Disability and Rehabilitation

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/idre20

The diagnostic value of Red Flags in thoracolumbar pain: a systematic review

Filippo Maselli , Michael Palladino , Valerio Barbari , Lorenzo Storari , Giacomo Rossettini & Marco Testa

To cite this article: Filippo Maselli , Michael Palladino , Valerio Barbari , Lorenzo Storari , Giacomo Rossettini & Marco Testa (2020): The diagnostic value of Red Flags in thoracolumbar pain: a systematic review, Disability and Rehabilitation, DOI: 10.1080/09638288.2020.1804626

To link to this article: https://doi.org/10.1080/09638288.2020.1804626

DISABILITY AND REHABILITATION
https://doi.org/10.1080/09638288.2020.1804626

Taylor & Francis
Taylor & Francis Group

REVIEW ARTICLE

The diagnostic value of Red Flags in thoracolumbar pain: a systematic review

Filippo Maselli[a,b], Michael Palladino[a,c], Valerio Barbari[a,d], Lorenzo Storari[a,e], Giacomo Rossettini[a,f]* and Marco Testa[a]*

[a]Department of Neurosciences, Rehabilitation, Ophthalmology, Genetic and Maternal Infantile Sciences (DINOGMI), University of Genoa – Campus of Savona, Savona, Italy; [b]Sovrintendenza Sanitaria Regionale Puglia INAIL, Bari, Italy; [c]Private Practice, Torino, Italy; [d]Private Practice, Rimini, Italy; [e]Private Practice, "Centro Retrain", Verona, Italy; [f]School of Physiotherapy, University of Verona, Verona, Italy

ABSTRACT

Purpose: Red Flags (RFs) are signs and symptoms related to the screening of serious underlying pathologies mimicking a musculoskeletal pain. The current literature wonders about the usefulness of RFs, due to high false-positive rates and low diagnostic accuracy. The aims of this systematic review are: (a) to identify and (b) to evaluate the most important RFs that could be found by a health care professional during the assessment of patients with low and upper back pain (named as thoracolumbar pain (TLP)) to screen serious pathologies.

Materials and methods: A systematic review of the literature was conducted. Searches were performed on seven databases (Pubmed, Web of Science, Cochrane Library, Pedro, Scielo, CINAHL, and Google Scholar) between March 2019 and June 2020, using a search string which included synonyms of low back pain (LBP), chest pain (CP), differential diagnosis, RF, and serious disease. Only observational studies enrolling patients with LBP or CP were included. Risk of bias was assessed with the Newcastle Ottawa Scale and inter-rater agreement between authors for full-text selection was evaluated with Cohen's Kappa. Where possible the diagnostic accuracy was recorded for sensitivity (Sn), specificity (Sp), and positive/negative likelihood ratio (LR+/LR–).

Results: Forty full-texts were included. Most of the included observational studies were judged as low risk of bias, and Cohen's Kappa was good (=0.78). The identified RFs were: advanced age; neurological signs; history of trauma; malignancy; female gender; corticosteroids use; night pain; unintentional weight loss; bladder or bowel dysfunction; loss of anal sphincter tone; saddle anaesthesia; constant pain; recent infection; family or personal history of heart or pulmonary diseases; dyspnoea; fever; postprandial CP; typical reflux symptoms; haemoptysis; sweating; pain radiated to upper limbs; hypotension; retrosternal pain; exertional pain; diaphoresis; and tachycardia. The diagnostic accuracy of RFs as self-contained screening tool was low, while the combination of multiple RFs showed to increase the probability to identify serious pathologies.

Conclusions: Despite the use of single RF should not be recommended for the screening process in clinical practice, the combination of multiple RFs to enhance diagnostic accuracy is promising. Moreover, the identified RFs could be a baseline to develop a screening tool for patients with TLP.

➤ IMPLICATIONS FOR REHABILITATION

- Differential diagnosis and screening for referral are mandatory skills for each healthcare professional in direct access clinical settings, and should be the primary step for an appropriate management of a patient with signs and symptoms mimicking serious pathologies in thoracolumbar region.
- Clinical reasoning and decision-making processes are essential throughout all phases of a patient's pathway of care. By which, the use of single Red Flag (RF) as a self-contained screening tool should not be recommended. The combination of multiple RFs promises to increase diagnostic accuracy and could grow into an excellent screening tool for thoracolumbar pain.

ARTICLE HISTORY
Received 1 January 2020
Revised 28 July 2020
Accepted 29 July 2020

KEYWORDS
Low back pain; chest pain; differential diagnosis; referral and consultation; Red Flag

Introduction

Rationale

Low back pain (LBP) is a common condition worldwide and it is experienced by 50–80% of adults during the lifetime [1,2]. In about 90% of cases, LBP is defined as non-specific due to the unrecognisable musculoskeletal causes [1,3]. Conversely, between 1 and 4% of cases, LBP is related to specific conditions such as fractures, malignancy, infections, or cauda equina syndrome (CES) [4]. Upper back pain (UBP) is defined as pain in the area of the spine above the base of the rib cage and below the neck, particularly, in the region of the thoracic spine, between the boundaries of T1–T12 and across the posterior aspect of the trunk [5,6]. Given that the one-year prevalence for UBP ranges from 15 to 34.8% in

CONTACT Filippo Maselli masellifilippo76@gmail.com Department of Neurosciences, Rehabilitation, Ophthalmology, Genetic and Maternal Infantile Sciences (DINOGMI), University of Genoa – Campus of Savona, Savona, Italy
*These authors contributed equally to this work.

Supplemental data for this article can be accessed here.

the adult population it is an ordinary presentation in direct access clinical settings [6].

Chest pain (CP) represents the second reason for access to the Emergency Department (ED) with a prevalence rate of 25% [7]. Although only in 20% of cases CP has a non-musculoskeletal cause, this condition continues to be a challenge for professionals; in fact, 50–80% of patients are discharged from ED without clear diagnosis [7]. The most significant serious pathologies mimicking CP are pulmonary, cardiovascular, or gastrointestinal disorders and it would be estimated that less than 35% of subjects with CP need a healthcare intervention such as physiotherapy and rehabilitation [8].

Considered together, LBP, UBP, and CP represent the painful areas between the front trunk and the back (from chest to the back between T1 vertebra and the inferior gluteal fold) [9,10]. In the context of this systematic review, we refer to thoracolumbar pain (TLP) as pain experienced in the region of lumbar and sacral spine, including L1-S4 segments to the lower back down to the inferior gluteal fold, and thoracic spine between T1 and T12 and across the posterior aspect of the trunk. Furthermore, we included the area of the chest wall, from breastbone to the costal region on the anterior aspect of the trunk.

According to the definition, Red Flags (RFs) consist of signs and symptoms that are related to a serious underlying pathology, and may indicate more diagnostic testing is necessary before the appropriate care can be delivered [11,12] challenging and RFs may be detected by the healthcare professionals (i.e., physicians, nurses, physiotherapists, chiropractors, etc.). Four clinical guidelines suggest taking a careful patient's history and physical examination as the optimal approach aimed to recognise RFs [11,13–15]. For a long time, RFs were considered a reference point in direct access clinical settings, aimed to reduce the risk of a contraindicated intervention and identify as soon as possible an appropriate treatment for the patient [11]. Over the last years, six systematic and narrative reviews concerning spinal pain were published proving that RFs have low diagnostic accuracy and several rates of false-positive, especially when used self-contained screening tool [12,14,16–19]. Nevertheless, the recently published positional statement by Finucane et al. has reported that RFs remain the best tools at the clinician's disposal to raise suspicion of serious pathologies, and their combined use improves the diagnostic accuracy [20].

In current literature, there are no published systematic reviews concerning RFs for TLP. Conversely, six systematic reviews regarding vertebral fracture and/or malignancy namely for LBP have already been published [13,14,17,21–23], but the association with CP has not been proven and the included studies are mostly dated. Particularly, the RFs with the highest post-test probability to detect vertebral fracture are: older age; use of corticosteroids; major trauma; distracting painful injury; tenderness; female gender; and presence of contusion or abrasion [4,14,17,21,23]. Concerning malignancy, the most important RF is the previous history of cancer. Up to now, there are only narrative reviews regarding CP and the main RFs are: previous history of cardiovascular disorders, dyspnoea, pain radiating to upper limbs and cough [24–26].

Regarding the methodological quality of the current systematic and narrative reviews on TLP, only four systematic reviews [14,17,21,23] stated the establishment of the methods prior to the commencement of the review and only three [13,17,21] used a comprehensive literature search strategy. In details, Downie and Williams et al. [17,21] thoroughly explained the selection of study designs for inclusion in the review and provided a list of excluded studies with reasons for the exclusion. In a low number of reviews, the data extraction process was performed by at least two authors and the majority of the studies did not use an adequate tool for quality assessment of the included primary studies. The study of Verhagen et al. [4] in 2016 compared the RFs included in the major international guidelines, observing that only three out of the 16 guidelines have any evidence to support RF screening. Most guidelines did not endorse the same set of RFs and most recommendations were not supported by reliable diagnostic accuracy data [4]. However, the use of RFs is still advocated [4,22,23]. Finally, the methodological quality of the studies concerning CP was generally low and based upon experts' opinion without references to any primary study [27].

According to recent literature [4,21–23,28], there is a strong need to investigate the most-commonly adopted RFs for TLP and their diagnostic accuracy to improve the screening process of serious pathologies by healthcare professionals in direct access clinical settings.

Objectives of the study

The aims of this systematic review are: (a) to identify; and (b) to evaluate the most important RFs to take into account during the assessment of patients with TLP by healthcare professionals in direct access clinical settings.

Materials and methods

Protocol

This systematic review was conducted according to the guidelines of the PRISMA statement [29]. The panel of the authors of this systematic review: (a) has extensive experience in performing systematic reviews and (b) presented specific clinical expertise and training in the screening of patients with TLP. Overall, Cohen's Kappa (*K*) was used to quantify the inter-rater agreement between the two authors (FM, MP) for full-text selection. Cohen's *K* was interpreted according to Altman's definition [30]: $k < 2$ poor, $0.2 < k < 0.4$ fair, $0.41 < k < 0.60$ moderate, $0.61 < k < 0.80$ good, $0.81 < k < 1.00$ excellent [31]. A confidence interval of 95% was calculated (95% CI). Disagreements were solved by a third reviewer (MT) not involved in the data extraction process.

Search strategy

An electronic search was performed independently by two reviewers (FM and MP) under the supervision of a third author (MT) between September 2018 and June 2020 on Pubmed, Web of Science, Cochrane Library, Pedro, Scielo, CINAHL, and Google Scholar. The search strings were developed according to the PI(C)O model of clinical question (participants, interventions, and outcomes) [29]. To make the search strategies sensitive, we did not insert key words for comparisons. The PI(C)O model and the full search strategy for all databases are available in the Supplementary Materials (Supplementary Files 1 and 2). Where possible MeSH (Medical Subject Headings) terms were used and combined with Boolean operators (AND, OR, NOT). Additionally, we conducted a manual search of all bibliographies of the studies assessed for the subsequent full-text selection. In addition, grey literature was screened (i.e., thesis, conference reports, expert opinions, and books) via web.

Eligibility criteria

To be eligible, full-texts had to be observational studies published in English or Italian language. Publication date was restricted from 01 January 1999 to nowadays. No age of participants restrictions was applied.

Study selection

All titles and abstracts were screened independently by two reviewers (FM and MP) under the supervision of a third author (MT). All the studies investigating the identification and/or the evaluation of RFs or specific sign and symptoms of serious pathologies in TLP patients were included. Where appropriate authors were contacted in order to obtain the full-text paper. Finally, full-texts were independently screened and assessed for eligibility by two reviewers (FM, MP).

Inter-rater agreement

Cohen's Kappa (*K*) was used to quantify the inter-rater agreement between two authors (FM, MP) for full-text selection, between three authors (FM, MP, and GR) for the data extraction and between four authors (FM, MP, VB, and LS) for the quality assessment. Cohen's *K* was interpreted according to Altman's definition [32]: $k < 2$ poor, $0.2 < k < 0.4$ fair, $0.41 < k < 0.60$ moderate, $0.61 < k < 0.80$ good, $0.81 < k < 1.00$ excellent.

Data extraction

Three authors (FM, MP, and GR) individually extracted data using a data extraction form developed in line with the PI(C)O model of the clinical question and adapted from the Cochrane Collaboration guidelines (Cochrane Handbook 5-1) [29,71]. Data extraction was organised as follows: (a) authors; (b) publication year; (c) study design; (d) study population; (e) study objectives; (f) RFs identified; and (g) diagnostic accuracy levels for each RF. All RFs for serious pathologies were identified and where possible diagnostic accuracy was analysed: positive likelihood ratio (LR+); negative likelihood ratio (LR–); sensibility (Sn); and specificity (Sp). The LRs use the sensitivity and specificity of the test to determine whether a test result usefully changes the probability that a condition (such as a disease state) exists. Reference values for LR are reported in Supplementary Materials (Supplementary File 3) [72]. The results were screened using the Ryyan software for the management of systematic reviews [73]. Bibliography was handled using the Mendeley software [74]. The interpretation for LR data and the full-text exclusion with reason are reported in Supplementary Materials (Supplementary Files 3 and 4).

