

Red flags for potential serious pathologies in people with neck pain: a systematic review of clinical practice guidelines

Daniel Feller^{1,2}, Alessandro Chiarotto², Bart Koes^{2,3}, Filippo Maselli^{4,5}, Firas Mourad^{6,7}

¹Provincial Agency for Health of the Autonomous Province of Trento, Trento - Italy

²Department of General Practice, Erasmus MC, University Medical Centre, Rotterdam - The Netherlands

³Research Unit of General Practice, Department of Public Health and Center for Muscle and Joint Health, University of Southern Denmark, Odense - Denmark

⁴Department of Human Neurosciences, Sapienza University of Rome, Rome - Italy

⁵Sovrintendenza Sanitaria Regionale Puglia, INAIL, Bari - Italy

⁶Department of Health, LUNEX University of Applied Sciences, Differdange - Luxembourg

⁷Luxembourg Health and Sport Sciences Research Institute A.s.b.l., Differdange - Luxembourg

ABSTRACT

Introduction: We conducted a systematic review of clinical practice guidelines to identify red flags for serious pathologies in neck pain mentioned in clinical practice guidelines, to evaluate agreement in red flag recommendations across guidelines, and to investigate the level of evidence including what study type the recommendations are based on.

Methods: We searched for guidelines focusing on specific and nonspecific neck pain in MEDLINE, EMBASE, and PEDro up to June 9, 2023. Additionally, we searched for guidelines through citation tracking strategies, by consulting experts in the field, and by checking guideline organization databases.

Results: We included 29 guidelines, 12 of which provided a total of 114 red flags for fracture (n = 17), cancer (n = 21), spinal infection (n = 14), myelopathy (n = 15), injury to the spinal cord (n = 1), artery dissection (n = 7), intracranial pathology (n = 3), inflammatory arthritis (n = 2), other systemic disease (n = 6), or unrelated to a specific condition (n = 19). Overall, there is very little agreement (median Fleiss' kappa of 0) between guidelines on the red flags to screen for serious pathologies.

Conclusion: Red flags were mainly supported by expert opinions. We also observed a general lack of consensus among guidelines regarding which red flags to endorse. Considering the current limitations of the evidence, specific recommendations on which red flags to use cannot be provided, except for using the Canadian C-Spine rule for screening posttraumatic fractures.

Keywords: Differential diagnosis, Guidelines, Neck pain, Red flags

What is already known

- *Triaging serious cervical conditions mimicking musculoskeletal neck pain is a mainstay in primary care. Although identifying these pathologies can be challenging for clinicians, their recognition is relevant to determine which patients need to be referred to ensure safe and effective patient care.*

What does this study add

- *Almost all the red flags were only based on mechanism reasoning (Level of evidence 5). Diagnostic accuracy values for red flags were not reported, except for the Canadian C-spine rules. Therefore, clinicians should rely on red flags cautiously, integrating them with sound clinical reasoning.*

Received: August 4, 2024

Accepted: November 11, 2024

Published online: December 4, 2024

This article includes supplementary material

Corresponding author:

Daniel Feller

email: d.feller@erasmusmc.nl

Introduction

Neck pain is a complex biopsychosocial disorder estimated to be the eighth leading cause of Years Lived with Disability globally (1-3). Although benign in the large majority of patients, it is estimated that 1% of neck pain can be caused by underlying serious pathologies, such as malignancy, cervical arterial pathology, myelopathy, congenital craniovertebral anomalies, infection, or fracture (4-7). Screening for

serious pathology masquerading as nonspecific musculoskeletal neck pain is a real challenge for clinicians, particularly in a direct access setting (5,8-10). It has been estimated that a delayed diagnosis of serious cervical pathologies ranges from 5% to 20% of all cases accessing the emergency department with neck pain, with potentially life-threatening consequences in the worst-case scenario (11,12). Therefore, the early recognition of serious cervical pathologies is a mainstay for safe physiotherapy practice and allows clinicians to identify those patients who require referral to another healthcare professional for optimal management and best possible outcomes (13).

As standard practice, red flags have been used to guide physiotherapists in identifying serious cervical pathology (14). Red flags are cues from a patient's medical history and clinical examination potentially associated with a higher risk of serious conditions (15). As practical examples, a past history of cancer is considered a red flag for spinal malignancy, urinary incontinence associated with back pain raises suspicion for a cauda equina syndrome, and pulse changes during palpation of a peripheral artery with neuropathic-like pain in the lower extremities (namely, radicular pain) may suggest the presence of peripheral arterial disease (13,16,17).

The recently released International Federation of Orthopaedic Manipulative Physical Therapists Cervical Framework highlights the need for physiotherapists to use a differential diagnosis tool for informed and safe management of the cervical spine (4,18). Therefore, investigating red flags for neck pain remains a priority for an informed practice and the patient's safety (14). To the best of the authors' knowledge, no systematic review has been published investigating the recommended red flags for neck pain in clinical practice guidelines for their scientific validity. Furthermore, knowledge on the level of evidence red flag recommendations were based on (e.g., systematic reviews of diagnostic test accuracy studies, cross-sectional studies, mechanism-based reasoning) may help clinicians to value the recommendations' strength. Therefore, we aimed to: (1) identify red flags to triage serious pathologies recommended in clinical practice guidelines for neck pain, (2) evaluate the agreement in red flag recommendations across guidelines, and (3) investigate the level of evidence on which the red flag recommendations are based.

Methods

We used the "Preferred Reporting Items for Systematic Reviews and Meta-analyses" (PRISMA) checklist for the reporting of the present manuscript (19). The study protocol was registered on MedRxiv (20).

