing conceptually unique information beyond intensity. In addition, we theorized that a non-numeric item may be easier to integrate into the work flow and pose fewer fidelity problems. We evaluated combinations using both Current and Average NRS scales as the second stage item. We chose to present a cutoff value of 4 or greater for both Current and Average NRS scales because this provided a good balance of sensitivity and specificity and because it is the most commonly used cutoff in practice.

Because Pain Bother was highly sensitive, the LR for important unrelieved pain among those with negative Pain Bother was very low (0.09, 95% CI 0.04, 0.15). Following this item with Current NRS or Average NRS resulted in a higher LR for those with both tests positive than did positive Pain Bother alone. Discordant results (i.e., Pain Bother positive, NRS negative) were less informa-

Table 2 Test characteristics of alternate single measure pain screening strategies for important unrelieved pain*

	N positive (%)	Sensitivity	Specificity	LR+ (95% CI) [†]	LR- (95% CI)
Current NRS[‡] (N = 527)					
3+	264 (50.1)	0.77	0.71	2.63 (2.17, 3.18)	0.33 (0.25, 0.42)
4+	225 (42.7)	0.67	0.76	2.86 (2.29, 3.57)	0.43 (0.35, 0.52)
5+	194 (36.8)	0.61	0.82	3.35 (2.58, 4.35)	0.48 (0.40, 0.57)
Average NRS (N = 527)					
3+	343 (65.1)	0.95	0.58	2.25 (1.97, 2.58)	0.09 (0.05, 0.16)
4+	287 (54.4)	0.87	0.71	3.02 (2.51, 3.63)	0.18 (0.13, 0.25)
5+	241 (45.7)	0.78	0.79	3.73 (2.96, 4.70)	0.28 (0.22, 0.36)
Pain Bother[§] (N = 524)					
Somewhat+	349 (66.7)	0.95	0.56	2.14 (1.88, 2.44)	0.09 (0.05, 0.15)
Quite a bit+	258 (49.2)	0.85	0.78	3.90 (3.12, 4.88)	0.20 (0.14, 0.27)

* Important unrelieved pain defined as BPI interference scale ≥ 5 .

[†] Current NRS is rating of current pain on a 0–10 scale. Average NRS is rating of average pain over the last week on a 0–10 scale.

[‡] "A positive LR is the ratio of those with a positive test who truly have the condition to those with a positive test who do not have the condition. A negative LR is the ratio of those with a negative test who truly don't have the condition to those with a negative test who do have the condition."

[§] "How much did overall pain distress or bother you during the past week?" (response set: not at all, a little bit, somewhat, quite a bit, very much).

BPI = Brief Pain Interference; NRS = numeric rating scale; LR = likelihood ratio; CI = confidence interval.

tive. For the sequential combination of Pain Bother and Average NRS, given a pretest probability of 40%, the post-test probabilities of important unrelieved pain are 5%, 26%, and 68% for negative Pain Bother, positive Pain Bother with negative Average NRS, and positive Pain Bother with positive Average NRS results, respectively. The performance of this combination is had better discriminatory ability than the combination of Pain Bother and Current NRS.

Comparing Strategies among Population Subgroups

Overall, performance of the single item strategies in the three vulnerable population subgroups was similar to performance in the full sample (Table 4). The Current NRS had AUC values of 0.78, 0.76, 0.84, and 0.79 for the full sample and poor health, substance abuse diagnosis, and substance misuse

subgroups, respectively (Table 4). The discriminatory ability of the Average NRS was significantly better than that of the Current NRS in the full sample ($P < 0.001$) and in the poor health ($P = 0.03$) and substance abuse diagnosis ($P = 0.02$) subgroups. Pain Bother had significantly better accuracy than the Current NRS in the full sample ($P < 0.001$) and poor health subgroup ($P = 0.03$). Pain Bother and Average NRS did not differ in their discriminatory ability in the full sample or any subgroup.