Quality assessment

Four authors (FM, MP, VB, and LS) performed the quality assessment of the included studies using the Newcastle-Ottawa Quality Assessment for observational studies [75], in accordance with the recommendations reported in the Cochrane Collaboration guidelines (Cochrane Handbook 5-1) [71].

Results

Study selection

The electronic database searches delivered 1840 results. Grey literature provided additional 19 studies. After removal of 348 duplicates, we excluded 1459 records because they did not meet the inclusion criteria, leaving 52 studies eligible for full-text assessment. Twelve out of 52 records were excluded after full-text reading. In total, 40 observational studies were included [27,28,33–37,39,40,42–70,76]. The full search process is reported in Figure 1. The agreement between the authors was good (Cohen's *K*: 0.790; 95% CI: 0.686–0.893) for the screening process.

Study characteristics

Forty primary observational studies were included in this systematic review [27,28,33,34,36,37,39,40,42–46,48–58,60–62,64–70,76]. Details of each paper are reported in Table 1.

The total number of patients recruited were 49 422, ranging between 11 and 91 years of age, with an equal F/M ratio. Twenty-one studies concerned patients with LBP and thoracic pain [28,33,37–43,45,46,50,53,54,58,63,68,69] and 19 were focused on CP [27,34–36,40,47,48,50,53,55–57,59,61,62,64,65,67,70]; all patients were screened by a physician in a clinical setting or at the ED, through a physical assessment [40,44,50,51,65] or by a questionnaire [27,28,41,45,49].

Risk of bias

Most of the studies were judged as low risk of bias. The main methodological limitations concerned non-representative cohort of patients, uncomplete reporting of follow-up data, lack of statistics about lost to follow-up patients, set of RFs for identical pathologies were differently analysed, differences in reference standards for RFs among studies. Only three studies [28,50,68] were judged as high methodological quality. Agreement between authors was good (Cohen's *K*: 0.777; 95% CI: 0.571–0.983). Methodological quality assessment is reported in Table 2.

Synthesis of results

The extraction process shows a substantial agreement between authors. Serious pathologies identified through the RFs screening as a cause of TLP were: vertebral fracture [28,41,43,45,50–52,60,63,68,69]; malignancy [28,39–41,43,53,63,68,69] infection [28,38,46,50,54,63,68]; CES [28,33,37,42,45,50,58,63,68]; cardiovascular disorders [27,34,35,47,48,56,59,61,62,65]; pulmonary disorders [36,40,57,67,70]; gastroesophageal disorders [53,55,64]; and inflammatory disorders [50]. For each of these, we listed RFs and relative diagnostic accuracy data in table form (Tables 3–7). Frequency of RFs among studies is reported in Figure 2.

Vertebral fracture

Vertebral fracture represents the most frequently reported serious pathologies in LBP patients with a prevalence between 0.4% and 5.6 [28,50].

The most useful RFs to identify a vertebral fracture were: advanced age; history of trauma; midline tenderness; female gender; neurological signs; pain Numeric Rating Scale >7; prolonged use of corticosteroids; osteoporosis; osteoarthritis; and distracting painful injury [28,41,43,45,50–52,60,63,68,69]. In the physical examination, is it important to pay attention to: gait abnormalities; palpable midline step; bruising; positive percussion test; decrease in height and neurological signs [51,52,60].

Eight out of 12 studies (65%) identified history of trauma, advanced age, prolonged corticosteroids use and female gender as key RFs for vertebral fractures [28,45,50,63]. Henschke et al. [50] investigated the combination of these four features obtaining

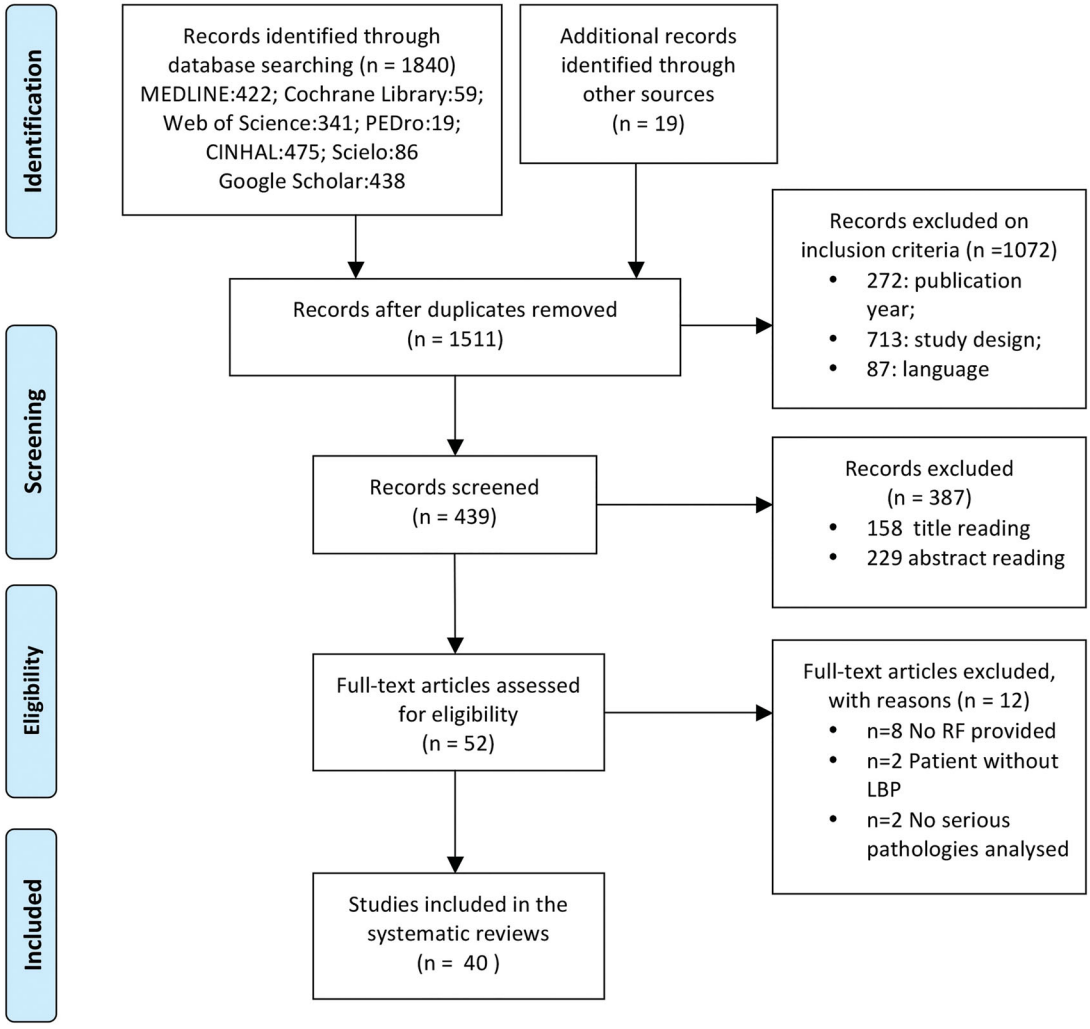

Figure 1. Flowchart.

promising results with a post-test probability to detect fracture increased up to 52%: one positive RF (Sn: 88%/Sp: 50%/LR+: 1.8); two positive RFs (Sn: 63%/Sp: 96%/LR+: 15.5); three or more positive RFs (Sn: 38%/Sp: 100%/LR+: 218.3). Enthoven et al. [45] identified a diagnostic prediction model combining multiple RFs: one positive RF (Sn 0.88; Sp 0.42; LR+ 1.5; LR– 0.3); two positive RFs (Sn 0.70; Sp 0.81; LR+ 3.6; LR– 0.4); ≥3 positive RFs (Sn 0.30 ; Sp 0.95; LR+ 5.8; LR– 1.0). Even Roman et al. [60] combined three RFs, improving diagnostic accuracy data (Sn 76%; Sp 68%; LR+ 2.5; LR– 0.34). According to Premkumar et al. [28], increasing diagnostic accuracy was achieved associating a history of recent trauma to age > 50 years (LR+ 2.54) or >70 years (LR+ 4.35), respectively, by 13.1% and 20.5%. Finally, higher reliability was observed when multiple RFs have been combined, but only five primary studies supported these data [28,45,50,60,68]. The overall diagnostic accuracy data are reported in Table 3.

Spinal malignancy

The prevalence of malignancy in LBP ranged from 0.1% to 1.6% and the spine was the most common site for bony metastases, affecting up to 30–70% of patients with cancer [39,50]. Early detection was the most important screening aimed to prevent the spread of any metastatic disease and development of further complications such as spinal cord compression [28,50].

The most reliable RFs to identify a malignancy were: previous history of cancer; older age; unexplained weight loss; absence of improvements in conservative treatments; positive clinical judgment; night pain and steady back pain unrelieved by rest. The previous history of cancer was the RF with highest diagnostic accuracy (LR+ 7.25) and carries a 10.6% probability of having a vertebral malignancy [28]. During the physical examination, absence of pain during any lumbar movements raises the probability to rule-out malignancy (Sn 100%; LR– 0.01) [39].

Data of diagnostic accuracy are reported in six primary studies [28,39,43,63,68,69], and the results are presented in Table 4. According to Henschke et al. [50], the RFs with the highest number of false positives (>10%) were: unrestful sleep; insidious onset; and age <20 years or >55 years. Two studies identified night pain as a false-positive in almost 85% of subjects with malignancy, when no other RF symptoms are present [28,49]. Premkumar et al. [28] evaluated the combination of multiple RFs which increased the probability of a serious pathology up to 14.3%. (LR+ 10.25), in case of personal history of malignancy and unexplained weight loss.

Spinal infection

Tuberculosis (TB) and *Staphylococcus aureus* represented the main source of spinal infection reported among the included studies. Early diagnosis is essential as the infection could hesitate to an extradural spinal abscess (ESA) with neurological signs and symptoms [54]. Spinal infection occurs between 0.3% and 1.2% [28,50] and thoracolumbar joint is the most common vertebral sites affected (62–65 %) [11,38].

Table 1. Characteristics of the included papers.

Author	Objective	Methods	Results
Ahad 2015 [33]	To underline clinical signs could predict the presence of CES	Retrospective study of 79 consecutive patients undergoing MRI of the spine	RFs: decreased anal tone, faecal incontinence, urinary retention, bladder, incontinence, constipation, and saddle anaesthesia
Albarran 2002 [34]	To investigate whether there were differences in pain radiation between those with and without MI and according to gender	Prospective study of 541 patients presenting with CP	Radiation to neck, back, and upper limbs are common features in MI. women with MI described more radiation to the right arm, upper right region than those without MI
Aydin 2019 [35]	To create a chest pain score in determining whether or not the chest pain was related to ischaemic heart disease (IHD)	Cross-sectional study of 484 patients presented to the cardiology clinics or ED complaining of CP	Chest pain score (5 questions, overall score: 10)>4.5 showed Sn: 90%/Sp: 95.83%/AUC: 0.967 to screen IHD
Bernard Bagattini 2004 [36]	To evaluate clinical characteristics that allow to predicting alternative diagnoses other than PE by ruling out venous thromboembolism	Retrospective study of 1090 consecutive patients admitted for clinically suspected of PE	RFs: tachycardia, recent immobilisation, dyspnoea, age > 40, and haemoptysis
Balasubramanian 2010 [37]	To evaluate the efficacy of clinical assessment in the diagnosis of CES	Retrospective cohort study of 80 patients who presented with clinical features of CES	Saddle anaesthesia is the only clinical feature with a statistically significant association with MRI positive CES. Other RFs: unilateral or bilateral leg pain, bladder or bowel dysfunction
Body 2010 [27]	To assess the value of individual historical and examination findings for diagnosing AMI and other cardiac events	Prospective observational study of 796 ED patients with CP	RFs for AMI: pain radiating to the right arm or both arms, vomiting, central CP, and sweating. The presence of rest pain or pain radiating to the left arm did not significantly modify the probability of AMI
Broderick 2018 [38]	To determine the demographics, presentation and investigation of patients with a TB infection	Retrospective observational study of 31 patients with positive TB cultures	Main RFs: pain and swelling. Fever, sweats, and weight loss are uncommon
Cook 2011 [39]	To investigate the diagnostic accuracy of lumbar movement restrictions and pain in patients with metastatic bone cancer	Retrospective cohort study of 1109 patients with LBP	Pain-free lumbar movements rule-out malignancy: Sn 55%, Sp 59% LR+ 1.3, LR– 0.8
Courtney 2010 [40]	To measure the predictive value of variables for pulmonary embolism	Prospective observational study of 7940 patients with pulmonary embolism	RFs: patient history of pulmonary embolism or deep venous thrombosis or thrombophilia, unilateral leg swelling, recent surgery, oestrogen use, hypoxemia, active cancer, pleuritic or substernal CP, female gender and oxygen saturation less than 95%
de Schepper 2015 [41]	To investigate the prevalence of spinal pathology	Cross-sectional, cohort study of 2975 patients presenting for an MRI lumbar examination	RFs for vertebral fracture are: age > 70 years, history of trauma, and female gender. The RFs for malignancy are: age at onset over 50 years, continuous back pain, back pain at night, history of malignancy, unexplained weight loss
Domen 2009 [42]	To overlooking a potential diagnosis of CES	Retrospectively studied 58 consecutive cases of suspected CES	RFs: bilateral sciatica, subjective urinary retention, or rectal incontinence symptoms
Donner-Banzhoff 2006 [43]	To evaluate the diagnostic validity of a simple heuristic based on the patient's view of the familiarity of LBP	Cross-sectional diagnostic study of 1378 patients presenting with LBP	Diagnostic validity of a simple heuristic based on patient's view is Sn: 50%; Sp: 83%; LR+ 2.95; LR– 0.6
Dugas 2011 [44]	To evaluate the presenting signs and symptoms of SCC and CES and to determine the incidence of Emergency Department (ED) misdiagnosis	Retrospective study of 1231 patients who had visited the ED for a related complaint	The main RFs are: pain, difficulty ambulating, weakness, motor, or sensory deficits
Enthoven 2016 [45]	To identify the prevalence of back pain and to assess associations between RF and vertebral fractures	Prospective cohort study of 669 patients with back pain	RFs for vertebral fracture: age > 70 years, female gender, corticosteroids use, history of trauma, reduction in height, positive percussion test, great disability, painful injury, NRS > 7/10, hip or knee osteoarthritis, and CP
Everden 2018 [46]	To reviewed the epidemiology, management and outcome of all cases of bone and joint TB (BJTB)	Retrospective study of 21 cases of BJTB	Thoracic and lumbar spine are the most common sites affected (62 %). RFs: localised pain, fever, and weight loss
Gesuete 2020 [47]	To evaluate the risk of recurrence, ED re-admissions, level of impairment, and school absenteeism in pediatric patients with CP	Retrospective observational study of 761 patients presented to the pediatric ED with CP	RF: fever; history of cardiac disease; school absenteeism; exertional pain
Goodacre 2002 [48]	To measure the predictive value and diagnostic performance of clinical features used to diagnose coronary syndromes	Prospective, observational cohort study of 893 patients presenting at the ED with acute CP	RFs for AMI: pain radiating to shoulders or both arms; exertional pain; absence of chest wall tenderness. RFs for ACS: pain radiating to shoulders, left arms or both arms; exertional pain
Harding 2005 [49]	To assess the importance of the symptom of night pain	Prospective longitudinal study of 482 patients attending a back pain triage clinic with night pain	A total of 213 (44%) patients had night pain, with 90 having pain every night. No serious pathology was identified. The presence of night is not a specific sign to detect serious pathologies

