RESEARCH

Red flags to screen for malignancy and fracture in patients with low back pain: systematic review

OPEN ACCESS

Aron Downie *PhD student*[1][2], Christopher M Williams *honorary research fellow*[1], Nicholas Henschke *research fellow*[1][3], Mark J Hancock *senior lecturer*[4], Raymond W J G Ostelo *professor*[5], Henrica C W de Vet *professor of clinimetrics*[6], Petra Macaskill *professor of biostatistics*[7], Les Irwig *professor of epidemiology*[8], Maurits W van Tulder *professor*[9], Bart W Koes *professor*[10], Christopher G Maher *director*[1]

[1]George Institute for Global Health, University of Sydney, Sydney, NSW, 2050, Australia; [2]Faculty of Science, Macquarie University, Sydney, Australia; [3]Institute of Public Health, University of Heidelberg, Germany; [4]Faculty of Human Sciences, Macquarie University, Sydney, Australia; [5]Department of Health Sciences, EMGO Institute for Health and Care Research, VU University, Amsterdam, Netherlands; [6]Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University Medical Centre, Amsterdam; [7]Screening and Test Evaluation Program (STEP), School of Public Health, Sydney; [8]School of Public Health, University of Sydney, Sydney, Australia; [9]Department of Health Sciences, Faculty of Earth and Life Sciences, VU University, Amsterdam, Netherlands; [10]Department of General Practice, Erasmus Medical Centre, Rotterdam, Netherlands

Abstract

Objective To review the evidence on diagnostic accuracy of red flag signs and symptoms to screen for fracture or malignancy in patients presenting with low back pain to primary, secondary, or tertiary care.

Design Systematic review.

Data sources Medline, OldMedline, Embase, and CINAHL from earliest available up to 1 October 2013.

Inclusion criteria Primary diagnostic studies comparing red flags for fracture or malignancy to an acceptable reference standard, published in any language.

Review methods Assessment of study quality and extraction of data was conducted by three independent assessors. Diagnostic accuracy statistics and post-test probabilities were generated for each red flag.

Results We included 14 studies (eight from primary care, two from secondary care, four from tertiary care) evaluating 53 red flags; only five studies evaluated combinations of red flags. Pooling of data was not possible because of index test heterogeneity. Many red flags in current guidelines provide virtually no change in probability of fracture or malignancy or have untested diagnostic accuracy. The red flags with the highest post-test probability for detection of fracture were older age (9%, 95% confidence interval 3% to 25%), prolonged use of corticosteroid drugs (33%, 10% to 67%), severe trauma (11%, 8% to 16%), and presence of a contusion or abrasion (62%, 49% to 74%). Probability of spinal fracture was higher when multiple red flags were present (90%, 34% to 99%). The red flag with the highest post-test probability for detection of spinal malignancy was history of malignancy (33%, 22% to 46%).

Conclusions While several red flags are endorsed in guidelines to screen for fracture or malignancy, only a small subset of these have evidence that they are indeed informative. These findings suggest a need for revision of many current guidelines.

Introduction

Low back pain is a major cause of disability,[1] leading to considerable healthcare expenditure around the world, especially in high income countries.[2] The difficulty in providing a definitive diagnosis for most presentations of back pain has given rise to the term "non-specific low back pain," which is generally considered to be benign and can be managed in a primary care setting.[3] Some patients, however, present with low back pain as the initial manifestation of a more serious pathology, such as malignancy, spinal fracture, infection, or cauda equina syndrome. Spinal fracture and malignancy are the most common serious pathologies affecting the spine. In patients with low back pain presenting to primary care, between 1% and 4% will have a spinal fracture[4] and in less than 1% malignancy, whether primary tumour or metastasis, will be the underlying cause.[5]

Correspondence to: A Downie, George Institute for Global Health, University of Sydney, PO Box M201, Camperdown, Sydney, NSW, 2050, Australia adownie@georgeinstitute.org.au

Identification of serious pathologies, when they exist, is important in the clinical assessment and further assessment and specific treatment is usually required, particularly for malignancy.[6 7] For instance, early detection of spinal malignancy could prevent further spread of metastatic disease.[8] Identification of spinal fracture will prevent the prescription of treatment such as manual therapy, which is contraindicated,[9] as well as progress the patient towards further testing and treatment of underlying disease (such as osteoporosis). Despite the potential consequences of a late or missed diagnosis of these serious pathologies, their low prevalence in primary care settings does not justify routine ancillary testing of patients presenting with low back pain. For this reason, accurate screening tools to aid clinical decisions about when to refer for further testing are paramount.

Most clinical practice guidelines for back pain recommend the use of red flags to help identify those patients with a higher likelihood of spinal fracture or malignancy who then become candidates for more extensive diagnostic investigations. There is confusion, however, as the guidelines have produced different lists of red flags to screen for spinal fracture and malignancy. Eight of the guidelines[6 10-16] investigated by Koes and colleagues in their review of back pain guidelines,[3] endorsed 26 red flags for fracture and 27 for malignancy. None of the eight guidelines endorsed the same set of red flags, for either condition, so it is unclear what clinicians should use in clinical care. Additionally, guidelines generally provide no information on diagnostic accuracy of the endorsed red flags, which limits their value in clinical decision making. Adding to the uncertainty, the same agency can provide inconsistent information on red flags. For example, the National Institute for Health and Care Excellence clinical guideline on the early management of persistent non-specific low back pain[4] does not endorse red flags, whereas the group's clinical knowledge summary for the management of low back pain does.[17]

To resolve the uncertainty around application of red flags in clinical practice, we conducted two Cochrane diagnostic test accuracy reviews assessing the accuracy of red flags to screen for the most common forms of serious pathology—spinal fracture and malignancy—in patients with low back pain.[4 5] We have provided a distilled summary of both reviews to help guide clinical decision making.

Methods

Data sources

For the Cochrane diagnostic test accuracy reviews, we searched Medline, OldMedline, Embase, and CINAHL (Ebsco) for eligible studies from the earliest record up to 7 March 2012. The search was updated to include studies up to 1 October 2013. The searches used combinations of terms related to the patient population, history taking, physical examination, and the target condition and was developed in collaboration with a medical information specialist. Forward and backward citation searches were completed.