Eligibility criteria

According to the Classification of Neck Pain and Associated Disorders (NAD) (21), we included guidelines focusing on specific (NAD III) and nonspecific neck pain (NAD I/II). We excluded guidelines for serious neck pain (NAD IV) because we expected them to only address managing these conditions, not identifying them in patients presenting with

musculoskeletal neck pain. Also, we excluded guidelines not explicitly focused on neck pain, such as guidelines in which neck pain is only briefly mentioned in the context of other disorders or a more complex topic (e.g., management of chronic pain in general). A document was considered as a clinical practice guideline if it fulfilled the following criteria (adapted from the PEDro criteria for evidence-based clinical practice guidelines (22)): it was produced under the auspices of a health professional association or society, public or private organization, healthcare organization or plan, or government agency; a systematic literature search and review of existing scientific evidence was performed during the guideline development; the guideline was based on published systematic reviews; and the guideline contained systematically developed statements that included recommendations, strategies, or information to guide decisions about appropriate healthcare (22).

We did not apply any restrictions regarding publication date and language. Non-English and non-Italian guidelines were translated using "DeepL Translate" ([Online](#)). In addition, we only included the most up-to-date version if multiple versions of the same guideline were present.

Study selection process

Without time restriction, we searched for guidelines in MEDLINE (via PubMed), EMBASE, and PEDro electronic databases on 09/06/2023. Supplementary Material 1 reports the full search strategy for these databases.

Guidelines were also searched through forward and backward citation tracking strategies (Web of Science on 12/07/2023), by consulting experts in the field (top 10 experts on neck pain according to ExpertScape.com on 15/07/2023), and by checking guideline organization databases. The following guideline organization databases were searched: the "Canadian Medical Association Infobase of clinical practice guidelines" ([Online](#)), the "Istituto Superiore Sanità – Sistema Nazionale LineeGuida" ([Online](#)), the "Guidelines International Network" ([Online](#)), the "National Institute for Clinical Excellence – NICE" ([Online](#)), the "OPTIMa collaboration" ([Online](#)), the "Guideline Central" ([Online](#)), the "Scottish Intercollegiate Guidelines Network – SIGN" ([Online](#)), and the "Agency for Healthcare Research and Quality" ([Online](#)). In addition, we screened the references of two recently published systematic reviews on guidelines for neck pain (23,24).

Duplicates were eliminated using the Deduplicator function of "Systematic Review Accelerator" (25). We used the online electronic systematic review software package (Rayyan QCRI) to organize and track the selection process (26). Two researchers independently performed the study selection process by title/abstract (DF and FMo, or DF and AC) and then by full text (DF and FMo). Any disagreement was resolved by consensus or by the decision of a third author (AC).

Data extraction process

Two reviewers (DF and FMa) performed the data extraction process independently using a standardized Excel form. The data extraction form was piloted on three included guidelines.



Any discrepancies were resolved with a consensus between the two authors and eventually by a third author's decision (AC).

We extracted the following data from each guideline: publication year, language of publication, association(s) or society(ies) which generated the guideline, serious pathologies considered (e.g., malignancy, fracture, infection, congenital craniovertebral anomalies, cervical arteries dysfunctions), reported red flags, if these red flags are presented for individual pathologies or in a more general sense (i.e., not tied to any specific pathology), level of the evidence of each red flag, how red flags were supported (study design, consensus of the guideline committee, or not reported), and, when available, the diagnostic accuracy underpinning each recommendation. We determined the level of evidence for each red flag recommended in the guidelines by extracting the citations provided in each source. The level of evidence was determined using the 2011 Levels of Evidence framework from the Oxford Centre for Evidence-Based Medicine (27). This classification system ranks evidence based on study design, with systematic reviews of cross-sectional studies representing the highest level and mechanism-based reasoning representing the lowest (Tab. 1). Two researchers independently determined the level of the evidence (DF and FMO). Any disagreement was resolved by consensus or by the decision of a third author (AC).

TABLE 1 - Level of evidence for diagnostic questions according to the 2011 framework by the Oxford Centre for Evidence-Based Medicine

Level of evidence	Description
Level 1	Systematic review of cross-sectional studies with consistently applied reference standard and blinding
Level 2	Individual cross-sectional studies with consistently applied reference standard and blinding
Level 3	Non-consecutive studies, or studies without consistently applied reference standards
Level 4	Case-control studies, or poor or non-independent reference standard
Level 5	Mechanism-based reasoning

Data synthesis

We calculated Fleiss' kappa to evaluate the agreement among guidelines recommendations (poor agreement <0.00, slight agreement 0.00–0.20, fair agreement 0.21–0.40, moderate agreement 0.41–0.60, substantial agreement 0.61–0.80, almost perfect agreement 0.81–1.00) (28). Additionally, to summarize the recommendations to triage serious pathologies and the study designs to support recommendations, we computed descriptive statistics (absolute and relative frequencies) and reported the results narratively.

Deviations from the protocol

Deviations from the published protocol were implemented in response to reviewers' requests. Specifically, we determined the level of evidence for each red flag

recommended in the guidelines to enhance the rigor of our findings and provide a clearer interpretation of the results in relation to the existing literature.

Equity, diversity, and inclusion statement

The group of authors involved in this study comprises five males from two high-income countries, Italy and the Netherlands. Among these authors, three are physical therapists (AC, FM, and FMO), one is both a physical therapist and a statistician (DF), and the fifth is an epidemiologist (BK). The group maintains a balance in terms of junior, mid-career, and senior researchers. At the time of submission, DF is a first-year PhD student, AC is an assistant professor, while BK is a full professor. FM holds a PhD, and FMO is an assistant professor with clinical and research experience focused on neck pain. Both FM and FMO teach a postgraduate course in screening for referrals for physical therapists in Italy. All the authors have experience in conducting systematic reviews. Additionally, all the authors have attended multiple courses on planning and conducting literature reviews. It is worth noting that our search strategy and data extraction process were not biased toward any specific gender, race, culture, or socioeconomic level.

Results

We retrieved 4,431 records from database investigations, 532 of which were duplicates. Titles and abstracts screening was performed on the remaining 3,899 records; we also retrieved six records from expert consultations, three from guideline organization databases, and seven from citation tracking strategies. In total, 59 reports were selected for full-text analysis. Ultimately, 29 guidelines met the inclusion criteria and were included in the present systematic review (Fig. 1). Supplementary material 2 contains the references to the included guidelines.