(*continued*)

Table 1. Continued.

Author	Objective	Methods	Results
Henschke 2009 [50]	To determine the prevalence of serious pathology in patients with acute LBP, and to evaluate their diagnostic accuracy	Cohort study of 1172 consecutive patients receiving primary care for acute LBP	Only 3 of the red flags for fracture were informative: corticosteroids use, major trauma, and age > 70 years. Clinical judgement had a very good Sp. The combination of multiple RFs increase the diagnostic accuracy
Holmes 2003 [51]	To analyse if clinical screening criteria can identify all patients with TL spine injuries	Prospective, observational cohort study of 2404 patients undergoing TL spine radiographs following blunt trauma	RFs: complaints of TL spine pain; TL spine tenderness on midline palpation; decreased level of consciousness; abnormal peripheral neurologic examination; distracting painful injury; evidence of intoxication with ethanol or drugs
Hsu 2003 [52]	To determine a clinical diagnostic pathway for the imaging of the thoracolumbar (TL) spine	Retrospective study of 200 traumatic patients	RFs for TL fracture are: back pain/midline, palpable midline step, back bruising, abnormal neurological signs, and history of trauma
Karlaftis 2013 [53]	To determine clinical characteristics that could identify GERD in patients with NCCP	Observational study of 52 patients with NCCP	RFs: typical reflux symptoms, postprandial CP, and use of anti-reflux drugs for pain relief
Kempthorne 2009 [54]	To define the presentation, findings and prognosis of extradural spinal abscesses (ESA)	Retrospective study of 42 patients diagnosed with ESA	RFs: severe back pain, not relieved by rest or sleep, patient's clinical history, and examination findings (nature and duration of their back pain)
Ko 2012 [55]	To examine GERD in young patients with NCCP and to evaluate their symptomatic characteristics	Observational study of 118 patients with NCCP	RFs: heartburn and acid regurgitation. In young NCCP patients, the prevalence of GERD was relatively low compared to average-aged
Milner 2002 [56]	To evaluate typical and atypical symptoms to detect ACS	Observational study of 246 women and 276 men seen in the ED with symptoms suggestive of ACS	RFs: CP, discomfort, dyspnoea, diaphoresis, upper limbs pain, diaphoresis (women), dizziness, or faintness (men)
Premkumar 2018 [28]	To examine the effectiveness of RF questions as a screening tool	Retrospective observational study of 9940 patients with LBP	RFs for vertebral fracture: age > 50/70 years, history of recent trauma; RF for malignancy: unexplained weight loss, history of cancer. RFs for infection: fever, chills and sweating, recent infection. RFs for CES: loss of bladder or bowel control. The combination of multiple RFs increase the diagnostic accuracy data
Punukollu 2005 [57]	To evaluate the clinical characteristics and outcome of acute PE in elderly patients	Observational study of 136 patients with a confirmed diagnosis of acute PE	RFs: shortness of breath, pleuritic CP, syncope, active cancer, immobilisation, and tachycardia
Raison 2014 [58]	To assess the effectiveness of RF used in the ED to identify spinal cord compression and CES	Retrospective cohort study of 206 patients with back pain attending the ED	RFs: saddle anaesthesia and bladder or bowel dysfunction. The combination of the RFs increase the diagnostic accuracy
Reuter 2019 [59]	To screen for ACS in male and female with CP	Prospective cohort study of 3727 patients whom called the Emergency Medical Communication Center (EMCC) because of non-traumatic CP	ACS diagnosed in 647 patients (33%). RF in male population: age, tobacco use, severe and permanent pain; retrosternal, breathing non-related and radiating pain; and additional symptoms (AUC: 0.76) RF in female: age ≥ 60 years, personal history of coronary artery disease, and breathing non-related and radiating pain (AUC: 0.79). Only the accuracy of the male model was validated in a validation dataset
Roman 2010 [60]	To identify clinical characteristics associated with a diagnosis of osteoporotic vertebral compression fracture (OVCF)	Retrospective, cross-sectional study of 1448 consecutive patients seen at a spine surgery centre	RFs for OVCF: age > 52 years; BMI < 22; female gender; no gait abnormality; does not exercise regularly; sitting decreases pain; concomitant osteoarthritis; no leg or buttock pain. Combining more RFs there is a better diagnostic accuracy with a post-test probability increasing up to 20.4%
Sánchez 2007 [61]	To establish a triage flowchart to rule out ACS	Prospective observational study of 1000 consecutive patients with CP on an ED	Predictor variables of not having an ACS: age < 40 years, absence of diabetes, no previously known coronary artery disease, no oppressive pain, and no retrosternal pain
Schillinger 2004 [62]	To investigate the predictive value of MI atypical characteristics for the exclusion of acute or subacute coronary events	Prospective study of 1288 consecutive patients presenting with acute CP at a non-trauma ED	RFs: left-sided or substernal chest pain defined as squeezing or crushing, burning; radiation of CP to the left or both arms, neck, or back; exercise-induced , undulating, relieved with rest or nitroglycerine; dyspnoea, nausea; diaphoresis; personal or family history of cardiac disorders; smoking ; obesity; hypertension; diabetes; hyperlipidaemia
Shaw 2020 [63]	To determine the frequency of RF in patients with serious pathologies	Retrospective observational study of 1000 patients with back pain to the ED	RF with LR+ >5 for spinal/non spinal pathologies: fever; tuberculosis history; known nephrolithiasis/ abdominal aortic aneurysm; unexplained weight loss; writhing pain; urinary symptoms; flank pain. RF with LR+ >5 for spinal pathologies: saddle anaesthesia; tuberculosis history; intravenous drug use; acute onset urinary retention; anal tone loss
Park 2015 [64]			RFs: obesity, smoking, and diabetes

(*continued*)

Table 1. Continued.

Author	Objective	Methods	Results
	To evaluate the risk factors and prevalence of gastroesophageal reflux diseases (GERD) in NNCP	Retrospective non-interventional observational study of 904 consecutive patients with NCCP	
Svensson 2003 [65]	To evaluate factors which predict the development of ACS or AMI	Prospective observational study of 538 patients who called for an ambulance due to CP and suspected ACS	RFs: history of myocardial infarction or angina pectoris or hypertension ECG changes and the elevation of serum markers (CKMB)
Thirugana-sambanda-moorthy 2014 [66]	To identify risk factors associated with serious pathology	Observational study of 329 patients with nontraumatic LBP	RFs: anticoagulant use, decreased sensation on physical examination, pain that is worse at night and pain that persists despite appropriate treatment
Timmons 2003 [67]	To compare the clinical presentation of younger and older patients with acute pulmonary embolism	Retrospective study of 60 consecutive patients with PE	RFs: collapse, dyspnoea, cough, haemoptysis, palpitations, hypotension, tachycardia, tachypnoea, fever, cyanosis, abnormal lower limb examination, and pleural rub
Tsiang 2019 [68]	To quantify the sensitivity and specificity of patient-reported RF	Retrospective case-control study of 500 patients with LBP	Patient-reported RFs malignancy: history of cancer; fracture: corticosteroids use, history of trauma; infection: fever; for CES: bladder or bowel dysfunction. Provider-reported red flags. Malignancy: history of cancer; fracture: osteoporosis and history of trauma; infection: fever; CES: bladder dysfunction and lower limbs weakness.
Van den Bosch 2004 [69]	To evaluate the prevalence of abnormalities findings on radiographic by age	Retrospective study of 2007 radiographic reports of patients referred with low back pain for lumbar spine radiography	The prevalence of reported degenerative changes in case of fracture, malignancy, and infection increased with age in patients > 55 years and older (62%)
Wells 2000 [70]	To determine a scoring system, that combined with D-dimer results, exclude PE	Prospective cohort study of 1211 patients	RFs: symptoms of DVT, no alternate diagnosis more likely than PE, heart rate > 100 beats per minute, immobilisation or surgery in past 4 weeks, previous objectively diagnosed DVT or PE, haemoptysis and malignancy

LR+: likelihood ratio positive; LR–: likelihood ratio negative; Sn: sensitivity; Sp: specificity; RFs: Red Flags; LBP: low back pain; CP: chest pain; TLP: thoracolumbar pain; ED: Emergency Department; CES: Cauda Equina Syndrome; NRS: Numeric Rating Scale; BMI: body mass index; AMI: acute myocardial infarction; ACS: acute coronary syndrome; NCCP: non cardiac chest pain; GERD: gastroesophageal reflux diseases; PE: pulmonary embolism.

The main RFs valuable to detect an infection were: fever; sweating; chills; recent infection; night pain; unexplainable weight loss; neurological signs; and constant pain not relieved by rest. Six out of eight studies (75%) reported the presence of fever as the most important RF, showing a good Sp (97–99%) [28,68]. Diagnostic accuracy was evaluated in four papers [28,63,68,69] and data are presented in Table 5. According to Premkumar et al. [28], a recent history of infection determines a 10.2% probability of having a spinal infection (LR+ 9.31); the presence of weight loss would increase the probability of 3%; and the history of fever, chills or sweating, if present alone, would lead to an increase of 2%. The same authors [28] assessed that combination of fever, chills, and sweating, associated with a recent infection, increases the probability to detect spinal infection (LR+ 13.15/LR– 0.93), with post-test probability of 13.8%. Night-awaking caused by pain is a false-positive in more than 96% of cases, when no other RFs are present [28]. Henschke et al. [50] highlighted that many RFs have a false- positive prevalence up to 10% such as: systemically unwell; constant pain; neurological signs; and recent infection.

Cauda equina syndrome

CES, along with fractures, malignancy, and infections, is considered the most frequent serious pathology of TLP, with a prevalence lower than 1% [28,50]. The most helpful RFs to recognise CES were: bladder or bowel dysfunction; neurological signs; saddle anaesthesia and loss of anal sphincter tone.

The associated presence of bladder or bowel dysfunction was underpinned by 100% of the studies increasing the probability of CES by 1.2% (Sn 8.3%; Sp 97.2%; LR+ 3.0; LR– 0.94) [28]. Raison and coworkers [58] reported as main RFs saddle anaesthesia and bladder/bowel dysfunction. Also they observed that combination of three RFs determines an improvement of the diagnostic accuracy (LR+ 3.46/Sn: 0.27/Sp: 0.92). Tsiang et al. [68] analysed dysregulation of bladder or bowel, identifying an Sn of 50% and a Sp of 86.5%. The diagnostic accuracy is reported in five papers [28,58,63,68,69] and presented in Table 6.