Discussion

Using the BPI interference as the reference standard, we found that alternative screening strategies assessing bothersomeness of pain or average pain intensity over a 1-week period, alone or in combination, improved accuracy substantially over the current fifth vital sign. Our measure of

Table 3 Likelihood of important unrelieved pain* using sequential pain screening strategies

Pain Screening Strategy Combinations	LR (95% CI)	Post-Test Probability [†]
Combination 1		
Pain Bother negative [‡]	0.09 (0.04, 0.15)	0.05
Pain Bother positive and Current NRS < 4 [§]	1.24 (0.93, 1.65)	0.45
Pain Bother positive and Current NRS ≥ 4	3.18 (2.50, 4.05)	0.68
Combination 2		
Pain Bother negative	0.09 (0.04, 0.15)	0.05
Pain Bother positive and Average NRS < 4	0.52 (0.33, 0.82)	0.26
Pain Bother positive and Average NRS ≥ 4	3.34 (2.73, 4.07)	0.69

* Important unrelieved pain defined as BPI interference ≥ 5 . A positive LR is the ratio of those with a positive test who truly have the condition to those with a positive test who do not have the condition.

[†] Calculated given a pretest probability of 0.40.

[‡] Positive value for Pain Bother is "somewhat" or greater to the question: "How much did overall pain distress or bother you during the past week?" (response set: not at all, a little bit, somewhat, quite a bit, very much).

[§] Current NRS is rating of current pain on a 0–10 scale. Average NRS is rating of average pain over the last week on a 0–10 scale.

BPI = Brief Pain Interference; NRS = numeric rating scale; LR = likelihood ratio; CI = confidence interval.

Table 4 Receiver operating characteristic area under the curve (AUC) for preferred single item pain screening approaches among population subgroups*

Subgroup	N	Current NRS [†]	Average NRS	Pain Bother [‡]
Full sample	522	0.78	0.86 [§]	0.86 [§]
Poor health	267	0.76	0.81 [§]	0.82 [§]
Substance abuse diagnosis	214	0.84	0.89 [§]	0.87
Substance misuse for pain	132	0.79	0.81	0.80

* Poor health = self-rated health poor or fair; substance abuse diagnosis = prior diagnosis of alcoholism or substance abuse in medical record; substance misuse for pain = self report of alcohol, drugs, or other prescription medications to manage pain.

[†] Current NRS is rating of current pain on a 0–10 scale. Average NRS is rating of average pain over the last week on a 0–10 scale.

[‡] "How much did overall pain distress or bother you during the past week?" (response set: not at all, a little bit, somewhat, quite a bit, very much).

NRS = numeric rating scale; LR = likelihood ratio; CI = confidence interval.

[§] $P = 0.05$.

unacceptable pain was highly sensitive, but numeric response sets did not perform as expected among a significant minority of veterans, probably because they are too complex [29]. Other inferior strategies included asking about the importance of talking to the doctor about pain and the acceptability of current medication.

We evaluated brief single-item and two-item combination strategies for pain screening because we are mindful of the complex challenge of practice implementation. Taking into account the primary goal of improving clinical relevance, minimizing burden for patients and those who administer pain screening is key. The burden of capturing any particular information seems small, but must be considered in the context of expanding patient-reported preventive screening [30]. Both nurses and physicians are more likely to comply with clinically relevant, time efficient tasks [31–33].

Another consideration is that the best screening strategies we identified would change the screening task from identifying current pain intensity to characterizing the impact of pain (i.e., bother) or intensity of pain over time (i.e., average pain over the past week); both may be more relevant than current pain intensity for patients with chronic and episodic pain. The current fifth vital sign approach may not be equally relevant in all clinical settings; specifically, its orientation toward current pain may make it less useful in settings where chronic persistent and episodic pain are more common [34,35]. Although capturing the same fifth vital sign measure across settings may facilitate goals such as quality assessment, varying screening approaches among settings with different patient characteristics may be more clinically useful [36].

We consider it a strength of this study that it was conducted among a diverse group of chroni-

cally ill veterans; however, this does affect the generalizability of our findings. The biggest difference between our cohort and the national veterans' population under care is the extent to which minority veterans were represented. About 80% of veterans nationally are white, whereas nearly half of our veterans were African American, Hispanic, or other minority ethnicity [37].