Study selection

Primary diagnostic studies were considered if they looked at the results of history taking or physical examination compared with those of an acceptable reference standard, to identify spinal fracture or spinal malignancy in adult patients presenting with low back pain. Examples of reference standards include diagnostic imaging to confirm the presence of spinal fracture, primary malignancy, or metastases in the spine. Long term follow-up (more than six months) of patients after the initial

consultation was also considered an appropriate reference standard for both fracture and malignancy if suspected cases identified during the follow-up period were confirmed by medical review. Cohort and cross sectional studies of a consecutive series of patients that presented sufficient data to allow estimates of diagnostic accuracy (such as sensitivity and specificity) were considered for eligibility. We considered studies reported in abstracts or conference proceedings and full journal publications in all languages and excluded studies in which "global clinician judgment" was the only red flag investigated. Four review authors (CMW, NH, and AD for fracture; NH, RWJGO, and AD for malignancy) independently applied the selection criteria to retrieved citations. Final selection was based on a review of full publications. Disagreements were resolved by consensus or by consulting other review authors in cases of persisting disagreement.

Data extraction and quality assessment

Three review authors (CMW and NH for fracture; NH and RWJGO for malignancy) independently extracted data on selection of participants, the index test, reference test, flow of participants, timing of the study, and diagnostic accuracy. Review authors independently assessed the risk of bias of each study using the QUality Assessment of Diagnostic Accuracy Studies (QUADAS) checklist.[18] We used the 11 item version of the QUADAS recommended by the Cochrane Diagnostic Test Accuracy Working Group.[19] This checklist (see table A in appendix) was used to classify each item as "yes" (adequately dealt with); "no" (inadequately dealt with); or "unclear" (inadequate detail presented to allow a judgment to be made). Disagreements were resolved by consensus. Authors of the current review who were involved in the original study were not involved in data extraction or quality assessment of that study.

Data synthesis and analysis

We generated diagnostic 2×2 tables to characterise diagnostic accuracy. Because of the heterogeneity of tests, study settings, and methods we could not pool results. We used Review Manager 5.2[20] to calculate study specific estimates of sensitivity and specificity with 95% confidence intervals and MetaDisc 1.4[21] to calculate likelihood ratios with 95% confidence intervals. Post-test probability was determined with standard methods[22 23] with the 95% confidence interval for post-test probability determined with point estimate of pre-test probability and the 95% confidence interval of the likelihood ratio.

We compared the diagnostic accuracy of red flags against the clinical practice guideline developed by Chou and colleagues[6 24] for the American College of Physicians and American Pain Society. The guideline provides advice on diagnostic investigation based on various patterns of risk factors (red flags).

Results

Search results

The electronic database search and citation tracking identified 13 669 unique titles for fracture and 3092 for malignancy (fig 1⇓). After screening of titles and abstracts, we retrieved full text copies of 68 articles relating to fracture and 70 full text articles relating to malignancy. Fourteen discrete studies were included in the reviews, of which eight related to fracture and nine related to malignancy (three studies dealt with both conditions).

For each included study, details on the design, setting, population, reference standard, index tests used, and definition

of the target condition are provided in tables B and C in the appendix. Most studies (eight) were set in primary care,[7 25-31] two in secondary care,[32 33] and the four remaining in tertiary care (three in emergency departments[34-36] and one in a spinal surgical unit[37]). Three studies investigated red flags for both fracture and malignancy.[7 27 34] Five of the eight primary care studies were prospective in design, and three were retrospective chart reviews (two for fracture, one for malignancy). Standard radiographs were the most common reference standard for diagnosing fracture, and the most common reference standards used for malignancy were magnetic resonance imaging or long term clinical follow-up.

Quality assessment

Appendix table D shows the individual results of quality assessment for included studies. The items that were not adequately dealt with were: an acceptable delay between index and reference tests, partial verification, differential verification, reference standard blinding, reporting uninterpretable results, and explaining withdrawals. Four criteria (acceptable delay between tests, differential verification, reference blinding, and explaining withdrawals) were often scored as "unclear," indicating that studies provided insufficient information. The specific details and criteria used for imaging reference standards to diagnose either condition (such as views taken or reporting procedure), though present, were usually poorly described.

Diagnostic accuracy

Prevalence varied among studies for both fracture (primary care: median 3.6%, interquartile range 1.8-4.3%; secondary/tertiary care: 6.5%, 2.9-9.1%), and malignancy (primary care: 0.2%, 0.1-0.7%; secondary care: one study 7%; tertiary care: two studies 1.5% and 5.9%). Point prevalence used to calculate post-test probability in this review was determined by extracting prevalence from a reduced set of methodologically robust studies for fracture[7 33 35] and cancer,[7 27] and by considering a value that could be readily applied in the clinical setting (fracture: 1% for primary care, 5% for secondary and tertiary care; malignancy: 0.5% for primary care, 1.5% for secondary and tertiary care).

The reviews identified 29 red flags to screen for fracture[7 25-27 33-36] and 24 to screen for malignancy,[7 27-32 34 37] with 13 (25%) evaluated in more than one study. Only six studies reported on the accuracy of combinations of red flags,[7 25 33 35-37] and few studies provided precise definitions for each red flag. Figures 2-5⇓⇓⇓⇓ show the positive and negative likelihood ratios and associated post-test probabilities for each red flag. Figures 2 and 4 are confined to the red flags endorsed in the American College of Physicians clinical practice guideline with additional red flags identified in our review reported in figures 3 and 5. The full diagnostic accuracy data, both raw data and calculated statistics, are provided in appendix tables E-G.

Red flags for spinal fracture

Figure 2 ⇓ presents the information on diagnostic accuracy data for red flags recommended in the American College of Physicians guideline. The first column contains the specific wording of the red flag from the guideline and the next column the specific wording of similar red flags evaluated in the diagnostic studies. For each evaluated red flag we have provided the positive and negative likelihood ratios, typical prevalence, and post-test probability if the red flag is present or absent. We found inconsistent descriptions between the red flags evaluated in primary studies and those presented in the guideline. For example, the American College of Physicians guideline red flag

"older age" (men aged >65, women aged >75) is compared with the similar red flag (age >64, van den Bosch and colleagues[25]) that had a positive likelihood ratio of 2.5 (95% confidence interval 2.2. to 2.8) and negative likelihood ratio of 0.3 (0.2 to 0.5). As this study was conducted in primary care, the typical prevalence was set at 1% and is represented by the vertical line in the figure. The probability of fracture given the presence of the red flag is 2% (95% confidence interval 2% to 3%). A second study using the same red flag (age >64, Henschke and colleagues[7]) found the probability of fracture given presence of red flag to be 7% (4% to 13%).