Cardiovascular disorders

Patients with CP can reach more than 5% of ED visits and early identification is essential to avoid life-threatening disorders such as acute coronary syndrome and acute myocardial ischemia [61].

The main RFs useful to identify cardiovascular disorders were: exertional pain; personal or family history of cardiac diseases; diaphoresis; dyspnoea; sweating; hypotension; pain radiating to upper limbs, neck, or back, and CP described as squeezing, burning, oppressive, crashing, severe, permanent, or retrosternal. The main RF is the radiated pain to upper limbs which were included by 100% of the studies, showing high potential to detect a serious heart disorder. Only three papers [27,48,62] included diagnostic accuracy data, and the latter are reported in Table 7. Schillinger et al. [62] evaluated the combination of multiple RFs on increasing the probability to identify cardiovascular disorders, but their diagnostic accuracy data were weak (LR+ 1.15–1.85; LR– 1.05–3.0). Evaluating associated symptoms, type, duration, onset, and localisation of pain, Aydin et al. [35] created a CP score in order to determine whether or not the pain was caused by ischaemic heart disease. An overall score > 4.5/8 showed good diagnostic accuracy data (Sn: 90%/Sp: 95.83%).

Table 2. Risk of bias of the included studies.

Included studies	Selection	Comparability	Outcome	Total score (out of 9)
Ahad 2015 [33]	3	1	3	7
Albarran 2002 [34]	3	1	2	6
Aydin 2019 [35]	3	2	3	8
Bernard Bagattini 2004 [36]	3	2	2	7
Balasubramanian 2010 [37]	3	1	2	6
Body 2010 [27]	3	2	3	8
Broderick 2018 [38]	2	2	3	7
Cook 2011 [39]	3	1	2	6
Courtney 2010 [40]	3	2	1	6
de Schepper 2015 [41]	3	2	2	7
Domen 2009 [42]	2	2	2	6
Donner-Banzhoff 2006 [43]	3	2	3	8
Dugas 2011 [44]	3	2	2	7
Enthoven 2016 [45]	3	2	3	8
Everden 2018 [46]	2	2	3	7
Gesuete 2020 [47]	3	2	3	8
Goodacre 2002 [48]	2	1	3	6
Harding 2005 [49]	3	1	3	7
Henschke 2009 [50]	4	2	3	9
Holmes 2003 [51]	3	1	1	5
Hsu 2003 [52]	2	2	1	5
Karlaftis 2013 [53]	2	2	1	5
Kempthorne 2009 [54]	3	1	2	6
Ko 2012 [55]	2	2	1	5
Milner 2002 [56]	2	1	2	5
Park SH 2015 [64]	3	2	3	8
Premkumar 2018 [28]	4	2	3	9
Punukollu 2005 [57]	2	1	2	5
Raison 2014 [58]	3	2	2	7
Reuter 2019 [59]	3	2	2	7
Roman 2010 [60]	3	2	2	7
Sanchez 2007 [61]	3	2	2	7
Schillinger 2004 [62]	3	2	2	7
Shaw 2020 [63]	4	2	3	9
Svensson 2003 [65]	2	1	1	4
Thiruganasambandamoorthy 2014 [66]	3	2	2	7
Timmons 2003 [67]	3	1	2	6
Tsiang 2019 [68]	4	2	2	8
Van den Bosch 2004 [69]	3	1	1	5
Wells 2000 [70]	3	2	1	6

Newcastle-Ottawa Quality Assessment for observational studies: good quality: 3 or 4 points in selection domain AND 1 or 2 points in comparability domain AND 2 or 3 points in outcome/exposure domain; fair quality: 2 stars in selection domain AND 1 or 2 points in comparability domain AND 2 or 3 points in outcome/exposure domain; poor quality: 0 or 1 point in selection domain OR 0 point in comparability domain OR 0 or 1 point in outcome/exposure domain.
Shades are representing the score of application of risk of bias for each study.

Pulmonary disorders

Pulmonary disorders had a prevalence of 10% of CP representing the most frequent and potentially life-threatening diseases. Notably, pulmonary embolism (PE) and deep venous thrombosis remain a diagnostic challenge due to low sensitivity and specificity of their signs and symptoms [36].

The main RFs able to recognise pulmonary disorders were: history of pulmonary pathologies; recent surgery or immobilisation; active cancer; female gender; haemoptysis; cough; syncope; dyspnoea; tachycardia; leg swelling; oxygen saturation less than 95%; palpitations; fever; tachypnoea; cyanosis; abnormal lower limb examination; shortness of breath; age >40; pleural rub; hypoxemia; and pleuritic or substernal CP. Patients reported CP (72%) and dyspnoea (70%) as the most common presenting symptoms [40]. Only Wells et al. [70] evaluated the diagnostic accuracy data, elaborating a scoring system to detect PE: the combination of multiple RFs show an LR + value between 0.13 and 6.75.

Gastroesophageal disorders

Non-cardiac chest pain (NCCP) is defined as a recurrent retrosternal angina-like pain in patients with normal cardiac evaluations with a prevalence between 25% and 35% [64]. Gastroesophageal disorders are the most common causes of NCCP, followed by chest wall syndromes and psychosomatic disorders [55]. It is a benign condition without an impact on mortality, though it decreases quality of life. Therefore, screening for the referral process may be required [53]. The most common pathologies are gastroesophageal reflux diseases, peptic ulcers diseases, and gastritis [64].

The main RFs suitable to identify gastroesophageal disorders were: typical reflux symptoms; postprandial CP; use of anti-reflux drugs for pain relief; obesity; smoking; diabetes; heartburn; and acid regurgitation. None study evaluated the diagnostic accuracy data or the combination of multiple RFs.

Inflammatory disorders

Only Henschke et al. [50] analysed inflammatory disorders as a cause of CP, which has a prevalence of 0.2% (0.1–0.6). The RFs identified by the authors were: gradual onset before age 40 years; tired sleep without relief; insidious onset; systemically unwell; constant, progressive, nonmechanical pain; morning back stiffness lasting >30 min; peripheral joint involvement; persisting limitation of spinal movements in all directions; iritis; skin rashes (psoriasis); colitis; urethral discharge; family history of arthritis or osteoporosis; pain

Table 3. Fracture and diagnostic accuracy.

Author	Red Flag	Sn	Sp	LR+	LR–
Enthoven 2016 [45]	Age > 75	45% (0.28–0.62)	85% (0.82–0.88)	3.1 (2.0–4.7)	0.6 (0.5–0.9)
Enthoven 2016 [45]	Female gender	67% (51–83%)	41% (37–44%)	1.1 (0.9–1.4)	0.8 (0.5–1.3)
Enthoven 2016 [45]	Osteoarthritis	16% (03–28%)	69% (65–72%)	0.50 (0.2–1.1)	1.2 (1.0–1.4)
Enthoven 2016 [45]	Corticosteroids use	18% (05–31%)	93% (91–95%)	2.5 (1.1–5.3)	0.90 (0.8–1.0)
Enthoven 2016 [45]	History of trauma	21% (07–35%)	97% (95–98%)	6.2 (2.8–13.5)	0.80 (0.5–1.3)
Enthoven 2016 [45]	Sudden decrease in height	9% (01–19%)	97% (95–98%)	2.9 (0.9–9.4)	0.90 (0.8–1.0)
Enthoven 2016 [45]	Percussion tenderness of the spine	21% (07–35%)	81% (78–84%)	1.1 (0.6–2.2)	1.0 (0.8–1.2)
Enthoven 2016 [45]	Severe disability	30% (14–46%)	87% (84–90%)	2.3 (1.3–4.2)	0.8 (0.6–1.0)
Enthoven 2016 [45]	Numeric Rating Scale > 7	67% (51–83%)	63% (59–67%)	1.8 (1.4–2.3)	0.50 (0.3–0.9)
Enthoven 2016 [45]	Painful injury	30% (15–46%)	64% (60–68%)	0.8 (0.5–1.4)	1.1 (0.9–1.4)
Enthoven 2016 [45]	Thoracic back pain	42% (26–59%)	78% (75–81%)	1.9 (1.3–3.0)	0.7 (0.5–1.0)
Henschke 2009 [50]	Age > 70	50%	96%	11.0 (4.65–19.48)	0.52 (0.23–0.82)
Henschke 2009 [50]	Corticosteroids use	25%	100%	48.5 (11.62–165.22)	0.75 (0.41–0.93)
Henschke 2009 [50]	History of trauma	25%	98%	10.0 (2.76–26.36)	0.77 (0.42–0.95)
Hsu 2003 [52]	Midline tenderness	62.1%	91.5%	–	–
Hsu 2003 [52]	Palpable midline step	13.8%	100%	–	–
Hsu 2003 [52]	Back bruising	6.9%	98.6%	–	–
Hsu 2003 [52]	Abnormal neurological signs	41.4%	95.8%	–	–
Premkumar 2018 [28]	Age > 50	74%	32.9%	1.1 (1.05–1.16)	0.79 (0.69–0.91)
Premkumar 2018 [28]	Age > 70	39%	80%	1.55 (1.36–1.76)	0.86 (0.82–0.91)
Premkumar 2018 [28]	History of trauma	24.7%	88.6%	2.17 (1.86–2.54)	0.84 (0.81–0.89)
Roman 2010 [60]	Age > 50	95% (83–95%)	39% (38–40%)	1.5 (1.3–1.5)	0.14 (0.03–0.45)
Roman 2010 [60]	Body mass index < 22	38% (24–55%)	83% (82–84%)	2.3 (1.4–3.4)	0.74 (0.54–0.91)
Roman 2010 [60]	Female gender	90% (76–96%)	41% (41–42%)	1.5 (1.3–1.6)	0.26 (0.10–0.60)
Roman 2010 [60]	Gait abnormality	66% (50–79%)	23% (22–23%)	0.86 (0.65–1–02)	1.5 (0.91–2.2)
Roman 2010 [60]	No regular exercise	81% (65–91%)	44% (43–45%)	1.5 (1.2–1.6)	0.43 (0.20–0.80)
Roman 2010 [60]	Sitting decrease pain	29% (27–32%)	81% (79–83%)	1.6 (1.2–1.9)	0.87 (0.82–0.92)
Roman 2010 [60]	Osteoarthritis	50% (35–65%)	52% (51–52%)	1.1 (0.70–1.4)	0.97 (0.67–1.3)
Roman 2010 [60]	No leg pain	31% (16–49%)	86% (85–87%)	2.2 (1.2–3.6)	0.81 (0.58–0.97)
Tsiang 2019 [68]	Corticosteroids patient reported	64.8%	58.5%	–	–
Tsiang 2019 [68]	History of trauma	81.1%	79.1%	–	–
Tsiang 2019 [68]	Trauma patient reported	64.8%	58.5%	–	–
Tsiang 2019 [68]	Osteoporosis	81.1%	79.1%	–	–
Van den Bosch 2004 [69]	Age > 55	–	–	1.5–8	–

Sn: sensibility; Sp: specificity; LR+: likelihood ratio positive; LR–: likelihood ratio negative.

improvement with exercise; and clinically diagnosed inflammatory disorders. No diagnostic accuracy data were provided and many RFs have a false-positive prevalence rate higher than 10% such as tired sleep without relief; insidious onset; morning back stiffness lasting >30 min; family history of arthritis or osteoporosis; and pain improvement with exercise.

Serious spinal pathologies

Shaw et al. [63] evaluated the prevalence of RFs in patients presenting to the ED with self-reported back pain. The RFs screening has been performed to identify an increased risk of serious spinal conditions not related to a musculoskeletal cause. They found 39 RFs, of which only 36 had a complete reporting of Sn and Sp values. For the complete list of them, see Table 8.

Discussion

Summary of evidence

To the best of the authors' knowledge, this is the first systematic review concerning RF aimed to guide healthcare professionals during the evaluation of patients with TLP. In this regard, it represents a new perspective due to (a) the absence of systematic reviews related to CP or LBP and the concurrent presence of serious pathologies, with the exception of fracture, malignancy and CES [13,14,17,21–23,77]; and (b) the authors' proposal of TLP as a new label for such patients.

Our systematic review provides the first list of RFs with highly reliable Sn and Sp values (Table 9) [78], useful for the screening of patients with TLP presenting with non-musculoskeletal signs and symptoms related to cardiovascular, pulmonary or gastroesophageal disorders [27,28,33–37,39,40,42–62,64–70,76]. In case of the latter, the assessment of the patients remains a challenge for healthcare professionals even due to the high risk of misdiagnosis [62] and the paucity of primary studies on the topic. The sparse evidence on the screening process is related to only four published papers [27,48,62,70] which evaluated the diagnostic accuracy data of RFs.

