A systematic review identifies five "red flags" to screen for vertebral fracture in patients with low back pain

Henschke N, Maher CG, Refshauge KM

CRD summary
This review identified five clinical features that should improve the diagnostic accuracy of vertebral fracture in patients with low back pain. However, the evidence presented suggested that none of the signs evaluated had sufficient accuracy to greatly impact on the probability of fracture, which was not reflected in the authors' conclusions.

Authors' objectives
To determine the diagnostic accuracy of clinical examination in identifying vertebral fracture in patients with low back pain.

Searching
MEDLINE, EMBASE and CINAHL were searched from inception to February 2007 for articles in any language (non-English articles were excluded from analysis if translation was not possible). Search terms were reported. References of relevant reviews and articles were searched manually. Eligible articles were entered into Science Citation Index to identify publications in which they had been cited. Experts in the field were contacted for unpublished data or articles that may have been missed.

Study selection
Diagnostic cohort studies that included patients with low back pain or trauma were eligible for inclusion if they evaluated the accuracy of clinical tests available to primary care practitioners and used diagnostic imaging as the reference standard. Eligible studies were required to report sufficient data to allow the construction of a 2x2 table of test performance.

Most included studies were set in accident and emergency departments; other settings included a university hospital walk-in clinic, a hospital radiology department and hospital inpatients and outpatients. Most studies involved blunt trauma patients. The range in prevalence of fracture varied among studies from 2.1% to 29%. Studies assessed a total of 50 clinical features (categorised as age >50 years, gender, trauma, corticosteroid use, altered consciousness, other injury, pain or tenderness, altered neurological signs and deformity) and global clinician judgement. All studies included X-ray as the reference standard; some studies included additional imaging procedures, such as computed tomography.

Two reviewers screened studies for relevance. It was unclear how discrepancies were resolved.

Assessment of study quality
Two reviewers assessed methodological quality using 13 items on the 14-item QUADAS tool. Discrepancies were resolved by consensus.

Data extraction
Two authors independently extracted data on true-positive, true-negative, false-positive and false-negative results for each clinical sign. Sensitivity, specificity, and positive and negative likelihood ratios (LR+ and LR-) were calculated for each set of 2x2 data, together with their 95% confidence intervals (CIs). A value of 0.5 was added to all cells to avoid computational problems. Disagreements were resolved through consensus.

Methods of synthesis
Due to significant heterogeneity among studies, data were presented as a narrative synthesis and in tables, grouped by clinical signs (categorised as age, gender, trauma, corticosteroid use, altered consciousness, other injury, pain/tenderness, altered neurological signs and deformity).
Results of the review

Twelve cohort studies, eight prospective (three also consecutive) and four retrospective chart reviews (n=7,147) were included. Sample size ranged between 71 and 2,404 patients. Included studies met between four and seven items using the QUADAS tool. Reporting of details on the index tests and the reference standard was poor. Most studies provided index test results only for those patients who received the reference standard.

None of the signs or symptoms evaluated showed consistently good performance for ruling in or ruling out fracture.

Age showed some association with fracture: age over 50 had some potential to rule in vertebral fracture (LR+ 2.2, 95% CI 1.4 to 2.8 and 1.7, 95% CI 1.5 to 1.9; two studies) and had some potential to rule out fracture (LR- 0.34, 95% CI 0.12 to 0.75 and 0.35, 95% CI 0.22 to 0.54). Women were also at slightly increased risk of fracture, but this was insufficient to be used to rule in (LR+ 1.3, 95% CI 1.1 to 1.4 and 2.3, 95% CI 1.1 to 4.3) or rule out fracture (LR- 0.65, 95% CI 0.45 to 0.90 and 0.67, 95% CI 0.37 to 0.97; two studies).

Studies on trauma reported inconsistent results that suggested slight to good potential for ruling in fracture (LR+ ranged from 1.7 to 14.4) and poor to excellent performance in ruling out fracture (LR- ranged from 0.00 to 0.78). Only two studies found that features related to trauma could be used to rule in a diagnosis of fracture (LR+ 12.8 and 14.4); one of these studies reported some potential for ruling out fracture (LR- 0.37) and the other suggested poor performance (LR- 0.73).

The presence of pain and/or tenderness showed inconsistent performance for ruling in or ruling out fracture. Use of corticosteroids, altered consciousness and presence of other injuries also showed poor performance both for ruling in and ruling out a diagnosis of fracture.

Sensitivity and specificity were reported in the review.

Authors’ conclusions

The presence or absence of five clinical features, which included age greater than 50 years, female gender, major trauma, pain and tenderness, and a distracting painful injury should improve the diagnostic accuracy of vertebral fracture. Further high-quality research was required.

CRD commentary

The review question was clear and was supported by appropriate inclusion criteria. A comprehensive literature search was conducted that used various relevant sources and included attempts to locate unpublished data, which reduced the risk that potentially relevant papers were missed. Attempts were made to include articles in any language, thus the risk of language bias was minimised. Study quality was assessed using appropriate criteria and was found to be poor. Attempts were made to reduce the potential for reviewer error and bias at each stage of the review. Due to significant heterogeneity between studies, a narrative synthesis was appropriate. No attempts were made to try and investigate the factors that influenced heterogeneity. Study quality was not taken into consideration in the synthesis. Further limitations included the classification of positive and negative likelihood ratios as significant or not significant, which was deemed inappropriate. The evidence presented suggested that none of the signs evaluated had sufficient accuracy to greatly impact on the probability of fracture and did not appear to reflect the authors’ conclusions.

Implications of the review for practice and research

Practice: The authors stated that the results must be considered with respect to the setting in which the features were assessed and the relevance of the clinical populations included. They also stated that as no clinical guidelines recommend the routine use of X-rays, clinicians should be able to exclude patients without vertebral fracture to avoid patients undergoing unnecessary diagnostic testing without missing fractures.

Research: The authors stated that there was a clear need for future high-quality studies to assess screening for vertebral fracture in low back pain patients.

Funding
Two reviewers received financial support from the National Health and Medical Research Council of Australia.

**Bibliographic details**

**PubMedID**
18177783

**DOI**
10.1016/j.jclinepi.2007.04.013

**Original Paper URL**
http://www.jclinepi.com/article/S0895-4356(07)00151-5/abstract

**Indexing Status**
Subject indexing assigned by NLM

**MeSH**
Acute Disease; Aged; Evidence-Based Medicine /standards; Female; Humans; Low Back Pain /etiology; Male; Mass Screening /methods; Middle Aged; Practice Guidelines as Topic; Psychometrics; Quality Assurance, Health Care /standards; Risk Factors; Spinal Fractures /complications /diagnosis

**AccessionNumber**
12008000211

**Date bibliographic record published**
02/03/2009

**Date abstract record published**
16/12/2009

**Record Status**
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.