Characteristics of the included guidelines

Of the 29 guidelines included in the study, 12 (41%) provided information on red flags for screening serious pathologies. Among the remaining guidelines, 10 (35%) contained recommendations for diagnosing neck pain but did not mention any signs or symptoms to screen for serious pathologies, while 7 (24%) did not provide any diagnostic recommendation. Supplementary material 3 reports the characteristics of the guidelines that do not report red flags.

Of the guidelines reporting red flags, 3 (25%) were developed for patients who suffered from whiplash-associated disorders (29-31), 5 (42%) for patients with NAD grade I to III (32-36), and 4 (33%) for mixed populations (e.g., whiplash and NAD) (1,37-39). Most studies mentioned red flags for specific pathologies (e.g., fracture, cancer, infection), while 3 (25%) described red flags unrelated to a particular disease (e.g., Whalen et al (33) did not specify any particular pathology but identified fever, alongside other signs and symptoms, as a warning sign for serious conditions) (29,33,40). Table 2 reports the complete characteristics of the 12 guidelines reporting on red flags.

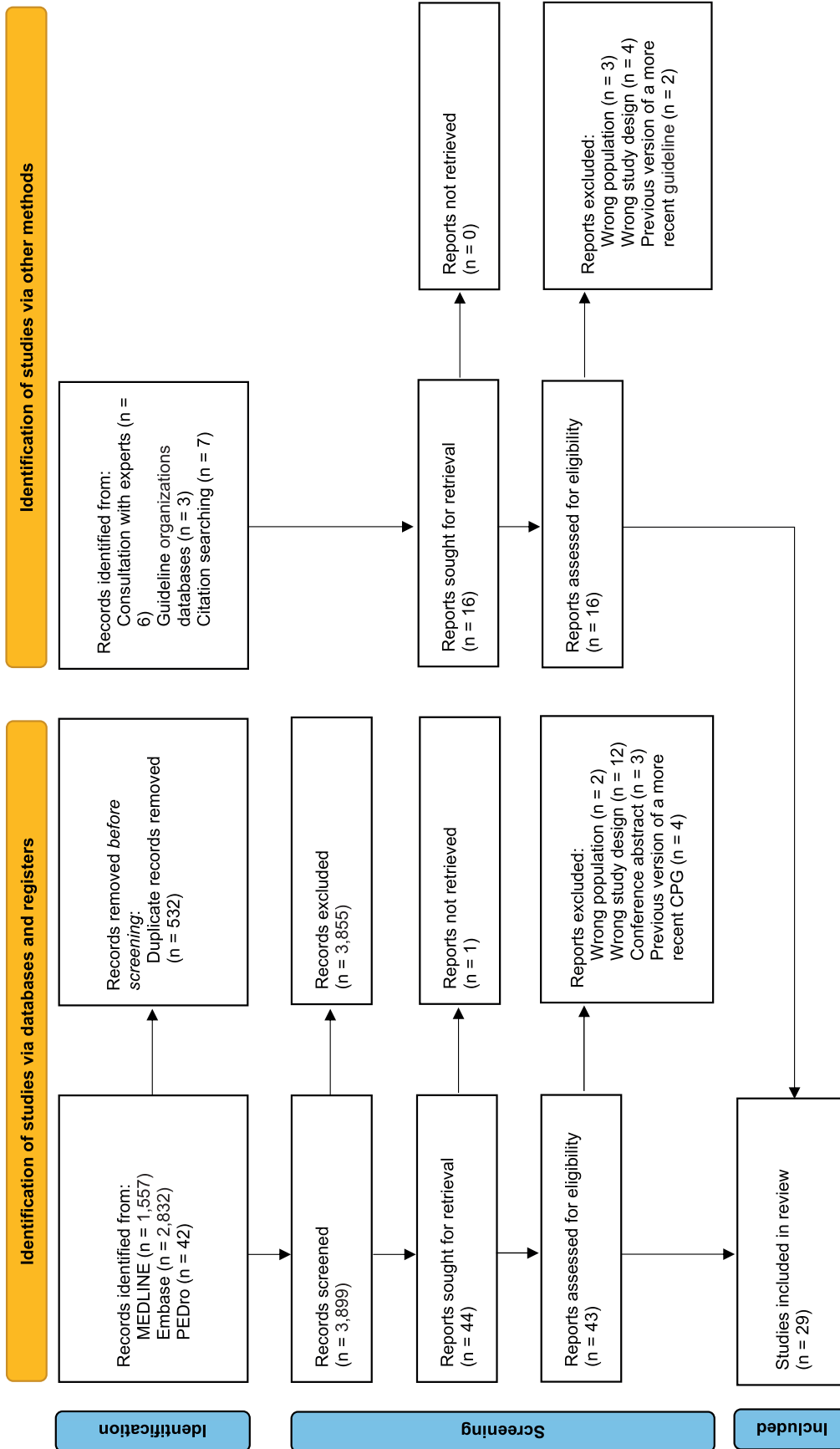


FIGURE 1 - PRISMA flowchart.



