Red flag screening for low back pain: nothing to see here, move along: a narrative review

Chad E Cook,[1,2] Steven Z George,[2,3] Michael P Reiman[2,4]

[1]Division of Physical Therapy, Duke MSK, Duke Clinical Research Institute, Durham, North Carolina, USA
[2]Department of Orthopaedic Surgery, Duke University, Durham, North Carolina, USA
[3]Duke MSK, Duke Clinical Research Institute, Durham, North Carolina, USA
[4]Division of Physical Therapy, Duke University, Durham, North Carolina, USA

Correspondence to
Dr Chad E Cook, Department of Physical Therapy, Duke University, Durham, NC 27708, USA; chad.cook@duke.edu

ABSTRACT

Screening for red flags in individuals with low back pain (LBP) has been a historical hallmark of musculoskeletal management. Red flag screening is endorsed by most LBP clinical practice guidelines, despite a lack of support for their diagnostic capacity. We share four major reasons why red flag screening is not consistent with best practice in LBP management: (1) clinicians do not actually screen for red flags, they manage the findings; (2) red flag symptomology negates the utility of clinical findings; (3) the tests lack the negative likelihood ratio to serve as a screen; and (4) clinical practice guidelines do not include specific processes that aid decision-making. Based on these findings, we propose that clinicians consider: (1) the importance of watchful waiting; (2) the value-based care does not support clinical examination driven by red flag symptoms; and (3) the recognition that red flag symptoms may have a stronger relationship with prognosis than diagnosis.

BACKGROUND

Despite intense focus and increased research funding, the self-reported levels of disability in individuals with low back pain (LBP) have not improved in the last decade.[1] Worsening disability has propagated, despite the presence of numerous classification schemes designed to lead to patient-specific treatments. The care provided to patients frequently does not meet professionally recommended standards. We can do much better. We *have* to do better.

Diagnosis is one of many necessary components during the clinical decision-making process. Characteristically, diagnosis is performed early in the management process of the patient and involves both soft skills (clinical reasoning) and highly valid quantitative clinical testing methods. All medical professionals who manage LBP, regardless of training, background or philosophy, use differential diagnostic methods to improve the likelihood of providing the right care to the appropriate patient, and reducing risks associated with delayed management. By its nature, differential diagnosis is a systematic process used to identify the proper diagnosis from a competing set of possible diagnoses.[2 3] For example, a practitioner would perform more extensive diagnostic testing for an individual with LBP that also has clinical indication of cancer or infection compared with one without those indicators.

In one of the earliest works on differential diagnosis for LBP, published in 1924, John M Berry[4] wrote prophetically: '*Confronted with such a multiplicity of symptoms and causes, diagnosis usually is difficult or uncertain, and consequently treatment is unsatisfactory.*' Berry identified a number of conditions that led to LBP that were not affiliated with muscle or bone, and suggested the importance in classifying potentially serious pathologies for proper care. Berry[4] espoused a 'screening' process for *red flags*, which is a modern term used to describe signs or symptoms that are related to a serious underlying pathology, and may indicate more diagnostic testing is necessary before the appropriate care can be delivered. Astonishingly, it has taken more than 90 years to conclude that screening for red flags associated with LBP does not work. We discuss four primary reasons why red flag screening does not work and provide alternatives to consider in future LBP management models.

Reason 1: red flag symptoms neither rule out nor identify serious pathology

In a meta-analysis by Downie and colleagues,[5] nearly all patient history, physical examination findings and diagnostic clinical tests associated with 'screening' for LBP exhibited negative likelihood ratios (−LR) near 1.0, suggesting that none were truly sufficient in ruling out a condition with a negative finding. A −LR is calculated from the sensitivity and specificity values of a contingency table. A −LR is calculated by taking the probability that a person who has the disease and tests negative, and dividing by the probability of a person who does not have the disease and tests negative. In other words, a lower numerator over a higher denominator and more closely approximating zero provides you with a lower and 'better' −LR. Low −LRs allow a clinician to have confidence that a negative finding definitely indicates that the underlying condition assessed (eg, cancer, fracture) is indeed absent.