Of the four red flags endorsed in the American College of Physicians guideline we could find data only on older age, trauma, and corticosteroid use and not on history of osteoporosis. In general, the presence of these red flags increased the likelihood of fracture by up to 15% (older age, trauma). The exception was the red flag "prolonged corticosteroid use," which, when present, suggested a post-test probability of 33% (95% confidence interval 10% to 67%) when positive.[7] By contrast, the American College of Physicians guideline uses the red flag "corticosteroid use," as studied by Deyo and Diehl[27], which, when present, results in a post-test probability of 4% (0% to 44%).[27]

Figure 3⇓ shows the red flags not endorsed in the American College of Physicians guideline. In general, when present, these produce trivial increases in probability of fracture; few produced a precise estimate beyond the pre-test prevalence. The exceptions are contusion or abrasion reported by Patrick and colleagues (62%, 95% confidence interval 49% to 74%, post-test probability),[35] the combination of red flags (any four of leg or buttock pain, female, older age, BMI <23, gait abnormality, no regular exercise, sitting pain, osteoarthritis) reported by Roman and colleagues (34%, 24% to 45%, post-test probability),[33] the combination of red flags (any three of female, age >70, severe trauma, prolonged use of corticosteroids) reported by Henschke and colleagues (90%, 34% to 99%, post-test probability),[7] and trauma with neurological signs reported by Gibson and Zoltie (43%, 11% to 82%, post-test probability).[36]

Red flags for spinal malignancy

Figure 4⇓ shows that of the red flags classified as "major risk" in the American College of Physicians guideline, only data on the red flag "history of cancer" are available. When present, "history of cancer" suggests a post-test probability of 7% (95% confidence interval 3% to 16%) in primary care,[31] and 33% (22% to 46%) in the emergency setting.[34] Red flags classified as "minor risk" in the American College of Physicians guideline (older age, unexplained weight loss, and failure to improve after one month) have post-test probability point estimates below 3%.[27 29 31 32 37] Figure 5⇓ shows red flags not endorsed in the American College of Physicians guideline; all seem uninformative.[30 31 37] The exception is neurological symptoms, for which there are contradictory data.[28 31]

Discussion

The informativeness of the large number of red flags endorsed in guidelines for the management of non-specific low back pain in primary care[3] varies substantially and many have poor or untested diagnostic accuracy. Of the red flags for fracture, older age, prolonged steroid use, severe trauma, and contusion or abrasion increased the probability of fracture to between 10% and 33%, while the presence of multiple red flags increased the probability of fracture to between 42% and 90%. Of the red flags for malignancy, "history of cancer" increased the

probability of malignancy to between 7% and 33% while older age, unexplained weight loss, and failure to improve after one month have post-test probabilities below 3%.

Our results support the approach taken in the American College of Physicians guideline, which provides a more focused list of red flags than other guidelines, and emphasises consideration of the low probability of disease (given the specific red flags present) when making decisions about the need for, and timing of, further diagnostic investigation. While older age, steroid use, and severe trauma are endorsed by the guideline to screen for fracture, the combined red flags prolonged steroid use and contusion or abrasion are absent. Our results suggest consideration of their inclusion when the guideline is revised. Of red flags endorsed by the guideline to screen for malignancy, a history of malignancy was the only red flag that was found to increase the chance of malignancy to greater than 7%.[7 31] Our data show that all red flags endorsed in the guideline as "minor risk" for cancer had post-test probabilities below 3%, supporting the guideline's distinction between major risk and minor risk red flags for cancer.

Many guidelines contain no information on diagnostic accuracy for individual red flags.[12-14 16 38] For example, the European guideline for the management of chronic non-specific low back pain[12] endorses 10 red flags for conditions including fracture and malignancy: patient aged <20 or >55, non-mechanical pain, thoracic pain, history of cancer, steroid use, structural changes, general unwellness, loss of weight, and diffuse neurological deficit. Of these, we found no evidence for the red flags age <20 or general unwellness. The red flags age >55, thoracic pain, non-mechanical pain (with movement), structural change (scoliosis, kyphosis), and loss of weight were all uninformative, with contradictory data on the red flag "diffuse neurological deficit." Only history of cancer and steroid use (prolonged) were found to be informative.

In addition to endorsing red flags with low or no diagnostic accuracy, we found guidelines that recommended immediate referral to imaging if any red flag was present.[13 14 39] For example the West Australian Diagnostic Imaging Pathways Guideline[39] advocates a larger list of red flags and directs clinicians to image patients when even a single red flag is present. If this advice was followed it would lead to substantial and arguably unwarranted referrals for imaging as some of the endorsed red flags are common (such as age <20 or >55)[9] or uninformative (such as thoracic pain, which has a positive likelihood ratio of about 1 in screening for malignancy).[31]

Strengths and weaknesses of the review

We combined two previous reviews[4 5] that followed published pre-specified protocols[40 41] and adopted the methods endorsed by the Cochrane Collaboration to search for, appraise, and summarise the evidence (http://srdta.cochrane.org/handbook-dta-reviews). By illustrating the pre- and post-test probabilities in figures we provide a simple method to describe the utility of red flags to inform clinical decisions about their use in practice. Also, the ability to view recommendations of guidelines in conjunction with our diagnostic accuracy data provides insight into the lack of standard for what constitutes a red flag.

A challenge in applying the results of diagnostic research in practice is that some key statistics might be misunderstood by clinicians.[42 43] Provision of a list of red flags with sensitivity and specificity values might therefore not be optimal. We have graphically portrayed the post-test probability and 95% confidence intervals for investigated red flags. The figures enable clinicians to easily interpret the informativeness of red

flags to screen for spinal fracture and malignancy.[44] A limitation of this approach is that prevalence of fracture and malignancy varied considerably between studies (fracture from 0.7% to 11.0%; malignancy from 0% to 7.0%) and depended on study methods and setting. Clinical and artefactual variability could also have contributed to the differences in prevalence.[45] Because of study heterogeneity and variability in quality, prevalence per care setting was determined by using only the most methodologically robust studies. Therefore values for prevalence and post-test probability in our review might not generalise to every setting. We have included sensitivities, specificities, and likelihood ratios in the appendix for all red flags to enable calculation of post-test probability for different prevalence rates. There is still work to be done in determining the best approach to bringing this information to the point of care.