TABLE 2 - Characteristics of the guidelines with red flag recommendations

Author and year	Nation	Language	Society or body issuing the guidelines	Population (as reported in the guideline)	Cited pathologies	Strength of recommendation for screening for serious pathologies as reported in the guideline
Papic, 2023	Australia	English	New South Wales State Insurance Regulatory Authority	Acute or chronic WAD (grades I to III)	– Fracture	Strong
Whalen, 2019	USA	English	Scientific Council of the Clinical Compass	Neck pain of any duration	Red flags unrelated to specific disease	Not reported
Bier, 2018	Netherlands	English	Royal Dutch Society for Physical Therapy	NAD grades I to III, irrespective of the duration	– Fracture – Cancer – Vertebral infection – Cervical myelopathy – Spinal cord injury – Vertebral artery dissection – Systemic disease	“Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate”
Blanpied, 2017	USA	English	American Physical Therapy Association	Neck pain, including WAD, headache, and radicular pain	– Fracture – Cancer* – Vertebral infection* – Cervical myelopathy – Arterial insufficiency – Systemic disease* – Intracranial pathology – Systemic disease* – Cardiac involvement* – Unexplained cranial nerve dysfunction*	Strong (“one or more systematic reviews support the recommendation, providing evidence for a strong magnitude of effect”)
Karppinen, 2017	Finland	Finnish	Finnish Medical Association Duodecim	Neck pain, including WAD and radiculopathy	– Fracture* – Cancer – Vertebral infection* – Cervical myelopathy – Cervical or carotid dissection – Systemic inflammatory disease*	Not reported
Lemeunier, 2017	France	French	Associations Française de Chiropratique and Institut Franco-Européen de Chiropraxie	Neck pain of any duration	– Fracture – Osteoporotic fracture – Cancer – Vertebral infection – Cervical myelopathy – Carotid/vertebral artery dissection – Intracranial pathology – Inflammatory arthritis	Not reported

(Cont.)



TABLE 2 - Continued

Author and year	Nation	Language	Society or body issuing the guidelines	Population (as reported in the guideline)	Cited pathologies	Strength of recommendation for screening for serious pathologies as reported in the guideline
Bussières, 2016	Canada	English	Canadian Chiropractic Guideline Initiative	NAD grade I to III and WAD grade I to III	<ul style="list-style-type: none"> – Major structural pathologies* – Other pathologies rather than NAD or WAD* 	Not reported
Côté, 2016	Canada	English	Ontario Protocol for Traffic Injury Management Collaboration	NAD grades I-III of less than 6 months duration, and WAD grades I-III of less than 6 months duration	<ul style="list-style-type: none"> – Fracture/dislocation – Osteoporotic fracture – Cancer – Vertebral infection – Cervical myelopathy – Carotid/vertebral artery dissection – Intracranial pathology – Inflammatory arthritis 	Not reported
Scherer, 2016	Germany	German	German Society for General Medicine and Family Medicine	Neck pain of any duration	Red flags unrelated to specific disease	
Monticone, 2013	Italy	English	Italian Society of Physical and Rehabilitation Medicine	Neck pain with or without limb involvement and/or headache	<ul style="list-style-type: none"> – Fracture – Cancer – Vertebral infection – Cervical myelopathy – Systemic disease 	Strongly recommended
Sterling, 2008	Australia	English	South Australian Centre for Trauma and Injury Recovery	WAD grades I to IV (both acute and chronic stages)	– Fracture	“Body of evidence can be trusted to guide practice in most situations”
Moore, 2005	United Kingdom	English	Chartered Society of Physiotherapy	WAD	Red flags unrelated to specific disease	“Evidence obtained from expert committee reports or opinions and/or clinical experience of respected authorities e.g. from the Delphi questionnaire”

NAD = neck pain and associated disorders; WAD = whiplash and associated disorders.

*No red flags provided.



Red flags

Supplementary material 4 summarizes the 114 red flags reported in the guidelines for fracture (number of guidelines = 8, red flags = 17), cancer (number of guidelines = 5, red flags = 21), spinal infection (number of guidelines = 4, red flags = 14), myelopathy (number of guidelines = 5, red flags = 15), injury to the spinal cord (number of guidelines = 1, red flags = 1), cervical artery dissection (number of guidelines = 4, red flags = 7), intracranial pathology (number of guidelines = 3, red flags = 3), inflammatory arthritis (number of guidelines = 2, red flags = 2), other systemic disease (number of guidelines = 2, red flags = 6), and unrelated to a specific condition (number of guidelines = 2, red flags = 19). Additionally, Supplementary material 5 provides the reference to external documents cited by Blanpied et al (37) for the reported red flags. Many red flags ($n = 77$, 67.5%) were reported only by a minority of the guidelines. As an example, only Bier et al (32) suggested that dysphagia could be a possible red flag for cancer, and only one out of four guidelines that considered spinal infection as a serious pathology mentioned “HIV positivity” as a red flag (36). Furthermore, only a few red flags ($n = 7$, 6.1%) were recommended by most of the guidelines: five out of seven guidelines (71.4%) mentioning red flags for fractures recommended the Canadian C-spine rule as a screening tool; both the two guidelines reporting red flags for osteoporotic fractures agreed in recommending “history of osteoporosis,” “use of corticosteroids,” and an “older age” as red flags; all guidelines mentioning cancer as a serious condition recommended a “history of cancer” and an “unexplained weight loss” as flags for this condition; and three out of four guidelines (75%) considering spinal infection reported the presence of fever and a “history of recent infection” as red flags for an infection (Supplementary material 4).

Agreement in red flags recommendations

Overall, there is very little agreement between guidelines on the red flags to screen for serious pathologies (Tab. 3). Notably, for all the pathologies, we found a poor agreement (Fleiss' kappa < 0), except for cancer (slight agreement with a Fleiss' kappa of 0.15) and osteoporotic fractures (perfect agreement with a Fleiss' kappa of 1).

TABLE 3 - Fleiss' kappa values

Pathology	Fleiss' kappa
Fracture	0
Osteoporotic fracture	1
Cancer	0.15
Vertebral infection	-0.14
Cervical myelopathy	-0.17
Arterial dissection	-0.28
Intracranial pathology	-0.50
Inflammatory arthritis	-0.33
Systemic disease	-1

Level of evidence on which the red flag recommendations are based

The Canadian C-spine rules were supported by Level 1 evidence, while the National Emergency X-Radiography Utilization Study (NEXUS) criteria had Level 2 evidence. Ten red flags, such as spasticity for cervical myelopathy and swelling in multiple joints for inflammatory arthritis, did not have any reference to determine their level of evidence. The remaining red flags (102, 89.5%) were based on mechanism-based reasoning, corresponding to Level 5 evidence.

