The Spine Journal 19 (2019) 293–300

Clinical Study

Sensitivity and specificity of patient-entered red flags for lower back pain

John T. Tsiang, BS[a,b,*], Tyler G. Kinzy, MS[c,d], Nicolas Thompson, MS[c,d], Joseph E. Tanenbaum, BS[a,b], Nitya L. Thakore[c], Tagreed Khalaf, MD[a], Irene L. Katzan, MD, MS[c]

[a] *Center for Spine Health, Neurological Institute, Cleveland Clinic, Cleveland, OH 44195, USA*
[b] *School of Medicine, Case Western Reserve University, Cleveland, Ohio, USA*
[c] *Center for Outcomes Research and Evaluation, Neurological Institute, Cleveland Clinic, Cleveland, Ohio, USA*
[d] *Quantitative Health Sciences, Cleveland Clinic, Cleveland, Ohio, USA*
Received 16 March 2018; revised 12 June 2018; accepted 13 June 2018

Abstract

BACKGROUND CONTEXT: Red flags are questions typically ascertained by providers to screen for serious underlying spinal pathologies. The utility of patient-reported red flags in guiding clinical decision-making for spine care, however, has not been studied.
PURPOSE: The aim of this study was to quantify the sensitivity and specificity of patient-reported red flags in predicting the presence of serious spinal pathologies.
STUDY DESIGN: This was a retrospective nested case-control study.
PATIENT SAMPLE: This study consisted of 120 patients with International Classification of Diseases, Ninth Revision, Clinical Modification codes for spinal pathologies and 380 randomly selected patients, from a population of 4,313 patients seen at a large tertiary care spine clinic between October 9, 2013 and June 30, 2014.
OUTCOME MEASURES: The presence of patient-reported red flags and red flags obtained from medical records was verified for chart review. The spinal pathology (ie, malignancy, fractures, infections, or cauda equina syndrome) was noted for each patient.
METHODS: The sensitivity and specificity of patient-reported red flags for detecting serious spinal pathologies were calculated from data obtained from the 500 patients. Youden's J was used to rank performance. Agreement between patient-reported red flags and those obtained from medical record review was assessed via Cohen's kappa statistic.
RESULTS: "History of cancer" was the best performing patient-reported red flag to identify malignancy (sensitivity=0.75 [95% confidence intervals, CI 0.53–0.90], specificity=0.79 [95% CI 0.75–0.82]). The best performing patient-reported red flag for fractures was the presence of at least one of the following: "Osteoporosis," "Steroid use," and "Trauma" (sensitivity=0.59 [95% CI 0.44–0.72], specificity=0.65 [95% CI 0.60–0.69]). The prevalence of infection and cauda equina diagnoses was insufficient to gauge sensitivity and specificity. Red flags from medical records had better performance than patient-reported red flags. There was poor agreement between patient red flags and those obtained from medical record review.
CONCLUSIONS: Patient-reported red flags had low sensitivity and specificity for identification of serious pathologies. They should not be used in insolation to make treatment decisions, although they may be useful to prompt further probing to determine if additional investigation is warranted.

Keywords: Low back pain; Patient reported; Red flags; Sensitivity; Specificity; Spine

FDA device/drug status: Not applicable.

Author disclosures: ***JTT***: Nothing to disclose. ***TGK:*** Nothing to disclose. ***NT:*** Nothing to disclose. ***JET***: Nothing to disclose. ***NLT***: Nothing to disclose. ***TK:*** Nothing to disclose. ***ILK***: Nothing to disclose.

* Corresponding author. Center for Spine Health, the Cleveland Clinic, 9500 Euclid Ave, S-80, Cleveland, OH 44195, USA. Tel.: 6507730916.
E-mail address: john.tsiang@case.edu (J.T. Tsiang).

https://doi.org/10.1016/j.spinee.2018.06.342
1529-9430/

Introduction

Low back pain is one of most common complaints among patients in developed countries [1], with an estimated 84.1% of adults developing low back pain at some point in their lives [2]. In the United States, low back pain is the leading cause of years lived with disability and is the second most common reason for physician visits [3], with 1.3% of office visits annually attributable to low back pain [4]. These visits amount to a direct medical cost of 34 billion USD [5].

Prior studies demonstrated that over 90% of patients presenting for low back pain experience resolution of pain symptoms within a month [6,7], and less than 1% of these patients have a serious underlying pathology [8]. "Red flags" are a series of questions used to screen low back pain patients for potentially serious underlying pathologies such as malignancy, vertebral fractures, spinal infections, and cauda equina syndrome [9]. Patients that screen positive for red flags warrant further clinical investigation, particularly with imaging or consultation with a specialist [3,10,11]. Indeed, the American College of Radiology Appropriateness Criteria for low back pain uses the presence of red flags to determine if patients need advanced imaging [3]. Delaying treatment for these serious and complex pathologies can result in adverse outcomes for patients, and it is therefore imperative to quickly and correctly identify patients that require further workup and treatment [12–14].

