Prevalence of and Screening for Serious Spinal Pathology in Patients Presenting to Primary Care Settings With Acute Low Back Pain

Nicholas Henschke,[1] Christopher G. Maher,[1] Kathryn M. Refshauge,[1] Robert D. Herbert,[1] Robert G. Cumming,[1] Jane Bleasel,[2] John York,[2] Anurina Das,[1] and James H. McAuley[1]

Objective. **To determine the prevalence of serious pathology in patients presenting to primary care settings with acute low back pain, and to evaluate the diagnostic accuracy of recommended "red flag" screening questions.**

Methods. **An inception cohort of 1,172 consecutive patients receiving primary care for acute low back pain was recruited from primary care clinics in Sydney, Australia. At the initial consultation, clinicians recorded responses to 25 red flag questions and then provided an initial diagnosis. The reference standard was a 12-month followup supplemented with a specialist review of a random subsample of participants.**

Results. **There were 11 cases (0.9%) of serious pathology, including 8 cases of fracture. Despite the low prevalence of serious pathology, most patients (80.4%) had at least 1 red flag (median 2, interquartile range 1–3). Only 3 of the red flags for fracture recommended for use in clinical guidelines were informative: prolonged use of corticosteroids, age >70 years, and significant trauma. Clinicians identified 5 of the 11 cases of serious pathology at the initial consultation and made 6 false-positive diagnoses. The status of a diagnostic**

prediction rule containing 4 features (female sex, age >70 years, significant trauma, and prolonged use of corticosteroids) was moderately associated with the presence of fracture (the area under the curve for the rule score was 0.834 [95% confidence interval 0.654–1.014]; *P* = 0.001).

Conclusion. **In patients presenting to a primary care provider with back pain, previously undiagnosed serious pathology is rare. The most common serious pathology observed was vertebral fracture. Approximately half of the cases of serious pathology were identified at the initial consultation. Some red flags have very high false-positive rates, indicating that, when used in isolation, they have little diagnostic value in the primary care setting.**

Low back pain imposes a considerable social and economic burden on the community and is one of the most common reasons for presentations to primary health care providers (1,2). Because pathology responsible for most cases of low back pain cannot be determined, most cases are classified as nonspecific low back pain (2,3). Guidelines (4) advise that specific diagnoses can be made for a few pathologies, several of which (such as infection, inflammatory disease, and cancer) are labeled as "serious," because they typically are not managed in a primary care setting and require referral for further assessment and specific treatment (5,6).

It is generally recommended that the assessment of patients with acute low back pain in primary care should include screening to identify potential cases of serious pathology (6). However, 2 recent systematic reviews on screening for cancer (7) and fracture (8) demonstrated considerable uncertainty regarding the prevalence of these serious pathologies in patients with low back pain. As an illustration, the reviews showed a

prevalence of cancer ranging from 0.1% to 3.5% (7) and a prevalence of vertebral fracture ranging from 3% to 29% (8). Most of the studies included in the reviews were not conducted in a primary care setting (7,8), and thus the prevalence estimates may have been inflated by referral filter bias (9). A less biased estimate of the prevalence of serious pathology would be obtained from consecutive cases of low back pain in primary care. Without accurate data on the prevalence of serious pathology, the usefulness of the clinical assessment to identify those patients with potentially serious pathology cannot be assessed with confidence.

The use of "red flag" questions (such as questions about unexplained weight loss) has been advocated to screen for serious pathology in patients with low back pain who present to a primary care setting (10). Despite several similar red flag questions being recommended in a variety of clinical guidelines for the management of low back pain, their usefulness in clinical practice remains uncertain. The main problem is that most red flags have been infrequently studied. For example, of the 20 clinical features considered in our recent cancer review (7), 1 feature had been evaluated in 4 studies, 2 features had been evaluated in 2 studies, and the remainder had been evaluated in only a single study. That review highlighted the fact that many of the recommended red flags have not been externally validated by studies performed in the primary care setting. The sparse data on diagnostic accuracy also need to be interpreted with caution, because they principally arise from studies that are of low methodologic quality (7,8).

We sought to determine the proportion of individuals presenting to a primary care provider with low back pain caused by serious pathology and the diagnostic accuracy of red flag questions recommended in practice guidelines. We also investigated whether clinician judgment at the initial consultation was useful to diagnose serious pathology.

PATIENTS AND METHODS

This study is part of a larger project studying low back pain in primary care. A cohort of 1,346 consecutive patients with acute low back pain were recruited from a socioeconomically diverse region in the Sydney metropolitan area. Data from the 2001 Australian Census (11) was used to classify the socioeconomic levels within the region. In a previous publication (12), we reported the prognosis of acute nonspecific low back pain based on the 973 participants eligible for the prognosis arm of the project. In this study, we report on the 1,172 participants eligible for the diagnosis arm of the project.