Of all the red flags identified, 36 (31.6%) were supported by systematic reviews in the low back pain field or systematic reviews that did not provide direct information on the diagnostic values of specific signs and symptoms for identifying serious conditions in patients with neck pain. Notably, 10 (8.8%) red flags lacked a reference. A combination of narrative reviews, case series, and guidelines for patients with low back pain supported the remaining red flags ($n = 68$, 59.6%). Only the Canadian C-spine rules as a screening tool for fractures were supported by systematic reviews and observational studies providing direct information on their diagnostic accuracy (Supplementary material 4). Four guidelines (33%) described the literature used to support the reported red flags. Côté et al (1) reported that the red flags were based on the existing literature on low back pain, Sterling (31) reported that the red flags were supported by one or two primary studies with a low risk of bias, and Lemenuier et al (34) reported that the red flags were supported by studies with an intermediate level of evidence, such as low-powered randomized controlled trials, well-conducted nonrandomized comparative studies, and cohort studies. Lastly, Monticone et al (36) reported that experts' opinions supported their red flags.

Discussion

This review aimed to systematically collect the red flags recommended by the guidelines to screen for serious pathologies masquerading as neck pain. We identified 29 guidelines, 12 of which made recommendations for screening serious pathologies with a total of 114 red flags. Notably, 17 guidelines (59%) did not include screening for serious pathology recommendations, indicating that this topic is overlooked in more than half of the current guidelines. Our analysis showed that only a few red flags were consistently mentioned by the 12 guidelines that reported recommendations for screening serious pathologies, with many red flags (59.6%) reported only by a minority of the guidelines. The agreement between guidelines on the red flags for screening serious pathologies was generally poor, as measured by Fleiss' kappa. Among all the red flags, only the Canadian C-spine rules were well referenced (Level 1 evidence) and had diagnostic value as a screening tool for fractures in patients with neck pain after trauma. All the other red flags were either not referenced or suggested by mechanism-based reasoning (Level 5 evidence).

There are three main reasons for the heterogeneity in the recommended red flags. First, there is a lack of secondary studies, such as systematic reviews, specifically conducted to identify red flags for neck pain. Except for the Canadian C-spine rules, all the red flags were supported by primary

studies or systematic reviews that did not aim to summarize the diagnostic values of red flags for neck pain or were not supported at all. For example, the guideline by Côté et al (1) reported that “as there is a paucity of literature on red flags for neck pain, the list of red flags was informed by the low back pain literature.” Most of the included guidelines cited Nordin et al’s (41) review as a reference to support the recommended red flags. However, this review does not contain results on the red flags for which it is used as a reference. Notably, there is no strong evidence for most of the red flags for neck pain, and, therefore, the guidelines mainly relied on studies conducted in other fields and expert opinions to make their recommendations, resulting in high variability in the red flags provided in each guideline. Second, guidelines frequently presented the same red flags but offered a different cutoff or definition due to the absence of a universally agreed definition or a different healthcare system. As an example, four guidelines agreed on older age as a red flag for cancer. However, three guidelines reported “age above 60” (1,32,34), while one reported “age above 50” (36). Thus, the heterogeneity in the red flags can also be attributed to a lack of an agreed definition for almost all red flags. Third, the guidelines are customized to align with the specific health policies of the countries where they are created. For instance, the way patients can see a physiotherapist varies between countries, with some allowing direct access and others requiring a physician’s referral. These disparities may have led to heterogeneity in the suggested red flags. Our results also highlight that certain serious medical conditions have received less attention in the guidelines. As an example, only three guidelines reported red flags for intracranial pathologies, and only two reported red flags for inflammatory arthritis. This lack of knowledge of clinical predictors may reflect the diagnostic delay in certain pathologies, such as axial spondyloarthritis (42).

Our review also aimed to gather data on the diagnostic accuracy of the red flags. Several guidelines have presented the diagnostic accuracy of the Canadian C-spine rule, revealing its accuracy as a screening tool for fractures with a sensitivity of almost 100%. Papic (30) also highlighted that a positive Canadian C-spine rule reduces unnecessary imaging by 44% by mentioning preliminary results of a Cochrane review (43). The Canadian C-spine rule is a decision tool that combines several red flags with a high sensitivity. Accordingly, the combination of red flags of serious lower back pathologies was found to increase their diagnostic accuracy positively (44). Notably, in our review, the diagnostic accuracy for all other red flags was not reported. Therefore, it is unclear how these signs and symptoms may affect the likelihood of a serious condition. In addition, their combination could not be investigated. This indicates that the clinical influence of these red flags remains, at best, uncertain.

Implication for practice

Clinicians are responsible for screening for underlying serious conditions when managing patients with neck pain. Of the 29 included clinical practice guidelines, only 12 recommended screening for serious non-musculoskeletal disorders. This recommendation consistently received a “strong”

indication in favor whenever the strength of the recommendation was provided. However, there seems to be a lack of consensus on which red flags to use, almost all red flags are merely based on mechanism-based reasoning (Level 5 of evidence), and a report or reference to their diagnostic accuracy is often lacking. For these reasons, specific recommendations on which red flags to use cannot be provided, except for using the Canadian C-Spine rule for screening posttraumatic fractures. In fact, this rule is recommended by multiple guidelines based on systematic reviews of the literature (Level 1 evidence). Additionally, we have access to diagnostic accuracy values that support the Canadian C-Spine rules as an excellent screening tool, with sensitivity approaching 100%.

It is important to consider that the absence of clear red flags does not rule out the presence of a serious underlying condition. In addition, due to the rarity of many serious pathologies, one of the difficulties in differential diagnosis and in investigating the diagnostic accuracy of red flags is that some of these conditions may be present but clinically unmanifested (6). Although red flag testing remains the best tool to screen for serious cervical pathology, red flags when used in isolation are often uninformative (45,46). However, when combined within a broad clinical reasoning framework to determine the level of suspicion about serious pathology, they may help clinicians make the best judgment on the appropriate clinical action (e.g., further investigation or referral) in a continuous monitoring process (46,47). Within this reasoning pathway, the evidence to support red flags should be considered in the context of the patient’s health profile (e.g., risk factors, medications, comorbidities, age, and gender) (47).