Possible explanations and implications for clinicians and policymakers

Unnecessary imaging is a concern in many settings, with adverse consequences for the patient and society.[6 7 46-48] Our review shows that the advice, provided in some guidelines, to refer all patients with a single positive red flag for imaging is unwise. Given that 80% of patients in primary care might have at least one positive red flag,[7] such a course would mean that most patients with low back pain would receive diagnostic imaging. In contrast, our review is consistent with the approach in the guidelines of the American College of Physicians guideline and others[7 27 46] to first of all consider the probability of serious disease (given the specific red flags present) when making decisions about the need for and timing of imaging. Our results suggest that many guidelines will need revision so that they provide advice for practice that is more firmly grounded in the relevant diagnostic research.

Unanswered questions and future research

Our review has highlighted the need for more high quality diagnostic research on the topic. Much of the existing research has evaluated single clinical features and our results suggest it might be more useful to evaluate combinations of clinical features, as is done with decision rules such as the Ottawa ankle rule[49] and Canadian C-Spine Rule.[50 51] While some decision rules are available,[7 33] all are at the derivation stage and await external validation. There is a need for systematic reviews of red flags to screen for other conditions such as infection and ankylosing spondylitis. Finally, consideration of the optimal way to provide this information so that it is available to the clinician at the point of care, in a readily understood format, is an important topic for future research.

What is already known on this topic

Most clinical practice guidelines provide "red flags" to suggest a need to screen for spinal fracture or malignancy in patients presenting with low back pain but do not agree on which ones to use

The total number of red flags endorsed in clinical guidelines is large

What this study adds

We identified the evidence for diagnostic accuracy of red flags for spinal fractures and spinal malignancies

Older age, prolonged corticosteroid use, severe trauma, and presence of a contusion or abrasion increased the likelihood of spinal fracture; likelihood was higher with multiple red flags