Useful screening tools should provide high values LR– and Sn (e.g., LR– <1.0 and Sn >85%), nevertheless in agreement with Premkumar et al. [28] almost no RF has an Sn greater than 75% and most of them have values below 60% [28]. According to the results emerged from our findings, only the 45% (*n*: 18/40) of the included studies evaluated the diagnostic accuracy of the identified RFs. Also, at least one RFs were found in the majority of patients without serious pathology, showing a very high false-positive rate [50]. To increase the diagnostic accuracy levels of RFs, it is recommended to combine them during the assessment process, in order to help the healthcare professionals to screen specific pathologies [28,50,55]. However, this strategy was found to be applied only in 12 out of 40 (about the 30%) of the included studies, which investigated different sets of RFs for the same pathologies.

The most frequent life-threatening pathologies were cardiovascular disorders and three of the included studies [27,48,62] identified a total of eight RFs, with poor diagnostic accuracy throughout the assessment process of such conditions. Radiated pain to the upper limbs, exertional and retrosternal pain were the most frequent and showed a good probability to rule-in a serious heart disorder [27,48]. However, the higher value of Sn was 55.6%, with an average below 30% [27,48]. Particularly, Aydin et al. [35] have created a score to examine the relationship between CP and ischaemic heart disease, that demonstrated good

Table 4. Spinal malignancy and diagnostic accuracy.

Author	Red Flag	Sn	Sp	LR+	LR–
Premkumar 2018 [28]	Age > 50	71.7%	32.6%	1.06 (0.96–1.17)	0.87 (0.68–1.11)
Van den Bosch 2004 [69]	Age > 55		–	1.5–8.0	–
Premkumar 2018 [28]	Age > 70	22.6%	79.5%	1.1 (0.82–1.47)	0.97 (0.9–1.06)
Premkumar 2018 [28]	Night pain	55.4%	41.8%	0.85 (0.83–1.1)	1.07 (0.9–1.27)
Premkumar 2018 [28]	Unexplained weight loss	8.2%	95.6%	1.87 (1.1–3.17)	0.96 (0.92–1.01)
Premkumar 2018 [28]	History of malignancy	32%	95.6%	7.25 (5.65–9.3)	0.71 (0.64–0.79)
Tsiang 2019 [68]	History of malignancy	91.7%	77.8%	–	–
Tsiang 2019 [68]	Malignancy patient reported	75%	78.7%	–	–
Cook 2011 [39]	No pain during flexion	67% (55–77)	41% (40–42)	1.1 (0.9–1.3)	0.80 (0.56–1.1)
Cook 2011 [39]	No pain during extension	65% (54–75)	50% (49–51)	1.3 (1.1–1.5)	0.70 (0.49–0.94)
Cook 2011 [39]	No pain during right side flexion	96% (88–98)	4% (04–05)	0.94 (0.85–1.0)	1.0 (0.92–1.0)
Cook 2011 [39]	No pain during left side flexion	96% (88–98)	4% (03–04)	0.99 (0.91–1.0)	1.1 (0.37–3.2)
Donner-Banzhoff 2006 [43]	Clinical judgment	50%	83%	2.95	0.6

Sn: sensibility; Sp: specificity; LR+: likelihood ratio positive; LR–: likelihood ratio negative.

Table 5. Spinal infection and diagnostic accuracy.

Author	Red Flag	Sn	Sp	LR+	LR–
Premkumar 2018 [28]	Fever	11.7%	93.2%	1.71 (1.04–2.81)	0.95 (0.89–1.01)
Tsiang 2019 [68]	Fever	12.5%	99.6%	–	–
Tsiang 2019 [68]	Fever patient reported	25%	97.6%	–	–
Premkumar 2018 [28]	Chills	11.7%	93.2%	1.71 (1.04–2.81)	0.95 (0.89–1.01)
Premkumar 2018 [28]	Night pain	57.5%	41.8%	0.99	1.02
Premkumar 2018 [28]	Sweating	11.7%	93.2%	1.71 (1.04–2.81)	0.95 (0.89–1.01)
Premkumar 2018 [28]	Persistent night sweating	17.5%	86.1%	1.26 (0.85–1.86)	0.96 (0.88–1.04)
Premkumar 2018 [28]	Recent infection	24.2%	97.4%	9.31 (6.63–13.07)	0.78 (0.7–0.86)
Van den Bosch 2004 [69]	Age > 55	–	–	1.5–8.0	–

Sn: sensibility; Sp: specificity; LR+: likelihood ratio positive; LR–: likelihood ratio negative.

Table 6. CES and diagnostic accuracy.

Author	Red Flag	Sn	Sp	LR+	LR–
Premkumar 2018 [28]	Bladder dysfunction	22.2%	90.4%	2.31 (1.25–4.27)	0.86 (0.72–1.03)
Raison 2014 [58]	Bladder dysfunction	65% (44–82%)	73% (66–80%)	2.45	–
Tsiang 2019 [68]	Bladder dysfunction	100%	76.9%	–	–
Tsiang 2019 [68]	Bladder dysfunction patient rep	50%	86.5%	–	–
Premkumar 2018 [28]	Bowel dysfunction	13.9%	95%	2.78 (1.23–6.3)	0.91 (0.8–1.03)
Raison 2014 [58]	Bowel dysfunction	65% (44–82%)	73% (66–80%)	2.45	–
Tsiang 2019 [68]	Bowel dysfunction patient rep	50%	86.5%	–	–
Raison 2014 [58]	Saddle dysfunction	27% (12–48%)	87% (81–92%)	2.11	–
Tsiang 2019 [68]	Lower limbs weakness	100%	76.9%	–	–
Van den Bosch 2004 [69]	Age > 55	–	–	1.5–8.0	–

Sn: sensibility; Sp: specificity; LR+: likelihood ratio positive; LR–: likelihood ratio negative.

Table 7. Cardiovascular disorders and diagnostic accuracy.

Author	Red Flag	Sn	Sp	LR+	LR–
Goodacre 2002 [48]	Pain radiating to upper limbs – AMI	38.2% (23.9–55.0%)	90.6% (88.5–92.4%)	4.07 (2.53–6.54)	0.68 (0.52–0.89)
Goodacre 2002 [48]	Pain radiating to upper limbs – ACS	55.6% (44.7–65.9%)	65.6% (62.3–68.8%)	1.62 (1.30–2.01)	0.68 (0.53, 0.87)
Body 2010 [27]	Pain radiating to upper limbs	13.5% (8.2–17.5%)	94.8% (93.2–96.8%)	2.58 (1.55–4.29)	0.91 (0.87–0.97)
Body 2010 [27]	Pain radiated to the right arm	18.9% (12.3–22.8%)	91.8% (89.8–94.3%)	2.31 (1.47–3.34)	0.88 (0.84–0.96)
Goodacre 2002 [48]	Exertional pain – AMI	35.3% (21.5–52.1%)	85% (82.4–87.2%)	2.35 (1.45–3.80)	0.76 (0.59–0.98)
Goodacre 2002 [48]	Exertional pain – ACS	29.6% (20.8–40.3%)	85.6% (83.0–87.8%)	2.06 (1.41–2.99)	0.82 (0.71–0.95)
Goodacre 2002 [48]	Absence of chest wall syndrome	91.7% (74.2–97.7%)	27.8% (24.6–31.2%)	1.27 (1.12–1.44)	0.30 (0.08–1.14)
Body 2010 [27]	Vomiting	16.2% (9.8–19.7%)	94.8% (93.2–96.8%)	3.09 (1.82–4.85)	0.88 (0.85–0.95)
Body 2010 [27]	Sweating	59.5% (49.0–62.7%)	54.3% (50.4–58.6%)	1.30 (1.06–1.43)	0.75 (0.68–0.96)
Body 2010 [27]	Sweating observed	36.5% (22.0–34.5%)	94.3% (92.4–96.2%)	6.39 (3.42–7.63)	0.67 (0.70–0.83)
Body 2010 [27]	Central CP	85.1%	34.1%	1.29	0.44
Body 2010 [27]	Pain left anterior	11.5%	68.2%	0.36	1.30
Body 2010 [27]	Duration >1 h	77% (65.0–77.5%)	44.9% (41.2–49.4%)	1.40 (1.17–1.46)	0.51 (0.50–0.79)
Body 2010 [27]	Hypotension	6.8% (4.4–12.0%)	97.7% (97.1–99.3%)	2.92 (2.21–10.98)	0.95 (0.90–0.98)
Body 2010 [27]	Basal crackles	16.2% (11.8–22.3%)	90.6% (88.9–93.6%)	1.72 (1.30–2.90)	0.92 (0.85–0.97)
Body 2010 [27]	Acute ischaemic ECG changes	71% (51.3–65.0%)	81.3% (79.1–85.4%)	3.80 (2.69–4.08)	0.36 (0.43–0.60)
Body 2010 [27]	Similar to previous ischemia	22.3%	69.4	0.73	1.12

Sn: sensibility; Sp: specificity; LR+: likelihood ratio positive; LR–: likelihood ratio negative; AMI: acute myocardial infarction; ACS: acute coronary syndrome; ECG: electrocardiogram.

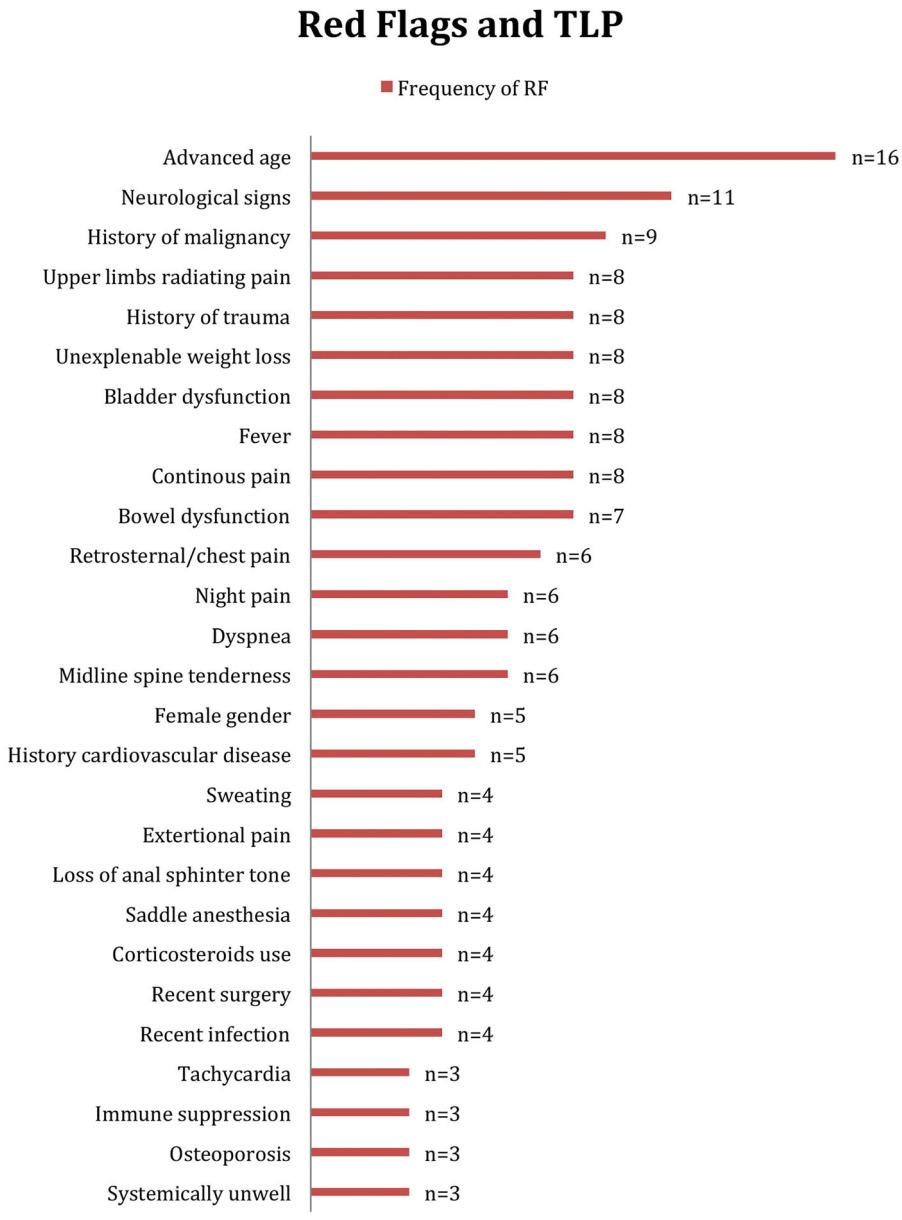

Figure 2. Frequency of RFs in TLP.

diagnostic accuracy and can be easily applied by healthcare professionals. Finally, the association of multiple RFs was evaluated in only one study, without any reported increase of diagnostic accuracy [62].

Concerning pulmonary disorders, several RFs from three studies were identified [36,40,70]. Among them, only a scoring system related to the screening for PE has confirmed a moderate likelihood to diagnose the pathology [70]. Relating to gastroesophageal and inflammatory disorders the diagnostic accuracy data were unavailable; therefore, we listed the identified RFs, without reporting any further information [50,55,64]. Thus, their applicability in healthcare settings is limited.