It is also important to consider that not all red flags masquerade severe medical conditions and that not all conditions and their stage require an emergency referral. Based on the level of concern, the decision might be: to begin a trial of therapy keeping an alert to clinical features that change unexpectedly in patients with no concerning features; begin a trial of therapy with watchful waiting in patients with few concerning features; urgent referral in patients with some concerning features – such as suspected myelopathy with long-lasting symptoms; or emergency referral in patients with some concerning features that might benefit from early specialized intervention – such as suspected myelopathy with new-onset neurological signs or symptoms. After evaluating the presence of red flags and considering the patient’s clinical profile, clinicians must use their clinical reasoning to thoughtfully weigh the risks and benefits when deciding whether to refer the patient or not. For a deeper discussion on integrating red flags in clinical reasoning, we invite readers to refer to Finucane et al (13), Rushton et al (14), de Best et al (48), and Kranenburg et al (47).

Implication for future research

Future research should focus on conducting secondary studies like scoping and systematic reviews to map and/or summarize all the evidence regarding using red flags in people with neck pain. Primary studies should also be conducted to determine red flags’ diagnostic accuracy and identify



additional signs and symptoms that could indicate less considered pathologies in the current guidelines, such as intracranial pathologies and inflammatory arthritis. Since serious pathologies are rare in patients with neck pain, conducting cross-sectional and prospective cohort studies is challenging. Hence, it would be better to rely on retrospective studies like case-control observational studies, even though they might have a higher risk of bias (49). Additionally, it would be helpful to study the diagnostic value in terms of discrimination and calibration of clusters of red flags, such as diagnostic predictive models (50). Finally, it would be beneficial to establish a clear and agreed definition for the most frequently reported red flags in the literature to prevent any future research wastage. As an example, the literature could define the duration and dosage of corticosteroid usage or establish a standard age threshold for identifying a person at risk of cancer. Such standardizations would ensure that the red flags are consistently and accurately reported across various studies, leading to more reliable and comparable research outcomes.

Comparison with the low back pain field

In line with our findings, it has been observed that there is high heterogeneity in the red flags presented in the guidelines for individuals with low back pain. Verhagen et al (51,52) found no agreement between guidelines on which red flags should be recommended, paucity of diagnostic accuracy, and insufficient empirical support for most red flags. However, in contrast to neck pain, a significant amount of research has recently been conducted regarding red flags for the low back pain field. For instance, in 2020, the IFOMPT released a framework to clarify the role of red flags in identifying serious pathology (47). Additionally, the Cochrane Collaboration published two systematic reviews of red flags to screen for cancer and fractures in patients with low back pain (45,46).

Strengths and limitations of the present systematic review

This study followed a rigorous methodology. Notably, we published a protocol with the study's objectives, the search strategy was comprehensive, including the consultation with experts in the neck pain field, and all the phases were performed independently by two authors. Nonetheless, this study has some limitations. First, we translated non-English and non-Italian guidelines using "DeepL Translate." DeepL is a software based on artificial intelligence that is highly precise in translating scientific papers (53). However, the translation would have probably been more accurate with the help of a human native speaker. Second, determining whether a paper should be classified as a guideline can be challenging. To decide if a document had to be considered a guideline, we employed the PEDro criteria for evidence-based clinical practice guidelines. However, even with these criteria and even though we consulted the top experts in the neck pain field asking them for additional guidelines we did not retrieve with our initial search, there is still a possibility that a guideline may have been misjudged as not being a guideline. Third, we determined the level of evidence on which the red flags recommendations are based using the 2011 Oxford Centre for Evidence-Based Medicine Levels of Evidence (27).

This classification system primarily focuses on study design rather than the quality or applicability of the evidence to clinical practice. As a result, we may have overlooked important nuances, particularly in the case of red flags based on lower-level or mechanism-based reasoning. Fourth, we assessed the strength of recommendations for screening for serious pathologies by referring directly to the descriptions in the guidelines (see Tab. 2). In some cases, such as the Bier et al. guidelines (32), the description of the strength of recommendation (e.g., "Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate") was unclear, as the guideline did not provide an effect estimate (i.e., diagnostic accuracy values). This shows that some guidelines have imprecise reporting, and the strength of the recommendation is often based on a general statement.

Conclusions

Our review observed significant heterogeneity in the red flags recommended in guidelines for neck pain, with a general lack of consensus between guidelines for which red flags to endorse. Most red flags were not supported by a reference or were supported only by mechanism-based reasoning. Also, evidence for the accuracy of recommended red flags was lacking, except for the Canadian C-spine rule for fractures. Addressing the gaps in the current literature is a mainstay for future research. This includes conducting secondary studies to systematically summarize the available red flags and primary studies to determine the diagnostic accuracy of signs and symptoms that may suggest a serious medical condition. According to the current limitations of the evidence, specific recommendations on which red flags to use cannot be provided, except for using the Canadian C-Spine rule for screening posttraumatic fractures. Therefore, clinicians should use the red flags mentioned in the guidelines cautiously and integrate them into a sound clinical reasoning process.

Disclosures

Conflict of interest: The authors declare no conflict of interest.

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

Authors contribution: DF: Conceptualization, Study selection, Data extraction, Data analysis, Writing; AC: Conceptualization, Study selection, Writing; BK: Conceptualization, Writing; FMa: Conceptualization, Data extraction, Writing; FMO: Conceptualization, Study selection, Writing.

Data availability statement: Not applicable. The data presented in this study are available as supplementary material to this article.