Only a history of malignancy increased the likelihood of spinal malignancy

1   Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;380:2197-223.
2   Hoy D, Brooks P, Blyth F, Buchbinder R. The epidemiology of low back pain. *Best Pract Res Clin Rheumatol* 2010;24:769-81.
3   Koes BW, van Tulder M, Lin C-WC, 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.
4   Williams CM, Henschke N, Maher CG, van Tulder MW, Koes BW, Macaskill P, et al. Red flags to screen for vertebral fracture in patients presenting with low-back pain. *Cochrane Database Syst Rev* 2013;1:CD008643.
5   Henschke N, Maher CG, Ostelo RW, de Vet HCW, Macaskill P, Irwig L. Red flags to screen for malignancy in patients with low-back pain. *Cochrane Database Syst Rev* 2013;2:CD008686.
6   Chou R, Qaseem A, Owens DK, Shekelle P. Clinical guideline diagnostic imaging for low back pain: advice for high-value health care from the American College of Physicians. *Ann Intern Med* 2011;154:181-9.
7   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.
8   Loblaw DA, Perry J, Chambers A, Laperriere NJ. Systematic review of the diagnosis and management of malignant extradural spinal cord compression: the Cancer Care Ontario Practice Guidelines Initiative's Neuro-Oncology Disease Site Group. *J Clin Oncol* 2005;23:2028-37.
9   Waddell G. The back pain revolution. 2nd ed. Churchill Livingstone, 2004.
10 Friedrich M, Likar R. [Evidence and consensus based Austrian guidelines for management of acute and chronic nonspecific backache]. *Wien Klin Wochenschr* 2007;119:189-97.
11 Laerum E, Storheim K, Brox JI. [New clinical guidelines for low back pain]. *Tidsskr Nor Laegeforen* 2007;127:2706.
12 Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, et al. Chapter 4 European guidelines for the management of chronic nonspecific low back pain. *Eur Spine J* 2006;15:s192-300.
13 Negrini S, Giovannoni S, Minozzi S, Barneschi G, Bonaiuti D, Bussotti A, et al. Diagnostic therapeutic flow-charts for low back pain patients: the Italian clinical guidelines. *Eura Medicophys* 2006;42:151-70.
14 National Health and Medical Research Council. Evidence-based management of acute musculoskeletal pain. NHMRC, 2004.
15 National Institute for Health and Care Excellence. NICE Clinical Knowledge Summaries. Back pain—low (without radiculopathy) 2009. http://cks.nice.org.uk/back-pain-low-without-radiculopathy.
16 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;15(suppl 2):S169-91.
17 National Institute for Health and Care Excellence. Low back pain: early management of persistent non-specific low back pain. NICE clinical guideline 88. Jul 2013. http://guidance.nice.org.uk/cg88.
18 Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, 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.
19 Reitsma JB, Rutjes AWS, Whiting P, Vlassov VV, Leeflang MMG, Deeks JJ. Chapter 9: assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Version 1.0.0. Cochrane Collaboration, 2009. http://srdta.cochrane.org/.
20 Review Manager (RevMan). Version 5.2. Nordic Cochrane Centre, Cochrane Collaboration, 2012.
21 Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol* 2006;6:31.
22 Akobeng AK. Understanding diagnostic tests 2: likelihood ratios, pre- and post-test probabilities and their use in clinical practice. *Acta Paediatr* 2007;96:487-91.
23 Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ* 2004;329:168-9.
24 Chou R, Qaseem A, Snow V, Casey D, Cross Jr JT, Shekelle 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.
25 Van den Bosch MAAJ, Hollingworth W, Kinmonth AL, Dixon AK. Evidence against the use of lumbar spine radiography for low back pain. *Clin Radiol* 2004;59:69-76.
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 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.
28 Khoo LA, Heron C, Patel U, Given-Wilson R, Grundy A, Khaw KT, et al. The diagnostic contribution of the frontal lumbar spine radiograph in community referred low back pain—a prospective study of 1030 patients. *Clin Radiol* 2003;58:606-9.
29 Frazier LM, Carey TS, Lyles MF, Khayrallah MA, McGaghie WC. Selective criteria may increase lumbosacral spine roentgenogram use in acute low-back pain. *Arch Intern Med* 1989;149:47-50.
30 Donner-Banzhoff N, Roth T, Sonnichsen AC, Luckmann J, Leonhardt C, Chenot JF, et al. Evaluating the accuracy of a simple heuristic to identify serious causes of low back pain. *Fam Pract* 2006;23:682-6.
31 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.
32 Jacobson AF. Musculoskeletal pain as an indicator of occult malignancy. Yield of bone scintigraphy. *Arch Intern Med* 1997;157:105-9.
33 Roman M, Brown C, Richardson W, Isaacs R, Howes C, Cook C. The development of a clinical decision making algorithm for detection of osteoporotic vertebral compression fracture or wedge deformity. *J Manipulative Physiol Ther* 2010;18:44-9.
34 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.
35 Patrick JD, Doris PE, Mills ML, Friedman J, Johnston C. Lumbar spine x-rays: a multihospital study. *Ann Emerg Med* 1983;12:84-7.
36 Gibson M, Zoltie N. Radiography for back pain presenting to accident and emergency departments. *Arch Emerg Med* 1992;9:28-31.
37 Cook C, Ross MD, Isaacs R, Hegedus E. Investigation of nonmechanical findings during spinal movement screening for identifying and/or ruling out metastatic cancer. *Pain Pract* 2012;12:426-33.
38 Poitras S, Rossignol M, Dionne C, Tousignant M, Truchon M, Arsenault B, et al. An interdisciplinary clinical practice model for the management of low-back pain in primary care: the CLIP project. *BMC Musculoskelet Disord* 2008;9:54.
39 Government of Western Australia Department of Health. Diagnostic imaging pathways—low back pain. May 2013. www.imagingpathways.health.wa.gov.au/index.php/imaging-pathways/musculoskeletal-trauma/musculoskeletal/low-back-pain#pathway.
40 Henschke N, Williams CM, Maher CG, van Tulder MW, Koes BW, Macaskill P, et al. Red flags to screen for vertebral fracture in patients presenting with low-back pain (Protocol). *Cochrane Database Syst Rev* 2010:CD008643.
41 Henschke N, Maher CG, Ostelo RWJG, de Vet HCW, Macaskill P, Irwig L. Red flags to screen for malignancy in patients with low-back pain (Protocol). *Cochrane Database Syst Rev* 2010;9:CD008686.
42 Steurer J, Fischer JE, Bachmann LM, Koller M, Riet Gt. Communicating accuracy of tests to general practitioners: a controlled study. *BMJ* 2002;324:824-6.
43 Bachmann LM, Steurer J, Riet Gt. Simple presentation of test accuracy may lead to inflated disease probabilities. *BMJ* 2003;326:393-4.
44 Deeks JJ. Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001;323:157-62.
45 Leeflang MM, Bossuyt PM, Irwig L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. *J Clin Epidemiol* 2009;62:5-12.
46 Jarvik JG, Deyo RA. Diagnostic evaluation of low back pain with emphasis on imaging. *Ann Intern Med* 2002;137:586-97.
47 Rhodes LA, McPhillips-Tangum CA, Markham C, Klenk R. The power of the visible: the meaning of diagnostic tests in chronic back pain. *Soc Sci Med* 1999;48:1189-203.
48 Deyo RA. Real help and red herrings in spinal imaging. *N Engl J Med* 2013;368:1056-8.
49 Bachmann LM, Kolb E, Koller MT, Steurer J, Riet Gt. Accuracy of Ottawa ankle rules to exclude fractures of the ankle and mid-foot: systematic review. *BMJ* 2003;326:417.
50 Stiell IG, Wells GA, Vandemheen KL, Clement CM, Lesiuk H, De Maio VJ, et al. The Canadian C-spine rule for radiography in alert and stable trauma patients. *JAMA* 2001;286:1841-8.
51 Michaleff ZA, Maher CG, Verhagen AP, Rebbeck T, Lin C-WC. Accuracy of the Canadian C-spine rule and NEXUS to screen for clinically important cervical spine injury in patients following blunt trauma: a systematic review. *Can Med Assoc J* 2012;184:E867-76.

Cite this as: *BMJ* 2013;347:f7095

Figures
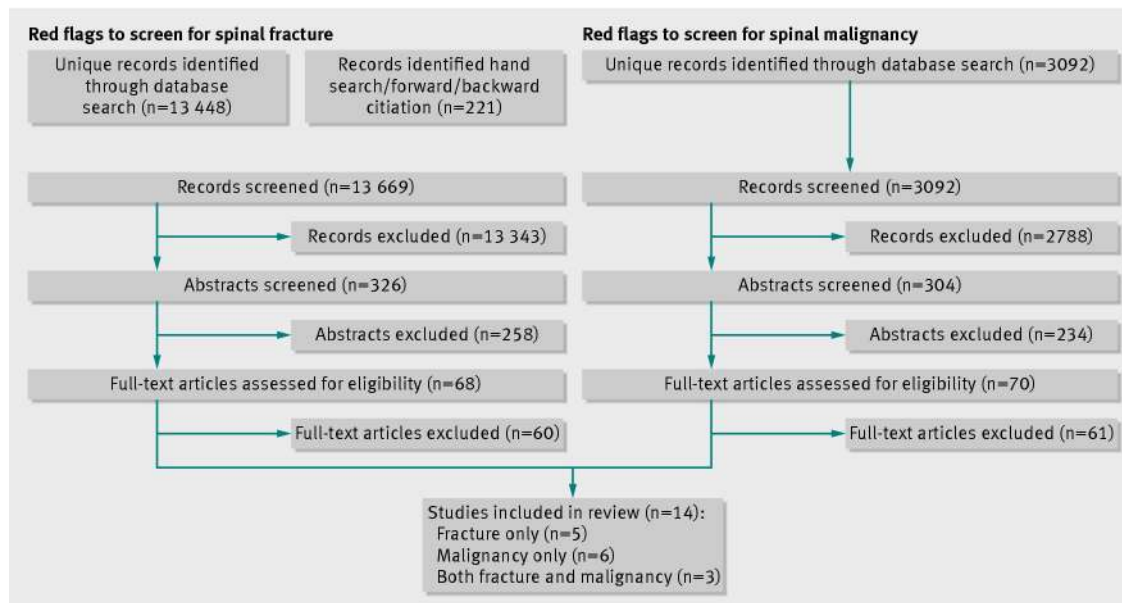
Fig 1 Flow diagram of search strategies for studies evaluating red flags to screen for spinal fracture and malignancy in patients presenting with low back pain