Consecutive patients with acute low back pain who presented to a primary care provider were assessed for the presence of red flags. The presence of serious pathology was assessed, with clinical followup conducted over the subsequent year. Based on previous estimates of the prevalence of serious pathology (4–8), it was thought that a sample size of 1,000 would be sufficient to identify the 5 most common serious pathologies (vertebral fracture, cancer, infection, cauda equina syndrome, and inflammatory arthritis) and assess the diagnostic accuracy of specific screening questions for these pathologies. The study received ethics approval from the University of Sydney Human Research Ethics Committee.

In Australia, the majority of primary care management for low back pain is provided by general medical practitioners, physiotherapists, and chiropractors (13). All clinicians in these 3 professions with clinics in the study region were invited to participate. Clinicians who volunteered to participate in the study were trained either in small groups or individually. Training was performed by 1 or 2 members of the study personnel, who are experienced researchers and clinicians with a background in physiotherapy (CGM, KMR, NH), medicine (AD), or rheumatology (JY, JB). Training involved an explanation of the purpose and methods of the study, instruction on the clinical examination, and how to triage patients into 1 of 3 categories: simple backache, nerve root compromise, or suspected serious spinal pathology. The triage process was aided by a 1-page decision guide (Appendix A), which was developed from previous clinical guidelines (14,15). Clinicians were also given a copy of current guidelines for the management of acute low back pain (6) and asked to follow them when appropriate.

Clinicians were asked to screen all patients with the primary symptom of low back pain who presented to their clinics. We defined an episode of acute low back pain as pain in the area bounded superiorly by T12 and inferiorly by the buttock crease (16), lasting for more than 24 hours but less than 6 weeks, and preceded by a period of at least 1 month without back pain (17,18). Patients remained eligible if they also had pain that referred beyond this region. Only patients presenting for the first consultation for their current episode of acute low back pain were included. Other inclusion criteria were that participants must have been at least 14 years old, provided written consent to participate in the study, and were able to speak and read English. Potential participants were excluded if serious pathology had been diagnosed prior to the consultation, and the serious pathology was considered to be the cause of the current episode of low back pain.

The index tests that were evaluated included 25 red flag questions (Appendix A) derived from clinical practice guidelines (4,6,14,15) and discussion with experts in the field. Each red flag was specific to at least 1 of the 5 serious pathologies being investigated. In addition, the clinician's triage decision (nonspecific low back pain, nerve root pain, or suspected serious pathology) was evaluated for diagnostic accuracy. The reference standard consisted of close followup for 12 months. Participants were contacted by telephone 6 weeks, 3 months, and 12 months after the initial consultation. At each followup contact, participants were asked the following question: "Low back pain is occasionally the result of a fracture, infection, arthritis, or cancer. Has a health care provider said that your back pain is caused by one of these rare diseases?" Participants were also prompted to provide any further details of a diagnosis or explanation for their low back pain that had been provided to them. All patients with
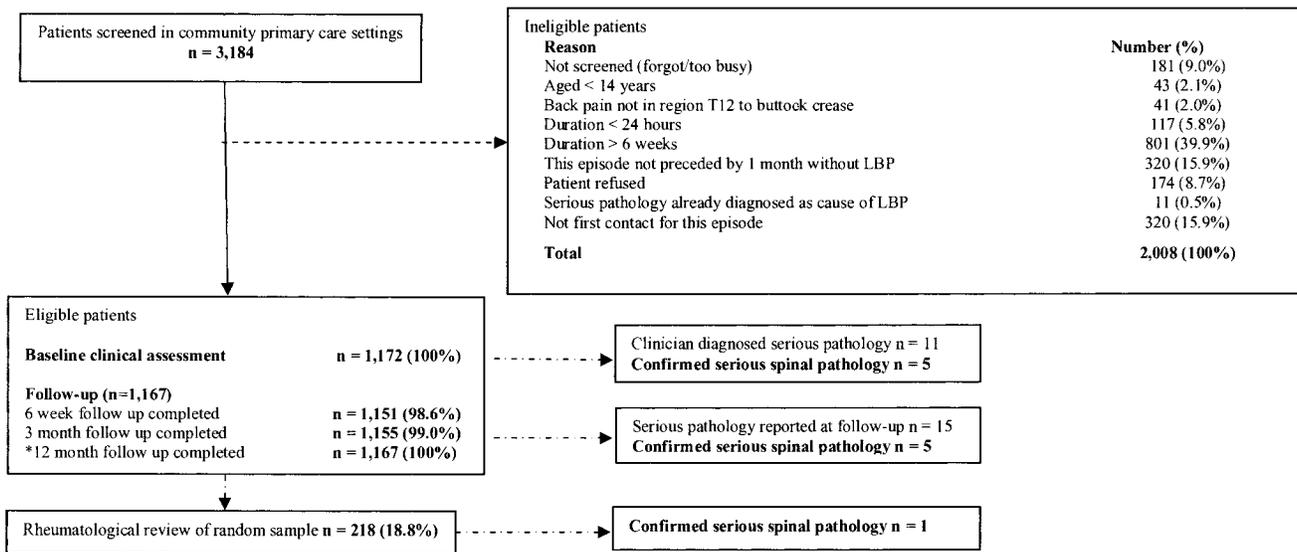
Figure 1. Flow of patients with low back pain (LBP) through the study, including followup rates and the number (%) of patients with confirmed serious spinal pathology. * = In 12 cases, the care provider, not the patient, was contacted at 12 months.