References

1. Côté P, Wong JJ, Sutton D, et al. Management of neck pain and associated disorders: a clinical practice guideline from the Ontario Protocol for Traffic Injury Management (OPTIMA) Collaboration. *Eur Spine J.* 2016;25(7):2000-2022. [CrossRef](#) [PubMed](#)
2. Safiri S, Kolahi AA, Hoy D, et al. Global, regional, and national burden of neck pain in the general population, 1990-2017:



- systematic analysis of the Global Burden of Disease Study 2017. *BMJ*. 2020;368:m791. [CrossRef PubMed](#)
3. Cieza A, Causey K, Kamenov K, Hanson SW, Chatterji S, Vos T. Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet*. 2021;396(10267):2006-2017. [CrossRef PubMed](#)
 4. Rushton AB, Verra ML, Emms A, et al. Development and validation of two clinical prediction models to inform clinical decision-making for lumbar spinal fusion surgery for degenerative disorders and rehabilitation following surgery: protocol for a prospective observational study. *BMJ Open*. 05 22;8(5):e021078. [CrossRef](#)
 5. Mourad F, Giovannico G, Maselli F, Bonetti F, Fernández de las Peñas C, Dunning J. Basilar impression presenting as intermittent mechanical neck pain: a rare case report. *BMC Musculoskelet Disord*. 2016;17(1):7. [CrossRef PubMed](#)
 6. Mourad F, Giudice A, Maritati G, et al. A guide to identify cervical autonomic dysfunctions (and associated conditions) in patients with musculoskeletal disorders in physical therapy practice. *Braz J Phys Ther*. 2023;27(2):100495. [CrossRef PubMed](#)
 7. Faletra A, Bellin G, Dunning J, et al. Assessing cardiovascular parameters and risk factors in physical therapy practice: findings from a cross-sectional national survey and implication for clinical practice. *BMC Musculoskelet Disord*. 2022;23(1):749. [CrossRef PubMed](#)
 8. Maselli F, Piano L, Cecchetto S, Storari L, Rossetti G, Mourad F. Direct access to physical therapy: should Italy move forward? *Int J Environ Res Public Health*. 2022;19(1):555. [CrossRef PubMed](#)
 9. Maselli F, Storari L, Mourad F, Barbari V, Signorini M, Signorelli F. Headache, loss of smell, and visual disturbances: symptoms of SARS-CoV-2 infection? A case report. *Phys Ther*. 2023;103(4):pzad017. [CrossRef PubMed](#)
 10. Mourad F, Cataldi F, Patuzzo A, et al. Craniopharyngioma in a young woman with symptoms presenting as mechanical neck pain associated with cervicogenic headache: a case report. *Physiother Theory Pract*. 2021;37(4):549-558. [CrossRef PubMed](#)
 11. Platzer P, Hauswirth N, Jandl M, Chatwani S, Vecsei V, Gaebler C. Delayed or missed diagnosis of cervical spine injuries. *J Trauma*. 2006;61(1):150-155. [CrossRef PubMed](#)
 12. Sizer PS Jr, Brismée JM, Cook C. Medical screening for red flags in the diagnosis and management of musculoskeletal spine pain. *Pain Pract*. 2007;7(1):53-71. [CrossRef PubMed](#)
 13. 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. [CrossRef PubMed](#)
 14. Rushton A, Carlesso LC, Flynn T, et al. International framework for examination of the cervical region for potential of vascular pathologies of the neck prior to musculoskeletal intervention: International IFOMPT Cervical Framework. *J Orthop Sports Phys Ther*. 2023;53(1):7-22. [CrossRef PubMed](#)
 15. Heick J, Lazaro RT, eds. Goodman and Snyder's differential diagnosis for physical therapists: screening for referral. 7th ed. Elsevier; 2022.
 16. Feller D, Giudice A, Faletra A, et al. Identifying peripheral arterial diseases or flow limitations of the lower limb: important aspects for cardiovascular screening for referral in physiotherapy. *Musculoskelet Sci Pract*. 2022;61:102611. [CrossRef PubMed](#)
 17. Feller D, Giudice A, Maritati G, et al. Physiotherapy screening for referral of a patient with peripheral arterial disease masquerading as sciatica: a case report. *Healthcare (Basel)*. 2023;11(11):1527. [CrossRef PubMed](#)
 18. Mourad F, Lopez G, Cataldi F, et al. Assessing cranial nerves in physical therapy practice: findings from a cross-sectional survey and implication for clinical practice. *Healthcare (Basel)*. 2021;9(10):1262. [CrossRef PubMed](#)
 19. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *The BMJ*. [Online](#). (Accessed August 2024)
 20. Feller D, Chiarotto A, Koes B, Maselli F, Mourad F. Red flags for potential serious pathologies masquerading as musculoskeletal neck pain: a protocol for a systematic review of clinical practice guidelines. 2023.06.20.23291691. [CrossRef](#)
 21. Guzman J, Hurwitz EL, Carroll LJ, et al; Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. A new conceptual model of neck pain: linking onset, course, and care. *Spine*. 2008;33(4)(suppl):S14-S23. [CrossRef PubMed](#)
 22. PEDro. Physiotherapy evidence database. Indexing criteria and codes. [Online](#). (Accessed August 2024)
 23. Parikh P, Santaguida P, Macdermid J, Gross A, Eshtiahi A. Comparison of CPG's for the diagnosis, prognosis and management of non-specific neck pain: a systematic review. *BMC Musculoskeletal Disorders*. 2019;20(1):N.PAG-N.PAG. [CrossRef](#)
 24. Corp N, Mansell G, Stynes S, et al. Evidence-based treatment recommendations for neck and low back pain across Europe: a systematic review of guidelines. *Eur J Pain*. 2021;25(2):275-295. [CrossRef PubMed](#)
 25. Guimarães NS, Ferreira AJF, Ribeiro Silva RC, et al. Deduplicating records in systematic reviews: there are free, accurate automated ways to do so. *J Clin Epidemiol*. 2022;152:110-115. [CrossRef PubMed](#)
 26. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan – a web and mobile app for systematic reviews. *Syst Rev*. 2016;5(1):210. [CrossRef PubMed](#)
 27. Howick J, Chalmers I, Glasziou P, et al. Explanation of the 2011 Oxford Centre for Evidence-Based Medicine (OCEBM) Levels of Evidence (Background Document). [Online](#). (Accessed August 2024)
 28. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull*. 1971;76(5):378-382. [CrossRef](#)
 29. Moore A, Jackson A, Jordan J, et al. Clinical guidelines for the physiotherapy management of whiplash associated disorder. Chartered Society of Physiotherapy; 2005.
 30. Papic. Australian Clinical Guidelines for Health Professionals Managing People with Whiplash-Associated Disorders, 4th ed. [Online](#) (Accessed August 2024)
 31. Sterling M, TRACsa: Trauma and Injury Recovery. Clinical guidelines for best practice management of acute and chronic whiplash-associated disorders – UQ eSpace. 2008. [Online](#) (Accessed August 2024)
 32. Bier JD, Scholten-Peeters WGM, Staal JB, et al. Clinical practice guideline for physical therapy assessment and treatment in patients with nonspecific neck pain. *Phys Ther*. 2018;98(3):162-171. [CrossRef PubMed](#)
 33. Whalen W, Farabaugh RJ, Hawk C, et al. Best-practice recommendations for chiropractic management of patients with neck pain. *J Manipulative Physiol Ther*. 2019;42(9):635-650. [CrossRef PubMed](#)
 34. Haute Autorité de Santé. Label de la HAS - Évaluation du patient atteint de cervicalgie et prise de décision thérapeutique en chiropraxie. [Online](#). (Accessed August 2024)
 35. Scherer M, Chenot JF. Pragmatic and effective treatment of painful neck. *Deutsche Gesellschaft für Allgemeinmedizin*. *MMW Fortschr Med und Familienmedizin*. 2006. [CrossRef](#)
 36. Monticone M, Iovine R, de Sena G, et al; Italian Society of Physical and Rehabilitation Medicine (SIMFER). The Italian Society of Physical and Rehabilitation Medicine (SIMFER) recommendations for neck pain. *G Ital Med Lav Ergon*. 2013;35(1):36-50. [PubMed](#)