Downie and associates'[5] meta-analysis exhibited that the tests evaluated in their review failed to exhibit a low −LR (close to zero) needed to rule out, and failed to exhibit a high positive likelihood ratio (+LR) necessary to 'rule in' a condition. High +LR is generally used to confirm the presence of a condition once a comprehensive clinical examination is used to guide hypothesis testing. Also calculated from sensitivity and specificity, a +LR is the probability that a person who has the disease and tests positive divided by the probability of a person who does not have the disease and tests positive. In other words, a higher numerator divided by a lower denominator provides you with a higher and 'better' +LR. Based on the findings that both the −LR and +LR of the clinical tests were of limited clinical utility not diagnostic, the authors[5] warned of unwarranted referrals

Table 1 Post-test probabilities of vertebral fracture (radiograph is the reference standard)

Test	Pretest probability	−LR (95% CI)	Post-test probability
Age >74 years	0.5%[7] and 4%[8]	0.77 (0.52 to 1.15)	0.38%–3.10%
Significant trauma	0.5%[7] and 4%[8]	0.77 (0.52 to 1.15)	0.38%–3.10%
History of corticosteroid use	0.5%[7] and 4%[8]	0.75 (0.51 to 1.13)	0.37%–3.03%

−LR, negative likelihood ratios.

for imaging, as some of the endorsed red flag tests were very commonly positive.

In an effect to provide a clinical example, we propose the following scenario using data from the meta-analysis of Williams and colleagues,[6] who examined red flag screening for vertebral fracture in patients with LBP. A vertebral fracture is more common in older individuals (0.5%–4% of all patients with LBP) and clinicians are taught to screen by examining age, history of trauma and history of corticosteroid use. Supposing a pretest prevalence of 0.5%–4%[7][8] and using the values reported by Williams *et al*,[6] the post-test probabilities for ruling out improve by *less than 1%* when a negative finding occurs (table 1); a change in post-test probability that has marginal clinical utility. In addition, the 95% CIs cross 1.0 in sensibility and motor testing, further questioning their discriminative capacity.

Reason 2: variability in definitions for red flag symptoms greatly limits research and clinical progress in this area

In research, there is no reliable way to compare red flag symptom rates across different cohorts so we know very little about overall prevalence. Studies such as this are extremely difficult to conduct, as very large sample size is required based on overall low prevalence estimates. The influence variability has on clinical practice is that it continues to perpetuate inconsistency in assessment of red flag symptoms.[9]

The extent of variability was highlighted during the development of a standardised self-report tool for assessment of red flag symptomology.[10] A systematic literature search was performed to identify the types of red flag symptomatology items used in prior studies as part of the creation of the OSPRO-ROS (Optimal Screening for Prediction of Referral and Outcome-Review of Systems) tool.[10] The OSPRO-ROS tool was developed with data from a physical therapy cohort receiving care for a variety of musculoskeletal pain conditions (including LBP). The search yielded 97 unique items representing symptoms from eight body systems (eg, cardiovascular, pulmonary or gastrointestinal). Subsequent analyses focused on the number of items needed to identify a 'red flag symptom responder' (operationally defined as a positive response to *any* of the 97 items) and 10 items had 94.7% accuracy while 23 items were needed to reach 100% accuracy. The development of the OSPRO-ROS tool was the first attempt we are aware of to standardise assessment of red flag symptomology, and it suggests that the variation in this process can be greatly reduced, making standardisation a potentially important first step in gaining better consistency on inquiry and reporting for research and clinical purposes.

Reason 3: LBP guidelines do not help

Underwood and Buchbinder[11] suggest that screening for red flags in patients with LBP is 'a popular idea that didn't work and should be removed from guidelines'. The vast majority of current clinical practice guidelines for back pain recommend the use of

red flags for determining potential presence of spinal fracture or malignancy.[12] Those that then present with these suggested red flags are recommended to undergo more extensive, costly and potentially harmful diagnostic testing. As mentioned earlier, extensive diagnostic testing has a tendency to lend to higher costs, and to potential overtreatment, elicitation of unnecessary fear and concern on part of the patient, as well as a significant potential for being cost prohibitive.

An overview of clinical guidelines for primary care non-specific LBP management[13] revealed eight different guidelines endorsing 27 red flags for malignancy and 26 for fracture. More distressing was the fact that none of these eight guidelines endorsed the same set of red flags for either condition! The lack of consistency does not provide the clinician with a stable set of 'rules' to follow to perform a screening test adequately for their patients, and is therefore detrimental to good practice. Additionally, despite clinicians generally defining red flags for back pain in accordance with guidelines, there is little consensus on consistency of inquiry and reporting red flags.[9]