potentially serious pathology were subsequently examined by a study rheumatologist. The rationale for this reference standard is that most serious pathologies will become more clinically obvious over time due to the clinical course (e.g., cancer), or because the condition is likely to be exacerbated by usual care for nonspecific low back pain (e.g., fracture) (19). This approach has been used in previous studies identifying serious pathology in patients with low back pain (20,21). At each followup contact, participants were also questioned to establish whether they had recovered from the episode of low back pain. Recovery was defined as 1 month with no pain, no interference with function due to pain, and return to previous work status for 1 month. The recovery was assumed to have occurred at the beginning of that month.

Patients suspected by their primary care clinician of having a serious spinal pathology and those who reported having a serious spinal pathology during the followup period were referred immediately to 1 of 2 study rheumatologists for a clinical assessment. The rheumatologists had 35 and 15 years, respectively, of specialist practice at a major teaching hospital, and their role was to confirm whether or not the patient had a serious pathology. Within 2 weeks from the time of referral, the rheumatologists examined each patient in their clinics and were additionally provided with the complete medical histories and all test results. The study rheumatologists were able to contact the referring clinician if further information about a patient was needed. In addition, the study rheumatologists performed a clinical assessment of a random sample of ~20% of the patients in whom serious pathology had not been diagnosed 12 months after the baseline assessment. This random sample was selected using an urn-sampling-without-replacement procedure. The purpose of this clinical assess-

Table 1. Baseline characteristics of the 1,172 patients with acute low back pain (LBP)*

Age, mean ± SD years	43.97 ± 15.1
Male sex	626 (53.4)
Primary care clinician consulted	
Medical practitioner	267 (22.8)
Physiotherapist	851 (72.6)
Chiropractor	54 (4.6)
Born in Australia	807 (68.9)
Socioeconomic status of place of residence below national mean	207 (17.7)
Smoker	194 (16.6)
Previous episode of LBP	888 (75.8)
Previous sick leave due to LBP	435 (37.1)
Previous surgery for LBP	29 (2.5)
Currently taking medication for LBP	424 (36.2)
Duration of LBP	
Less than 1 week	696 (59.4)
1–2 weeks	145 (12.4)
2–3 weeks	174 (14.8)
3–4 weeks	73 (6.2)
4–5 weeks	30 (2.6)
5–6 weeks	54 (4.6)
Days forced to cut down on usual activities due to LBP, mean ± SD	3.6 ± 4.7
Interference with function due to LBP, median (range)†	4 (0–5)
LBP intensity, median (range)‡	4 (1–6)
Leg pain	295 (25.2)
Working preinjury	892 (76.1)

* Except where indicated otherwise, values are the number (%) of patients.
† 0 = not at all, 4 = quite a bit, 5 = extreme.
‡ 0 = none, 4 = severe, 6 = very severe.

Table 2. Prevalence of serious spinal pathology among the 1,172 patients with acute low back pain presenting to a primary care setting

Pathology	No. of cases of confirmed pathology	Prevalence (95% CI)*
Spinal fracture	8	0.7 (0.4–1.3)
Cancer	0	0.0 (0.0–0.3)
Infection	0	0.0 (0.0–0.3)
Cauda equina syndrome	1	0.1 (0.0–0.5)
Inflammatory disorder	2	0.2 (0.1–0.6)
Total	11	0.9 (0.5–1.7)

* 95% CI = 95% confidence interval.

ment was to test the veracity of the reference standard and to provide an estimate of how many cases of serious spinal pathology, if any, had been missed.

Confidence intervals (CIs) for the prevalence of serious pathologies and for the sensitivity, specificity, and positive and negative likelihood ratios (LRs) of individual red flags and clinician diagnosis were calculated using Wilson score methods (22) with Confidence Interval Analysis 2.0.0 software (BMJ Books, London, UK).

The primary analysis of diagnostic accuracy assumed that the followup procedure correctly identified all patients with serious pathology. However, it is possible that some cases could have been missed. To test the robustness of the findings, we calculated the upper limit of the 95% CI for the proportion of missed cases in the random sample of patients in whom serious pathology had not been diagnosed. This proportion was applied to the remaining sample to obtain a "worst case" estimate of the number of missing cases. The worst case number of missing cases was added to the observed number of cases to obtain worst case estimates of the prevalence of serious spinal pathology and revised estimates of the sensitivity, specificity, and positive and negative LRs of red flags and clinician diagnosis.

