Brazilian Journal of Physical Therapy

https://www.journals.elsevier.com/brazilian-journal-of-physical-therapy

MASTERCLASS

Higher order thinking about differential diagnosis

Chad E. Cook [a],[*], Simon Décary [b]

[a] *Division of Physical Therapy, Duke Clinical Research Institute, Duke University, Durham, NC, United States*
[b] *Canada Research Chair in Shared Decision Making and Knowledge Translation, Université Laval, Faculty of Medicine, Québec, Canada*

Abstract
Background: Differential diagnosis is a systematic process used to identify the proper diagnosis from a set of possible competing diagnoses.
Methods: The goal of this masterclass is to discuss the higher order thinking components of differential diagnosis.
Conclusions: For healthcare providers, diagnosis is one of many necessary components during the clinical decision making process and it is hallmarked by differentiation of competing structures for a definitive understanding of the underlying condition. The diagnostic process involves identifying or determining the etiology of a disease or condition through evaluation of patient history, physical examination, and review of laboratory data or diagnostic imaging; and the subsequent descriptive title of that finding. Whereas differential diagnosis is a varied skill set among all healthcare providers, the concept of a diagnosis is equally germane, regardless of one's background. In theory, a diagnosis improves the use of classification tools, improves clarity and communication, provides a trajectory of treatment, improves understanding of a person's prognosis, and in some cases, may be useful for preventative treatments. To achieve these improvements, one must have an understanding of relation of the clinical utility of tests and measures with diagnosis, and how to best implement these findings in clinical practice. This requires a deeper understanding (higher order thinking) of the role of diagnosis in the management of patients.
© 2019 Associação Brasileira de Pesquisa e Pós-Graduação em Fisioterapia. Published by Elsevier Editora Ltda. All rights reserved.

* Corresponding author at: 2200 W. Main Street, Suite B230, Durham, NC 27710, United States.
  E-mail: chad.cook@duke.edu (C.E. Cook).

The where, when and why of diagnosis

Background

The diagnostic process involves identifying or determining the etiology of a disease or condition through evaluation of patient history, physical examination, and review of laboratory data or diagnostic imaging; and the subsequent descriptive title of that finding.[1] Diagnoses are used to improve communication with patients, across providers, for payers, and with health systems. According to Walker,[2] within 3000 years, there have been a number of crucial progressions that have shaped how diagnoses are used today. These include the use of a rational basis for establishing medicine as a profession, the development of equipment to aide in diagnosis, the use of autopsy for confirmation of diagnostic findings, the dissection of human bodies for educational purposes, the growth of physical and laboratory examination findings, and the classification of diagnostic commonalities.

In 1893, the first international classification edition, which was known as the International List of Causes of Death (ICD),[3] was adopted by the International Statistical Institute. In May of 2018, the World Health Organization (WHO) released the 11th Revision to the ICD system. This system aims to homogenously define diseases, disorders, injuries and health conditions. The ICD classification organizes information into standard groupings of diseases, which allows for: easy storage, retrieval and analysis of health information for evidenced-based decision-making; sharing and comparing health information between hospitals, regions, settings and countries; and data comparisons in the same location across different time periods. Further, the coding system allows a higher level of specificity and clinical detail, improving the ability to document a patient encounter and compare outcomes at a larger, system-based level. At its core, the ICD diagnostic system improves communication among providers of healthcare and should be considered a minimum competency for healthcare diagnosticians.

Improved communication and a common language describing categories of diseases are a useful feature in differential diagnosis. However, understanding how to use the disease categories and recognizing the shortcomings of diagnostic labeling requires *higher order thinking*. Higher order thinking is posited on the concept that some forms of learning require greater cognitive processing and indirectly require different forms of learning beyond memorization, facts and, concepts. Higher order cognitive skills include conceptualization, analysis, and evaluation, and involves ordered levels of reasoning containing productive thinking, or reasoning, versus learned, or reproductive, thinking.[4] Fundamental skills involved with higher order thinking includes analogical and logical reasoning.[5] Analogical reasoning involves any type of thinking that relies upon an analogy and involves analyzing sameness. Logical reasoning involves drawing on prior knowledge to make inferences and solve problems. Critical thinking is a component of higher order thinking.

A basic understanding of diagnoses is a complex, iterative and necessary process. In this masterclass, we argue that higher order thinking goes well beyond memorization of tests and measures, sensitivity and specificity, and a laundry list of ICD-based diagnoses. Specifically, we argue that for higher order differential diagnostic reasoning, a clinician requires careful attention to: (1) *how test metrics can be misleading*; (2) *how diagnostic label may overcomplicate care*; and (3) *how using different methods of classifying diagnoses could improve management*.