A prior study found that only 5% of primary care physicians regularly screen for the presence of red flags in low back pain patients [15]. Di Iorio et al. found that physicians will only regularly inquire about two out of the seven red flags included in their study [16]. In contrast, patient-entered data do not rely on physician involvement and may therefore represent a more consistent and reliable alternative to screen for the presence of red flags [17–19]. For example, the Quality Outcomes Database, formerly known as the National Neurosurgical Quality Outcomes Database, already collects patient-reported outcomes to analyze the quality of spinal neurosurgical care [20]. This method of collecting red flag information has the capacity to circumvent the characteristically poor documentation of red flags by physicians and could improve the quality of the office visit by reducing the time needed by the physician to review red flags. Although previous studies examined the effectiveness of provider-reported red flags [8,21–23], the effectiveness of patient-reported red flags has not been studied. Therefore, the objective of this study was to evaluate the sensitivity and specificity of patient-reported red flags in a spine clinic setting to identify serious underlying pathology, and whether the use of patient-reported red flags in isolation is able to identify the need for further workup.

Materials and methods and/or case material

Study design

To investigate the sensitivity and specificity of patient-reported red flags, the authors reviewed the medical records of all patients seen in a spine clinic at a large, academic tertiary care medical center between October 9, 2013 and June 30, 2014.

Eligibility criteria for this study included patients who received their primary care within the health system, and had an office visit in the spine center within the study period, and who completed a red flags questionnaire on the specified office visit date. The criterion of receiving primary care within the same health system helps to increase the likelihood that the diagnosis of serious pathology would be recorded within the electronic health record (EHR).

A nested case-control design [24] was employed to increase the proportion of patients in the study sample who had one of following serious pathologies, allowing more accurate sensitivity and specificity estimates (Fig. 1). All patients meeting the eligibility criteria with an encounter diagnosis of International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes corresponding spinal metastasis (198.5), multiple myeloma (203.00), discitis (722.90), osteomyelitis (730.28), vertebral compression fracture (805.8), or cauda equina syndrome (344.60/344.61) in a 1-year window around the specified office visit date were included in this sample. The remainder of the sample population was randomly selected from the remaining pool of patients that did not have ICD-9-CM codes matching the above pathologies.

The study was approved by the Institutional Review Board. Patient consent was not required as this was a deidentified retrospective case-control study.

Patient-reported red flags

As part of routine care, both patient and clinician-reported data are collected in our institution through the

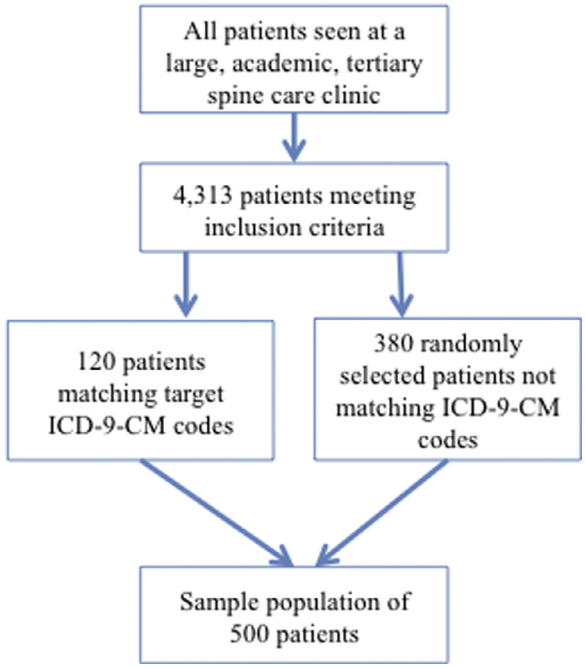

Fig. 1. Flow diagram of study subjects.

Table 1
Red flags questionnaire for patients

Question	Red flag	Target pathology
Have you ever had any type of cancer?	History of cancer	Malignancy
Do you have pain at night that causes you to be unable to fall back asleep?	Night pain	Malignancy, Infection
Does bed rest provide any relief of your back pain?	Pain at rest	Malignancy, Infection
Have you had a recent onset of inability to urinate at all even when you have a full bladder?	Urinary retention	Malignancy, Cauda Equina
Have you recently had an unexplained fever?	Unexplained fever	Infection
Have you ever been diagnosed with osteoporosis?	Osteoporosis	Fracture
Have you ever used corticosteroids for a prolonged duration?	Steroid use	Fracture
Within the past 30 days have you been involved in significant trauma (a fall or a car accident)?	Trauma	Fracture
Have you had a recent onset of urinary or bowel incontinence?	Bowel/Bladder incontinence	Cauda Equina
Have you had a recent onset of numbness and/or tingling in your genitalia, anus, or perineum?	Saddle numbness	Cauda Equina
Do you have progressive weakness in one or both of your legs?	Weakness in limbs	Cauda Equina

electronic platform, the Knowledge Program (KP) [25], for systematic collection of patient-reported information. Questions are administered on tablets at the time of a patient's ambulatory visit or through the EHR patient portal (MyChart, Epic Systems, Verona, Wisconsin) before a patient's appointment. The KP data are immediately available to the provider through the EHR.