Guidelines describing variable and numerous red flags without reported diagnostic accuracy are undermined by guidelines recommending immediate referral for imaging if any red flag is present.[14][15] A list of potential red flags, such as 'age <45 years, night/morning pain, slow beginning, rigidity duration over 3 months, family history of spondyloarthritis, ulcerating colitis, Crohn's disease, psoriasis' would lend to significant overutilisation of imaging, and to inappropriate clinician clinical reasoning.[14] For example, in the aforementioned study of developing the OSPRO-ROS tool, 393/431 participants (91.2%) reported at least one red flag symptom.[5] Red flags of 'age <22 or >55 years'[15] should not necessitate the clinical concern that 'history of malignancy or immune compromise'[15] is present based on face value alone. Additionally, when red flags are present, the recommendation that plain radiographs alone can explain the 'cause for pain found' is simply not the type of clinical reasoning to use when such decision-making processes are not universally supported by the literature.[16]

Let us not confuse the use of technology associated with imaging and lab work with an improvement in differential diagnosis, since there is notable overuse of inappropriate imaging and identification of false positives in those with LBP. No clinical guidelines for LBP support the use of early imaging, and as stated previously, the use of imaging when a red flag is present has led to notable overimaging.[17] In a study of 1003 original patients with LBP referred from four primary care clinics, 110 had at least one red flag (75% of which had a single red flag). These 110 underwent advanced imaging (eg, CT, MRI, and so on). Twenty-four of these individuals had a non-benign spinal disorder. Among these 24 individuals, 50% had correlating clinical neurological findings reported, leaving only 12 cases of highest priority for advanced imaging. Additionally, in the 893 patients who did not have advanced imaging done and were followed for approximately 1 year, none had a non-benign spinal disorder. Unfortunately, most had radiographs prior to implementation of study. The prevalence of many serious low back pathologies is extremely low, including the reported prevalence of <1% for spinal malignancy in patients with LBP presenting to a primary care physician.[18]

Reason 4: clinicians do not actually screen for red flags; they manage LBP conditions they see

Medical screening is a process in which a disease or a condition is assessed in an asymptomatic population who may or may

not have early disease or disease precursors.[19] Test results are then used to guide whether or not a diagnostic test should be offered. Medical screening is typically performed in the preclinical phase of a condition, a time frame in which there are no outward symptoms. Beneficial medical screening tests are able to identify findings that are hallmarks of a serious pathology within the preclinical phase.

In contrast, diagnostic testing involves clinical testing that is designed to aid in the diagnosis or detection of a suspected disease or condition when symptoms exist. In an ideal diagnostic scheme, the symptoms help to guide identification of the underlying pathology. In the case of LBP, symptoms offer little to no guidance in detecting underlying pathology.

Although recognised to encompass a set of diagnoses within the WHO disease classification,[20] LBP is actually a symptom; and typically, a symptom associated with an unknown underlying condition. LBP not affiliated with a serious pathology will often exhibit symptoms that are similar to competing diagnoses such as fracture, cancer and other red flags. Many of the red flags associated with LBP are more prevalent in older individuals,[21] a subset of individuals who will also frequently have concomitant orthopaedic-related LBP.[22] In a study designed to identify movement examination features unique to metastatic spinal cancer, 61 of 66 individuals with LBP were diagnosed with metastatic bone cancer and (concomitantly) were diagnosed with a condition such as lumbar stenosis, spondylosis or degenerative disc disease.[22] No unique movements discriminated metastatic spinal cancer from other age-related conditions. *Rarely does the literature outline a definitive set of signs or symptoms that are unique to serious pathology of the low back—for either the screening OR the diagnostic phase.*

To recapitulate, clinicians managing patients with LBP do not actually screen for red flags. Medical screening is typically performed in the preclinical phase of a condition, and red flag symptoms for LBP are captured in the clinical phase. For clinicians to truly understand and change their clinical reasoning processes, educational curricula need to be revised. Similar suggestions have also been provided related to changing one's emphases on diagnosis to understanding the importance of prognosis.[23]

SUMMARY AND THREE RECOMMENDATIONS
For those who fear that a philosophical change in screening for LBP may result in detrimental patient outcomes (harms) or marginal benefits, we offer the following options: (1) watchful waiting as an alternative to diagnostic testing, (2) enhancement of value-based care when cost is considered, and (3) linking red flag symptomology directly with outcomes.

