BMJ 2013;347:f7095 doi: 10.1136/bmj.f7095 (Published 11 December 2013)



Page 1 of 9

RESEARCH

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

© © © OPEN ACCESS

Aron Downie *PhD student*¹², Christopher M Williams *honorary research fellow*¹, Nicholas Henschke *research fellow*¹³, 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

¹George Institute for Global Health, University of Sydney, Sydney, NSW, 2050, Australia; ²Faculty of Science, Macquarie University, Sydney, Australia; ³Institute of Public Health, University of Heidelberg, Germany; ⁴Faculty of Human Sciences, Macquarie University, Sydney, Australia; ⁵Department of Health Sciences, EMGO Institute for Health and Care Research, VU University, Amsterdam, Netherlands; ⁶Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University Medical Centre, Amsterdam; ⁷Screening and Test Evaluation Program (STEP), School of Public Health, Sydney; ⁸School of Public Health, University of Sydney, Sydney, Australia; ⁹Department of Health Sciences, Faculty of Earth and Life Sciences, VU University, Amsterdam, Netherlands; ¹⁰Department 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 fracture⁴ and in less than 1% malignancy, whether primary tumour or metastasis, will be the underlying cause.⁵

Extra material supplied by the author (see http://www.bmj.com/content/347/bmj.f7095?tab=related#webextra) Appendix: Supplementary tables A-G

No commercial reuse: See rights and reprints http://www.bmj.com/permissions

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.⁶⁷ 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 pain⁴ 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 $1 \downarrow$). 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 departments³⁴⁻³⁶ and one in a spinal surgical unit³⁷). 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 carean for medianeous $\sqrt{27-33} = 34 \sqrt{37}$ with 12 (25%)

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 2-51,11,11,11 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 Upresents 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 colleagues²⁵) 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 colleagues⁷) 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 Diehl²⁷, which, when present, results in a post-test probability of 4% (0% to 44%).²⁷

Figure 3 \downarrow 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 \Downarrow 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 \Downarrow 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 care³ 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 pain¹² 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 Guideline³⁹ 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/handbookdta-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 rule⁴⁹ 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.

This review is adapted from two Cochrane reviews.45

Contributors: Conception and design: AD, CGM, MJH, CMW, NH. Analysis and interpretation of the data: AD, CGM, MJH, CMW, NH, RWO, HCWdV, PM, LI. Drafting of the article: AD, CGM, MJH, CMW, NH. All authors critically revised the article for important intellectual content and approved the final article. Statistical expertise: RWO, HCWdV, PM, LI, CGM. Administrative, technical, or logistic support: CGM, MJH. Collection and assembly of data: AD, CGM, MJH, CMW, NH. CGM is guarantor.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the

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 us

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

submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Not required.

Declaration of transparency: The lead author (Aron Downie) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

Data sharing: No additional available.

- Murray CJL, Vos T, Lozano R, Naghayi M, Flaxman AD, Michaud C, et al. 1 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.
- Hoy D, Brooks P, Blyth F, Buchbinder R. The epidemiology of low back pain. Best Pract 2 Res Clin Rheumatol 2010;24:769-81.
- Koes BW, van Tulder M, Lin C-WC, Macedo LG, McAuley J, Maher C. An updated overview 3 of clinical guidelines for the management of non-specific low back pain in primary care. Eur Spine J 2010;19:2075-94.
- 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.
- 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 Re 2013;2:CD008686
- Chou R, Qaseem A, Owens DK, Shekelle P. Clinical guideline diagnostic imaging for low 6 back pain: advice for high-value health care from the American College of Physicians Ann Intern Med 2011;154:181-9.
- 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.
- Loblaw DA, Perry J, Chambers A, Laperriere NJ. Systematic review of the diagnosis and 8 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
- 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.
- Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, et al. 12 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
- National Health and Medical Research Council. Evidence-based management of acute 14 musculoskeletal pain. NHMRC. 2004.
- 15 National Institute for Health and Care Excelence. NICE Clinical Knowledge Summaries. Back pain-low (without radiculopathy) 2009. http://cks.nice.org.uk/back-pain-low-withoutradiculopathy
- Van Tulder M, Becker A, Bekkering T, Breen A, del Real MT, Hutchinson A, et al. Chapter 16 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 Excelence. Low back pain: early management of persistent non-specific low back pain. NICE clinical guideline 88. Jul 2013. http://guidance. ice.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.
- Reitsma JB, Rutjes AWS, Whiting P, Vlassov VV, Leeflang MMG, Deeks JJ. Chapter 9: 19 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.
- Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for 21 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.

- Chou R, Qaseem A, Snow V, Casey D, Cross Jr JT, Shekelle P, et al. Diagnosis and 24 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.
- Scavone JG, Latshaw RF, Rohrer GV. Use of lumbar spine films. Statistical evaluation 26 at a university teaching hospital. JAMA 1981;246:1105-8.
- Deyo RA, Diehl AK. Lumbar spine films in primary care: current use and effects of selective 27 ordering criteria. J Gen Intern Med 1986;1:20-5. Khoo LA, Heron C, Patel U, Given-Wilson R, Grundy A, Khaw KT, et al. The diagnostic
- 28 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.
- Jacobson AF. Musculoskeletal pain as an indicator of occult malignancy. Yield of bone 32 scintigraphy. Arch Intern Med 1997;157:105-9.
- Roman M, Brown C, Richardson W, Isaacs R, Howes C, Cook C. The development of a 33 clinical decision making algorithm for detection of osteoporotic vertebral compression fracture or wedge deformity. J Manipulative Physiol Ther 2010;18:44-9.
- 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. Patrick JD, Doris PE, Mills ML, Friedman J, Johnston C. Lumbar spine x-rays: a
- 35 multihospital study. Ann Emerg Med 1983;12:84-7.
- Gibson M, Zoltie N. Radiography for back pain presenting to accident and emergency 36 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.
- Poitras S, Rossignol M, Dionne C, Tousignant M, Truchon M, Arsenault B, et al. An 38 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 pathwayslow back pain. May 2013. www.imagingpathways.health.wa.gov.au/index.php/imagingpathways/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.
- Steurer J, Fischer JE, Bachmann LM, Koller M, Riet Gt. Communicating accuracy of tests 42 to general practitioners: a controlled study. BMJ 2002;324:824-6.
- Bachmann LM, Steurer J, Riet Gt. Simple presentation of test accuracy may lead to inflated disease probabilities. *BMJ* 2003;326:393-4. 43
- Deeks JJ. Systematic reviews in health care: systematic reviews of evaluations of 44 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.
- Rhodes LA, McPhillips-Tangum CA, Markham C, Klenk R. The power of the visible: the 47 meaning of diagnostic tests in chronic back pain. Soc Sci Med 1999;48:1189-203.
- Deyo RA. Real help and red herrings in spinal imaging. N Engl J Med 2013;368:1056-8. Bachmann LM, Kolb E, Koller MT, Steurer J, Riet Gt. Accuracy of Ottawa ankle rules to 49 exclude fractures of the ankle and mid-foot; systematic review. BMJ 2003;326:417
- Stiell IG, Wells GA, Vandemheen KL, Clement CM, Lesiuk H, De Maio VJ, et al. The 50 Canadian C-spine rule for radiography in alert and stable trauma patients. JAMA 2001:286:1841-8.
- Michaleff ZA, Maher CG, Verhagen AP, Rebbeck T, Lin C-WC. Accuracy of the Canadian 51 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.

Accepted: 18 November 2013

Cite this as: BMJ 2013;347:f7095

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works

on different terms, provided the original work is properly cited and the use is

non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/.

Figures

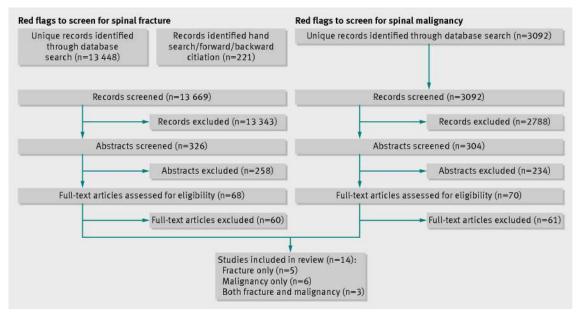


Fig 1 Flow diagram of search strategies for studies evaluating red flags to screen for spinal fracture and malignancy in patients presenting with low back pain

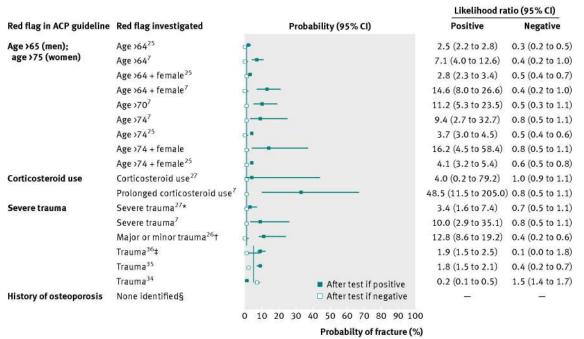


Fig 2 Diagnostic accuracy of red flags for spinal fracture included in American College of Physicians (ACP) guideline. Vertical line indicates prevalence of spinal fracture: 1% in primary care, 5% in secondary and tertiary care. *Severe trauma: trauma such as fall from height or motor vehicle crash (Deyo²⁷); trauma which is major in young patients and minor in elderly (Henschke⁷). †Minor trauma: in elderly women with osteoporosis (Scavone²⁶). ‡Trauma: history of direct trauma (Gibson³⁶); history of trauma (Patrick,³⁵ Reinus³⁴). §History of osteoporosis not evaluated by any included study