From the 25 red flag questions and the clinical and sociodemographic variables, the most plausible variables, based on previous literature and consultation with experts in the field, were considered for inclusion in a clinical diagnostic rule. Those variables that had a statistically significant association ($P < 0.1$) with a diagnosis of vertebral fracture were retained as candidate variables for the rule. The area under the curve (AUC) and the sensitivity, specificity, and LRs for each

score on the diagnostic rule were calculated. All analyses were performed using SPSS v.14.0 software (SPSS, Chicago, IL).

RESULTS

A total of 1,254 primary care clinicians in the study region were identified and contacted. Of these, 170 (73 general medical practitioners, 77 physiotherapists, and 20 chiropractors) agreed to participate and were trained. The trained clinicians screened a total of 3,184 consecutive patients with low back pain between November 2003 and July 2005. Of these, 1,172 patients had acute low back pain and were presenting for the first consultation for that episode and were thus eligible to participate (Figure 1). The reasons for ineligibility are shown in Figure 1. Eleven patients were ineligible because they presented with serious pathology that had already been diagnosed (8 vertebral fractures, 2 inflammatory arthritides, 1 metastatic bone cancer).

The baseline demographic and clinical features of participants are shown in Table 1. Eleven cases of previously undiagnosed serious pathology were confirmed (8 vertebral fractures, 2 inflammatory arthritides, 1 cauda equina syndrome) among the 1,172 participants (Table 2). The remainder were considered not to have serious pathology based on the following criteria: 1) the patient had a negative rheumatologic review at 12 months (n = 217), 2) in a patient reporting at 12 months who previously had no serious pathology, serious pathology had been diagnosed in the intervening year (n = 932), 3) the patient recovered at an earlier time point, and at 12 months the patient's care provider reported that no serious pathology had been diagnosed in the intervening period (n = 7), and 4) the care provider confirmed that no serious pathology had been diagnosed in the intervening year (n = 5). Thus, the prevalence of serious pathology in our cohort was 0.9% (95% CI 0.5–1.7%). As shown in Table 2, the prevalence of the individual pathologies ranged from 0.0% (95% CI 0.0–

Table 3. Diagnostic accuracy of recommended "red flag" questions for detecting spinal fracture in the 1,172 patients with acute low back pain*

Red flag question	No. (%) red flag positive	Sensitivity, %	Specificity, %	Positive LR (95% CI)	Negative LR (95% CI)
Age >70 years	56 (4.8)	50	96	11.19 (4.65–19.48)	0.52 (0.23–0.82)
Significant trauma (major in young, minor in elderly)	31 (2.6)	25	98	10.03 (2.76–26.36)	0.77 (0.42–0.95)
Prolonged use of corticosteroids	8 (0.7)	25	100	48.50 (11.62–165.22)	0.75 (0.41–0.93)
Sensory level (altered sensation from trunk down)	19 (1.6)	0	98	0.00 (0.00–21.01)	1.02 (1.02–1.03)
Clinician diagnosis of fracture	7 (0.6)	50	100	194.00 (52.10–653.61)	0.50 (0.22–0.79)

* LR = likelihood ratio; 95% CI = 95% confidence interval.

Table 4. False-positive rates of recommended red flag questions for detecting serious spinal pathologies*