Watchful waiting
Early use of diagnostic tests is often used to provide reassurance to patients. After careful evaluation of studies that have used early diagnostic testing methods, van Ravesteijn and colleagues[24] supported the practice of providing a clear explanation and watchful waiting. In medicine, watchful waiting is the act of close surveillance, but allowing time to pass before medical intervention or therapy is used. There is evidence that early intervention may actually be detrimental/harmful in patients with LBP.[25–27] Moreover, the incidence of several diseases that tend to be detected early is increasing with no corresponding reductions in mortality rate.[28] Hence, the surge in attention on early detection and the focus on benefits over harms do not seem justified.[29]

We propose *careful monitoring for changes in symptoms over time*. Careful monitoring, performed concurrently with open and timely communication with the patient regarding potential contributing factors to their symptoms, can not only be as (or more) effective than early testing. Watchful waiting may also improve patient–provider relationship, improve clinician clinical reasoning/decision-making, improve patient satisfaction and anxiety[30] and be one small step forward for improvement of exorbitant healthcare costs. Tangentially, a 'wait and see' approach has proven effective in having patients avoid major oncological surgery, averting permanent alterations in lifestyle without loss of oncological safety.[31]

Enhancement of value-based care
An often-overlooked consideration relative to screening in musculoskeletal pain is cost; and whether screening adds to the overall value of a particular care episode.[32] In a comparison of diagnostic strategies for diagnosis of cancer in patients with LBP, MRI used as a first-line diagnostic strategy was shown to cost 10 times as much as the conventional strategy (MRI if erythrocyte sedimentation rate elevated and radiographs were positive). Each extra patient with a spine malignancy in the MRI group exceeded US$625 000.[33] These costs are likely of greater disparity today.

Although (to our knowledge) no clinical practice guidelines address value-based care within their recommendations, we propose this should be an adopted consideration before ordering tests for red flags. We feel value-based care is a small but important step towards managing unnecessary and potentially unwarranted care for LBP, one that has strong support within the literature.

Linking red flag symptomology not with diagnostic testing but directly with health status
Early diagnostic imaging testing is very likely to overinterpret potential pathology in patients with LBP with as high as 94% of patients presenting to a general practice office presenting with 'abnormal' MRI findings, yet only 3% of these individuals had actual serious pathology.[34] The 3% rate was three times higher than another study of younger patients with more acute LBP presenting to general practice,[35] demonstrating that even in older patients with chronic LBP the prevalence is extremely low. Additionally, in both studies, the primary serious pathology presentation was vertebral fracture (3%) for which there are good clinical screening tests. Unwarranted diagnostic imaging leads to higher healthcare costs, unnecessary, potentially invasive interventions and, as discussed in reason 2, no improvement in diagnostic screening accuracy.

In light of our above proposal for enhanced value-based care and in place of using red flag symptoms to drive more diagnostic testing, perhaps a change paradigm is necessary and the symptoms themselves can be used to refine recommendations for care pathways and outcome prediction. For example, Roach et al[36] used red flag symptomology to predict whether LBP would result in surgical or conservative care and George et al[37] have used the OSPRO-ROS tool to predict change in 12-month comorbidity status. These alternative models of using red flags deserve further exploration and may be appropriate for a value-based era of musculoskeletal care.[32]

What are the findings

- ► Screening requires preclinical phase of a condition, a time frame in which there are no outward symptoms. As such, red flag screening for low back pain is actually a management strategy that does not involve screening methodology.
- ► Clinical tests for red flags do not exhibit low negative likelihood ratios, suggesting that they fail to 'rule out' a red flag when a negative finding is present with the test.
- ► Red flag findings are more closely affiliated with prognosis than they are to actual serious pathology such as cancer, fracture, and so on.
- ► Although low back pain clinical practice endorses assessing red flags, only a few outline which findings are useful for clinical practice.

Contributors CEC, SZG and MPR all worked together for the concept of the study, participated in the full writing and approved the final manuscript.