Primay care After test if positive Age: 50 ⁷ Age: 50 ⁷ Age: 50 ⁷ 2.2 (1.5 (5: 3.0) 0.3 (0.1 to 18 (1.1 to 3.2) Age: 54 ⁷ 2.2 (1.5 (5: 4.0) 0.5 (0.2 to 3.2 (1.5 (1.5 (1.0) Age: 54 ⁷ 2.6 (1.5 (1.5 (1.0) 0.3 (0.2 to 1.5 (1.5 (1.0) History 1.7 (1.5 (1.5 (1.0) 0.3 (0.2 to 3.0 (1.5 (1.5 (1.0) History 1.3 (1.1 to 1.1) 0.6 (0.5 to 1.3 (1.1 to 1.1) Statics ⁴⁶ 3.3 (0.4 to 2.7) 1.0 (0.8 to 2.2 (1.1 to 4.3) Statics ⁴⁶ 3.3 (0.2 to 5.9 (0.4 (1.5 (1.2 (1.5 (1.4 (1.5 (1.5 (1.5 (1.5 (1.5 (1.5 (1.5 (1.5			Likelihood ratio (95% CI)	
Demographic Age: 50 ⁷ <	Red flag (not in ACP guideline)	Probability (95% CI)	Positive	Negative
Age: 50 ⁷⁷ Age: 54 ⁷⁷ After test if "negative 2.2 (1.4 to 3.20) 0.3 (0.1 to 4.2) Age: 54 ⁷⁷ Female" 1.3 (1.1 to 1.2) 0.6 (0.2 to 4.2) 0.3 (0.2 to 4.2) Age: 54 ⁷⁷ Female" 1.3 (1.1 to 1.2) 0.6 (0.2 to 4.2) 0.3 (0.2 to 4.2) History 1.3 (1.1 to 1.2) 0.6 (0.2 to 4.2) 0.3 (0.2 to 4.2) 0.3 (0.2 to 4.2) History 1.3 (1.1 to 1.2) 0.6 (0.5 to 4.2) 0.3 (0.2 to 4.2) 0.3 (0.2 to 4.2) Spanning 1.3 (0.4 to 3.27) 1.0 (0.4 to 3.27) 1.0 (0.4 to 3.27) 1.0 (0.4 to 3.27) Sensation change* 3.3 (0.2 to 4.2) 3.3 (0.2 to 4.2) 0.4 (0.1 to 2.9) 1.1 (0.4 to 3.27) 1.0 (0.4 to 3.27) Dir abornality** 1.1 (0.4 to 3.27) 0.2 (0.5 (0.4) 0.4 (0.2 to 4.2) 0.4 (0.2 to 4.2) Age: 54.4 - Female** 4.4 (3.1 to 9.4) 0.4 (0.2 to 4.2) 0.4 (0.2 to 4.2) 0.4 (0.2 to 4.2) Henschle 2 positive*7 4.4 (3.1 to 9.2) 0.4 (0.2 to 4.2) 0.4 (0.2 to 4.2) 0.4 (0.2 to 4.2) Henschle 2 positive*7 4.4 (3.1 to 9.2) 0.4 (0.2 to 4.2) 0.4 (0.2 to 4.2) 0.4 (0.2 to 4.2) Deremale*10 4.4 (1.1 to 1.2)	Primary care			
Age: 50 ² Age: 54 ² 2.6 (1.5 to 4.8) 0.5 (0.2 to 4.8) Age: 54 ² 2.6 (1.5 to 4.8) 0.5 (0.2 to 4.8) 1.7 (1.5 to 1.9) 0.5 (0.2 to 4.8) Harding 1.3 (1.1 to 1.5) 0.6 (0.5 to 4.8) 1.3 (1.1 to 1.5) 0.6 (0.5 to 4.8) Histoy High/leg gain ¹⁶ 0.2 (0.0 to 3.4) 1.1 (0.0 to 3.0) 1.1 (0.0 to 3.0) Scattac ¹⁶ 0.4 (0.1 to 2.9) 1.1 (0.0 to 3.0) 1.3 (0.4 to 3.7) 1.0 (0.8 to 3.7) Tendeness ¹⁶ Sensation charge ¹⁶ 3.3 (0.2 to 5.00) 1.0 (0.8 to 3.7) 1.0 (0.8 to 3.7) Sensation charge ¹⁶ 2.2 (1.1 to 4.3) 0.9 (0.7 to 4.8) 2.2 (1.1 to 4.3) 0.9 (0.7 to 4.8) Age 54 - female ¹⁷ 4.4 (1.1 to 1.7) 1.1 (0.4 to 3.7) 0.0 (0.8 to 3.7) 0.9 (0.8 to 3.7) Age 54 - female ¹⁷ 1.1 (0.4 to 3.7) 0.0 (0.8 to 3.7) 0.9 (0.7 to 1.1) 0.9 (0.7 to 1.1) Age 52 ²⁷³ 1.1 (1.1 to 1.7) 0.4 (0.2 to 3.7) 1.6 (1.4 to 1.7) 0.4 (0.2 to 3.7) Reg 52 ²⁷³ 1.1 (0.1 to 3.1) 1.1 (0.4 to 3.1) 1.0 (0.8 to 3.1) 1.0 (0.8 to 3.1) No gail abnormality ¹¹ 0 0 0.9 (0.7 to 1.1) 1.5 (0.8 to	Demographic	After test if positive		
Age s947 2.6 (1,5 for 4.4) 0.5 (0, 2 for 4.4) 0.6 (0, 2 for 4.4) </td <td>Age >50²⁷</td> <td>- 🗆 After test if negative</td> <td>2.2 (1.6 to 3.0)</td> <td>0.3 (0.1 to 0.9)</td>	Age >50 ²⁷	- 🗆 After test if negative	2.2 (1.6 to 3.0)	0.3 (0.1 to 0.9)
Age 34 ³⁰ 1,7 (1,5 10.9) 0,3 (0,2 to 3,1 (1,1 (0,1 5)) 0,6 (0,5 to 3,1 (1,1 (0,1 5)) Hig/leg pair ⁸⁴ 0,2 (0,0 (0,3,4)) 1,1 (1,0 to 5,0 and ** 0,2 (0,0 (0,3,4)) 1,1 (1,0 to 5,0 and ** Scalate ²⁴ 0,2 (0,0 (0,3,4)) 1,1 (0,0 to 3,0 (2,1 to 2,2)) 1,1 (1,0 to 5,0 (3,0 to 2,2)) 1,1 (1,0 to 5,0 (3,0 to 2,2)) 1,1 (0,0 to 3,0 (2,1 to 2,4)) 0,0 (2,0 to 3,0 (2,1 to 4,3)) 0,0 (2,0 to 3,0 (2,1 t	Age >50 ⁷	·	1.8 (1.1 to 3.2)	0.6 (0.2 to 1.4)
Female ²³ 1,3 (1,1 to 1,5) 0,6 0,5 to Histoy 0,2 (0,0 to 3,4) 1,1 (1,0 to Statica ²⁶ 0,4 (0,1 to 2,7) 1,0 (0,8 to Spasm ²⁶ 1,3 (0,4 to 3,7) 1,0 (0,8 to Sensation change ²⁶ 3,3 (0,2 to 5,09) 1,0 (0,8 to Sensation change ²⁶ 2,2 (1,1 to 4,3) 0,8 (0,7 to Sensation change ²⁶ 3,3 (0,2 to 5,09) 1,0 (0,8 to Sensation change ²⁶ 2,2 (1,1 to 4,3) 0,0 (0,7 to DTR abnormality ²⁶ 2,2 (1,1 to 4,3) 0,0 (0,7 to Age 54, female ²⁷ 4 5,4 (3,1 to 9,4) 0,4 (0,2 to Age 534, female ²⁷ 4 4,6 (1,4 to 1,7) 0,1 (0,0 to Henschle 2 positive ²⁷ 1,5 (3,6 4 to 2,4) 0,4 (0,2 to 1,5 (3,4 to 1,7) 0,6 (0,4 to Secondary care 2,2 (1,4 to 3,4) 0,8 (0,6 to 2,2 (1,4 to 3,4) 0,8 (0,6 to BM (23 ²³) - - 1,5 (1,3 to 1,7) 0,4 (0,2 to Semale 7 ³ - - 1,5 (0,5 to 1,3) 1,0 (0,0 to BM (23 ²³) - - 1,5 (0,5 to 4,5) 1,0 (0,0 to 1,1) 1,5 (0,5 to 4,0)		a-•	2.6 (1.5 to 4.4)	0.5 (0.2 to 1.2)
History Hig/leg pair ³⁶ Solatica ⁷⁶ Solatica ⁷⁶ Solatica ⁷⁶ Spam ³⁶ Lamination Tendeness ³⁶ Sensation change ⁷⁶ Sensation change ⁷⁸ Age 54 - female ⁷³ Sensation change ⁷⁸ Sensation change ⁷⁸ Sensa	Age >54 ²⁵		1.7 (1.5 to 1.9)	0.3 (0.2 to 0.5)
Hip/reg pain** 0.2 0.0 to 3.4) 1.1 0 to to 3.0 Scatar4* 0.4 0.0 to 3.7) 1.0 0.8 to 3.7) Examination 1.9 (1.3 to 2.8) 0.7 0.5 to 3.7) Tendemess ³⁶ 3.3 (0.2 to 5.9) 1.0 0.8 to 3.7) Sensation change* 3.3 (0.2 to 5.9) 1.0 0.8 to 3.7) Sensation change* 3.3 (0.2 to 5.9) 1.0 0.8 to 3.7) Sensation change* 3.3 (0.2 to 5.9) 1.0 0.8 to 3.7) Motor deficit** 2.2 (1.1 to 4.3) 0.9 (7.5) Motor deficit** 2.0 (1.7 to 4.3) 0.9 (7.5) Motor deficit** 2.0 (1.7 to 4.3) 0.0 (0.5 to 4.7) Motor deficit** 2.0 (1.7 to 4.3) 0.0 (0.5 to 4.7) Motor deficit** 2.0 (1.7 to 4.7) 0.1 (0.6 to 4.7) Menscheiz positive*1 1.6 (1.4 to 1.7) 0.1 (0.6 to 4.7) Henscheiz positive*1 1.6 (1.4 to 1.7) 0.1 (0.6 to 4.7) Demographic Age: 52 ¹³ - 1.6 (1.4 to 1.7) 0.1 (0.6 to 4.7) Reg (32 ¹³) - - 1.6 (1.4 to 1.7) 0.1 (0.6 to 1.5) 0.6 (0.6 to 7.7) Spotitive ²¹ - - - - -	Female ²⁵	•	1.3 (1.1 to 1.5)	0.6 (0.5 to 0.9)
Scientaria 0.4 (0.110 2.9) 1.1 (0.10 Spann ²⁴ 1.3 (0.410 2.9) 1.0 (0.8 to Earnination 1.9 (1.3 to 2.0) 0.0 8 to Sensation change ²⁶ 3.3 (0.2 to 5.0.9) 0.0 8 to Sensation change ²⁶ 2.2 (1.1 to 4.3) 0.8 (0.7 to Motor deficit ²⁶ 2.2 (1.1 to 4.3) 0.9 (0.7 to DR abormality ²⁶ 2.2 (1.1 to 4.3) 0.9 (0.7 to Combination red flags 5.4 (3.1 to 9.4) 0.4 (0.2 to Age 54 - female ²⁷ 2.0 (1.7 to 2.4) 0.5 (0.6 to 6.0 to Henschke 1 positive ⁷ 1.6 (1.4 to 1.7) 0.1 (0.0 to Henschke 2 positive ⁷ 1.6 (1.4 to 1.7) 0.1 (0.0 to Berographic 1.5 (1.3 to 1.7) 0.3 (0.1 to Age 552 ³³ - - 1.6 (0.9 to 2.6) No regular exercise ³³ - - 1.6 (0.9 to 2.6) Ostoarthitis ³³ - 1.0 (0.0 to 1.1) 1.5 (0.3 to 1.7) Ostoarthitis ³³ - - 1.6 (1.4 to 1.7) 1.0 (0.7 to 1.1) Combination end flags - - 1.5 (0.1 to 1.1) 1.5 (0.5 to 4.0 2.0) Spositive ³¹ <	History			