Fig 2 Diagnostic accuracy of red flags for spinal fracture included in American College of Physicians (ACP) guideline. Vertical line indicates prevalence of spinal fracture: 1% in primary care, 5% in secondary and tertiary care. *Severe trauma: trauma such as fall from height or motor vehicle crash (Deyo[27]); trauma which is major in young patients and minor in elderly (Henschke[7]). †Minor trauma: in elderly women with osteoporosis (Scavone[26]). ‡Trauma: history of direct trauma (Gibson[36]); history of trauma (Patrick,[35] Reinus[34]). §History of osteoporosis not evaluated by any included study

Red flag (not in ACP guideline)	Probability (95% CI)	Likelihood ratio (95% CI)	
		Positive	Negative
Primary care			
Demographic			
Age >50[27]		2.2 (1.6 to 3.0)	0.3 (0.1 to 0.9)
Age >50[7]		1.8 (1.1 to 3.2)	0.6 (0.2 to 1.4)
Age >54[7]		2.6 (1.5 to 4.4)	0.5 (0.2 to 1.2)
Age >54[25]		1.7 (1.5 to 1.9)	0.3 (0.2 to 0.5)
Female[25]		1.3 (1.1 to 1.5)	0.6 (0.5 to 0.9)
History			
Hip/leg pain[26]		0.2 (0.0 to 3.4)	1.1 (1.0 to 1.1)
Sciatica[26]		0.4 (0.1 to 2.9)	1.1 (1.0 to 1.1)
Spasm[26]		1.3 (0.4 to 3.7)	1.0 (0.8 to 1.1)
Examination			
Tenderness[26]		1.9 (1.3 to 2.8)	0.7 (0.5 to 1.0)
Sensation change[7]		3.3 (0.2 to 50.9)	1.0 (0.8 to 1.1)
Sensation change[26]		2.2 (1.1 to 4.3)	0.8 (0.7 to 1.1)
Motor deficit[26]*		2.2 (1.1 to 4.5)	0.9 (0.7 to 1.1)
DTR abnormality[26]		1.1 (0.4 to 3.2)	1.0 (0.9 to 1.1)
Combination red flags			
Age >54 + female[7]		5.4 (3.1 to 9.4)	0.4 (0.2 to 1.0)
Age >54 + female[25]		2.0 (1.7 to 2.4)	0.5 (0.4 to 0.7)
Henschke 1 positive[7]†		1.8 (1.3 to 2.3)	0.3 (0.0 to 1.6)
Henschke 2 positive[7]		15.5 (8.4 to 28.4)	0.4 (0.2 to 1.0)
Henschke 3 positive[7]		906 (50 to 16300)	0.6 (0.4 to 1.0)
Secondary care			
Demographic			
Age >52[33]		1.6 (1.4 to 1.7)	0.1 (0.0 to 0.5)
History			
Female[33]		1.5 (1.3 to 1.7)	0.3 (0.1 to 0.7)
Absence of buttock/leg pain[33]		2.2 (1.4 to 3.6)	0.8 (0.6 to 1.0)
Decreased pain on sitting[33]		1.6 (0.9 to 2.6)	0.9 (0.7 to 1.1)
No regular exercise[33]		1.5 (1.3 to 1.7)	0.4 (0.2 to 0.8)
Examination			
BMI <23[33]		2.2 (1.4 to 3.4)	0.8 (0.6 to 1.0)
No gait abnormality[33]		0.9 (0.7 to 1.1)	1.5 (0.9 to 2.3)
Osteoarthritis[33]		1.1 (0.8 to 1.5)	1.0 (0.7 to 1.3)
Combination red flags			
1 positive[33]‡		1.0 (1.0 to 1.1)	0.4 (0.1 to 3.0)
2 positive[33]		1.4 (1.3 to 1.6)	0.2 (0.0 to 0.6)
3 positive[33]		2.4 (2.0 to 3.0)	0.3 (0.2 to 0.6)
4 positive[33]		9.6 (5.9 to 15.7)	0.7 (0.5 to 0.8)
5 positive[33]		7.6 (0.9 to 63.7)	1.0 (0.9 to 1.0)
Tertiary care			
Examination			
Contusion or abrasion[35]		31.1 (18.2 to 53.0)	0.2 (0.1 to 0.3)
Spasm[35]		1.5 (0.8 to 2.6)	0.9 (0.8 to 1.1)
Tenderness[35]		1.8 (1.4 to 2.2)	0.5 (0.3 to 0.8)
Sensation change[35]		1.4 (0.2 to 10.9)	1.0 (0.9 to 1.0)
Motor deficit[35]		3.1 (0.4 to 27.3)	1.0 (0.9 to 1.0)
DTR abnormality[35]		1.5 (0.5 to 4.9)	1.0 (0.9 to 1.1)
Straight leg raise <40°[35]		1.0 (0.5 to 2.0)	1.0 (0.9 to 1.2)
Neurological signs[36]§		2.4 (0.7 to 8.7)	0.8 (0.5 to 1.3)
Neurological deficit[34]¶		0.7 (0.2 to 2.2)	1.0 (1.0 to 1.1)
Combination red flags			
Trauma and neurological signs[36]		14.4 (2.4 to 87.6)	0.7 (0.5 to 1.2)
Multiple findings[35]		2.1 (1.4 to 3.1)	0.7 (0.6 to 0.9)