The red flags questionnaire consisted of 12 questions administered to the patient as part of the KP (Table 1). The questions were developed for standardized use across our institution through a workgroup of specialists reviewing the limited available literature and are thus based on the consensus of the experts in the screening of low back pain at our institution [26,27].

Electronic health record review

Chart review was manually performed to verify that the ICM-9-CM codes matched the clinical diagnoses of each patient in a 1-year window around the specified visit date. Information was extracted from clinical notes, imaging reports, and diagnoses entered in medical history section of the EHR. Patients with a diagnosis of spinal metastasis or multiple myeloma were classified as having "malignancy" pathology. Patients with benign vertebral compression fractures were classified as having "fracture" pathology. Patients with osteomyelitis or discitis were classified as having "infection" pathology. Last, patients with cauda equina syndrome were classified as having "cauda equina" pathology.

The EHR was also reviewed for documentation of red flags by health-care providers. Red flags were extracted from the physician note completed at the spine visit. If presence or absence of red flags was not documented in the note, the red flag was coded as "not mentioned" in the data set. The red flags "History of cancer," "Osteoporosis," and "Steroid use" were elicited by reviewing prior diagnoses and medications in the EHR. Any history of a neoplasm with malignant potential was recorded as a positive "History of cancer" red flag. A diagnosis of osteopenia was not recorded as a positive "Osteoporosis" red flag. Last, only oral or intravenous steroids were counted for a positive "Steroid use" red flag.

Analysis

Patient characteristics collected at the time of spine center visit were summarized using means and standard deviations for continuous variables, and frequencies and percentages for categorical variables.

The frequency and proportion of endorsed red flags were computed for both patient and provider medical record sources. If a red flag was coded as "not mentioned" in the provider medical record, it was considered as an absent red flag. Sensitivity and specificity were calculated with 95% confidence intervals (CI) using the exact binomial method as measures of diagnostic utility for corresponding spinal pathologies according to Table 1. Combinations of red flags, where one or more of the red flags in the combination were present, were also evaluated. Individual and combination red flags were then ranked by maximum Youden's J statistic (sensitivity+specificity−1) to determine the best possible individual or combination red flag(s) for each diagnosis group. On a receiver operating characteristic plot (Figure 2 a and b), the maximum Youden's J statistic (J=1) is the point closest to the upper left corner that is Cartesian coordinates (0,1). A Youden's J of zero corresponds to a test with zero utility when sensitivity and specificity are weighted equally.

Contingency tables were constructed comparing endorsement of spine red flag by source (patient or medical record) using the medical record as the gold standard. To measure statistical agreement, positive, negative, and overall agreement were calculated, and Cohen's kappa was used. Kappa notably provides biased estimates of agreement when marginal disagreement and agreement are imbalanced between two diagnostic tools. Therefore, the theoretical maximum of kappa, referred to as $\text{kappa}_{\text{max}}$, was calculated. We also reported the proportion of estimated kappa divided by $\text{kappa}_{\text{max}}$; this provides a more accurate measure of agreement when $\text{kappa}_{\text{max}} < 1$ [28]. The proportion can thus be interpreted similarly to Cohen's

kappa: 0.41–0.60, 0.61–0.80, and >0.81 for moderate, substantial, and excellent agreement, respectively [29].

All statistical analyses were completed using R 3.4.0 [30].

Results

Participants

Out of the 4,313 patients that met the inclusion criteria, one hundred and twenty patients matched the target ICD-9 codes. An additional 380 patients were chosen randomly from the remaining pool of patients, for a total of 500 patients in the sample population (Figure 1).

Out of the 500 patients in the final sample, 26 patients (5.2%) were removed caused by lack of follow-up care or not having presented for the specified office visit. The mean age of the remaining 474 patients was 66.8±12.5 years old. There were 353 (74.5%) women in the sample, and 406 (85.7%) patients were white (Table 2).

Patient-reported red flags

Sensitivities and specificities of single red flags are shown in Table 3 (please see Appendix, Tables 1, 2, 3, and 4 for additional information on the performance of various combinations of patient-entered red flags). The patient-reported red flag with the highest sensitivity and specificity for malignancy was "History of cancer" (sensitivity, 0.750 [95% CI, 0.533–0.902] and specificity, 0.787 [95% CI, 0.746–0.824]). The presence of one or more of the following patient-reported red flags—"Osteoporosis," "Steroid use," and "Trauma"—were better for detection of fracture than any other single red flag. This had a sensitivity of 0.585 (95% CI, 0.441–0.719) and a specificity of 0.648 (95% CI, 0.601–0.694). The single patient-reported red flag with the best performance for detection of infection was "Unexplained fever," which had a sensitivity of 0.250 (95% CI, 0.032–0.651) and specificity of 0.976 (95% CI, 0.958–0.988). No patient-reported red flag combination ranked higher in terms of Youden's J. The best patient-reported red flag for detection of cauda equina syndrome was "Bowel and/or Bladder incontinence," which had a sensitivity of 0.500 (95% CI, 0.118–0.882) and specificity of 0.865 (95% CI, 0.831–0.895). Figure 2a shows the receiver operating characteristic plot for the above red flag(s).