How test metrics can be misleading

Interpreting test metrics

In order to obtain a diagnosis for a patient (determine the presence or absence of a disorder), clinicians rely on diagnostic tests such as clinical examination or imaging. The main characteristic of a diagnostic study is the comparison of a test (or combination of tests) called the ''index test'' to a known ''reference standard''. This produces the test metrics.[6,7]

Table 1 provides the most commonly used test metrics in diagnostic assessment. Sensitivity (SN) and Specificity (SP) values are calculated only on defined population within a case-based, case control design. For example, SN is only calculated on patients who have the disease of interest. SP in only calculated on the population who does not have the disease of interest. Like SN and SP, Positive Predictive Value (PPV) and Negative Predictive Value (NPV) are only calculated from a proportion of the population within a case-based, case control design. PPV is calculated only from those who scored a ''positive'' on the test finding, whereas NPV is calculated only from those who score a ''negative'' on the test finding. SN, SP, PPV, and NPV are considered internal test metrics and should not to be used for post-test decision-making.

Likelihood ratios (LR) are calculated from the full population represented in the case-based, case control design thus are considered metrics that influence clinical utilities (the ability to make rational diagnostic decisions). A LR+ above 1.0 influences post-test probability with a positive finding, whereas a low LR— (a value close to 0) influences post-test probability with a negative finding.[8] Both values are wedded to pretest probability and can be used to determine the post-test probability of a given diagnosis; whether ruling in or ruling out. Benchmark values have been provided to give clinicians perspective of individual meaningful levels. For example, LR+ >5 and LR— <0.2 is stated to moderately increase positive or negative post-test probability by approximately 30%.[9] But in reality, each likelihood ratio is dependent on the pretest probability and should be considered individually to guide decision-making.

The current diagnostic system is based on interpreting these metrics for a given test to identify the most likely diagnostic label for a patient, but interpreting these metrics have many pitfalls. As previously stated, SN and SP and PPV and NPV are internal test metrics and are not used independently for decision-making. Indeed, individual values can be misleading because they do not represent the full population that one may experience in a clinic. The concepts of SPin and SNout are purported to rule in with high specificity and rule out with high sensitivity; however, this concept is

Table 1 Common test metrics for differential diagnosis.

Metric	Abbreviation	Definition
Sensitivity	SN	Percentage of people who test positive for a specific disease among a group of people who have the disorder.
Specificity	SP	Percentage of people who test negative for a specific disease among a group of people who do not have the disorder.
Positive Predictive Value	PPV	Probability that subjects with a positive test truly have the disorder.
Negative Predictive Value	NPV	Probability that subjects with a negative test truly don't have the disorder.
Positive Likelihood Ratio	LR+	The odds of a patient to have a disorder if the test is positive compared to the probability for someone who does not have the disorder.
Negative Likelihood Ratio	LR—	The odds of a patient not having the disorder if the test is negative compared to the probability for a patient who has the disorder.

outdated and can lead to interpretation errors.[9] For example, a study on the diagnosis of subacromial pain demonstrated that combining three clinical features reached 100% SP, but only 9% SN.[10] Even if a patient with three clinical features most likely have a diagnosis of sub-acromial pain, this cluster will only identify 9% of all patients with this diagnosis, making this ''perfect scheme'' rare in clinical practice. This may lead to biases if a clinician seeks to identify patients with this combination. For the concepts of SPin and SNout to work, the ''other'' metrics must be at a reasonable level to reduce the risk of errors in decision-making.

Likelihood ratios have proven useful since they inform on the magnitude of change in post-test probability, but this can also mislead decision-making processes. For example, the accuracy of the Lachman test for anterior cruciate ligament (ACL) tear was established in primary care and orthopedic cohorts.[11,12] However, in primary care cohorts, the prevalence (pretest probability) of ACL tears is approximately 4%, while it is closer to 20–25% in orthopedic clinics since primary care cohorts include a diversity of knees (e.g. contusions and sprains) that (often) do not make their way to secondary care clinics.[11] Even if this test demonstrates a high LR+,[13] the post-test probability when applied to cohorts with different prevalence may be significantly different, which may impact uncertainty and decisions about further imaging or surgical references. Prevalence can influence the assessment of red flags as well. Authors have recently questioned the utility of history elements to rule out serious cause of low back pain (poor negative LR)[14,15] because of the exceptionally low prevalence.

Study design and condition severity can influence outcomes

Interpretation of test metrics is also highly dependable on the quality of the evidence about the test. For example, the Thessaly test for meniscal tears was initially developed in a study with a low-quality design and the results were not replicated afterward.[6,16,17] Clinicians should look closely at study designs, reference standards and how the tests were described as these have most commonly influenced biases in diagnostic accuracy studies.[18] Further, the severity of the population group may influence outcomes. Conditions that are advanced and have high degrees of disability and pain will exhibit test results that are more sensitive and less specific. Conditions with low disability and pain will exhibit lower sensitivity and higher specificity.

Impact on decision-making processes

Decision-making models propose a balance between the analytical approach based on evidence (e.g., test metrics), and the intuitive approach that relies on the experience of the evaluator.[19] Clinicians face the daily challenge of avoiding pitfalls when interpreting diagnostic test results. All tests, clinical examination or imaging, have strength and weaknesses. Flaws in interpreting the accuracy of a test, poor understanding of probabilities and low-quality evidence can derail the analytical process.[20] The intuitive process can be overridden by verification or confirmation biases such as anchoring or premature closure in which a favored diagnosis is found and the clinician stop the diagnostic process too early when the scheme seems fitting.[20]

Ultimately, the results of diagnostic tests lead clinicians to initiate decisions about further tests and treatments. Therefore, clinical reasoning is of paramount importance to link the test results to an appropriate management plan in a complete care pathway. Higher order thinking demands that a clinician move beyond test metrics and reflect on misclassification costs and how the decision may influence downstream utilization of health care.