Limitations of our study include its focus on the outpatient setting, where populations and environmental factors that might influence pain screening performance differ substantially from those in inpatient, emergency, and long-term care. Different settings likely require specific evaluations of pain screening approaches. Our findings are likely most applicable to chronic pain because of the care setting and our chosen reference standard, which assesses functional interference due to pain over the past week. An important limitation is that our conclusions about how pain screening strategies might perform in screening practice are derived from a cross-sectional researcher-administered survey. Determining the clinical appropriateness of these strategies will require a longitudinal trial of actual screening strategies. Finally, we did not adjust secondary analyses of population subgroups for multiple comparisons and that for this reason, some of the findings in this regard might be considered preliminary.

In conclusion, our findings suggest that screening incorporating pain-related bother or average pain intensity over the past week is desirable for populations in poor health. In the strategy combining both average pain and pain-related bother, both contributed unique information.

Our analysis suggests promising strategies to evaluate in a randomized controlled trial of improved pain screening that includes clinically meaningful outcomes.

Acknowledgments

This study was supported by a grant from the VA Health Services Research and Development (HSRD) Service (IIR 03-150). Dr. Erin Krebs is supported by a VA HSRD Career Development Award.

Disclosures

The authors have no additional relevant financial disclosures to report.

References

- Department of Veterans Affairs. VHA directive 2003-021: Pain management. Vol. 2003. 2003. Available at: http://www1.va.gov/pain_management/docs/VHAPainDirective_03.pdf (accessed October 14, 2007).
- Chronic Pain Policy Coalition. Pain: The 5th vital sign. Available at: <http://www.paincoalition.org.uk/pain5.html> (accessed July 31, 2008).
- JCAHO standards for pain management. Available at: <http://www.jointcommission.org> (accessed July 31, 2008).
- Lucas CE, Vlahos AL, Ledgerwood AM. Kindness kills: The negative impact of pain as the fifth vital sign. *J Am Coll Surg* 2007;205:101-7.
- Kozol RA, Voytovich A. Misinterpretation of the fifth vital sign. *Arch Surg* 2007;142(5):417-19.
- Mularski RA, White-Chu F, Overbay D, et al. Measuring pain as the 5th vital sign does not improve quality of pain management. *J Gen Intern Med* 2006;21(6):607-12.
- Krebs EE, Carey TS, Weinberger M. Accuracy of the pain numeric rating scale as a screening test in primary care. *J Gen Intern Med* 2007;22(10):1453-8.
- Lorenz KA, Sherbourne CD, Shugarman LR, et al. Reliable is Pain as the 5th Vital Sign? *J Am Board Fam Med* 2009;22(3):291-8.
- Reid M, Carrington C, Crone KT, Otis J, Kerns RD. Differences in pain-related characteristics among younger and older veterans receiving primary care. *Pain Med* 2002;3(2):102-7.
- Krebs EE, Lorenz KA, Bair MJ, et al. Development and initial validation of the PEG, a 3-item scale assessing pain severity and functional interference. *J Gen Intern Med* 2009;24(6):733-8.
- Callahan CM, Unverzagt FW, Hui SL, Perkins AJ, Hendrie HC. Six-item screener to identify cognitive impairment among potential subjects for clinical research. *Med Care* 2002;40(9):771-81.
- Cleeland CS, Ryan KM. Pain assessment: Global use of the Brief Pain Inventory. *Ann Acad Med Singapore* 1994;23:129-38.
- Cleeland CS, Nakamura Y, Mendoza TR, et al. Dimensions of the impact of cancer pain in a four country sample: New information from multidimensional scaling. *Pain* 1996;67:267-73.
- Keller S, Bann DM, Dodd SL, et al. Validity of the Brief Pain Inventory for use in documenting the outcomes of patients with noncancer pain. *Clin J Pain* 2004;20:309-18.
- Tan G, Jensen MP, Thornby JL, Shanti BF. Validation of the Brief Pain Inventory for chronic nonmalignant pain. *J Pain* 2004;5:133-7.
- Cleeland CS, Ryan KM. Pain assessment: Global use of the Brief Pain Inventory. *Ann Acad Med Singapore* 1994;23(2):129-38.
- Williams VS, Smith MY, Fehnel SF. The validity and utility of the BPI interference measures for evaluating the impact of osteoarthritic pain. *J Pain Symptom Manage* 2006;31(1):48-57.
- Department of Veterans Affairs. Pain as the 5th vital sign toolkit. Available at: <http://www.vachronicpain.org/Downloads/pain5thvitalsignv2.pdf> (accessed October 17, 2007).
- Ganz PA, Goodwin PJ. Health-related quality of life measurement in symptom management trials. *J Natl Cancer Inst Monogr* 2007;37:47-52.
- Casarett D, Karlawish J, Sankar P, Hirschman K, Asch DA. Designing pain research from the patient's perspective: What trial endpoints are important to patients with chronic pain? *Pain Med* 2001;2(4):309-16.
- Melville MR, Lari MA, Brown N, Young T, Gray D. Quality of life assessment using the short form 12 questionnaire is as reliable and sensitive as the short form 36.
- Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: Validity of a two-item depression screener. *Med Care* 2003;41(11):1284-92.
- Kroenke K, Spitzer RL, Williams JB, Monahan PO, Lowe B. Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. *Ann Intern Med* 2007;146(5):317-25.
- Lang AJ, Stein MB. An abbreviated PTSD checklist for use as a screening instrument in primary care. *Beh Res Ther* 2005;43(5):584-94.
- Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York: Oxford University Press; 2004.
- Peirce JC, Cornell RG. Integrating stratum-specific likelihood ratios with the analysis of ROC curves. *Med Decis Making* 1993;13:141-51.
- Fletcher RH, Fletcher SW, Wagner EH. *Clinical Epidemiology: The Essentials*, 3rd edition. Baltimore: Williams & Wilkins; 1996.
- Sackett DL, Straus DE, Richardson WS, Rosenberg W, Haynes RB. *Evidence-Based Medicine: How to Practice and Teach EBM*, 2nd edition. London: Churchill Livingstone; 2000.
- Peters E, Hibbard J, Slovic P, Dieckmann N. Communication, comprehension, and use of risk-benefit information. *Health Affairs* 2007;26(3):741-8.