Table 2
Descriptive summary of red flags sample demographics (N=474)

Variable	N (%)*
Age (years), *mean (SD)*	66.8 (12.5)
Median income ($1,000s), *median (range)*	57.4 (17.9–127.0)
Women	353 (74.5%)
Race	
White	406 (85.7%)
Black	48 (10.1%)
Other	15 (3.2%)
Unknown and/or refusal	5 (1.1%)
Marital status	
Married	283 (59.7%)
Single	63 (13.3%)
Divorced	37 (7.8%)
Other	83 (17.5%)
Unknown and/or refusal	8 (1.7%)
Tobacco usage	
Current usage (primary or secondary)	46 (9.7%)
Past usage	175 (36.9%)
Never used	198 (41.8%)
Not reported	55 (11.6%)

* Reported as N (%) unless noted otherwise.

Provider-reported red flags

Sensitivities and specificities of single red flags obtained from medical record review for detecting target pathologies are shown in Table 3 (see Appendix, Tables 5, 6, 7, and 8 for additional information on the performance of various combinations of red flags collected from the medical health record). The red flag with the highest sensitivity (0.917 [95% CI 0.730–0.990]) and specificity [0.778 (95% CI 0.736–0.815]), for the detection of malignancy was the single red flag "History of cancer." The best performing red flags for the detection of fractures was the combination "Osteoporosis" and "Trauma"; the presence of one or both of these two red flags had a sensitivity of 0.811 (95% CI 0.680–0.906) and a specificity of 0.791 (95% CI 0.749–0.829) for detecting fractures. "Unexplained fever" had a sensitivity of 0.125 (95% CI 0.003–0.527) and a specificity of 0.996 (95% CI 0.985–0.999) for detecting infections. The combination of "Urinary retention" and/or "Limb weakness" had a sensitivity of 1.000 (95% CI 0.541–1.000) and a specificity of 0.769 (95% CI 0.728–0.807) for detecting cauda equina syndrome. Figure 2b shows the receiver operating characteristic plot for the above red flag(s).

Agreement

The prevalence of red flags varied between those reported by patients and those documented in the medical record (Table 4). Level of agreement between patient and medical record documentation of red flags, measured by Cohen's kappa coefficient, was highest for having "History of cancer" at 0.718 (95% CI, 0.645–0.791), second highest for "Osteoporosis" at 0.419 (95% CI, 0.323–0.515), followed by "Trauma" at 0.358 (95% CI 0.232–0.484). Overall, negative agreement ranged from 0.690 to 0.987, whereas positive agreement ranges from 0.071 to 0.788, with a majority falling below 0.500. When interpreting kappa over $kappa_{max}$, substantial agreement [28] is evident for "History of Cancer" (0.752), "Unexplained Fever" (0.657), "Night Pain" (0.657), and "Weakness in Limbs" (0.652). In the

Table 3
Sensitivities and specificities of individual red flags as reported by patients and as documented in the medical record

	Patient			Medical record	
Target	Red flag	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Malignancy	History of cancer	0.750 (0.533–0.902)	0.787 (0.746–0.824)	0.917 (0.730–0.990)	0.778 (0.736–0.815)
	Night pain	0.542 (0.328–0.744)	0.496 (0.448–0.543)	0.167 (0.0470–0.374)	0.878 (0.844–0.907)
	Pain at rest	0.250 (0.098–0.467)	0.698 (0.653–0.740)	0.042 (0.001–0.0211)	0.933 (0.906–0.955)
	Urinary retention	0.042 (0.001–0.211)	0.958 (0.935–0.974)	0 (0–0.142)	0.973 (0.954–0.986)
Infection	Night pain	0.375 (0.085–0.755)	0.491 (0.445–0.538)	0 (0–0.369)	0.873 (0.840–0.902)
	Pain at rest	0.125 (0.003–0.527)	0.697 (0.653–0.739)	0 (0–0.369)	0.933 (0.907–0.954)
	Fever	0.250 (0.032–0.651)	0.976 (0.958–0.988)	0.125 (0.003–0.527)	0.996 (0.985–0.999)
Fracture	Osteoporosis	0.415 (0.281–0.559)	0.762 (0.719–0.802)	0.585 (0.441–0.719)	0.879 (0.844–0.908)
	Steroid use	0.283 (0.168–0.423)	0.867 (0.831–0.898)	0.245 (0.138–0.383)	0.843 (0.805–0.877)
	Trauma	0.226 (0.123–0.362)	0.938 (0.911–0.959)	0.415 (0.281–0.559)	0.895 (0.862–0.923)
Cauda equina	Urinary retention	0.333 (0.043–0.777)	0.962 (0.940–0.977)	0.500 (0.118–0.882)	0.981 (0.964–0.991)
	Incontinence	0.500 (0.118–0.882)	0.865 (0.831–0.895)	0.667 (0.223–0.957)	0.940 (0.915–0.960)
	Saddle numbness	0 (0–0.459)	0.947 (0.922–0.965)	0 (0–0.459)	0.994 (0.981–0.999)
	Weakness in limbs	0.667 (0.223–0.957)	0.532 (0.486–0.578)	0.500 (0.118–0.882)	0.786 (0.746–0.823)

CI, confidence interval.

total cohort, 70 of 474 patients (14.8%) showed complete agreement between the patient-reported red flags and the red flags extracted from the EHR.