Vertebral fracture is the most analysed pathology and its relative RFs include advanced age; history of trauma; corticosteroids use; and female gender. In details, the use of corticosteroids has reported values of Sn ranging between 18% and 25%; therefore, it seems to be a nearly unreliable tool in clinical practice. Furthermore, the female gender usually is ordinarily just assessed as a risk factor and not a proper RF [79]. However, female gender showed a better diagnostic accuracy when combined with advanced age and history of trauma. History of trauma was not a necessary condition for having a fracture since only Tsiang et al. [68] provided an Sn of 81.1%, contrary to what was claimed to other three papers, which showed values between 21% and 25% [28,45,51].

With regard to vertebral fractures and malignancies, our results are consistent with current systematic reviews [13,14,17,21–23], identifying approximately the same set of RFs. Notably, analysing the diagnostic accuracy of each RF, low levels were identified for the most widespread signs and symptoms used in clinical settings. Thus, their reliability should be reconsidered if used as a self-contained screening tool.

All identifiable hallmarks to detect fractures during the physical examination indicated a low clinical utility; nevertheless, the combination of multiple RFs throughout the screening process increased up to 52% the post-test probability (Sn 88%; Sp 95–100%) [45,50].

Table 8. Serious spinal pathologies and diagnostic accuracy.

Author	Red Flag	Sn	Sp	LR+	LR–
Shaw 2020 [63]	Saddle anaesthesia	9.1%	99.2%	11 (3.1–39.6)	0.9 (0.8–1.0)
Shaw 2020 [63]	Tuberculosis history	3.0	99.7	9.8 (1.0–91.4)	1.0 (0.9–1.0)
Shaw 2020 [63]	Intravenous drug use	12.1	98.2	6.9 (2.5–19.4)	0.9 (0.8–1.0)
Shaw 2020 [63]	Acute urinary retention or overflow incontinence	15.2	97.6	6.4 (2.6–15.7)	0.9 (0.8–0.1)
Shaw 2020 [63]	Anal tone loss or faecal incontinence	9.1	98.6	6.3 (1.9–20.8)	0.9 (0.8–1)
Shaw 2020 [63]	Recent infection	18.2	96.0	4.5 (2.1–9.9)	0.9 (0.7–1)
Shaw 2020 [63]	Inflammatory arthritis or osteoporosis fracture	30.3	93.4	4.6 (2.6–8.1)	0.8 (0.6–0.9)
Shaw 2020 [63]	Constant progressive non-mechanical pain	3.0	99.3	4.2 (0.5–33.1)	1.0 (0.9–1)
Shaw 2020 [63]	Immune suppression	12.1	96.0	3.0 (1.1–7.9)	0.9 (0.8–1)
Shaw 2020 [63]	Prolonged use of corticosteroids	12.1	96.9	3.9 (1.5–10.5)	0.9 (0.8–1)
Shaw 2020 [63]	History of trauma (major in young/minor in elderly)	27.3	90.7	2.9 (1.6–5.3)	0.8 (0.7–1)
Shaw 2020 [63]	Urinary symptoms	27.3	88.7	2.4 (1.4–4.3)	0.8 (0.7–1)
Shaw 2020 [63]	Insidious onset	6.1	97.2	2.2 (0.5–8.8)	1.0 (0.9–1.1)
Shaw 2020 [63]	Anticoagulated	9.1	95.4	2.0 (0.6–6.0)	1.0 (0.9–1.1)
Shaw 2020 [63]	Fever	3.0	98.5	2.0 (0.3–14.4)	1.0 (0.9–1.1)
Shaw 2020 [63]	Central spine tenderness	18.2	90.7	2.0 (0.9–4.1)	0.9 (0.8–1.1)
Shaw 2020 [63]	History of cancer	15.2	92.1	1.9 (0.8–4.5)	0.9 (0.8–1.1)
Shaw 2020 [63]	Progressive motor weakness in legs or gait disturbances	3.0	98.4	1.8 (0.3–13.4)	1.0 (0.9–1.1)
Shaw 2020 [63]	Known nephrolithiasis or abdominal aortic aneurysm	6.1	95.6	1.4 (0.3–5.4)	1.0 (0.9–1.1)
Shaw 2020 [63]	Age $<$ 20 and over 55	3.0	97.5	1.2 (0.2–8.8)	1.0 (0.9–1.1)
Shaw 2020 [63]	Spinal instrumentation	3.0	97.2	1.1 (0.2–7.8)	1.0 (0.9–1.1)
Shaw 2020 [63]	Recent spinal procedure	3.0	97.3	1.1 (0.2–8.1)	1.0 (0.9–1.1)
Shaw 2020 [63]	Flank pain	9.1	87.1	0.7 (0.2–2.1)	1.0 (0.9–1.2)
Shaw 2020 [63]	Thoracic/chest/abdominal pain	3.0	94.1	0.5 (0.1–3.6)	1.0 (1.0–1.1)
Shaw 2020 [63]	Unexplained weight loss	1.1	99.9	9.2 (0.8–100.6)	1.0 (1.0–1.0)
Shaw 2020 [63]	Writhing in pain	5.0	99.3	6.9 (2.5–19.9)	1.0 (0.9–1.0)
Shaw 2020 [63]	Iritis, skin rashes (psoriasis), colitis, urethral discharge	1.7	99.6	4.6 (0.9–22.6)	1.0 (1.0–1.0)
Shaw 2020 [63]	Systemically unwell	62.0	84.2	2.8 (0.7–11.4)	1.0 (1.0–1.0)
Shaw 2020 [63]	Pregnancy	0.6	99.8	2.3 (0.2–25.2)	1.0 (1.0–1.0)
Shaw 2020 [63]	Age over 70 years	34.6	76.0	1.4 (1.1–1.8)	0.9 (0.8–1.0)
Shaw 2020 [63]	Sensory level (altered sensation trunk down)	0.6	97.9	0.3 (0.0–2.0)	1.0 (1.0–1.0)
Shaw 2020 [63]	Known spinal Paget's disease	0	99.9	0.0	1.0
Shaw 2020 [63]	Herpes zoster rash	1.1	100	–	1.0 (1.0–1.0)
Shaw 2020 [63]	Gradual onset before age 40	0	99.6	0.0	1.0
Shaw 2020 [63]	Tried bed rest but no relief	0	99.8	0.0	1.0
Shaw 2020 [63]	Morning back stiffness $\geq$ 30 minutes	0	99.8	0.0	1.0

Sn: sensibility; Sp: specificity; LR+: likelihood ratio positive; LR–: likelihood ratio negative.

In case of spinal infection, the presence of fever, sweat, chills, and recent infection resulted to be important screening RFs and showed high Sp values (93.2–99.6%), but a very low Sn (11.7–24.2%) [28,68]. Few authors [80–82] supported as important RFs for spinal infection the presence of unintentional weight loss; neurological signs; and constant pain not relieved by rest. However, no primary study has analysed their diagnostic accuracy. Besides, different authors identified RFs with high false-positive rate such as neurological signs; constant pain [50]; and the presence of night pain [28].

Performing a screening for malignancy, a history of prior malignancy is the RF with the best diagnostic accuracy [28]. Particularly, the latter and other circumstances as an unexplained weight loss; advanced age; and night pain represent the most suggestive RFs of malignancy [28,39–41,43,53,68,69]. However, the absence of one or two of such RFs did not significantly enhance the LR–. Indeed, the 64% of the subjects with spinal malignancy denied any RFs during the medical interview [28]. Additionally, the unexplained weight loss, investigated by a few narrative reviews [80,81], showed a very low Sn (8.3%) [28]. Combining a history of malignancy and unintentional weight loss increased the probability to rule-in the pathology up to 14.3% (LR+ 10.25) [28]. Sadly, although certain elements were reported as RFs have been shown poor reliability on this matter. In fact, night pain was reported by more than 55% of patients, but nonetheless it is a false positive in the 85% of patients [28,49]. Also, the age $<$20 years or $>$55 years displays a high false-positive rate.

Eight narrative reviews examined the presence of a hidden infection as a cause for the TLP symptoms of the patients [80–86]; nevertheless, the low methodological quality and the opinion-based statements have reduced their usefulness in clinical settings.

The screening process of CES has included the following RFs: bladder or bowel dysfunction; neurological signs; saddle anaesthesia; and loss of anal sphincter tone. Despite the loss of anal sphincter tone has been reported as one of the most frequently RF, none primary study evaluated its diagnostic accuracy. The occurrence of saddle anaesthesia was analysed in two studies and showed an Sn of 24% [58] and 9.1% [63], respectively. For this reason, its clinical utility is low [58]. Bladder or bowel dysfunction are cited by all the included papers [28,33,37,42,45,50,58,63,68] but the reported diagnostic accuracy is very conflicting (range of Sn between 13.9% and 100%). Particularly noteworthy, albeit it has proved unhelpful in the rule-out process (Sn 8.3–27%), that pooling of several RFs improved the probability to rule-in CES to remarkable values (Sp 92–97.2). Furthermore, only one systematic review [77] evaluated the diagnostic accuracy of RFs related to CES when compared to magnetic resonance imaging. Hence, in accordance with our results, the authors stated that RFs appear to be more specific than sensitive. Regarding the evaluation of RFs for serious spinal pathologies [63], although some of these have reported interesting diagnostic accuracy values (see Table 8), it must be highlighted that the authors did not give a clear association between each RFs and the concomitant

Table 9. Red Flags with more reliable diagnostic accuracy values.