Spann ²⁴ 1.3 (0.4 to 3.7) 1.0 (0.8 to Sensation change ⁷⁶ Sensation change ⁷⁶ 3.3 (0.2 to 5.09) 1.0 (0.8 to Sensation change ⁷⁶ Sensation change ⁷⁶ 3.3 (0.2 to 5.09) 1.0 (0.8 to 2.2 (1.1 to 4.5) Motor deficit ⁷⁶ 2.2 (1.1 to 4.5) 0.9 (0.7 to 2.2 (1.1 to 4.5) DTR abnormality ⁷⁶ 5.4 (3.1 to 9.4) 0.4 (0.2 to 2.0 (1.7 to 2.4) Age 54 - female ⁷⁵ 1.8 (1.3 to 2.3) 0.3 (0.0 to 1.5 (5.8 to 2.3) Hensche 1 positive ⁷¹ 1.4 (0.4 to 3.2) 0.0 (0.4 to 2.0 (1.7 to 2.3) Beographic Age 524 - female ⁷³ 1.6 (1.4 to 1.7) 0.1 (0.0 to 1.5 (5.8 to 2.8) Decreased plan on sitting ⁷³ 0 0 0.8 (0.6 to 2.0 (1.0 500) No regular exercise ¹³ 0 0 0.9 (0.7 to 1.5 (0.9 to 1.5) Examination 1.9 (0.10 to 1.5) 0.0 (0.7 to 1.5 (0.9 to 1.5) 0.0 (0.7 to 1.5 (0.9 to 1.5) Osteoarthitis ¹³ 0 0 0.9 (0.7 to 1.5 (0.9 to 1.5 (0.9 to 1.5) 0.0 (0.7 to 1.5 (0.9 to 1.5) Spasitive ⁷¹ 0 0 0.9 (0.7 to 1.5 (0.9 to 1.5 (0.9 to 1.5) 0.0 (0.7 to 1.5 (0.9 to 1.5) Spasitive ⁷¹ 0 0 0 0.9 (0.7 to 1.5 (0.9 to 1.5 (0.9 to 1.5) 0.0 (0.7 to 1		÷	0.2 (0.0 to 3.4)	1.1 (1.0 to 1.1)
Examination 1.9 (1.3 to 2.8) 0.7 (0.5 to 3.3 (0.2 to 5.0.9) Sensation change ²⁶ 3.3 (0.2 to 5.0.9) 1.0 (0.8 to 3.3 (0.2 to 5.0.9) Sensation change ²⁶ 2.2 (1.1 to 4.3) 0.8 (0.7 to 3.1 to 4.3 (0.8 (0.7 to 4.3 to	Sciatica ²⁶	÷	0.4 (0.1 to 2.9)	1.1 (1.0 to 1.1)
Tendemess ²⁶ 1.9 (1.3 to 2.8) 0.7 (0.5 to 3.3 (0.2 to 50.9) 1.0 (0.8 to 3.3 (0.2 to 50.9) 1.0 (0.8 to 3.3 (0.2 to 2.2 (1.1 to 4.3) 0.9 (0.7 to 2.2 (1.1 to 4.3) 0.9	Spasm ²⁶	÷-	1.3 (0.4 to 3.7)	1.0 (0.8 to 1.1)
Sensation change ¹⁶ 3.3 (0.2 to 50.9) 1.0 (0.8 to 2.2 (1.1 to 4.3) 0.8 (0.7 to 2.0 (0.7 to 2.4) OffR abnormality ²⁶ 1.1 (0.4 to 3.2) 1.0 (0.9 to 2.2 (1.1 to 4.3) 0.8 (0.7 to 2.0 (0.7 to 2.4) Age 54 + female ² 4 2.4 (3.1 to 9.4) 0.4 (0.2 to 3.3 (0.2 to 50.9) 1.0 (0.4 to 2.2 (1.1 to 4.3) Age 54 + female ² 4 2.0 (1.7 to 2.4) 0.5 (0.4 to 1.8 (1.3 to 2.3) 0.3 (0.0 to 3.3 (0.1 to 2.6 (1.4 to 1.7) 0.4 (0.2 to 3.3 (0.2 to 50.9) Bendgraphic Age 554 5.4 (3.1 to 9.4) 0.4 (0.2 to 3.0 (0.0 to 1.8 (1.3 to 1.7) 0.3 (0.1 to 3.0 (0.1 to 2.2 (1.4 to 3.4) 0.8 (0.6 to 3.0 (0.1 to 2.2 (1.4 to 3.4) 0.8 (0.6 to 3.0 (0.7 to 1.5 (1.3 to 1.7) 0.3 (0.1 to 3.0 (0.1 to 2.2 (1.4 to 3.4) 0.8 (0.6 to 3.0 (0.7 to 3.1 (1.6 2 to 3.7) 0.4 (0.2 to 3.0 (0.7 to 3.1 (1.6 2 to 3.7) 0.4 (0.2 to 3.1 (0.4 to 2.7)				
Sensation change ³⁶ 2.2 (1.1 to 4.3) 0.8 (0.7 to 2.0) Motor deficit ¹⁶⁴ D1R abornality ³⁶ 2.2 (1.1 to 4.3) 0.8 (0.7 to 2.0) Combination red flags 2.2 (1.1 to 4.3) 0.8 (0.7 to 2.0) 0.7 to 2.0 (1.7 to 2.4) 0.9 (0.7 to 1.1 (0.4 to 3.2) 0.0 0.9 to 2.0 (1.7 to 2.4) Age 54 + female ²⁶ Age 54 + female ²⁶ 5.4 (3.1 to 9.4) 0.4 (0.2 to 2.4) 0.6 (0.2 to 2.4) 0.6 (0.4 to 3.2) 0.9 (0.7 to 1.1 (0.4 to 3.2) 0.8 (0.6 to 3.4) 0.6 (0.2 to 2.4) 0.6 (0.4 to 3.2) 0.9 (0.7 to 1.1 (0.4 to 1.2)) 0.6 (0.4 to 3.2) 0.9 (0.7 to 1.1 (0.4 to 1.2)) 0.6 (0.4 to 3.2) 0.9 (0.7 to 1.1 (0.4 to 1.2)) 0.6 (0.4 to 4.2 to 3.4) 0.6 (0.4 to 2.5) 0.8 (0.6 to 2.5) 0.8 (0.6 to 2.5) 0.9 (0.7 to 1.5) 0.6 (0.7 to 1.1) 0.4 (0.2 to 2.5) 0.9 (0.7 to 1.1 (0.4 to 1.1) 0.8 (0.6 to 2.5) 0.9 (0.7 to 1.1 (0.4 to 1.2)) 0.6 (0.6 to 2.5) 0.9 (0.7 to 1.1 (0.4 to 1.2)) 0.6 (0.7 to 1.1) 0.4 (0.2 to 2.5) 0.9 (0.7 to 1.1 (0.4 to 1.2)) 0.8 (0.6 to 2.5) 0.9 (0.7 to 1.1 (0.4 to 1.2)) 0.9 (0.7 to 1.1 (0.4 to 1.2)) 0.9 (0.7 to 1.	Tenderness ²⁶	p	1.9 (1.3 to 2.8)	0.7 (0.5 to 1.0)
Motor deficit ²⁶⁴ 2.2 (1.1 to 4.5) 0.9 (0.7 to 1.0 (0.4 to 3.2) 1.0 (0.9 to 7.0 (0.9	Sensation change ⁷	÷	3.3 (0.2 to 50.9)	1.0 (0.8 to 1.1)
DTR abnormality ²⁶ Combination red flags Age 54 + female ²⁵ Age 54 + female ²⁶ Age 524 + female ²⁶ Age 524 + female ²⁷ Henschke 2 positive ⁷ History Female ³³ Absence of buttock/leg pain ⁷³ Combination red flags Age 54 + female ²⁸ Age 52 ³ Combination red flags Age 54 + female ²⁸ Age 52 ³ Combination red flags Age 54 + female ²⁸ Combination red flags Combinati	Sensation change ²⁶	÷	2.2 (1.1 to 4.3)	0.8 (0.7 to 1.1)
Combination red flags Age 554 + female ⁷ 5.4 (3.1 to 9.4) 0.4 (0.2 to 2.4) 0.4 (0.2 to 2.4) 0.4 (0.2 to 2.4) 0.4 (0.2 to 1.5) 0.6 (0.4 to 1.5)	Motor deficit ²⁶ *	*	2.2 (1.1 to 4.5)	0.9 (0.7 to 1.1)
Age 54 + female ² 5.4 (3.1 to 9.4) 0.4 (0.2 to 2.0 (1.7 to 2.4)) 0.5 (0.4 to 2.3) Henschke 2 positive ⁷ 0.5 (0.4 to 2.4) 0.5 (0.4 to 2.4) 0.5 (0.4 to 2.4) 0.6 (0.2 to 2.3) Bemographic Age 552 ³² 0 115, (1.3 to 1.7) 0.3 (0.1 to 2.3) 0.4 (0.2 to 2.3) 0.3 (0.1 to 2.3) 0.3 (0.1 to 2.3) 0.3 (0.1 to 2.3) 0.3 (0.1 to 2.3) 0.4 (0.2 to 2.3)	DTR abnormality ²⁶	÷	1.1 (0.4 to 3.2)	1.0 (0.9 to 1.1)
Age 554 + female ²⁵ 2.0 (1.7 to 2.4) 0.5 (0.4 to 1.8 (1.3 to 2.3) Henschke 1 positive ⁷ 906 (50 to 16300) 0.6 (0.4 to 1.7 (0.2 to 2.4)) 0.5 (0.4 to 1.8 (1.3 to 2.3)) Henschke 2 positive ⁷ 906 (50 to 16300) 0.6 (0.4 to 1.7 (0.2 to 2.4)) 0.5 (0.4 to 1.8 (1.3 to 2.3)) 0.4 (0.2 to 2.4)) Henschke 3 positive ⁷ 906 (50 to 16300) 0.6 (0.4 to 1.7 (0.2 to 2.4)) 0.6 (0.4 to 2.8 (0.4 to 1.7 (0.5 to 2.4)) 0.6 (0.4 to 2.8 (0.4 to 1.7 (0.4 to 1.7 (0.4 to 2.4))) 0.6 (0.4 to 2.6 (0.6 to 0.6 (0.4 to 2.6)) 0.6 (0.4 to 2.6 (0.6 to 0.6 (0.4 to 2.6)) 0.6 (0.6 to 0.6 (0.4 to 2.6)) 0.6 (0.6 to 0.6 (0.4 to 2.6)) 0.6 (0.4 to 2.6) 0.6 (0.6 to 0.6 (0.4 to 2.6)) 0.6 (0.6 to 0.6 (0.4 to 2.6)) 0.6 (0.6 to 0	그는 것이 같은 것이 같은 것이 있는 것이 같은 것이 같은 것이 같이			
Henschke 1 positive ⁷ t I.8 (1.3 to 2.3) 0.3 (0.0 to 155, (6.4 to 28.4) 0.4 (0.2 to 95, (0.4 to 28.4) 0.4 (0.4 to 17.7) 0.3 (0.1 to 22, (0.4 to 38.4) 0.4 (0.4 to 17.7) 0.3 (0.1 to 22, (0.4 to 38.4) 0.8 (0.6 to 10.6) 0.9 (0.7 to 1.1) 1.5 (0.9 to 2.6) 0.9 (0.7 to 1.1) 0.5 (0.9 to 03.7) 0.4 (0.4 to 17.7) 0.4 (0.2 to 95, 0.3 to 12.5) 0.9 (0.7 to 1.1) 0.5 (0.9 to 03.7) 0.9 (0.7 to 1.1) 0.5 (0.9 to 03.7) 0.9 (0.7 to 1.1) 1.5 (0.9 to 3.7) 0.9 (0.7 to 1.1) 0.5 (0.9 to 03.7) 1.0 (0.9 to 13.7) 0.9 (0.7 to 1.1) 1.5 (0.9 to 0.6) 0.9 (0.5 to 1.6) 1.0 (0.5 to 1.6) 0.9 (0.5 to 1.6) 1.1 (0.8 to 1.5) 1.0 (0.5 to 1.7)	Age >54 + female ⁷	o	5.4 (3.1 to 9.4)	0.4 (0.2 to 1.0)
Henschke 2 positive ⁷ Image: 15,5 (8,4 to 28,4) 0,4 (0,2 to 900 (50 to 16300) 0,6 (0,4 to 28,4) Secondary care 906 (50 to 16300) 0,6 (0,4 to 28,4) 906 (50 to 16300) 0,6 (0,4 to 28,4) Bemographic Age >52 ²³ Image: 1,6 (1,4 to 1,7) 0,1 (0,0 to 28,4) 0,4 (0,2 to 28,4) 0,4 (0,2 to 28,4) History Female ³³ Image: 1,5 (1,3 to 1,7) 0,3 (0,1 to 20,4) 0,4 (0,2 to 28,4) 0,4 (0,4 to 1,7) 0,1 (0,0 to 4,4) 0,4 (0,4 to 1,7) 0,1 (0,0 to 4,4) 0,4 (0,4 to 28,4) 0,4 (0,4 to 38,4) 0,4 (0,4 to 38,4) 0,4 (0,4 t		¢	2.0 (1.7 to 2.4)	0.5 (0.4 to 0.7)