■ After test if positive
□ After test if negative

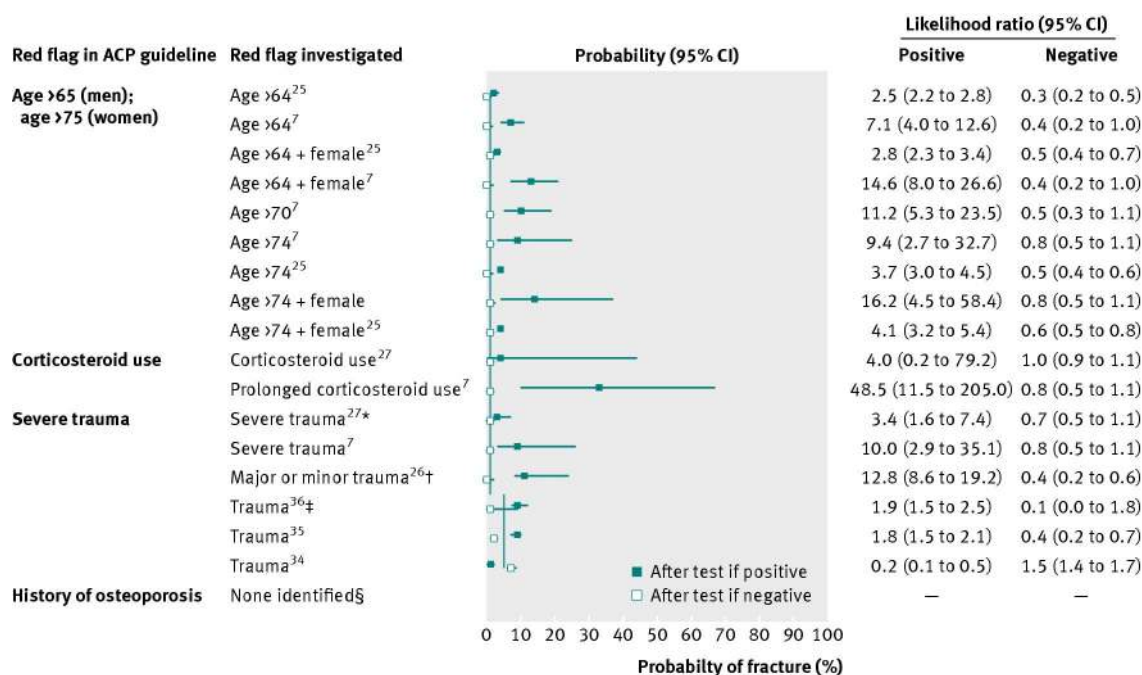0 10 20 30 40 50 60 70 80 90 100
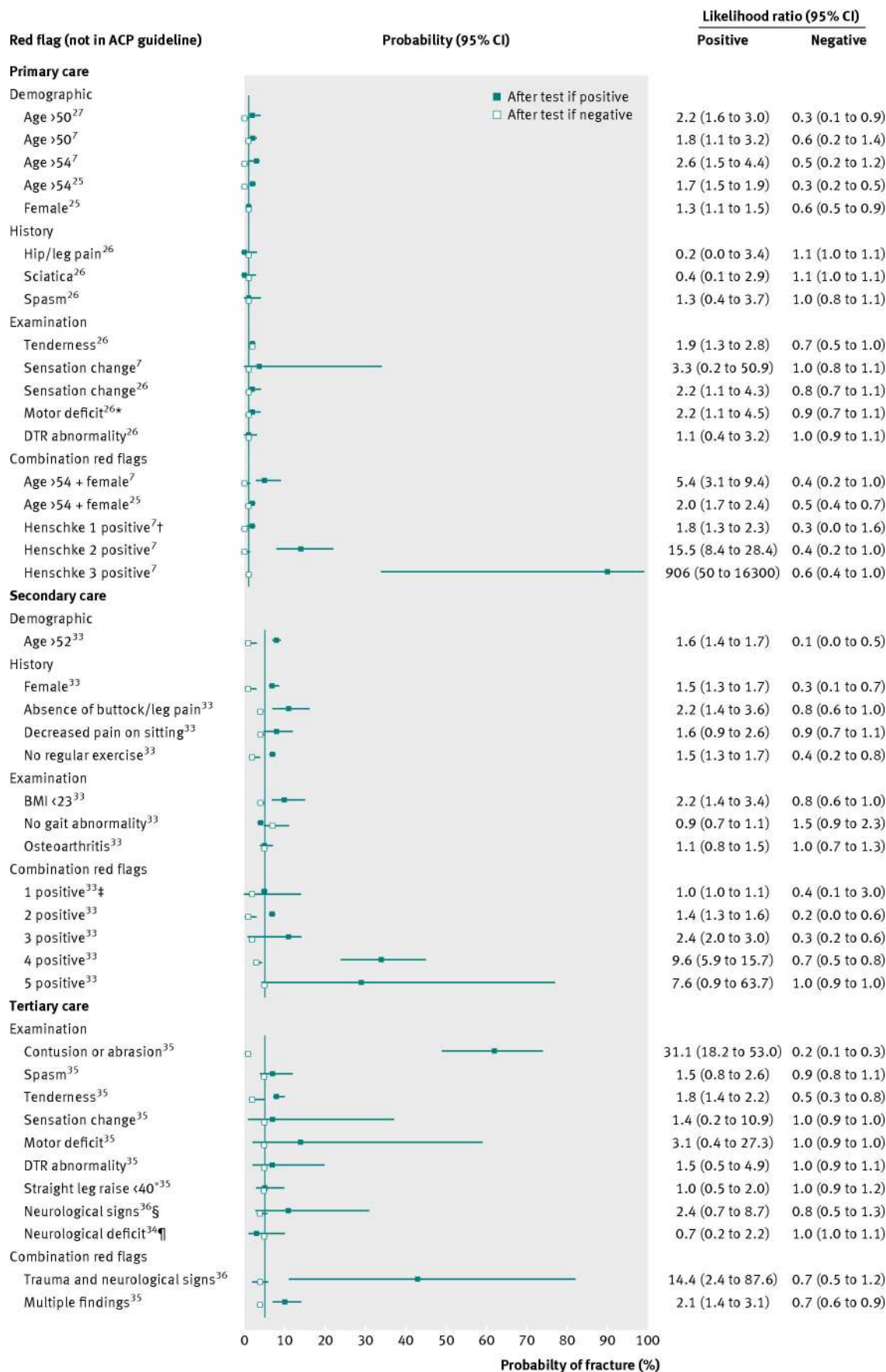
Probabilty of fracture (%)