Discussion

To the authors' knowledge, this study is the first to evaluate the utility of patient reports to screen low back pain patients for potential red flags of serious underlying pathologies. Patient-reported red flags, both alone and in combination, had low sensitivity and specificity for identification of infection, malignancy, vertebral fracture, and cauda equina syndrome as defined by ICD-9-CM codes. Despite the difficulty in accurately collecting patient-reported red flags, we suggest the use of patient-reported red flags as an initial screen to alert providers to their possible presence and need for further inquiry.

Even with the lower sensitivity and specificity, the results of this study on patient-reported red flags were consistent with the results of previous studies on provider-reported red flags. The patient-reported "History of cancer" red flag performed with the best sensitivity and specificity in detecting malignant pathology, which is consistent with the literature of provider-reported red flags among low back pain patients [9,31]. Additionally, the combination of patient-reported "Steroid use," "Osteoporosis," and "Trauma" was the best predictor for fracture pathology. This finding is supported by studies that found that among the combinations of one or more of the provider-reported red flags, "Steroid use" and "Trauma" provided the best screen for detection of spinal fractures among low back pain patients ("Osteoporosis" was not reviewed) [31]. The prevalence of infection [32] and cauda equina [10] pathologies in the sample were too low to effectively draw conclusions on the sensitivity and specificity. Because of the low incidence of infection and cauda equina syndrome in the general population, evaluating screening tools for these diagnoses in a non-emergent setting is challenging [8].

Although determining the utility of provider-reported red flags was not a primary focus of this study, chart review of corresponding cases of low back pain was performed to allow for comparison between the two methods of red flag ascertainment. Providers did not systematically document the presence or absence of red flags in the EHR, and extraction of red flags from the EHR was comprised of a combination of provider documentation in visit notes and data available in the medical history and medication sections of the EHR. Overall, patient-reported red flags had worse performance than the performance of red flags documented in the medical record (Fig. 2a and b). Despite the lack of systematic assessment, the superior performance of red flag information documented in the medical record suggests that provider ascertainment of red flag may be a superior approach.

Agreement between patient-reported red flags and red flags documented by medical record was moderate overall, with kappas ranging from 0.03 (Pain at rest) to 0.72 (History of cancer). There was the greatest agreement in red flags pertaining to medical history: "History of cancer," "Osteoporosis," and "Trauma". It is possible that providers make an assessment of clinical applicability of symptom-based red flags, only documenting the presence of red flags in the medical record that they felt might be clinically relevant. In addition, the low prevalence of these pathologies created a bias in the kappa scores. Because Cohen's kappa will significantly underestimate agreement in these situations, it may be more informative to look at positive, negative, and overall agreement and the proportion of kappa divided by kappa$_{max}$ (Table 4). However, even when adjusting for maximum kappa, agreement was only substantial at best, with the majority of red flags having modest or