Henschke 3 positive ² 906 (50 to 16300) 0.6 (b.4 to 72) Secondary care I.6 (1.4 to 1.7) 0.1 (0.0 to 71) Age 552 ²² I.6 (1.4 to 1.7) 0.1 (0.0 to 71) History I.5 (1.3 to 1.7) 0.3 (0.1 to 72) Absence of buttock/leg pain ³³ I.6 (0.9 to 2.6) 0.9 (0.7 to 1.1) Decreased pain on sitting ³² I.5 (1.3 to 1.7) 0.3 (0.1 to 71) No regular exercise ³³ I.5 (1.3 to 1.7) 0.4 (0.2 to 71) BMI (23) ³¹ I.9 (0.9 to 2.6) 0.9 (0.7 to 1.1) No gait abnormality ³³ I.9 (0.7 to 1.1) 0.4 (0.1 to 71) Osteoarthritis ³³ I.1 (0.8 to 1.5) 1.0 (0.7 to 7.0 (0.5 to 72) Combination red flags I.1 (0.1 to 1.1) 0.4 (0.1 to 72) 1 positive ³³ I.1 (0.8 to 1.5) 0.2 (0.1 to 72) 3 positive ³³ I.1 (0.8 to 1.5) 0.9 (0.8 to 1.6) Spasm ⁵⁵ I.4 (0.2 to 1.3) 0.2 (0.1 to 72) Spasm ⁵⁵ I.1 (0.8 to 2.6) 0.9 (0.8 to 1.6) Straight leg raise (40 ⁷³⁵) I.4 (0.2 to 1.2) 1.0 (0.5 to 2.9) Neurological sign ³⁶ I.4 (0.2 to 1.2) 0.8 (0.5 to 1.5 to 1.9) Neurological sign	Henschke 1 positive ⁷ †	d •	1.8 (1.3 to 2.3)	0.3 (0.0 to 1.6)
Secondary care Demographic Age >52 2 ³³ I.6 (1.4 to 1.7) 0.1 (0.0 to 1.1) History I.5 (1.3 to 1.7) 0.3 (0.1 to 2.0 (0.9 to 1.0 to 1.1)) Female ³³ I.5 (1.3 to 1.7) 0.3 (0.1 to 2.0 (0.9 to 1.0 to 1.1)) 0.4 (0.9 to 2.6 (0.9 (0.7 to 1.1)) Abscnee of buttock/leg pain ³³ I.6 (1.4 to 3.7) 0.3 (0.1 to 1.6 (0.9 to 2.6 (0.9 (0.7 to 1.1))) 0.4 (0.0 to 1.1) 0.4 (0.9 to 2.6 (0.9 (0.7 to 1.1))) No regular exercise ³³ I.6 (1.4 to 3.7) 0.3 (0.6 to 0.9 (0.7 to 1.1)) 0.6 (0.9 to 0.7 to 1.1) 0.9 (0.7 to 1.1) 0.6 (0.6 to 0.9 (0.7 to 1.1)) 1.6 (0.9 to 2.6 (0.9 to 0.7)) 0.6 (0.6 to 0.9 (0.7 to 1.1)) 0.6 (0.7 to 1.1) 0.6 (0.7 to 1.1) 0.6 (0.6 to 0.9 (0.7 to 1.1)) 0.6 (0.7 to 1.1) 0.7 (0.5 to 2.1) 0.7 (0.5 to 2.		o ——	15.5 (8.4 to 28.4)	0.4 (0.2 to 1.0)
Demographic Age 55 ²⁷³ History Female ³³ Absence of buttock/leg pain ³³ Decreased pain on sitting ³³ BMI (23 ³³ BMI (23 ³³) Combination reaf flags 1 positive ³³ 4 1 positive ³³ 4 1 positive ³³ 4 2 positive ³³ 2 posi	Henschke 3 positive ⁷	۰ • • • • • • • • •	906 (50 to 16300)	0.6 (0.4 to 1.0)
Age 52 ³³ - + 1.6 (1.4 to 1.7) 0.1 (0.0 to 1.1) History - + 1.5 (1.3 to 1.7) 0.3 (0.1 to 2.0 (0.	Secondary care			
History Female ³³ Absence of buttock/leg pain ³³ Decreased pain on slitting ³³ No regular exercise ³³ BMI (23 ³³ Osteoarthrittis ³² Combination red flags 1 positive ³³ 2 positive ³³ 3 positive ³³ 5 positive ³³ Combination or abrasion ³⁵ 5 positive ³³ Combination or abrasion ³⁵ Sensation change ³⁵ Motor deficit ²⁵ Trendemess ³⁵ Combination red flags 1 positive ³³ Combination or abrasion ³⁵ Sensation change ³⁵ Motor deficit ²⁵ Trendemess ³⁵ Combination red flags 1 positive ³³ Combination or abrasion ³⁵ Sensation change ³⁵ Motor deficit ²⁵ Trendemess ³⁵ Combination red flags Trendemess ³⁶ Combination red flags Trendemess ³⁶ Trendemess ³⁶ Trendemess ³⁶ Trendemess ³⁶ Trendemess ³⁶ Trendemess ³⁶ Trendemess ³⁶ Combination red flags Trauma and neurological signs ³⁶ O 0 10 20 30 40 50 60 70 80 90 100	Demographic			
Female ³³ 1.5 (1.3 to 1.7) 0.3 (0.1 to 2.2 (1.4 to 3.6) Absence of buttock/leg pain ³³ 2.2 (1.4 to 3.6) 0.8 (0.6 to 0.9 (0.7 to 1.5 (1.3 to 1.7) No regular exercise ³³ 0 1.6 (0.9 to 2.6) 0.9 (0.7 to 1.5 (1.3 to 1.7) Examination 2.2 (1.4 to 3.4) 0.8 (0.6 to 0.9 (0.7 to 1.1) 1.5 (0.3 to 1.7) BMI (23 ³³ 0 0 0 0.9 (0.7 to 1.1) 1.5 (0.9 to 2.6) No gait abnormality ²³ 0.9 (0.7 to 1.1) 1.5 (0.9 to 0.9 (0.7 to 1.1) 1.5 (0.9 to 0.9 (0.7 to 1.1) 1 positive ³¹ 4 1.0 (1.0 to 1.1) 0.4 (0.1 to 2 positive ³¹ 1.4 (1.3 to 1.6) 0.2 (0.0 to 3 go 3.0 (0.9 to 3.7) Tertlary care 2 1.5 (1.4 to 2.2) 0.9 (0.8 to 1.5) 1.0 (0.9 to 7.6 (0.9 to 3.7) 1.0 (0.9 to 7.6 (0.9 to 3.7) Sensation change ³⁵ 0 0 1.5 (0.5 to 4.9) 1.0 (0.9 to 3.1 (0.4 to 2.2)	Age >52 ³³	c- *	1.6 (1.4 to 1.7)	0.1 (0.0 to 0.5)
Absence of buttock/leg pain ³³ 0 0 2.2 (1.4 to 3.6) 0.8 (0.6 to 1.6 (0.9 to 2.6) 0.9 (0.7 to 1.0 (0.9 to 3.7) 0.0 (0.9 to 3.1 (0.4 to 2.7) 0.9 (0.8 to 2.6) 0.7 (0.2 to 2.2) </td <td>History</td> <td></td> <td></td> <td></td>	History			
Decreased pain on sitting ³³ 1.6 (0.9 to 2.6) 0.9 (0.7 to 1.5) No regular exercise ³³ 1.5 (1.3 to 1.7) 0.4 (0.2 to 2.6) Examination BMI (23 ³³) 2.2 (1.4 to 3.4) 0.8 (0.6 to 0.9 (0.7 to 1.1) No agit abnormality ³³ 0.9 (0.7 to 1.1) 1.5 (0.9 to 0.9 (0.7 to 1.1) 1.5 (0.9 to 0.9 (0.7 to 1.1) Osteoarthritis ³³ 0.9 (0.7 to 1.1) 1.0 (0.8 to 1.5) 1.0 (0.8 to 1.5) 1.0 (0.7 to 1.1) Combination red flags 1.1 (0.8 to 1.5) 0.4 (0.1 to 1.1) 0.4 (0.2 to 1.0) 0.2 (0.1 to 1.1) 0.4 (0.2 to 1.0) 0.2 (0.1 to 1.1) 0.4 (0.1 to 1.1) 0.4 (0.2 to 1.0) 0.2 (0.1 to 1.1) 0.5 (0.3 to 1.1) 0.5	Female ³³	- ·	1.5 (1.3 to 1.7)	0.3 (0.1 to 0.7)
No regular exercise ³³ - 1.5 (1.3 to 1.7) 0.4 (0.2 to 50) Examination 2.2 (1.4 to 3.4) 0.8 (0.6 to 0.9 (0.7 to 1.1) 1.5 (0.9 to 0.9 (0.7 to 1.1) 1.6 (0.9 to 6.3.7) 1.0 (1.0 to 1.1) 0.4 (0.2 to 1.1) 0.4 (0.2 to 1.1) 0.9 (0.7 to 1.1) 1.5 (0.9 to 0.9 (0.7 to 1.1) 0.4 (0.2 to 0.9 to 3.1 to 0.9 (0.7 to 1.1) 0.4 (0.2 to 1.1) 0.9 (0.7 to 1.1) 0.9 (0.7 to 1.1) 0.9 (0.7 to 1.1) 0.4 (0.2 to 1.1) 0.9 (0.7 to 1.1) 0.4 (0.2 to 1.9) 0.9 (0.7 to 1.1) 0.4 (0.2 to 1.9) 0.9 (0.7 to 1.1) 0.4 (0.2 to 1.9) 0.9 (0.7 to 1.1)	Absence of buttock/leg pain ³³	•	2.2 (1.4 to 3.6)	0.8 (0.6 to 1.0)
Examination BMI (23 ³³) 2.2 (1.4 to 3.4) 0.8 (0.6 to 0.9 (0.7 to 1.1) 1.5 (0.9 to 1.1) 1.6 (1.0 to 1.1) 0.4 (0.1 to 1.1) 0.4 (0.2 to 1.3)			1.6 (0.9 to 2.6)	0.9 (0.7 to 1.1)
BMI (23 ³³ 0 0 0 0.8 (0.6 to 0.9 (0.7 to 1.1) 1.5 (0.9 to 0.9 (0.7 to 1.1) 1.6 (1.0 to 1.1) 0.4 (0.1 to 1.4 (1.3 to 1.6) 0.2 (0.1 to 1.4 (1.3 to 1.6) 0.2 (0.1 to 0.9 (0.7 to 1.1) 0.4 (0.1 to 0.9 (0.7 to 1.1) 0.9 (0.8 to 1.4 (1.4 to 2.2) 0.9 (0.8 to 1.5 (0.5 to 4.9) 0.9 (0.8 to 1.1 (0.4 to 2.2) 0.1 (0.9 to 1.1 (0.9 to 1.1 (0.9 to 1.1 (0.9 to 1.1 (0.9 to 1.1 (0.9 to 2.2) 0.1 (0.9 to 1.1 (0.9 to 1.1 (0.9 to 2.2) 0.1 (0.9 to 1.1 (0.9 to 1.1 (0.9 to 1.1 (0.9 to 2.2) 0.1 (0.9 to 1.1 (0.9 to 1.1 (0.9 to 2.2) 0.4 (0.7 to 8.7) 0.8 (0.5 to 0.7 (No regular exercise ³³	o- •	1.5 (1.3 to 1.7)	0.4 (0.2 to 0.8)