Fig 3 Diagnostic accuracy of red flags for spinal fracture excluded from American College of Physicians (ACP) guideline. Vertical line indicates prevalence of spinal fracture: 1% in primary care, 5% in secondary and tertiary care. DTR=deep tendon reflex; BMI=body mass index. *Motor deficit: weakness or atrophy. †Henschke index tests: female, age >70, severe trauma, and prolonged use of corticosteroids. ‡Roman index tests: leg or buttock pain, sex, age, BMI, gait abnormality, no

regular exercise, sitting pain, osteoarthritis. §Neurological signs or straight leg raise <40º. ¶Neurological deficit: consistent with lumbar plexus distribution
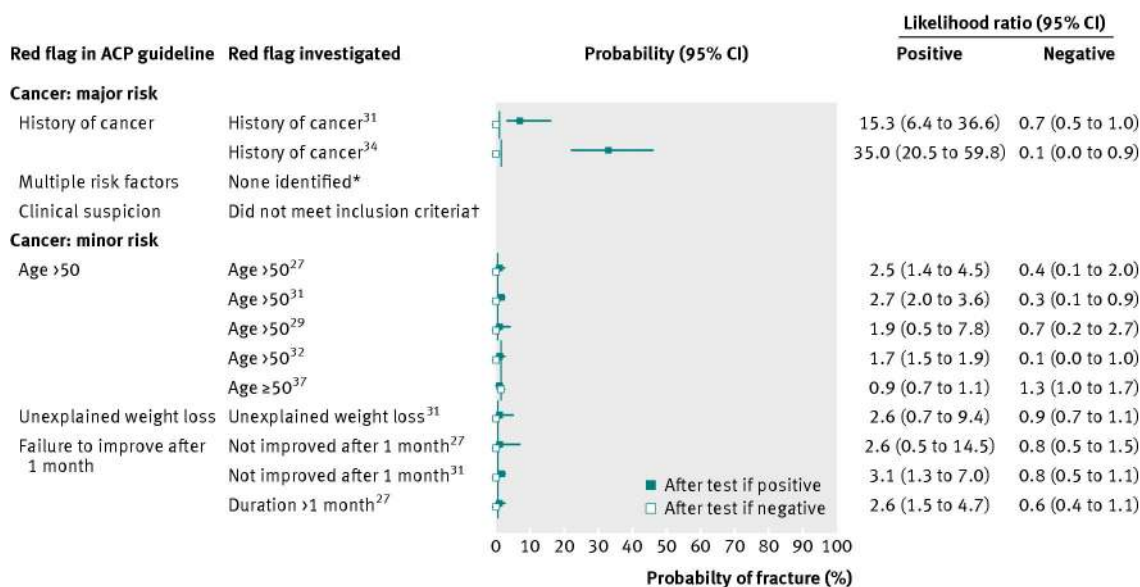

Fig 4 Diagnostic accuracy of red flags for spinal malignancy included in American College of Physicians (ACP) guideline. Vertical line indicates prevalence of spinal malignancy: 0.5% in primary care, 1.5% in secondary and tertiary care. *Included studies for spinal malignancy did not investigate combination red flags. †Red flag "clinical suspicion" did not meet inclusion criteria of either Cochrane review[4][5]
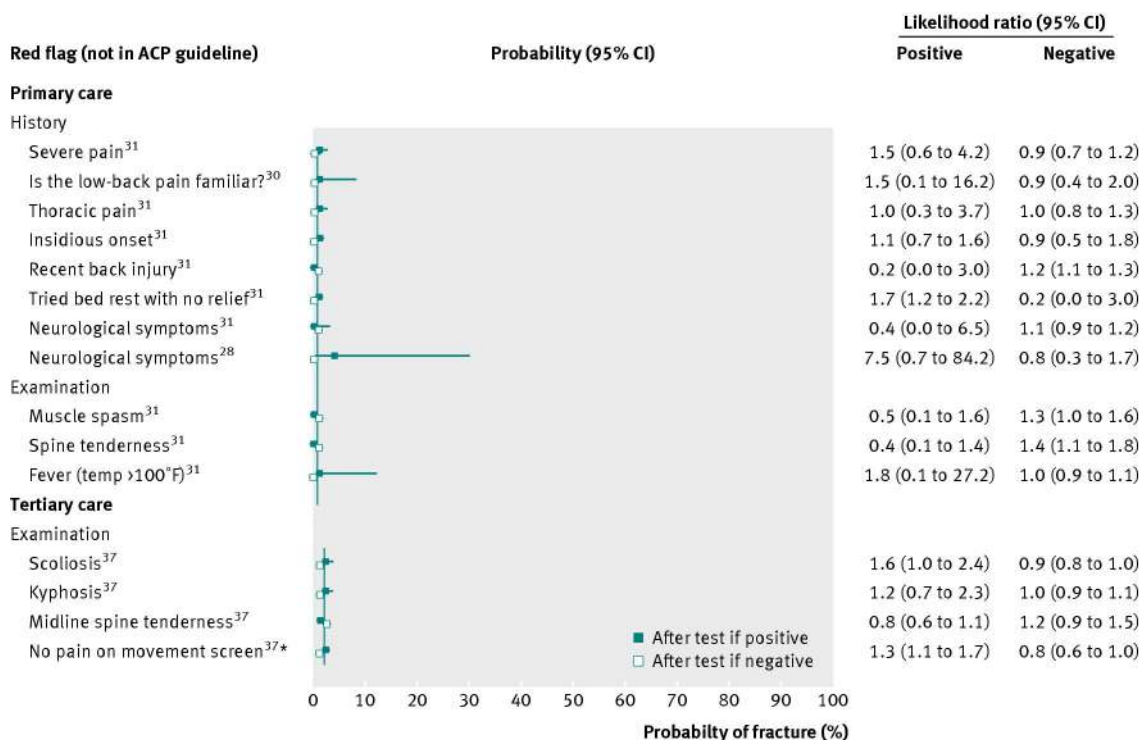

Fig 5 Diagnostic accuracy of red flags for spinal malignancy excluded from American College of Physicians (ACP) guideline. Vertical line indicates prevalence of spinal malignancy: 0.5% in primary care, 1.5% in secondary and tertiary care. *Absence of pain during combined movements of flexion, extension and lateral flexion