37. Blanpied PR, Gross AR, Elliott JM, et al. Neck pain: revision 2017. *J Orthop Sports Phys Ther.* 2017;47(7):A1-A83. [CrossRef](#) [PubMed](#)
38. Bussi eres AE, Stewart G, Al-Zoubi F, et al. The treatment of neck pain-associated disorders and whiplash-associated disorders: a clinical practice guideline. *J Manipulative Physiol Ther.* 2016;39(8):523-564.e27. [CrossRef](#) [PubMed](#)
39. Karpinnen et al. Liikunta (yll apito lopetettu). *Duodecim.* 2017. [Online](#) (Accessed August 2024)
40. Nordin M, Carragee EJ, Hogg-Johnson S, et al. Assessment of neck pain and its associated disorders: results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *Spine (Phila Pa 1976).* 2008;33(4 Suppl):S101-122. [CrossRef](#) [PubMed](#)
41. Nordin M, Carragee EJ, Hogg-Johnson S, et al; Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. Assessment of neck pain and its associated disorders: results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *Spine.* 2008;33(4)(suppl):S101-S122. [CrossRef](#) [PubMed](#)
42. Zhao SS, Pittam B, Harrison NL, Ahmed AE, Goodson NJ, Hughes DM. Diagnostic delay in axial spondyloarthritis: a systematic review and meta-analysis. *Rheumatology (Oxford).* 2021;60(4):1620-1628. [CrossRef](#) [PubMed](#)
43. Saragiotto BT, Maher CG, Lin CC, Verhagen AP, Goergen S, Michaleff ZA. Canadian C-spine rule and the National Emergency X-Radiography Utilization Study (NEXUS) for detecting clinically important cervical spine injury following blunt trauma. *Cochrane Libr.* 2018;2018(4):CD012989. [CrossRef](#)
44. Maselli F, Rossettini G, Viceconti A, Testa M. Importance of screening in physical therapy: vertebral fracture of thoracolumbar junction in a recreational runner. *BMJ Case Rep.* 2019;12(8):e229987. [CrossRef](#) [PubMed](#)
45. Han CS, Hancock MJ, Downie A, et al. Red flags to screen for vertebral fracture in people presenting with low back pain. *Cochrane Database Syst Rev.* 2023;8(8):CD014461. [CrossRef](#) [PubMed](#)
46. Henschke N, Maher CG, Ostelo RW, de Vet HC, Macaskill P, Irwig L. Red flags to screen for malignancy in patients with low-back pain. *Cochrane Database Syst Rev.* 2013;2013(2):CD008686. [CrossRef](#) [PubMed](#)
47. Kranenburg HA, Kerry R, Taylor A, Mourad F, Puentedura E, Hutting N. Correspondence re: de Best et al. *J Physiother.* 2024;70(1):78. [CrossRef](#) [PubMed](#)
48. de Best RF, Coppieters MW, van Trijffel E, et al. Risk assessment of vascular complications following manual therapy and exercise for the cervical region: diagnostic accuracy of the International Federation of Orthopaedic Manipulative Physical Therapists framework (The Go4Safe project). *J Physiother.* 2023;69(4):260-266. [CrossRef](#) [PubMed](#)
49. Rutjes AW, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PM. Case-control and two-gate designs in diagnostic accuracy studies. *Clin Chem.* 2005;51(8):1335-1341. [CrossRef](#) [PubMed](#)
50. van Smeden M, Reitsma JB, Riley RD, Collins GS, Moons KG. Clinical prediction models: diagnosis versus prognosis. *J Clin Epidemiol.* 2021;132:142-145. [CrossRef](#) [PubMed](#)
51. Verhagen AP, Downie A, Maher CG, Koes BW. Most red flags for malignancy in low back pain guidelines lack empirical support: a systematic review. *Pain.* 2017;158(10):1860-1868. [CrossRef](#) [PubMed](#)
52. Verhagen AP, Downie A, Popal N, Maher C, Koes BW. Red flags presented in current low back pain guidelines: a review. *Eur Spine J.* 2016;25(9):2788-2802. [CrossRef](#) [PubMed](#)
53. Takakusagi Y, Oike T, Shirai K, et al. Validation of the reliability of machine translation for a medical article from Japanese to English using DeepL translator. *Cureus.* 2021;13(9):e17778. [CrossRef](#)