No gait abnormality ³³ 0.9 (0.7 to 1.1) 1.5 (0.9 to 1.1) Osteoarthritis ³³ 1.1 (0.8 to 1.5) 1.0 (0.7 to 1.1) Combination red flags 1.0 (1.0 to 1.1) 0.4 (0.1 to 1.2) 1 positive ³³ 0.9 (0.7 to 1.1) 1.5 (0.9 to 1.2) 2 positive ³³ 0.9 (0.7 to 1.1) 1.5 (0.9 to 1.2) 3 positive ³³ 0.9 (0.7 to 1.1) 0.4 (0.1 to 1.1) 4 positive ³³ 0.9 (0.7 to 1.1) 0.4 (0.1 to 1.1) 5 positive ³³ 0.9 (0.7 to 1.1) 0.4 (0.1 to 1.1) Contusion or abrasion ³⁵ 0.9 (0.7 to 1.1) 0.4 (0.1 to 1.1) Spasm ³⁵ 0.9 (0.7 to 1.1) 0.4 (0.1 to 1.1) Contusion or abrasion ³⁵ 0.9 (0.7 to 1.1) 1.0 (1.0 to 1.1) Sensation change ³⁵ 0.9 (0.7 to 1.1) 1.5 (0.8 to 2.0) Motor deficit ²⁵ 0.9 (0.8 to 1.5) 1.5 (0.8 to 2.6) 0.9 (0.8 to 1.5) DTR abnormality ³⁵ 0.9 (0.7 to 1.1) 1.5 (0.9 to 3.1) 0.0 (0.9 to 1.5) Neurological deficit ²⁵ 0.9 (0.8 to 1.5) 0.7 (0.2 to 2.2) 1.0 (0.9 to 1.5) Neurological deficit ²⁴ ¶ 0.9 (0.7 to 5.1 to 2.5) 0.9 (0.8 to 1.5) 0.7 (0.2 to 2.2) 0.7 (0.5 to 2.5) <td>Examination</td> <td></td> <td></td> <td></td>	Examination			
Osteoarthritis ³³ 1.1 (0.8 to 1.5) 1.0 (0.7 to 7 Combination red flags 1.0 (1.0 to 1.1) 0.4 (0.1 to 7 2 positive ³³ 2.4 (2.0 to 3.0) 0.3 (0.2 to 7 3 positive ³³ 2.4 (2.0 to 3.0) 0.3 (0.2 to 7 4 positive ³³ 2.4 (2.0 to 3.0) 0.3 (0.2 to 7 5 positive ³³ 2.4 (2.0 to 3.0) 0.3 (0.2 to 7 Tertiary care 2.4 (2.0 to 3.0) 0.3 (0.2 to 7 Examination Contusion or abrasion ³⁵ 31.1 (18.2 to 53.0) 0.2 (0.1 to 7 Spasm ³⁵ 1.5 (0.8 to 2.6) 0.9 (0.8 to 1.5) 1.0 (0.9 to 6.3.7) 1.0 (0.9 to 7 Motor deficit ²⁵ 0.4 0.4 (0.2 to 10.9) 1.0 (0.9 to 7 1.0 (0.9 to 7 1.1 (0.8 to 2.2) 1.0 (0.9 to 7 Straight leg raise (40° ³⁵) 0.4 0.4 (0.2 to 10.9) 1.0 (0.9 to 7 1.0 (0.5 to 2.0) 1.0 (0.9 to 7 Neurological deficit ³⁴ 0.4 (0.2 to 3.0) 0.7 (0.5 to 7 0.8 (0.5 to 7) 0.8 (0.5 to 7) 0.7 (0.2 to 2.2) 1.0 (1.0 to 7) Combination red flags 0.4 (0.2 to 3.0) 0.4 (0.5 to 6.0) 0.7 (0.5 to 2.0) 0.7 (0.5 to 2.0) 0.7 (0.5 to 2.0) 0.7 (0.5 to 2.0) 0.7 (0.5 to 2.	BMI <23 ³³	o	2.2 (1.4 to 3.4)	0.8 (0.6 to 1.0)
Combination red flags 1 positive ³³ + 1.0 (1.0 to 1.1) 0.4 (0.1 to 2 positive ³³ 1.4 (1.3 to 1.6) 0.2 (0.0 to 3 positive ³³ 2.4 (2.0 to 3.0) 0.3 (0.2 to 4 positive ³³ 2.4 (2.0 to 3.0) 0.3 (0.2 to 5 positive ³³ 9.6 (5.9 to 15.7) 0.7 (0.5 to Tertiary care 7.6 (0.9 to 63.7) 1.0 (0.9 to Examination 1.4 (1.2 to 53.0) 0.2 (0.1 to Contusion or abrasion ³⁵ 1.5 (0.8 to 2.6) 0.9 (0.8 to Spasm ³⁵ 1.4 (0.2 to 10.9) 1.0 (0.9 to Tendemess ³⁵ 1.4 (0.2 to 10.9) 1.0 (0.9 to Motor deficit ³⁵ 1.0 (0.5 to 4.9) 1.0 (0.9 to DTR abnormality ³⁵ 1.0 (0.5 to 2.0) 1.0 (0.9 to Straight leg raise (40 ⁻³⁵) 1.0 (0.5 to 2.0) 1.0 (0.9 to Neurological deficit ³⁶ 1.4 (2.4 to 87.6) 0.7 (0.5 to Neurological signs ³⁶ 1.4 (2.4 to 87.6) 0.7 (0.5 to Neurological signs ³⁶ 1.4 (2.4 to 87.6) 0.7 (0.5 to Nultiple findings ³⁵ 0 0 0 0 0 0 0.7 (0.5 to 0	No gait abnormality ³³	•	0.9 (0.7 to 1.1)	1.5 (0.9 to 2.3)
1 positive ³³ 4 0 1.0 (1.0 to 1.1) 0.4 (0.1 to 1.2) 2 positive ³³ 0 1.4 (1.3 to 1.6) 0.2 (0.0 to 2.4 (2.0 to 3.0) 0.3 (0.2 to 2.4 (2.0 to 3.0)	Osteoarthritis ³³	+	1.1 (0.8 to 1.5)	1.0 (0.7 to 1.3)
2 positive ³³ 1.4 (1.3 to 1.6) 0.2 (0.0 to 2.4 (2.0 to 3.0) 3 positive ³³ 9.6 (5.9 to 15.7) 0.7 (0.5 to 7.6 (0.9 to 63.7) 4 positive ³³ 9.6 (5.9 to 15.7) 0.7 (0.5 to 7.6 (0.9 to 63.7) 5 positive ³³ 9.6 (5.9 to 15.7) 0.7 (0.5 to 7.6 (0.9 to 63.7) 7.6 (0.9 to 63.7) 1.0 (0.9 to 63.7) 1.0 (0.9 to 63.7) 5 positive ³³ 9.6 (5.9 to 15.7) 0.7 (0.5 to 7.6 (0.9 to 63.7)) 7.6 (0.9 to 63.7) 1.0 (0.9 to 63.7) 1.0 (0.9 to 63.7) 5 positive ³³ 9.6 (5.9 to 15.7) 0.7 (0.5 to 7.6 (0.9 to 63.7)) 7 renderness ³⁵ 9.6 (5.9 to 15.7) 0.7 (0.5 to 7.6 (0.9 to 63.7)) 8 to 1.8 (1.4 to 2.2) 0.5 (0.3 to 1.8 (1.4 to 2.2)) 0.5 (0.3 to 1.8 (1.4 to 2.2)) 9.7 Trauma and neurological signs ³⁶ 9.6 (0.9 to 63.7) 1.0 (0.9 to 1.5 (0.9 to 1.9 (0.	Combination red flags			
3 positive ³³ 2.4 (2.0 to 3.0) 0.3 (0.2 to 9.6 (5.9 to 15.7)) 4 positive ³³ 9.6 (5.9 to 15.7) 0.7 (0.5 to 7.6 (0.9 to 63.7)) Tertiary care 1.5 (0.8 to 2.6) 0.9 (0.8 to 15.7) Examination 1.5 (0.8 to 2.6) 0.9 (0.8 to 15.7) Spasm ³⁵ 1.8 (1.4 to 2.2) 0.5 (0.3 to 15.7) Tenderness ³⁵ 1.8 (1.4 to 2.2) 0.5 (0.3 to 15.7) Sensation change ³⁵ 1.4 (0.2 to 10.9) 1.0 (0.9 to 3.1) Motor deficit ²⁵ 1.4 (0.2 to 10.9) 1.0 (0.9 to 3.1) DTR abnormality ³⁵ 1.5 (0.5 to 4.9) 1.0 (0.9 to 3.1) Straight leg raise (40° ³⁵) 1.4 (0.2 to 10.9) 1.0 (0.9 to 3.1) Neurological signs ³⁶ 1.4 (0.2 to 10.9) 1.0 (0.9 to 3.1) Combination red flags 1.4 (0.2 to 10.9) 1.0 (0.9 to 3.1) Trauma and neurological signs ³⁶ 1.4 (0.2 to 10.9) 0.7 (0.2 to 2.2) 0 1.0 2.0 3.0 4.0 5.0 5.0 0 1.0 2.0 3.0 4.0 5.0 5.0 1.1 (18.2 to 53.0) 1.0 (0.9 to 2.1) 1.0 (0.9 to 2.1) 1.0 (0.9 to 2.1) 1.0 (0.9 to 2.1)	1 positive ³³ ‡		1.0 (1.0 to 1.1)	0.4 (0.1 to 3.0)
4 positive ³³ 9.6 (5.9 to 15.7) 0.7 (0.5 to 7) 5 positive ³³ 7.6 (0.9 to 63.7) 1.0 (0.9 to 63.7) Tertiary care Examination 31.1 (18.2 to 53.0) 0.2 (0.1 to 7) Spasm ³⁵ 1.5 (0.8 to 2.6) 0.9 (0.8 to 7) Tenderness ³⁵ 1.8 (1.4 to 2.2) 0.5 (0.3 to 7) Sensation change ³⁵ 1.4 (0.2 to 10.9) 1.0 (0.9 to 7) Motor deficit ³⁵ 1.5 (0.5 to 4.9) 1.0 (0.9 to 7) DTR abnormality ³⁵ 1.5 (0.5 to 4.9) 1.0 (0.9 to 7) Straight leg raise (40° ³⁵) 1.0 (0.9 to 7) 1.0 (0.9 to 7) Neurological signs ³⁶ (S) 1.4 (0.2 to 10.9) 1.0 (0.9 to 7) Neurological signs ³⁶ (S) 1.4 (0.2 to 10.9) 1.0 (0.9 to 7) Neurological signs ³⁶ (S) 1.4 (0.2 to 10.9) 1.0 (0.9 to 7) Neurological signs ³⁶ (S) 1.4 (0.7 to 8.7) 0.8 (0.5 to 7) Neurological signs ³⁶ (S) 1.4 (2.4 to 87.6) 0.7 (0.5 to 4.9) 0 10 20 30 40 50 60 70 80 90 10	2 positive ³³	p- •	1.4 (1.3 to 1.6)	0.2 (0.0 to 0.6
5 positive ³³ 7.6 (0.9 to 63.7) 1.0 (0.9 to 63.7) Tertiary care 5 positive ³³ 31.1 (18.2 to 53.0) 0.2 (0.1 to 1.5 (0.8 to 2.6) Spasm ³⁵ 1.5 (0.8 to 2.6) 0.9 (0.8 to 1.5 (0.8 to 2.6) 0.9 (0.8 to 1.5 (0.8 to 2.6) Sensation change ³⁵ 1.4 (0.2 to 10.9) 1.0 (0.9 to 1.5 (0.8 to 2.6) 0.9 (0.8 to 1.5 (0.8 to 2.6) Motor deficit ³⁵ 1.4 (0.2 to 10.9) 1.0 (0.9 to 1.5 (0.5 to 4.9) 1.0 (0.9 to 1.5 (0.5 to 4.9) DTR abnormality ³⁵ 1.5 (0.5 to 4.9) 1.0 (0.9 to 1.5 (0.5 to 2.0) 1.0 (0.9 to 1.5 (0.5 to 2.0) Straight leg raise <40° ³⁵ 1.0 (0.5 to 2.0) 1.0 (0.9 to 1.5 (0.5 to 2.0) 1.0 (0.9 to 1.5 (0.5 to 2.0) Neurological signs ³⁶ § 1.0 (0.5 to 2.0) 1.0 (0.9 to 1.5 (0.5 to 2.0) 1.0 (0.9 to 1.5 (0.5 to 2.0) Combination red flags 1.1 (1.4 to 3.1) 0.7 (0.5 to 1.5 (0.5 to 1.	3 positive ³³	u =	2.4 (2.0 to 3.0)	0.3 (0.2 to 0.6