- 30 Agency for Healthcare Research and Quality. Guide to clinical preventive services, 2008. AHRQ Publication No. 08-05122. Rockville, MD: Author. Available at: <http://www.ahrq.gov/clinic/pocketgd.htm> (accessed January 31, 2009).
- 31 Wu SJ, Lehto M, Yih Y, Saleem JJ, Doebbeling BN. Relationship of estimated resolution time and computerized clinical reminder adherence. *AMIA Annu Symp Proc* 2007;11:334–8.
- 32 Patterson ES, Doebbeling BN, Fung CH, et al. Identifying barriers to the effective use of clinical reminders: Bootstrapping multiple methods. *J Biomed Inform* 2005;38(3):189–99.
- 33 Saleem JJ, Patterson ES, Militello L, et al. Impact of clinical reminder redesign on learnability, efficiency, usability, and workload for ambulatory clinic nurses. *J Am Med Inform Assoc* 2007;14(5):632–40.
- 34 Brockerick JE, Stone AA, Calvanese P, Schwarz JE, Turk DC. Recalled pain ratings: A complex and poorly defined task. *J Pain* 2006;7(2):142–9.
- 35 Kazis LE, Miller DR, Skinner K, et al. Health-related quality of life in patients served by the Department of Veterans Affairs: Results from the Veterans Health Study. *Arch Intern Med* 1998;158(6):626–32.
- 36 National Institutes of Health. Patient-reported outcomes measurement information system. Available at: <http://www.nihpromis.org/default.aspx> (accessed February 5, 2009).
- 37 Department of Veterans Affairs. Demographics: Veteran population 2007. Available at: <http://www1.va.gov/vetdata/page.cfm?pg=15> (accessed April 14, 2009).