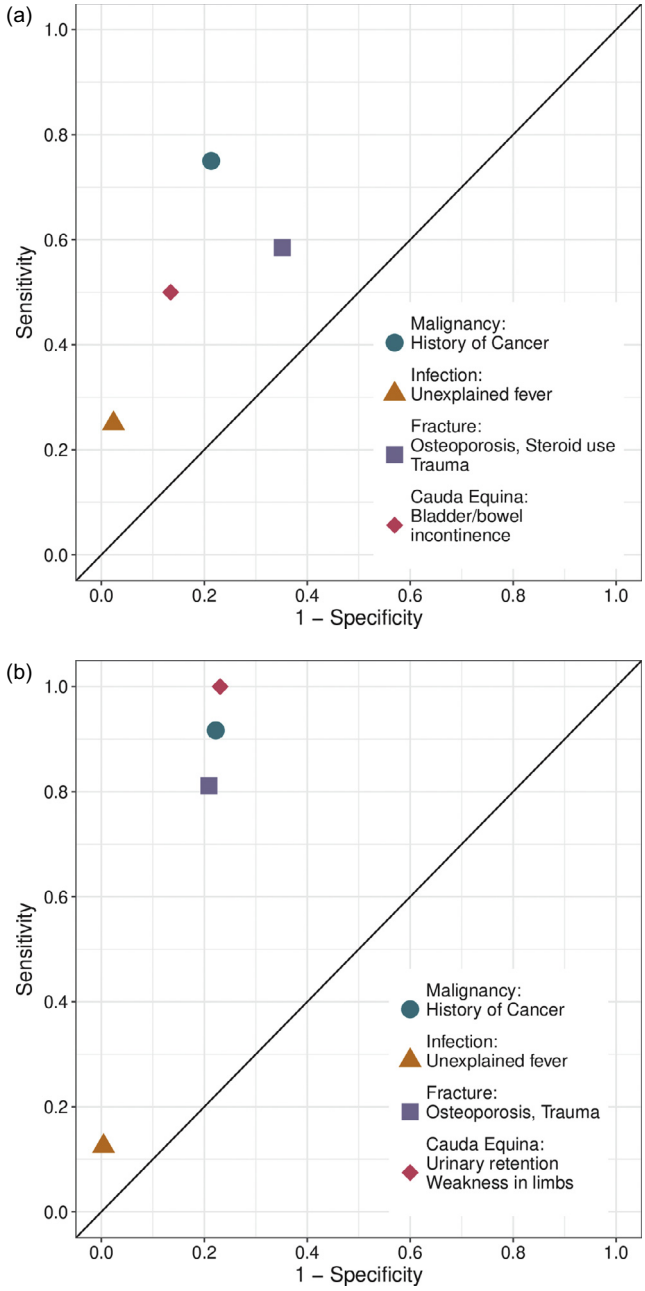

Fig. 2. Best performing (a) patient-reported red flags and (b) red flags as documented in the medical record for the corresponding pathologies. Points are plotted on an receiver operating characteristics plot; the upper left hand corner (coordinates [0,1]) corresponds to the optimal screening tool with a sensitivity and specificity of 100%.

poorer agreement [28]. Ultimately, only a small proportion of patients in our sample (14.77%) showed complete agreement on red flags between physician notes and patient forms.

Implications for practice

There has been significant debate over the efficacy of red flags in detecting serious spinal pathologies [33–35]. There is no general consensus as to which signs and symptoms should constitute a red flag [36]. Studies on the optimal set of questions that should be used during a medical interview to screen for underlying pathologies have yielded mixed results [37,38]. As such, the questions that comprise a red flag questionnaire vary greatly between institutions. Physician adherence to documentation of red flags during history-taking is also characteristically poor [16]. Although evaluation has been complicated by the variability of specific red flags used and the low prevalence of underlying serious pathologies [35], several studies have reported low or only modest sensitivity and specificity of red flags obtained by providers [33]. Our results of low sensitivity and specificity of patient-reported red flags are therefore not surprising. Guidelines for the management of low back pain [39–41], however, consistently recommend the inclusion of red flags during history-taking as a cost-reducing measure to prevent all patients with low back pain from receiving imaging [3,42]. Given the severity of possible sequelae in the event of a missed pathologic diagnosis, it is imperative that there is a high index of suspicion for serious spinal pathologies resulting in further workup. Ironically, given their low sensitivity and specificity, these red flags may result in an increase in unnecessary imaging and costs if used in isolation to make decisions on testing. These studies and recommendations combined thus support the notion that red flags may be more appropriate to use as an initial screen rather than as a decision tool for the use of imaging. The presence of red flags should cue providers to further explore symptoms and consider further investigation based on their clinical judgment. Current imaging guidelines that use red flags in isolation without considering a patient's entire clinical picture to determine if patients require further investigations may consequently lead to inappropriate imaging.

Study limitations

Strengths of this study include a thorough chart review that included review of the diagnoses received by patients in a 1-year window surrounding the office visit. Limitations of this study include that the study was done at a single institution in a specialty spine clinic rather than a primary care clinic and may skew the rates of documentation of red flags. Additional studies are needed of the utility of patient-reported red flags in the emergency department setting, which may have a higher incidence of serious pathologies. Additionally, the analyses only included the 11 red flags employed at the institution, which have since been modified from when the study was initially performed. The red flag questions used in this particular study did not include all red flags proposed in published guidelines and clinical care [31]. Thus, the results of the study can only be

Table 4
Rates of agreement between red flags reported by patients and as documented in the electronic health record

Red flag	Patient endorsed, N (%)	Medical record, N (%)	Kappa (95% CI)	$Kappa_{max}$*	Kappa /$Kappa_{max}$	Percent agreement		
						Negative	Positive	Overall
History of cancer	114 (24.1)	122 (25.7)	0.718 (0.645–0.791)	0.955	0.752	93.0	78.8	89.5
Night pain	240 (50.6)	59 (12.4)	0.160 (0.102–0.218)	0.243	0.657	69.0	32.8	57.6
Pain at rest	142 (30)	31 (6.5)	0.035 (−0.032 to 0.102)	0.281	.125	80.8	13.9	68.6
Urinary retention	20 (4.2)	12 (2.5)	0.225 (0.024–0.427)	0.742	0.304	97.4	25.0	94.9
Unexplained fever	13 (2.7)	3 (0.6)	0.242 (−0.036 to 0.521)	0.369	0.657	98.7	25.0	97.5
Osteoporosis	122 (25.7)	82 (17.3)	0.419 (0.323–0.515)	0.753	0.557	87.4	53.9	80.1
Steroid use	71 (15)	79 (16.7)	0.272 (0.161–0.383)	0.937	0.290	88.5	38.7	80.6
Trauma	38 (8)	66 (13.9)	0.358 (0.232–0.484)	0.700	0.511	92.9	42.3	87.3
Bladder or bowel incontinence	66 (13.9)	32 (6.8)	0.237 (0.114–0.360)	0.618	0.383	92.0	30.6	85.7
Saddle numbness	25 (5.3)	3 (0.6)	0.061 (−.068 to 0.189)	0.205	0.296	97.2	7.10	94.5
Weakness in limbs	223 (47)	103 (21.7)	0.310 (0.236–0.385)	0.476	0.652	74.6	51.5	66.7