Tertiary care Examination 31.1 (18.2 to 53.0) 0.2 (0.1 to 55.0) Contusion or abrasion ³⁵ 593m ³⁵ 1.5 (0.8 to 2.6) 0.9 (0.8 to 1.8 (1.4 to 2.2)) Tendemess ³⁵ 6 1.8 (1.4 to 2.2) 0.5 (0.3 to 1.4 (0.2 to 10.9)) 1.0 (0.9 to 1.4 (0.2 to 10.9)) Motor deficit ³⁵ 6 6 7 3.1 (0.4 to 27.3) 1.0 (0.9 to 1.4 (0.2 to 10.9)) DTR abnormality ³⁵ 7 7 7 1.5 (0.5 to 4.9) 1.0 (0.9 to 1.9 to 1.9 (0.9 to 1.9 to 1	4 positive ³³	•	9.6 (5.9 to 15.7)	0.7 (0.5 to 0.8
Examination Contusion or abrasion ³⁵ 31.1 (18.2 to 53.0) 0.2 (0.1 to Spasm ³⁵ Tendemess ³⁵ 1.5 (0.8 to 2.6) 0.9 (0.8 to Sensation change ³⁵ Motor deficit ³⁵ 1.8 (1.4 to 2.2) 0.5 (0.3 to DTR abnormality ³⁵ 1.6 (0.4 to 27.3) 1.0 (0.9 to Straight leg raise <40° ³⁵ 1.5 (0.5 to 4.9) 1.0 (0.9 to Neurological signs ³⁶ 1.0 (0.5 to 2.0) 1.0 (0.9 to Combination red flags 1.4 (2.4 to 87.6) 0.7 (0.2 to 2.2) 1.0 (1.0 to Trauma and neurological signs ³⁶ 14.4 (2.4 to 87.6) 0.7 (0.5 to 2.1 (1.4 to 3.1) 0.7 (0.6 to 0 10 20 30 40 50 60 70 80 90 100	5 positive ³³		7.6 (0.9 to 63.7)	1.0 (0.9 to 1.0)
Contusion or abrasion ³⁵ 31.1 (18.2 to 53.0) 0.2 (0.1 to Spasm ³⁵ Tendemess ³⁵ 1.5 (0.8 to 2.6) 0.9 (0.8 to Sensation change ³⁵ Image: Sensation change ³⁵ 1.8 (1.4 to 2.2) 0.5 (0.3 to Motor deficit ³⁵ Image: Sensation change ³⁵ Image: Sensation change ³⁵ 1.4 (0.2 to 10.9) 1.0 (0.9 to DTR abnormality ³⁵ Image: Sensation change ³⁶	Tertiary care			
Spasm ³⁵ 1.5 (0.8 to 2.6) 0.9 (0.8 to 1.5 (0.8 to 2.6) Tendemess ³⁵ Sensation change ³⁵ 1.4 (0.2 to 10.9) 1.0 (0.9 to 1.4 (0.2 to 10.9) Motor deficit ³⁵ DTR abnormality ³⁵ 1.5 (0.5 to 4.9) 1.0 (0.9 to 1.4 (0.2 to 10.9) Straight leg raise (40° ³⁵) Neurological signs ³⁶ § 1.6 (0.7 to 8.7) 0.8 (0.5 to 1.9) Neurological deficit ³⁴ ¶ 0.7 (0.2 to 2.2) 1.0 (1.0 to 1.0 to 1.0 (0.9 to 2.1) 0.7 (0.5 to 2.9) Combination red flags 1.4 (2.4 to 87.6) 0.7 (0.5 to 2.9) 1.0 (1.0 to 1.9) Multiple findings ³⁵ 0 1.0 20 30 40 50 60 70 80 90 100	Examination			
Tendemess ³⁵ 1.8 (1.4 to 2.2) 0.5 (0.3 to 3.5 (0.4 to 2.2)) Sensation change ³⁵ 1.4 (0.2 to 10.9) 1.0 (0.9 to 3.1 (0.4 to 27.3)) Motor deficit ³⁵ 1.5 (0.5 to 4.9) 1.0 (0.9 to 3.1 (0.4 to 27.3)) DTR abnormality ³⁵ 1.5 (0.5 to 4.9) 1.0 (0.9 to 3.1 (0.4 to 27.3)) Straight leg raise (40° ³⁵) 1.0 (0.5 to 2.0) 1.0 (0.9 to 3.1 (0.4 to 27.3)) Neurological signs ³⁶ § 1.4 (0.2 to 10.9) 1.0 (0.9 to 3.1 (0.4 to 27.3)) Neurological deficit ³⁴ ¶ 1.0 (0.5 to 2.0) 1.0 (0.9 to 3.1 (0.4 to 27.3)) Combination red flags 1.4 (0.2 to 10.9) 1.0 (0.9 to 3.1 (0.4 to 2.0)) Trauma and neurological signs ³⁶ 1.4 (0.2 to 8.7) 0.8 (0.5 to 4.9) 0 10 20 30 40 50 60 70 80 90 10 0 10 20 30 40 50 60 70 80 90 10	Contusion or abrasion ³⁵		31.1 (18.2 to 53.0)	0.2 (0.1 to 0.3
Sensation change ³⁵ 1.4 (0.2 to 10.9) 1.0 (0.9 to Motor deficit ³⁵ 1.5 (0.5 to 4.9) 1.0 (0.9 to DTR abnormality ³⁵ 1.5 (0.5 to 4.9) 1.0 (0.9 to Straight leg raise (40° ³⁵) 1.0 (0.9 to 1.5 (0.5 to 4.9) 1.0 (0.9 to Neurological signs ³⁶ § 1.0 (0.7 to 8.7) 0.8 (0.5 to 2.4 (0.7 to 8.7) 0.8 (0.5 to Combination red flags 1.4 (2.4 to 87.6) 0.7 (0.5 to 1.0 (1.0 to Multiple findings ³⁵ 0 1.0 20 30 40 50 60 70 80 90 100	Spasm ³⁵		1.5 (0.8 to 2.6)	0.9 (0.8 to 1.1)
Motor deficit ³⁵ 3.1 (0.4 to 27.3) 1.0 (0.9 to DTR abnormality ³⁵ Straight leg raise (40° ³⁵) 1.5 (0.5 to 4.9) 1.0 (0.9 to Straight leg raise (40° ³⁵) 1.0 (0.9 to Neurological signs ³⁶ § Combination red flags Multiple findings ³⁵ 0 10 20 30 40 50 60 70 80 90 100	Tenderness ³⁵	- ·	1.8 (1.4 to 2.2)	0.5 (0.3 to 0.8
DTR abnormality ³⁵ Straight leg raise (40° ³⁵ Neurological signs ³⁶ § Neurological deficit ³⁴ ¶ Combination red flags Trauma and neurological signs ³⁶ Multiple findings ³⁵ 0 10 20 30 40 50 60 70 80 90 100 1.5 (0.5 to 4.9) 1.0 (0.9 to 1.0 (0.5 to 2.0) 1.0 (0.9 to 2.4 (0.7 to 8.7) 0.8 (0.5 to 0.7 (0.2 to 2.2) 1.0 (1.0 to 2.1 (1.4 to 3.1) 0.7 (0.6 to 2.1 (1.4 to 3.1) 0.7 (0.6 to	Sensation change ³⁵		1.4 (0.2 to 10.9)	1.0 (0.9 to 1.0)
Straight leg raise <40° ³⁵ 1.0 (0.5 to 2.0) 1.0 (0.9 to 2.0) Neurological signs ³⁶ § 2.4 (0.7 to 8.7) 0.8 (0.5 to 2.0) Neurological deficit ³⁴ ¶ 0.7 (0.2 to 2.2) 1.0 (1.0 to 2.0) Combination red flags 1.4.4 (2.4 to 87.6) 0.7 (0.5 to 2.0) Multiple findings ³⁵ 0 10 20 30 40 50 60 70 80 90 100	Motor deficit ³⁵		3.1 (0.4 to 27.3)	1.0 (0.9 to 1.0)
Neurological signs ³⁶ § 2.4 (0.7 to 8.7) 0.8 (0.5 to 0.7 (0.2 to 2.2)) Neurological deficit ³⁴ ¶ 0.7 (0.2 to 2.2) 1.0 (1.0 to 0.7 (0.2 to 2.2)) Combination red flags 14.4 (2.4 to 87.6) 0.7 (0.5 to 2.1) Multiple findings ³⁵ 0 10 20 30 40 50 60 70 80 90 100	DTR abnormality ³⁵		1.5 (0.5 to 4.9)	1.0 (0.9 to 1.1)
Neurological deficit ³⁴ ¶ 0.7 (0.2 to 2.2) 1.0 (1.0 to 2000) Combination red flags 0.7 (0.2 to 2.2) 1.0 (1.0 to 2000) Trauma and neurological signs ³⁶ 0.7 (0.2 to 2.2) 1.0 (1.0 to 2000) Multiple findings ³⁵ 0.7 (0.2 to 2.2) 1.0 (1.0 to 2.2) 0 10 20 30 40 50 60 70 80 90 100	Straight leg raise <40°35	÷-	1.0 (0.5 to 2.0)	1.0 (0.9 to 1.2)
Combination red flags Trauma and neurological signs ³⁶ Multiple findings ³⁵ 0 10 20 30 40 50 60 70 80 90 100 14.4 (2.4 to 87.6) 0.7 (0.5 to 2.1 (1.4 to 3.1) 0.7 (0.6 to	Neurological signs ³⁶ §		2.4 (0.7 to 8.7)	0.8 (0.5 to 1.3
Trauma and neurological signs ³⁶	Neurological deficit ³⁴ ¶	+= <mark></mark>	0.7 (0.2 to 2.2)	1.0 (1.0 to 1.1)
Multiple findings ³⁵ 2.1 (1.4 to 3.1) 0.7 (0.6 to 0 10 20 30 40 50 60 70 80 90 100	Combination red flags			
Multiple findings ³⁵ 2.1 (1.4 to 3.1) 0.7 (0.6 to 0 10 20 30 40 50 60 70 80 90 100	Trauma and neurological signs ³⁶	-0	14.4 (2.4 to 87.6)	0.7 (0.5 to 1.2)
0 10 20 30 40 50 60 70 80 90 100		o		0.7 (0.6 to 0.9)
	n y en la freinne de la Balancie est a la CERTA (CONTROL ES CONTROL ES C		10080-00 8 000-00000000000000000000000000	
Probabilty of fracture (%)		0 10 20 30 40 50 60 70 80 90 100	0	
		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 <40°. ¶Neurological deficit: consistent with lumbar plexus distribution