REFERENCES

1 GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 2016;388:1545–602.
2 Cook C. Clinical assessment, moving from rote to rigorous. ed. *Brukner and khan's clinical sports medicine*. 5th. Australia: McGraw Hill, 2017.
3 Reiman MP. *Orthopedic Clinical Examination*. Champaign, IL: Human Kinetics, 2016.
4 Berry JM. Painful conditions in the lumbar, lumbosacral and sacro-iliac regions. *Arch Surg* 1925 archsurg.jamanetwork.com.
5 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.
6 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:CD008643.
7 Suarez-Almazor ME, Belseck E, Russell AS, *et al*. Use of lumbar radiographs for the early diagnosis of low back pain. Proposed guidelines would increase utilization. *JAMA* 1997;277:1782–6.
8 Chou R, Qaseem A, Snow V, *et al*. Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American college of physicians and the American pain society. *Ann Intern Med* 2007;147:478–91.
9 Ferguson FC, Morison S, Ryan CG. Physiotherapists' understanding of red flags for back pain. *Musculoskeletal Care* 2015;13:42–50.
10 George SZ, Beneciuk JM, Bialosky JE, *et al*. Development of a review-of-systems screening tool for orthopaedic physical therapists: results from the optimal screening for prediction of referral and outcome (OSPRO) cohort. *J Orthop Sports Phys Ther* 2015;45:512–26.
11 Underwood M, Buchbinder R. Red flags for back pain. *BMJ* 2013;347:f7432.
12 O'Connell NE, Cook CE, Wand BM, *et al*. Clinical guidelines for low back pain: A critical review of consensus and inconsistencies across three major guidelines. *Best Pract Res Clin Rheumatol*. In Press. 2017.
13 Koes BW, van Tulder M, Lin CW, *et al*. An updated overview of clinical guidelines for the management of non-specific low back pain in primary care. *Eur Spine J* 2010;19:2075–94.
14 Negrini S, Giovannoni S, Minozzi S, *et al*. Diagnostic therapeutic flow-charts for low back pain patients: the Italian clinical guidelines. *Eura Medicophys* 2006;42:151–70.
15 Health GoWADo. Diagnostic imaging pathways-low back pain. 2013 http://www.imagingpathways.health.wa.gov.au/index.php/imaging-pathways/musculoskeletal-trauma/musculoskeletal/low-back-pain#pathway (accessed 10 May 2017)
16 Brinjikji W, Luetmer PH, Comstock B, *et al*. Systematic literature review of imaging features of spinal degeneration in asymptomatic populations. *AJNR Am J Neuroradiol* 2015;36:811–6.
17 Ferrari R. Imaging studies in patients with spinal pain: practice audit evaluation of choosing wisely canada recommendations. *Can Fam Physician* 2016;62:e129–37.
18 Deyo RA, Diehl AK. Cancer as a cause of back pain: frequency, clinical presentation, and diagnostic strategies. *J Gen Intern Med* 1988;3:230–8.
19 Lewis G, Sheringham J, Kalim K, *et al*. *Mastering public health: a postgraduate guide to examinations and revalidation*. London: The royal society of medicine press limited, 2008.
20 World Health Organization. International classification of diseases (ICD) information sheet. http://www.who.int/classifications/icd/factsheet/en/ (accessed 13 May 2015)
21 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;96:305–12.
22 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;12:426–33.
23 Croft P, Dinant GJ, Coventry P, *et al*. Looking to the future: should 'prognosis' be heard as often as 'diagnosis' in medical education? *Educ Prim Care* 2015;26:367–71.
24 van Ravesteijn H, van Dijk I, Darmon D, *et al*. The reassuring value of diagnostic tests: a systematic review. *Patient Educ Couns* 2012;86:3–8.
25 GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the global burden of disease study 2013. *Lancet* 2015;385:117–71.
26 Fitzmaurice C, Allen C, Barber RM, *et al*. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: a systematic analysis for the global burden of disease study. *JAMA Oncol* 2017;3:524–48.
27 Welch HG, Schwartz L, Woloshin S. *Overdiagnosed: making people sick in the pursuit of heath*. Boston: Beacon Press, 2011.
28 Saquib N, Saquib J, Ioannidis JP. Does screening for disease save lives in asymptomatic adults? Systematic review of meta-analyses and randomized trials. *Int J Epidemiol* 2015;44:264–77.
29 Hofmann B, Skolbekken JA. Surge in publications on early detection. *BMJ* 2017;357:j2102.
30 Jung HP, Wensing M, Grol R. What makes a good general practitioner: do patients and doctors have different views? *Br J Gen Pract* 1997;47:805–9.
31 Peck BM, Asch DA, Goold SD, *et al*. Measuring patient expectations: does the instrument affect satisfaction or expectations? *Med Care* 2001;39:100–8.
32 Ross MD, Boissonnault WG. Red flags: to screen or not to screen? *J Orthop Sports Phys Ther* 2010;40:682–4.
33 Lentz TA, Harman JS, Marlow NM, *et al*. Application of a Value Model for the Prevention and Management of Chronic Musculoskeletal Pain by Physical Therapists. *Phys Ther* 2017;97:354–64.
34 Joines JD, McNutt RA, Carey TS, *et al*. Finding cancer in primary care outpatients with low back pain: a comparison of diagnostic strategies. *J Gen Intern Med* 2001;16:14–23.
35 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;60:3072–80.
36 Roach KE, Brown M, Ricker E, *et al*. The use of patient symptoms to screen for serious back problems. *J Orthop Sports Phys Ther* 1995;21:2–6.
37 George SZ, Beneciuk JM, Lentz TA, *et al*. The optimal screening for prediction of referral and outcome (OSPRO) in patients with musculoskeletal pain conditions: a longitudinal validation cohort from the USA. *BMJ Open* 2017;7:e015188.