Red flag question	Red flag–positive, no. (%) of patients	False-positive rate (95% CI)
Fracture		
Age >70 years	56 (4.8)	4.5 (3.4–5.8)
Significant trauma (major in young, minor in elderly)	31 (2.6)	2.5 (1.7–3.6)
Prolonged use of corticosteroids	8 (0.7)	0.5 (0.2–1.1)
Sensory level (altered sensation from trunk down)	19 (1.6)	1.6 (1.1–2.5)
Clinician diagnosis of fracture	7 (0.6)	0.3 (0.1–0.8)
Cancer		
Age at onset <20 years or >55 years	281 (24.0)	24.0 (21.6–26.5)
Unexplained weight loss (>4.5 kg in 6 months)	3 (0.3)	0.3 (0.1–0.8)
Previous history of cancer	46 (3.9)	3.9 (3.0–5.2)
Tried bed rest, but no relief	192 (16.4)	16.4 (14.4–18.6)
Insidious onset	202 (17.2)	17.2 (15.2–19.5)
Systemically unwell	27 (2.3)	2.3 (1.6–3.3)
Constant, progressive, nonmechanical pain	33 (2.8)	2.8 (2.0–3.9)
Sensory level (altered sensation from trunk down)	19 (1.6)	1.6 (1.0–2.5)
Clinician diagnosis of cancer	1 (0.1)	0.1 (0.0–0.3)
Infection		
Systemically unwell	27 (2.3)	2.3 (1.6–3.3)
Constant, progressive, nonmechanical pain	33 (2.8)	2.8 (2.0–3.9)
Recent bacterial infection, e.g., UTI or skin infection	27 (2.3)	2.3 (1.6–3.3)
Intravenous drug abuse	0	0.0 (0.0–0.3)
Immune suppression from steroids, transplant, or HIV	3 (0.3)	0.3 (0.1–0.8)
Sensory level (altered sensation from trunk down)	19 (1.6)	1.6 (1.0–2.5)
Clinician diagnosis of infection	0	0.0 (0.0–0.3)
Cauda equina syndrome		
Acute onset of urinary retention or overflow incontinence	5 (0.4)	0.4 (0.2–1.0)
Loss of anal sphincter tone or fecal incontinence	2 (0.2)	0.2 (0.1–0.6)
Saddle anesthesia about the anus, perineum, or genitals	3 (0.3)	0.3 (0.1–0.8)
Widespread (>1 nerve root) or progressive motor weakness in the legs or gait disturbances	5 (0.4)	0.4 (0.2–1.0)
Clinician diagnosis of cauda equina syndrome	0	0.0 (0.0–0.3)
Inflammatory disorder		
Gradual onset before age 40 years	102 (8.7)	8.7 (7.2–10.5)
Tried bed rest, but no relief	192 (16.4)	16.3 (14.3–18.6)
Insidious onset	202 (17.2)	17.1 (15.1–19.4)
Systemically unwell	27 (2.3)	2.3 (1.6–3.3)
Constant, progressive, nonmechanical pain	33 (2.8)	2.8 (2.0–3.9)
Morning back stiffness lasting ≥0.5 hours	325 (27.7)	27.8 (25.3–30.4)
Peripheral joint involvement	63 (5.4)	5.3 (4.2–6.7)
Persisting limitation of spinal movements in all directions	98 (8.4)	8.4 (6.9–10.1)
Iritis, skin rashes (psoriasis), colitis, urethral discharge	10 (0.9)	0.9 (0.4–1.6)
Family history of arthritis or osteoporosis	271 (23.1)	23.2 (20.8–25.7)
Pain improves with exercise	429 (36.6)	36.7 (34.0–39.5)
Clinician diagnosis of inflammatory disorder	1 (0.1)	0.1 (0.0–0.5)

* The false-positive rate is the number of false-positive results/number of patients without disease. 95% CI = 95% confidence interval; UTI = urinary tract infection; HIV = human immunodeficiency virus.

0.3%) to 0.7% (95% CI 0.4–1.3%). The primary care clinicians identified 5 of the 11 cases of serious pathology at the initial consultation and made 6 false-positive diagnoses (Figure 1).

Due to the low prevalence of serious pathology, it was meaningful to present only diagnostic accuracy data for detection of fracture. Three of the 4 red flags for fracture had informative positive LRs: prolonged use of corticosteroids (positive LR = 48.5), significant trauma (positive LR = 10), and age >70 years (positive LR = 11). The clinician's diagnosis had a positive LR of 194.0 for detecting fracture (Table 3).

For the remaining pathologies, the red flags were evaluated in terms of their false-positive rates (Table 4). The false-positive rate was calculated as the number of false-positive results for each red flag divided by the total number of patients without each disease. The patients had a median of 2 (interquartile range 1–3)

Table 5. Diagnostic rule to identify vertebral fracture*

	Criteria for a positive test		
	1 positive feature	≥2 positive features	≥3 positive features
Sensitivity, %	88	63	38
Specificity, %	50	96	100
Positive LR (95% CI)	1.8 (1.1–2.0)	15.5 (7.2–24.6)	218.3 (45.6–953.8)
Posttest probability of vertebral fracture, %			
Pretest probability 0.5%	1	7	52
Pretest probability 3%	5	32	87

* Four features were included in the rule: female sex, age >70 years, significant trauma (major in young patients, minor in elderly patients), and prolonged use of corticosteroids. LR = likelihood ratio; 95% CI = 95% confidence interval.

positive red flags. Of the 25 red flags evaluated, only 1 (reported intravenous drug abuse) was negative for all patients. The most common red flag (pain improves with exercise) relates to inflammatory arthritides and had a false-positive rate of 36.7% (95% CI 34.0–39.5%). Five of the 11 red flags for inflammatory arthritides had a false-positive rate >10%. For detecting cancer, 3 of the 8 red flags had a false-positive rate >10% (Table 4).

After the 12-month followup period, 518 randomly selected patients (44.2%) who did not have serious pathology according to the reference standard were invited to be reviewed by our study rheumatologists; 218 of these patients accepted the invitation. Of these, 1 patient (0.5% [95% CI 0.1–2.5%]) had a specific pathology diagnosed (a vertebral fracture) that went unreported throughout the followup period. The upper limit of the 95% CI of this proportion was used to generate a "worst case" estimate of the number of cases of serious pathology that could have been missed by the reference standard. The worst case estimate is that there were 24 missed cases of serious pathology within the cohort, which, when added to the number of confirmed cases, gives a total of 35 cases and an overall prevalence of 3.1%.