				Likelihood ratio (95% CI)	
Red flag in ACP guideline	Red flag investigated	Probability (95% CI)	Positive	Negative	
Cancer: major risk					
History of cancer	History of cancer ³¹	d		15.3 (6.4 to 36.6)	0.7 (0.5 to 1.0
	History of cancer ³⁴			35.0 (20.5 to 59.8)	0.1 (0.0 to 0.9
Multiple risk factors	None identified*				
Clinical suspicion	Did not meet inclusion criteria†				
Cancer: minor risk					
Age >50	Age >50 ²⁷	÷		2.5 (1.4 to 4.5)	0.4 (0.1 to 2.0
	Age >50 ³¹	4		2.7 (2.0 to 3.6)	0.3 (0.1 to 0.9
	Age >50 ²⁹	+		1.9 (0.5 to 7.8)	0.7 (0.2 to 2.7
	Age >50 ³²	4		1.7 (1.5 to 1.9)	0.1 (0.0 to 1.0
	Age ≥50 ³⁷	ę.		0.9 (0.7 to 1.1)	1.3 (1.0 to 1.7
Unexplained weight loss	Unexplained weight loss ³¹	+		2.6 (0.7 to 9.4)	0.9 (0.7 to 1.1
Failure to improve after 1 month	Not improved after 1 month ²⁷	-		2.6 (0.5 to 14.5)	0.8 (0.5 to 1.5
	Not improved after 1 month ³¹		After test if positive	3.1 (1.3 to 7.0)	0.8 (0.5 to 1.1
	Duration >1 month ²⁷	+	 After test if negative 	2.6 (1.5 to 4.7)	0.6 (0.4 to 1.1
		0 10 20	30 40 50 60 70 80 90 10	00	
				- × ·	