Author	Red Flag	Sn	Sp	LR+	LR–
Fracture					
Enthoven 2016 [45]	Age > 75	45% (0.28–0.62)	85% (0.82–0.88)	3.1 (2.0–4.7)	0.6 (0.5–0.9)
	Corticosteroids use	18% (05–31%)	93% (91–95%)	2.5 (1.1–5.3)	0.90 (0.8–1.0)
	Corticosteroids use	18% (05–31%)	93% (91–95%)	2.5 (1.1–5.3)	0.90 (0.8–1.0)
	History of trauma	21% (07–35%)	97% (95–98%)	6.2 (2.8–13.5)	0.80 (0.5–1.3)
	Percussion tenderness of the spine	21% (07–35%)	81% (78–84%)	1.1 (0.6–2.2)	1.0 (0.8–1.2)
	Severe disability	30% (14–46%)	87% (84–90%)	2.3 (1.3–4.2)	0.8 (0.6–1.0)
	Sudden decrease in height	9% (01–19%)	97% (95–98%)	2.9 (0.9–9.4)	0.90 (0.8–1.0)
Henschke 2009 [50]	Age > 70	50%	96%	11.0 (4.65–19.48)	0.52 (0.23–0.82)
	Corticosteroids use	25%	100%	48.5 (11.62–165.22)	0.75 (0.41–0.93)
	History of trauma	25%	98%	10.0 (2.76–26.36)	0.77 (0.42–0.95)
Hsu 2003 [52]	Abnormal neurological signs	41.4%	95.8%	–	–
	Back bruising	6.9%	98.6%	–	–
	Midline tenderness	62.1%	91.5%	–	–
	Palpable midline step	13.8%	100%	–	–
Premkumar 2018 [28]	Age > 70	39%	80%	1.55 (1.36–1.76)	0.86 (0.82–0.91)
	History of trauma	24.7%	88.6%	2.17 (1.86–2.54)	0.84 (0.81–0.89)
Roman 2010 [60]	Age > 50	95% (83–95%)	39% (38–40%)	1.5 (1.3–1.5)	0.14 (0.03–0.45)
	Body mass index < 22	38% (24–55%)	83% (82–84%)	2.3 (1.4–3.4)	0.74 (0.54–0.91)
	Female gender	90% (76–96%)	41% (41–42%)	1.5 (1.3–1.6)	0.26 (0.10–0.60)
	Female gender	90% (76–96%)	41% (41–42%)	1.5 (1.3–1.6)	0.26 (0.10–0.60)
	No regular exercise	81% (65–91%)	44% (43–45%)	1.5 (1.2–1.6)	0.43 (0.20–0.80)
	Sitting decrease pain	29% (27–32%)	81% (79–83%)	1.6 (1.2–1.9)	0.87 (0.82–0.92)
Tsiang 2019 [68]	History of trauma	81.1%	79.1%	–	–
	Osteoporosis	81.1%	79.1%	–	–
Serious spinal pathology					
Shaw 2020 [63]	Acute urinary retention or overflow incontinence	15.2	97.6	6.4 (2.6–15.7)	0.9 (0.8–0.1)
	Age < 20 and over 55	3.0	97.5	1.2 (0.2–8.8)	1.0 (0.9–1.1)
	Age over 70 years	34.6	76.0	1.4 (1.1–1.8)	0.9 (0.8–1.0)
	Anal tone loss or faecal incontinence	9.1	98.6	6.3 (1.9–20.8)	0.9 (0.8–1)
	Anticoagulated	9.1	95.4	2.0 (0.6–6.0)	1.0 (0.9–1.1)
	Central spine tenderness	18.2	90.7	2.0 (0.9–4.1)	0.9 (0.8–1.1)
	Constant progressive non-mechanical pain	3.0	99.3	4.2 (0.5–33.1)	1.0 (0.9–1)
	Fever	3.0	98.5	2.0 (0.3–14.4)	1.0 (0.9–1.1)
	Flank pain	9.1	87.1	0.7 (0.2–2.1)	1.0 (0.9–1.2)
	Gradual onset before age 40	0	99.6	0.0	1.0
	Herpes zoster rash	1.1	100	–	1.0 (1.0–1.0)
	History of cancer	15.2	92.1	1.9 (0.8–4.5)	0.9 (0.8–1.1)
	History of trauma (major in young/minor in elderly)	27.3	90.7	2.9 (1.6–5.3)	0.8 (0.7–1)
	Immune suppression	12.1	96.0	3.0 (1.1–7.9)	0.9 (0.8–1)
	Inflammatory arthritis or osteoporosis fracture	30.3	93.4	4.6 (2.6–8.1)	0.8 (0.6–0.9)
	Insidious onset	6.1	97.2	2.2 (0.5–8.8)	1.0 (0.9–1.1)
	Intravenous drug use	12.1	98.2	6.9 (2.5–19.4)	0.9 (0.8–1.0)
	Iritis, skin rashes (psoriasis), colitis, urethral discharge	1.7	99.6	4.6 (0.9–22.6)	1.0 (1.0–1.0)
	Known nephrolithiasis or abdominal aortic aneurysm	6.1	95.6	1.4 (0.3–5.4)	1.0 (0.9–1.1)
	Known spinal Paget's disease	0	99.9	0.0	1.0
	Morning back stiffness ≥ 30 minutes	0	99.8	0.0	1.0
	Pregnancy	0.6	99.8	2.3 (0.2–25.2)	1.0 (1.0–1.0)
	Progressive motor weakness in legs or gait disturbances	3.0	98.4	1.8 (0.3–13.4)	1.0 (0.9–1.1)
	Prolonged use of corticosteroids	12.1	96.9	3.9 (1.5–10.5)	0.9 (0.8–1)
	Recent infection	18.2	96.0	4.5 (2.1–9.9)	0.9 (0.7–1)
	Recent spinal procedure	3.0	97.3	1.1 (0.2–8.1)	1.0 (0.9–1.1)
	Saddle anaesthesia	9.1%	99.2%	11 (3.1–39.6)	0.9 (0.8–1.0)
	Sensory level (altered sensation trunk down)	0.6	97.9	0.3 (0.0–2.0)	1.0 (1.0–1.0)
	Spinal instrumentation	3.0	97.2	1.1 (0.2–7.8)	1.0 (0.9–1.1)
	Systemically unwell	62.0	84.2	2.8 (0.7–11.4)	1.0 (1.0–1.0)
	TB history	3.0	99.7	9.8 (1.0–91.4)	1.0 (0.9–1.0)
	Thoracic/chest/abdominal pain	3.0	94.1	0.5 (0.1–3.6)	1.0 (1.0–1.1)
	Tried bed rest but no relief	0	99.8	0.0	1.0
	Unexplained weight loss	1.1	99.9	9.2 (0.8–100.6)	1.0 (1.0–1.0)
	Urinary symptoms	27.3	88.7	2.4 (1.4–4.3)	0.8 (0.7–1)
	Writhing in pain	5.0	99.3	6.9 (2.5–19.9)	1.0 (0.9–1.0)
Spinal malignancy					
Premkumar 2018 [28]	Unexplained weight loss	8.2%	95.6%	1.87 (1.1–3.17)	0.96 (0.92–1.01)
	History of malignancy	32%	95.6%	7.25 (5.65–9.3)	0.71 (0.64–0.79)
Tsiang 2019 [68]	History of malignancy	91.7%	77.8%	–	–
Cook 2011 [39]	No pain during right side flexion	96% (88–98)	4% (04–05)	0.94 (0.85–1.0)	1.0 (0.92–1.0)
	No pain during left side flexion	96% (88–98)	4% (03–04)	0.99 (0.91–1.0)	1.1 (0.37–3.2)
Donner-Banzhoff 2006 [43]	Clinical judgment	50%	83%	2.95	0.6
Spinal infection					
Premkumar 2018 [28]	Chills	11.7%	93.2%	1.71 (1.04–2.81)	0.95 (0.89–1.01)
	Fever	11.7%	93.2%	1.71 (1.04–2.81)	0.95 (0.89–1.01)
	Persistent night sweating	17.5%	86.1%	1.26 (0.85–1.86)	0.96 (0.88–1.04)
	Recent infection	24.2%	97.4%	9.31 (6.63–13.07)	0.78 (0.7–0.86)
	Sweating	11.7%	93.2%	1.71 (1.04–2.81)	0.95 (0.89–1.01)
Tsiang 2019 [68]	Fever	12.5%	99.6%	–	–
	Fever patient reported	25%	97.6%	–	–

(*continued*)

Table 9. Continued.

Author	Red Flag	Sn	Sp	LR+	LR–
Cauda equina syndrome (CES)					
Premkumar 2018 [28]	Bladder dysfunction	22.2%	90.4%	2.31 (1.25–4.27)	0.86 (0.72–1.03)
	Bowel dysfunction	13.9%	95%	2.78 (1.23–6.3)	0.91 (0.8–1.03)
Raison 2014 [58]	Saddle dysfunction	27% (12–48%)	87% (81–92%)	2.11	–
Tsiang 2019 [68]	Bladder dysfunction patient rep	50%	86.5%	–	–
	Bowel dysfunction patient rep	50%	86.5%	–	–
Cardiovascular disorders					
Body 2010 [27]	Acute ischaemic ECG changes	71% (51.3–65.0%)	81.3% (79.1–85.4%)	3.80 (2.69–4.08)	0.36 (0.43–0.60)
	Basal crackles	16.2% (11.8–22.3%)	90.6% (88.9–93.6%)	1.72 (1.30–2.90)	0.92 (0.85–0.97)
	Central CP	85.1%	34.1%	1.29	0.44
	Hypotension	6.8% (4.4–12.0%)	97.7% (97.1–99.3%)	2.92 (2.21–10.98)	0.95 (0.90–0.98)
	Pain radiated to the right arm	18.9% (12.3–22.8%)	91.8% (89.8–94.3%)	2.31 (1.47–3.34)	0.88 (0.84–0.96)
	Pain radiating to upper limbs	13.5% (8.2–17.5%)	94.8% (93.2–96.8%)	2.58 (1.55–4.29)	0.91 (0.87–0.97)
	Sweating observed	36.5% (22.0–34.5%)	94.3% (92.4–96.2%)	6.39 (3.42–7.63)	0.67 (0.70–0.83)
	Vomiting	16.2% (9.8–19.7%)	94.8% (93.2–96.8%)	3.09 (1.82–4.85)	0.88 (0.85–0.95)
Goodacre 2002 [48]	Absence of chest wall syndrome	91.7% (74.2–97.7%)	27.8% (24.6–31.2%)	1.27 (1.12–1.44)	0.30 (0.08–1.14)
	Exertional pain – ACS	29.6% (20.8–40.3%)	85.6% (83.0–87.8%)	2.06 (1.41–2.99)	0.82 (0.71–0.95)
	Exertional pain – AMI	35.3% (21.5–52.1%)	85% (82.4–87.2%)	2.35 (1.45–3.80)	0.76 (0.59–0.98)
	Pain radiating to upper limbs – AMI	38.2% (23.9–55.0%)	90.6% (88.5–92.4%)	4.07 (2.53–6.54)	0.68 (0.52–0.89)

Sn: sensibility; Sp: specificity; LR+: likelihood ratio positive; LR–: likelihood ratio negative.

pathologies. This is because they have grouped into wide cluster several conditions (e.g., fracture, infection, cancer, inflammatory spinal pathology, visceral, or systemic pathology).

Concerning the internal validity of results, our systematic review has outlined a fairly good quality of the included studies, even if some methodological limitations were identified such as non-representative cohort of patients, uncomplete reporting of follow-ups data, heterogeneous set of RFs for the same pathologies, and miscellaneous reference standards for RFs among studies.

Upon the external validity, our results have been affected by two elements. First, the screening of RFs was performed mostly by physicians through a personal interview [40,44,50,51,65] or administering a questionnaire [27,28,41,45,49]; therefore, this is not lined with the aim of informing any type of healthcare professional as declared by our systematic review. Moreover, a limited number of studies have analysed the diagnostic accuracy of RFs, thus resizing their implementation in the clinical practice.

Limitations

This SR has several limits. First, we cannot rule-out a publication bias because the exclusion of those papers published before 01 January 1999 [87]. However, we have screened seven databases and "grey" literature in order to make our search strategy sensitive and to improve the probability to retrieve the higher number of studies [29]. Moreover, we created a review protocol before the commencement of the review, but we did not submit such protocol on the reference database (PROSPERO) [88]. Furthermore, we adopted the guidelines of the PRISMA statements to design the study and to guarantee a management methodology, relevance of results and clarity of optimal reporting [29]. Finally, only the 45% of primary studies have analysed the diagnostic accuracy of RFs due to the low prevalence of serious pathologies. Therefore, we decided to provide information concerning the clinical reliability of RFs only for signs and symptoms which diagnostic accuracy was reported.

Conclusions

Several clinical guidelines endorse RFs screening during the physical examination of the patients, notwithstanding the lack of evidence in supporting the reliability of the majority of RFs commonly used in clinical practice. Despite the application of RFs as a self-contained screening tool, it should not be supported because of the low informativeness in detecting a serious pathology, the combination of multiple RFs showed higher diagnostic accuracy, proving to be a promising screening tool. In our systematic review, we have analysed 40 primary studies concerning TLP identifying several RFs. Among them, only 30% have evaluated the combination of RFs. In order to develop a worthwhile diagnostic tool for TLP, we have evaluated every available RFs, reporting a list of their diagnostic accuracy values. Although many of these confirmed to be pointless, it is noteworthy highlighting the findings related to the scoring systems to detect PE, and the relationship between CP and cardiovascular diseases, which have demonstrated good reliability. Despite our findings are promising, further primary studies related to the combination of multiple RFs are needed.

Disclosure statement

The authors declare that they have no competing interests.

ORCID

Filippo Maselli http://orcid.org/0000-0001-9683-9975
Marco Testa http://orcid.org/0000-0001-8643-7200

Data availability statement

All data generated or analysed during this study are included in this published article. Other information of this study is available from the corresponding author on reasonable request.

References

[1] Oliveira CB, Maher CG, Pinto RZ, et al. Clinical practice guidelines for the management of non-specific low back pain in primary care: an updated overview. Eur Spine J. 2018;27(11):2791–2803.
[2] Koes BW, van Tulder M, Lin CC, et al. An updated overview of clinical guidelines for the management of non-specific low back pain in primary care. Eur Spine J. 2010;19(12):2075–2094.