A diagnostic rule was developed only for vertebral fracture, because it was the most prevalent serious pathology. The candidate variables included the red flags for vertebral fracture (Table 4) as well as sex. The rule contained 4 variables (Table 5): female sex, age >70 years, significant trauma (major in young patients, minor in elderly patients), and prolonged use of corticosteroids. When at least 1 of these features was positive (n = 584), the positive LR was 1.8 (95% CI 1.1–2.0). With at least 2 positive features (n = 52), the positive LR increased to 15.5 (95% CI 7.2–24.6), and with 3 positive features (n = 5) it increased to 218.3 (95% CI 45.6–953.8). No patient in the cohort was positive for all

4 features. The AUC value for the rule score was 0.834 (95% CI 0.654–1.014; $P = 0.001$).

DISCUSSION

In this study of 1,172 patients with undiagnosed acute low back pain who presented to a primary care provider, serious spinal pathology was rare. The most common serious pathology was vertebral fracture. A few serious spinal pathologies known to cause low back pain (e.g., cancer, infection) were not found in our cohort, so we were unable to evaluate screening questions for these pathologies. The majority of patients without serious pathology had more than 1 positive red flag, and some red flags had very high false-positive rates. Clinicians were able to identify about half of the cases of serious pathology at the initial consultation.

We observed a lower prevalence of specific pathology than previously reported for cancer (range 0.1–3.5%) (6) and fracture (range 3–29%) (7). Our estimate for the prevalence of previously undiagnosed serious pathology in patients presenting to primary care providers for low back pain was 0.9%; if we also included patients with a preexisting diagnosis of serious pathology, the figure would rise to 1.9%. The difference can potentially be explained by the context in which the studies were conducted: previous studies were not conducted in a primary care setting and did not always apply the reference standard to the entire cohort of patients with low back pain (7,8), which could have inflated the estimates of prevalence (9).

We enrolled a representative sample of patients in whom there was diagnostic uncertainty and applied the reference standard to all patients. Close followup over 12 months was considered to be the best possible reference standard against which to test the recommended red flags. The approach used here recognizes

that although some cases of serious disease can be missed at the first consultation, the disease manifestations become progressively more obvious if the disease is left untreated. We validated the reference standard with a randomly selected subsample of 218 patients and identified only 1 missed case of serious pathology (a fracture). Routine exhaustive testing of all patients at baseline with laboratory, imaging, and biopsy procedures would arguably provide a better reference standard, but the cost would be prohibitive, the procedures would pose unacceptable risks to patients, and false-positive results could generate inflated estimates of the prevalence of serious disease (23,24).

One possible concern is that we missed cases of serious pathology because such patients self-refer to medical specialists. This is unlikely, because in Australia, it is not possible to self-refer: patients need to have a referral from a general practitioner to see a medical specialist. If, in other countries, patients with serious undiagnosed pathologies chose to bypass primary care and directly consult a medical specialist, the prevalence of serious pathology in primary care might be lower still. However, we found a similar prevalence of serious pathology among the patients seeking medical and non-medical primary care, suggesting that patients with low back pain are not able to recognize that they have a more serious disease and do not preferentially seek medical care. Only 1 case of serious pathology was diagnosed after the 12-month followup period in the sample reviewed by our study rheumatologists. Because only 42% of the patients randomly selected to be reviewed accepted the offer, it is possible that our estimate of the prevalence of undiagnosed serious pathology is biased. However, it is likely that those patients who felt most at risk of having a serious pathology would be more likely to accept the invitation to a free specialist review. In that case, the rate of undiagnosed serious pathology might tend to be overestimated, and our estimate of the rate of undiagnosed serious pathology would be conservative.

Our estimates of diagnostic accuracy can be used to calculate the posttest probability of having a vertebral fracture. For example, a patient who has had prolonged use of corticosteroids (positive LR = 48.5) but is otherwise not atypical (pretest probability of 0.7%, the prevalence of vertebral fracture in this population) will have a (posttest) probability of having a vertebral fracture of 25.5%. Such calculations must, however, be interpreted with caution because of the wide CIs around the estimates of diagnostic accuracy caused by the low prevalence of vertebral fracture.

It is also possible that our estimates of diagnostic accuracy could be inflated by incorporation bias (25), because data used to establish the presence of serious disease at followup may have included information on red flags collected at baseline. However, most of the red flags had poor diagnostic accuracy, so it appears that estimates of diagnostic accuracy were not subject to serious incorporation bias. Another limitation of this study is that the diagnostic accuracy of clinician judgment may not be generalizable to clinicians who are not familiar with the diagnostic triage approach advocated in the guidelines and that we covered in our training. However, the data on prevalence and the accuracy of red flags would be unaffected by this issue.