Probabilty of fracture (%)

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 review⁴⁵

			Likelihood ratio (95% CI)		
Red flag (not in ACP guideline)		Probability (9	95% CI)	Positive	Negative
Primary care					
History					
Severe pain ³¹	9			1.5 (0.6 to 4.2)	0.9 (0.7 to 1.2)
Is the low-back pain familiar? ³⁰				1.5 (0.1 to 16.2)	0.9 (0.4 to 2.0)
Thoracic pain ³¹	÷			1.0 (0.3 to 3.7)	1.0 (0.8 to 1.3)
Insidious onset ³¹	c.			1.1 (0.7 to 1.6)	0.9 (0.5 to 1.8)
Recent back injury ³¹	4			0.2 (0.0 to 3.0)	1.2 (1.1 to 1.3)
Tried bed rest with no relief ³¹				1.7 (1.2 to 2.2)	0.2 (0.0 to 3.0)
Neurological symptoms ³¹	÷			0.4 (0.0 to 6.5)	1.1 (0.9 to 1.2)
Neurological symptoms ²⁸	-			7.5 (0.7 to 84.2)	0.8 (0.3 to 1.7)
Examination					
Muscle spasm ³¹	•			0.5 (0.1 to 1.6)	1.3 (1.0 to 1.6)
Spine tenderness ³¹				0.4 (0.1 to 1.4)	1.4 (1.1 to 1.8)
Fever (temp >100°F) ³¹				1.8 (0.1 to 27.2)	1.0 (0.9 to 1.1)
Tertiary care					
Examination					
Scoliosis ³⁷	d+			1.6 (1.0 to 2.4)	0.9 (0.8 to 1.0)
Kyphosis ³⁷	d e			1.2 (0.7 to 2.3)	1.0 (0.9 to 1.1)
Midline spine tenderness ³⁷	-			0.8 (0.6 to 1.1)	1.2 (0.9 to 1.5)
No pain on movement screen ³⁷ *	÷		 After test if positive After test if negative 	1.3 (1.1 to 1.7)	0.8 (0.6 to 1.0)
	0 10 20	30 40 50	60 70 80 90 100)	
			Probabilty of fracture (%))	

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