CI, confidence Interval.
* $Kappa_{max}$ is the theoretical maximum of Kappa given the marginal distribution of agreements and disagreements for the specified red flag.

applied to the listed red flags used in this study. As noted previously, low prevalence of certain diseases limited the ability to obtain meaningful estimates for sensitivity and specificity of red flags. Furthermore, providers had immediate access to patient responses to red flag questions and this may have impacted their documentation of red flags within the EHR. The modest agreement between patient-reported red flags and those documented in the medical record suggests that the patient-reported red flags were not copied into provider notes but may have been a reason for reduced provider documentation of red flags, with higher probability that discrepant red flags were documented in the provider note. Lastly, the authors assumed that a red flag that was not mentioned in the provider notes was equivalent to absence of that red flag. Therefore, only a minority of negative red flags from the analysis were specifically documented as negative in the medical record.

Conclusions

Patient-reported red flags had low sensitivity and specificity for the identification of serious pathologies in patients with low back pain seen in a large spine clinic. It may not be appropriate to use patient-reported red flags in isolation to trigger additional testing and/or referral, although they may be clinically useful to providers as a prompt for further questioning of a patient's symptoms to determine if further investigation is warranted. The results of this study suggest that the use of patient-reported red flags alone cannot replace the clinical judgment of health-care providers.

Acknowledgments

We thank Dr. Edward C. Benzel for his suggestions and critique of the manuscript. We would also like to thank Ms. Dawn Carabello for her assistance in the acquisition of the data. Last, we would like to acknowledge Sagar Vallabh, Brandon Chang, and Dr. Benjamin K.P. Woo for their assistance in preparing the manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.spinee.2018.06.342.

References

[1] Hart LG, Deyo RA, Cherkin DC. Physician office visits for low back pain. frequency, clinical evaluation, and treatment patterns from a U. S. national survey. Spine (Phila Pa 1976) 1995;20:11–9.
[2] Cassidy JD, Carroll LJ, Cote P. The saskatchewan health and back pain survey. the prevalence of low back pain and related disability in saskatchewan adults. Spine (Phila Pa 1976) 1998;23:6. discussion 1867
[3] Patel ND, Broderick DF, Burns J, Deshmukh TK, Fries IB, Harvey HB, et al. ACR appropriateness criteria low back pain. J Am Coll Radiol 2016;13:1069–78.
[4] National ambulatory medical care survey: 2010 summary tables. [Internet].; cited April 28, 2017]. Available from: http://www.cdc.gov/nchs/data/ahcd/namcs_summary/2010_namcs_web_tables.pdf.
[5] Gaskin DJ, Richard P. The economic costs of pain in the United States. J Pain 2012;13:715–24.
[6] Coste J, Delecoeuillerie G, Cohen de Lara A, Le Parc JM, Paolaggi JB. Clinical course and prognostic factors in acute low back pain: an inception cohort study in primary care practice. BMJ 1994;308:577–80.
[7] Bach SM, Holten KB. Guideline update: what's the best approach to acute low back pain. J Fam Pract 2009;58. E1.
[8] Henschke N, Maher CG, Refshauge KM, Herbert RD, Cumming RG, Bleasel J, et al. Prevalence of and screening for serious spinal pathology in patients presenting to primary care settings with acute low back pain. Arthritis Rheum 2009;60:3072–80.
[9] Verhagen AP, Downie A, Popal N, Maher C, Koes BW. Red flags presented in current low back pain guidelines: a review. Eur Spine J 2016;25:2788–802.
[10] Lurie JD. What diagnostic tests are useful for low back pain. Best Pract Res Clin Rheumatol 2005;19:557–75.
[11] Koes BW, van Tulder M, Lin CW, Macedo LG, McAuley J, Maher C. An updated overview of clinical guidelines for the management of non-specific low back pain in primary care. Eur Spine J 2010;19:2075–94.