[3] Fatoye F, Gebrye T, Odeyemi I. Real-world incidence and prevalence of low back pain using routinely collected data. Rheumatol Int. 2019;39(4):619–626.
[4] Verhagen AP, Downie A, Popal N, et al. Red flags presented in current low back pain guidelines: a review. Eur Spine J. 2016;25(9):2788–2802.
[5] Spencer L, McKenna L, Fary R, et al. Upper back pain in postmenopausal women and associated physical characteristics. PLoS One. 2019;14(7):e0220452.
[6] Briggs AM, Smith AJ, Straker LM, et al. Thoracic spine pain in the general population: prevalence, incidence and associated factors in children, adolescents and adults. A systematic review. BMC Musculoskelet Disord. 2009;10(1):77.
[7] Geyser M, Smith S. Chest pain prevalence, causes, and disposition in the emergency department of a regional hospital in Pretoria. Afr J Prim Health Care Fam Med. 2016;8(1): e1–e5.
[8] Buntinx F, Knockaert D, Bruyninckx R, et al. Chest pain in general practice or in the hospital emergency department: is it the same? Fam Pract. 2001;18(6):586–589.
[9] Hickam DH. Chapter 9. Chest pain or discomfort. In: Walker HK, Hall WD, Hurst JW, editors. Clinical methods: the history, physical, and laboratory examinations. 3rd ed. Boston (MA): Butterworths; 1990. p. 72–75.
[10] Burton AK, Balagué F, Cardon G, et al. Chapter 2: European guidelines for prevention in low back pain. Eur Spine J. 2006;15(S2):s136–s168.
[11] Goodman CC, Heick J, Lazaro RT. Differential diagnosis for physical therapist: screening for referral. 6th ed. Philadelphia (PA): Saunders; 2018.
[12] Cook CE, George SZ, Reiman MP. Red flag screening for low back pain: nothing to see here, move along: a narrative review. Br J Sports Med. 2018;52(8):493–496.
[13] Henschke N, Maher CG, Refshauge KM. Screening for malignancy in low back pain patients: a systematic review. Eur Spine J. 2007;16(10):1673–1679.
[14] Henschke N, Maher CG, Refshauge KM. A systematic review identifies five "red flags" to screen for vertebral fracture in patients with low back pain. J Clin Epidemiol. 2008;61(2): 110–118.e1.
[15] Koes BW, van Tulder MW, Ostelo R, et al. Clinical guidelines for the management of low back pain in primary care: an international comparison. Spine. 2001;26(22):2504–2513.
[16] Henschke N, Keuerleber J, Ferreira M, et al. The methodological quality of diagnostic test accuracy studies for musculoskeletal conditions can be improved. J Clin Epidemiol. 2014;67(4):416–424.
[17] Williams CM, Henschke N, Maher CG, et al. Red flags to screen for vertebral fracture in patients presenting with low-back pain. Cochrane Database Syst Rev. 2013;3(1): CD008643.
[18] Henschke N, Maher CG, Ostelo RW, et al. Red flags to screen for malignancy in patients with low-back pain. Cochrane Database Syst Rev. 2013;28(2):CD008686.
[19] Ramanayake RJC, Basnayake BT. Evaluation of red flags minimizes missing serious diseases in primary care. J Fam Med Prim Care. 2018;7(2):315–318.
[20] Finucane LM, Downie A, Mercer C, et al. International framework for Red Flags for potential serious spinal pathologies. J Orthop Sports Phys Ther. 2020;50(7):350–372.
[21] Downie A, Williams CM, Henschke N, et al. Red flags to screen for malignancy and fracture in patients with low back pain: systematic review. BMJ. 2013;347:f7095.
[22] Verhagen AP, Downie A, Maher CG, et al. Most red flags for malignancy in low back pain guidelines lack empirical support: a systematic review. Pain. 2017;158(10):1860–1868.
[23] Parreira PCS, Maher CG, Traeger AC, et al. Evaluation of guideline-endorsed red flags to screen for fracture in patients presenting with low back pain. Br J Sports Med. 2019;53(10):648–654.
[24] Yousef E, Abdulaziz A. Pediatric chest pain: the Red Flags? J Cardiol Cardiovasc Ther. 2017;7(4):555720.
[25] MacKnight JM, Mistry DJ. Chest pain in the athlete: differential diagnosis, evaluation, and treatment. Clin Rev Allergy Immunol. 2005;29(2):87–96.
[26] Hung C, Hou CJ, Yeh H, et al. Atypical chest pain in the elderly: prevalence, possible mechanisms and prognosis. Int J Gerontol. 2010;4(1):1–8.
[27] Body R, Carley S, Wibberley C, et al. The value of symptoms and signs in the emergent diagnosis of acute coronary syndromes. Resuscitation. 2010;81(3):281–286.
[28] Premkumar A, Godfrey W, Gottschalk MB, et al. Red Flags for low back pain are not always really red: a prospective evaluation of the clinical utility of commonly used screening questions for Low Back Pain. J Bone Joint Surg Am. 2018;100(5):368–374.
[29] Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol. 2009;62(10):e1–e34.
[30] Altman DG. Practical statistics for medical research. London (UK): Chapman and Hall; 1991.
[31] McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276–282.
[32] Higgins JPT, Thomas J, Chandler J, et al. Cochrane handbook for systematic reviews of interventions. 2nd ed. Chichester (UK): Wiley; 2019.
[33] Ahad A, Elsayed M, Tohid H. The accuracy of clinical symptoms in detecting cauda equina syndrome in patients undergoing acute MRI of the spine. Neuroradiol J. 2015; 28(4):438–442.
[34] Albarran J, Durham B, Gowers J, et al. Is the radiation of chest pain a useful indicator of myocardial infarction? A prospective study of 541 patients. Accid Emerg Nurs. 2002; 10(1):2–9.
[35] Aydin F, Aksit E, Yildirim OT, et al. Chest pain score: a novel and practical approach to angina pectoris. A diagnostic accuracy study. Sao Paulo Med J. 2019;137(1):54–59.
[36] Bernard Bagattini S, Bounameaux H, Perneger T, et al. Suspicion of pulmonary embolism in outpatients: nonspecific chest pain is the most frequent alternative diagnosis. J Intern Med. 2004;256(2):153–160.
[37] Balasubramanian K, Kalsi P, Greenough CG, et al. Reliability of clinical assessment in diagnosing cauda equina syndrome. Br J Neurosurg. 2010;24(4):383–386.
[38] Broderick C, Hopkins S, Mack DJF, et al. Delays in the diagnosis and treatment of bone and joint tuberculosis in the United Kingdom. Bone Joint J. 2018;100-B(1):119–124.
[39] Cook C, Ross MD, Isaacs R, et al. Investigation of nonmechanical findings during spinal movement screening for identifying and/or ruling out metastatic cancer. Pain Pract. 2012;12(6):426–433.
[40] Courtney DM, Kline JA, Kabrhel C, et al. Clinical features from the history and physical examination that predict the presence or absence of pulmonary embolism in symptomatic emergency department patients: results of a

prospective, multicenter study. Ann Emerg Med. 2010;55(4):307–315.e1.
[41] de Schepper EIT, Koes BW, Veldhuizen EFH, et al. Prevalence of spinal pathology in patients presenting for lumbar MRI as referred from general practice. Fam Pract. 2016;33(1):51–56.
[42] Domen PM, Hofman PA, van Santbrink H, et al. Predictive value of clinical characteristics in patients with suspected cauda equina syndrome. Eur J Neurol. 2009;16(3):416–419.
[43] Donner-Banzhoff N, Roth T, Sönnichsen AC, et al. Evaluating the accuracy of a simple heuristic to identify serious causes of low back pain. Fam Pract. 2006;23(6):682–686.
[44] Dugas AF, Lucas JM, Edlow JA. Diagnosis of spinal cord compression in nontrauma patients in the emergency department. Acad Emerg Med. 2011;18(7):719–725.
[45] Enthoven WT, Geuze J, Scheele J, et al. Prevalence and "Red Flags" regarding specified causes of back pain in older adults presenting in general practice. Phys Ther. 2016;96(3):305–312.
[46] Everden A, Mamo JP, Somasunderam D, et al. Bone and joint mycobacterial infection: a retrospective review of cases presenting to a UK district hospital. J Med Microbiol. 2018;67(12):1698–1705.
[47] Gesuete V, Fregolent D, Contorno S, et al. Follow-up study of patients admitted to the pediatric emergency department for chest pain. Eur J Pediatr. 2020;179(2):303–308.
[48] Goodacre S, Locker T, Morris F, et al. How useful are clinical features in the diagnosis of acute, undifferentiated chest pain? Acad Emerg Med. 2002;9(3):203–208.
[49] Harding IJ, Davies E, Buchanan E, et al. The symptom of night pain in a back pain triage clinic. Spine. 2005;30(17):1985–1988.
[50] Henschke N, Maher CG, Refshauge KM, 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(10):3072–3080.
[51] Holmes JF, Panacek EA, Miller PQ, et al. Prospective evaluation of criteria for obtaining thoracolumbar radiographs in trauma patients. J Emerg Med. 2003;24(1):1–7.
[52] Hsu JM, Joseph T, Ellis AM. Thoracolumbar fracture in blunt trauma patients: guidelines for diagnosis and imaging. Injury. 2003;34(6):426–433.
[53] Karlaftis A, Karamanolis G, Triantafyllou K, et al. Clinical characteristics in patients with non-cardiac chest pain could favor gastroesophageal reflux disease diagnosis. Ann Gastroenterol. 2013;26(4):314–318.
[54] Kempthorne JT, Pratt C, Smale EL, et al. Ten-year review of extradural spinal abscesses in a New Zealand tertiary referral centre. J Clin Neurosci. 2009;16(8):1038–1042.
[55] Ko SY, Kim SI, Kim JH, et al. Clinically distinct characteristics in patients younger than 40 years old with non-cardiac chest pain. J Gastroenterol Hepatol. 2012;27(9):1484–1489.
[56] Milner KA, Funk M, Arnold A, et al. Typical symptoms are predictive of acute coronary syndromes in women. Am Heart J. 2002;143(2):283–288.
[57] Punukollu H, Khan IA, Punukollu G, et al. Acute pulmonary embolism in elderly: clinical characteristics and outcome. Int J Cardiol. 2005;99(2):213–216.
[58] Tobias N, Raison J, Alwan W, et al. The reliability of Red Flags in spinal cord compression. Arch Trauma Res. 2014;3(1):e17850.
[59] Reuter P, Pradeau C, Huo Yung Kai S, et al. Predicting acute coronary syndrome in males and females with chest pain who call an emergency medical communication centre. Scand J Trauma Resusc Emerg Med. 2019;27(1):92.
[60] Roman M, Brown C, Richardson W, et al. The development of a clinical decision making algorithm for detection of osteoporotic vertebral compression fracture or wedge deformity. J Man Manip Ther. 2010;18(1):44–49.
[61] Sánchez M, López B, Bragulat E, et al. Triage flowchart to rule out acute coronary syndrome. Am J Emerg Med. 2007;25(8):865–872.
[62] Schillinger M, Sodeck G, Meron G, et al. Acute chest pain—identification of patients at low risk for coronary events. The impact of symptoms, medical history and risk factors. Wien Klin Wochenschr. 2004;116(3):83–89.
[63] Shaw B, Kinsella R, Henschke N, et al. Back pain "red flags": which are most predictive of serious pathology in the emergency department? Eur Spine J. 2020;48(8):27.
[64] Park SH, Choi JY, Park EJ, et al. Prevalence of gastrointestinal diseases and treatment status in noncardiac chest pain patients. Korean Circ J. 2015;45(6):469–472.
[65] Svensson L, Isaksson L, Axelsson C, et al. Predictors of myocardial damage prior to hospital admission among patients with acute chest pain or other symptoms raising a suspicion of acute coronary syndrome. Coron Artery Dis. 2003;14(3):225–231.
[66] Thiruganasambandamoorthy V, Turko E, Ansell D, et al. Risk factors for serious underlying pathology in adult emergency department nontraumatic low back pain patients. J Emerg Med. 2014;47(1):1–11.
[67] Timmons S, Kingston M, Hussain M, et al. Pulmonary embolism: differences in presentation between older and younger patients. Age Ageing. 2003;32(6):601–605.
[68] Tsiang JT, Kinzy TG, Thompson N, et al. Sensitivity and specificity of patient-entered red flags for lower back pain. Spine J. 2019;19(2):293–300.
[69] Van den Bosch MA, Hollingworth W, Kinmonth AL, et al. Evidence against the use of lumbar spine radiography for low back pain. Clin Radiol. 2004;59(1):69–76.
[70] Wells PS, Anderson DR, Rodger M, et al. Derivation of a simple clinical model to categorize patients probability of pulmonary embolism: increasing the models utility with the SimpliRED D-dimer. Thromb Haemost. 2000;83(03):416–420.
[71] Turner T, Green S, Tovey D, et al. Producing Cochrane systematic reviews—a qualitative study of current approaches and opportunities for innovation and improvement. Syst Rev. 2017;6(1):147.
[72] McGee S. Simplifying likelihood ratios. J Gen Intern Med. 2002;17(8):647–650.
[73] Ouzzani M, Hammady H, Fedorowicz Z, et al. Rayyan—a web and mobile app for systematic reviews. Syst Rev. 2016;5(1):210.
[74] Mendeley [Internet]. Philadelphia (PA): Saunders; 2020. Available from: https://www.mendeley.com/?interaction_required=true
[75] Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. Eur J Epidemiol. 2010;25(9):603–605.
[76] Patel ND, Broderick DF, Burns J, et al. ACR appropriateness criteria low back pain. J Am Coll Radiol. 2016;13(9):1069–1078.

[77] Dionne N, Adefolarin A, Kunzelman D, et al. What is the diagnostic accuracy of red flags related to cauda equina syndrome (CES), when compared to magnetic resonance imaging (MRI)? A systematic review. Musculoskelet Sci Pract. 2019;42:125–133.
[78] Šimundić AM. Measures of diagnostic accuracy: basic definitions. EJIFCC. 2009;19(4):203–211.
[79] Cawthon PM. Gender differences in osteoporosis and fractures. Clin Orthop Relat Res. 2011;469(7):1900–1905.
[80] Lurie JD. What diagnostic tests are useful for low back pain? Best Pract Res Clin Rheumatol. 2005;19(4):557–575.
[81] Casazza BA. Diagnosis and treatment of acute low back pain. Am Fam Physician. 2012;85(4):343–350.
[82] Arce D, Sass P, Abul-Khoudoud H. Recognizing spinal cord emergencies. Am Fam Physician. 2001;64(4):631–638.
[83] Bowers B. Recognising metastatic spinal cord compression. Br J Community Nurs. 2015;20(4):162–165.
[84] Della-Giustina D, Kilcline BA. Acute low back pain: a comprehensive review. Compr Ther. 2000;26(3):153–159.
[85] Della-Giustina DA. Emergency department evaluation and treatment of back pain. Emerg Med Clin North Am. 1999;17(4):877–893.
[86] Dudler J, Balague F. What is the rational diagnostic approach to spinal disorders? Best Pract Res Clin Rheumatol. 2002;16(1):43–57.
[87] Lin L, Chu H. Quantifying publication bias in meta-analysis. Biometrics. 2018;74(3):785–794.
[88] PROSPERO [Internet]. York (UK): Centre for Reviews and Dissemination University of York; 2020. Available from: https://www.crd.york.ac.uk/prospero