One finding with direct clinical implications is that several recommended red flags were found to have high false-positive rates. An approach in which any positive red flag is acted upon will mean that there will be a large number of unnecessary referrals and investigations of patients with acute low back pain presenting for primary care. A better approach may be achieved by using multivariate statistical methods to evaluate a combination of red flags that will identify potentially serious pathology while reducing the number of false-positive results. However, due to the low number of cases involving serious pathology identified in this study, it was feasible in this instance to develop a diagnostic prediction model only for detection of fracture (19). Our finding that clinician judgment and the 3 red flags of age >70 years, significant trauma, and prolonged use of corticosteroids were all associated with fracture is generally consistent with a systematic review of 12 primary studies evaluating screening for fracture (8). Older age (20,24) was predictive in 2 studies that tested this feature. Five studies (20,26–29) evaluated trauma, and all showed that this feature is predictive of fracture except when the trauma was only minor. The single study that evaluated the use of corticosteroids (20) concluded that this feature was not predictive. Two studies tested global clinician judgment (30,31), and only 1 of the studies (31) concluded that clinician judgment was predictive.

To help clinicians diagnose vertebral fracture in patients with acute low back pain, a diagnostic rule was developed. The rule (Table 5) suggests that when any 3 of the 4 features are positive, the posttest probability of vertebral fracture greatly increases (e.g., from ~0.5% to 52%). However, due to the low prevalence of vertebral fracture in this cohort, wide CIs were observed around the LR estimates. A more precise diagnostic rule may be

developed in a larger cohort with more cases of vertebral fracture.

Because the primary care setting plays a vital role in early detection of serious disease, it is there that reliable and accurate diagnostic information is needed. However, many of the recommended red flag questions could not be tested, because the prevalence of the relevant pathology was too low. Nevertheless, some red flags were positive for a large proportion of participants, most of whom did not have serious pathology. We could not evaluate the diagnostic accuracy of red flags for conditions such as inflammatory arthritis, cancer, cauda equina syndrome, and infection, because they were rare or were not present in our cohort of 1,172 patients. The diagnostic accuracy of red flags for all serious pathologies is an important, although challenging, topic for future research in low back pain.

AUTHOR CONTRIBUTIONS

REFERENCES

1. Atlas SJ, Deyo RA. Evaluating and managing acute low back pain in the primary care setting. J Gen Intern Med 2001;16:120–31.
2. Deyo RA, Weinstein JN. Low back pain. N Engl J Med 2001;344:363–70.
3. Hancock MJ, Maher CG, Latimer J, Spindler MF, McAuley JH, Laslett M, et al. Systematic review of tests to identify the disc, SIJ or facet joint as the source of low back pain. Eur Spine J 2007;16:1539–50.
4. Van Tulder M, Becker A, Bekkering T, Breen A, Gil de Real MT, Hutchinson A, et al, on behalf of the COST B13 Working Group on Guidelines for the Management of Acute Low Back Pain in Primary Care. Chapter 3. European guidelines for the management of acute nonspecific low back pain in primary care. Eur Spine J 2006;15 Suppl 2:S169–91.
5. Deyo RA, Rainville J, Kent DL. What can the history and physical examination tell us about low back pain? JAMA 1992;268:760–5.
6. Australian Acute Musculoskeletal Pain Guidelines Group. Evidence-based management of acute musculoskeletal pain. Bowen Hills (Australia): Australian Academic Press; 2003. Sponsored by the National Health and Medical Research Council of Australia. URL: http://www.nhmrc.gov.au/publications/synopses/_files/cp94.pdf.
7. Henschke N, Maher CG, Refshauge KM. Screening for malignancy in low back pain patients: a systematic review. Eur Spine J 2007;16:1673–9.
8. Henschke N, Maher CG, Refshauge KM. A systematic review

9. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999;282:1061–6.
10. Koes BW, van Tulder MW, Ostelo R, Burton A, Waddell G. Clinical guidelines for the management of low back pain in primary care: an international comparison. Spine 2001;26:2504–14.
11. Australian Bureau of Statistics. Socio-economic Indexes for Areas. Canberra: Australian Bureau of Statistics; 2004.
12. Henschke N, Maher CG, Refshauge KM, Herbert RD, Cumming RG, Bleasel J, et al. Prognosis in patients with recent onset low back pain in Australian primary care: inception cohort study. BMJ 2008;337:a171.
13. Walker BF, Muller R, Grant WD. Low back pain in Australian adults: health provider utilization and care seeking. J Manipulative Physiol Ther 2004;27:327–35.
14. Bigos SJ, Bowyer OR, Braen GR, Brown K, Deyo R, Haldeman S, et al. Acute low back problems in adults. Clinical practice guideline No. 14. Rockville (MD): Agency for Health Care Policy and Research, Public Health Service, US Department of Health and Human Services, 1994. AHCPR publication no. 95-0642.
15. Waddell G, Feder G, McIntosh A, Lewis M, Hutchison A. Low back pain evidence review. 1st ed. London: Royal College of General Practitioners; 1996.
16. Watson KD, Papageorgiou AC, Jones GT, Taylor S, Symmons DP, Silman AJ, et al. Low back pain in schoolchildren: the role of mechanical and psychosocial factors. Arch Dis Child 2003;88:12–7.
17. De Vet HC, Heymans MW, Dunn KM, Pope DP, van der Beek AJ, Macfarlane GJ, et al. Episodes of low back pain: a proposal for uniform definitions to be used in research. Spine 2002;27:2409–16.
18. Van Tulder MW, Furlan A, Bombardier C, Bouter L, and the Editorial Board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. Spine 2003;28:1290–9.
19. Knottnerus JA. Diagnostic prediction rules: principles, requirements and pitfalls. Prim Care 1995;22:341–63.
20. Deyo RA, Diehl AK. Lumbar spine films in primary care: current use and effects of selective ordering criteria. J Gen Intern Med 1986;1:20–5.
21. Deyo RA, Diehl AK. Cancer as a cause of back pain: frequency, clinical presentation and diagnostic strategies. J Gen Intern Med 1988;3:230–8.
22. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. Stat Med 1998;17:857–72.
23. Jarvik JG, Hollingworth W, Martin B, Emerson SS, Gray DT, Overman S, et al. Rapid magnetic resonance imaging vs radiographs for patients with low back pain: a randomized controlled trial. JAMA 2003;289:2810–8.
24. Van den Bosch MA, Hollingworth W, Kinmonth AL, Dixon AK. Evidence against the use of lumbar spine radiography for low back pain. Clin Radiol 2004;59:69–76.
25. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003;3:25.
26. Scavone JG, Latshaw RF, Rohrer GV. Use of lumbar spine films: statistical evaluation at a university teaching hospital. JAMA 1981;246:1105–8.
27. Reinus WR, Strome G, Zwemer FL Jr. Use of lumbosacral spine radiographs in a level II emergency department. AJR Am J Roentgenol 1998;170:443–7.

28. Holmes JF, Panacek EA, Miller PQ, Lapidis AD, Mower WR. Prospective evaluation of criteria for obtaining thoracolumbar radiographs in trauma patients. J Emerg Med 2003;24:1–7.
29. Gibson M, Zoltie N. Radiography for back pain presenting to accident and emergency departments. Arch Emerg Med 1992;9: 28–31.
30. Durham RM, Luchtefeld WB, Wibbenmeyer L, Maxwell P, Shapiro MJ, Mazuski JE. Evaluation of the thoracic and lumbar spine after blunt trauma. Am J Surg 1995;170:681–4.
31. Samuels LE, Kerstein MD. 'Routine' radiologic evaluation of the thoracolumbar spine in blunt trauma patients: a reappraisal. J Trauma 1993;34:85–9.

APPENDIX A: DECISION AID AND LIST OF "RED FLAGS" COMPLETED BY PRIMARY CARE PRACTITIONERS

1. Please complete the red flag checklist (circle Yes or No)

RED FLAG CHECK LIST			CANCER	INFECTION	SPINAL FRACTURE	INFLAMMATORY DISORDER	CAUDA EQUINA SYNDROME	OTHER
Age at onset less than 20 or over 55 years	Y	N	■					
Gradual onset before age 40	Y	N				■		
Age over 70 years	Y	N			■			
Unexplained weight loss (of more than 10 pounds [4.5 kg] in 6 months)	Y	N	■					
Previous history of cancer	Y	N	■					
Tried bed rest, but no relief	Y	N	■			■		
Insidious onset	Y	N	■			■		
Systemically unwell	Y	N	■	■				
Constant, progressive, non-mechanical pain	Y	N	■					
Recent bacterial infection, e.g., urinary tract or skin infection	Y	N		■				
Intravenous drug abuse	Y	N		■				
Immune suppression from steroids, transplant or HIV	Y	N		■				
Significant trauma (major in young, minor in elderly)	Y	N			■			
Prolonged use of corticosteroids	Y	N			■			
Morning back stiffness, 0.5 hours or more	Y	N				■		
Peripheral joint involvement	Y	N				■		
Persisting limitation of spinal movements in all directions	Y	N				■		
Iritis, skin rashes (psoriasis), colitis, urethral discharge	Y	N				■		
Family history of arthritis or osteoporosis	Y	N				■		
Pain improves with exercise	Y	N				■		
Acute onset of urinary retention or overflow incontinence	Y	N					■	
Loss of anal sphincter tone or fecal incontinence	Y	N					■	
Saddle anesthesia about the anus, perineum or genitals	Y	N					■	
Widespread (greater than 1 nerve root) or progressive motor weakness in the legs or gait disturbances	Y	N					■	
Sensory level (Altered sensation from trunk down)	Y	N	■	■	■			■

FYI: Highlighted disease corresponds to red flag ticked

2. Please choose one diagnosis based on today's full clinical assessment.