[12] Gardner A, Gardner E, Morley T. Cauda equina syndrome: a review of the current clinical and medico-legal position. Eur Spine J 2011;20:690–7.
[13] Raison NT, Alwan W, Abbot A, Farook M, Khaleel A. The reliability of red flags in spinal cord compression. Arch Trauma Res 2014;3: e17850.
[14] Nishisho T, Sakai T, Tezuka F, Higashino K, Takao S, Takata Y, et al. Delayed diagnosis of primary bone and soft tissue tumors initially treated as degenerative spinal disorders. J Med Invest 2016;63:274–7.
[15] Amorin-Woods LG, Beck RW, Parkin-Smith GF, Lougheed J, Bremner AP. Adherence to clinical practice guidelines among three primary contact professions: a best evidence synthesis of the literature for the management of acute and subacute low back pain. J Can Chiropr Assoc 2014;58:220–37.
[16] Di Iorio D, Henley E, Doughty A. A survey of primary care physician practice patterns and adherence to acute low back problem guidelines. Arch Fam Med 2000;9:1015–21.
[17] Wood WA, Bennett AV, Basch E. Emerging uses of patient generated health data in clinical research. Mol Oncol 2015 May 01;9:1018–24.
[18] Basch E. New frontiers in patient-reported outcomes: adverse event reporting, comparative effectiveness, and quality assessment. Annu Rev Med 2014;65:307–17.
[19] Broderick JE, DeWitt EM, Rothrock N, Crane PK, Forrest CB. Advances in patient-reported outcomes: the NIH PROMIS((R)) measures. EGEMS (Wash DC) 2013;1. 9214.1015. eCollection 2013.
[20] McGirt MJ, Bydon M, Archer KR, Devin CJ, Chotai S, Parker SL, et al. An analysis from the quality outcomes database, part 1. disability, quality of life, and pain outcomes following lumbar spine surgery: Predicting likely individual patient outcomes for shared decision-making. J Neurosurg Spine 2017;27:357–69.
[21] Ferguson F, Holdsworth L, Rafferty D. Low back pain and physiotherapy use of red flags: The evidence from scotland. Physiotherapy 2010;96:282–8.
[22] Henschke N, Maher CG, Ostelo RW, de Vet HC, Macaskill P, Irwig L. Red flags to screen for malignancy in patients with low-back pain. Cochrane Database Syst Rev 2013;28:CD0 08686.
[23] Patel AT, Ogle AA. Diagnosis and management of acute low back pain. Am Fam Physician 2000;61:90.
[24] Biesheuvel CJ, Vergouwe Y, Oudega R, Hoes AW, Grobbee DE, Moons KG. Advantages of the nested case-control design in diagnostic research. BMC Med Res Methodol 2008;8:48.
[25] Katzan I, Speck M, Dopler C, Urchek J, Bielawski K, Dunphy C, et al. The knowledge program: an innovative, comprehensive electronic data capture system and warehouse. AMIA Annu Symp Proc 2011;2011:683–92.
[26] Bhangle SD, Sapru S, Panush RS. Back pain made simple: an approach based on principles and evidence. Cleve Clin J Med 2009;76:393–9.
[27] Casazza BA. Diagnosis and treatment of acute low back pain. Am Fam Physician 2012;85:343–50.
[28] Brennan RL, Prediger DJ. Coefficient kappa: some uses, misuses, and alternatives. Educ Psychol Meas 1981;41:687–99.
[29] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
[30] R Core Team, R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.
[31] Downie A, Williams CM, Henschke N, Hancock MJ, Ostelo RW, de Vet HC, et al. Red flags to screen for malignancy and fracture in patients with low back pain: systematic review. BMJ 2013;347:f7095.
[32] Atlas SJ, Deyo RA. Evaluating and managing acute low back pain in the primary care setting. J Gen Intern Med 2001;16:120–31.
[33] Henschke N, Maher C. Red flags need more evaluation. Rheumatology (Oxford) 2006;45:920–1.
[34] Ross MD, Boissonnault WG. Red flags: to screen or not to screen. J Orthop Sports Phys Ther 2010;40:682–4.
[35] Underwood M. Diagnosing acute nonspecific low back pain: Time to lower the red flags. Arthritis Rheum 2009;60:2855–7.
[36] Leerar PJ, Boissonnault W, Domholdt E, Roddey T. Documentation of red flags by physical therapists for patients with low back pain. J Man Manip Ther 2007;15:42–9.
[37] Henschke N, Maher CG, Refshauge KM. Screening for malignancy in low back pain patients: a systematic review. Eur Spine J 2007;16:1673–9.
[38] Henschke N, Maher CG, Refshauge KM. A systematic review identifies five "red flags" to screen for vertebral fracture in patients with low back pain. J Clin Epidemiol 2008;61:110–8.
[39] Acute low back problems in adults: Assessment assessment and treatment. agency for health care policy and research. Clin Pract Guidel Quick Ref Guide Clin. 1994;14:1–25.
[40] van Tulder M, Becker A, Bekkering T, Breen A, del Real MT, Hutchinson A, et al. Chapter 3. European guidelines for the management of acute nonspecific low back pain in primary care. Eur Spine J 2006;2:169. 15 Suppl.
[41] Chou R, Qaseem A, Snow V, Casey D, Cross JT, Shekelle Jr. P, et al. Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American college of physicians and the American pain society. Ann Intern Med 2007;147:478–91.
[42] Grunau GL, Darlow B, Flynn T, O'Sullivan K, O'Sullivan PB, Forster BB. Red flags or red herrings? Redefining the role of red flags in low back pain to reduce overimaging. Br J Sports Med 2017;52: 488–9.