Take home message: Most test metrics are internal metrics and are not used to determine post-test probability. Metrics can be biased by study design and patient severity as well. Even those used for post-test probability such as likelihood ratios must be used in a full understanding of how pretest might influence the outcomes.

How a diagnostic label may overcomplicate care

The creation of diagnostic codes from a patho-anatomical viewpoint has led to focus on tissue based musculoskeletal

disorders. Here we argue that identifying and classifying patients based on this model can lead to overcomplicated or asymptomatic diagnostic labeling that may not translate into better patient outcomes.

Overuse of diagnostic tests and overdiagnosis in musculoskeletal disorders

Most fields in medicine have relied on diagnostic tests and metrics to inform clinical decision-making.[21] However, it is now recognized that this overreliance on diagnostic labeling in medicine is a driver of overuse of diagnostic tests and overdiagnosis. Overdiagnosis occurs when a patient receives a diagnostic label that may have never caused them harm,[22] such as when diagnostic tests identify abnormalities or risk factors that most often will not cause symptoms or impairments.[23] Thus, the core of overdiagnosis is strongly related to the definition of diagnostic labeling and interpretation of test metrics.

When a patient presents with symptoms of pain to the spine, knee, hip or shoulder, clinicians often initiate a cascade of history questions, physical examination tests, clinical measures and imaging tests in order to diagnose the sources of the symptoms.[24] Musculoskeletal care is plagued by overuse of diagnostic tests. Up to 50% of all imaging test referrals are considered inappropriate.[25] Musculoskeletal disorders are particularly prone to overdiagnosis given the documented high prevalence of asymptomatic structural deficits seen on imaging. Examples of such labels include ''lumbar degeneration'', ''disk bulges'',[26] ''disk herniation'',[27] ''degenerative meniscal tears'',[28] ''degenerative labral tears'',[29] ''subacromial bursal thickening''[30] or ''rotator cuff tendinosis''.[30]

From a clinical pathway perspective, overuse of diagnostic tests and overdiagnosis can trigger downstream utilization of potentially inappropriate treatments such as orthopedic surgery, opioids overprescription, or early and intensive rehabilitation protocols as first-line options.[24,31] Differentiating between specific patho-anatomic diagnoses may not be relevant to choose appropriate first-line options. We need to investigate whether diagnostic methods improve patients' outcomes.

Trials linking diagnostic tests to patients' outcomes

The evidence linking diagnostic tests to improve patient outcomes in musculoskeletal disorders is scarce. A meta-analysis explored the effect of routine diagnostic imaging on patient reported outcomes for patients with musculoskeletal disorders.[32] The authors found 11 trials for low back pain and knee complaints that provided moderate evidence that using routine diagnostic imaging was not beneficial to improve pain.[32] A trial showed that replacing spine radiographs by early magnetic resonance imaging (MRI) in primary care did not lead to improved back-related disabilities while increasing cost and possibly the number of spine surgery based on MRI findings.[27] A trial found that patients who receive early MRI for low back pain are more likely to be out of work with disabilities one year later.[24] Another trial found that adding MRI in primary care for younger patients with traumatic knee complaints did not improve knee-related function after one year.[33]

These studies show that adding imaging tests that are known to yield high rates of structural asymptomatic findings to the care pathway of musculoskeletal disorders do not translate into better patients' outcomes. It can contribute to overdiagnosis and overuse of subsequent treatments such as surgery. Future research should investigate if the implementation of current and new diagnostic methods (e.g. ultrasound), classification and biomechanical systems (e.g. McKenzie,[34] movement system[35]) or prediction algorithms (e.g. clinical prediction rules) improve the complete clinical pathway leading to improved patients' outcomes without exposing them to the harms of overdiagnosis. Said otherwise, knowing the exact structure or movement patterns may not change the downstream decisions of high-quality first-line options that are required to improve outcomes.

Prognosis-rarely used but as equally important

Prognosis is a method of classification which is designed to determine the likelihood something happening in the future.[36] Prognosis research asks whether a decision will influence the patient's future outcome. It has been argued that prognostic decision-making should be as much a focus as diagnostic research, since ''no care'' is as often a good choice as providing care for patients. Failure to incorporate prognostic research in clinical care may lead to harmful effects and overtreatment (see earlier discussion).

Much medical education is concerned with the principles of disease diagnosis and treatment. Indeed, a historical emphasis has been placed on informing the patient and public about new understandings about the causes and mechanisms of disease and how best to reach a diagnosis and prescribe effective treatments linked to the diagnosis. We argue that an equal emphasis on prognosis may reduce overdiagnosis and over-treatment of patients. For example, providing a ''watchful waiting'' approach to benign conditions that often improve with no formal management, we reduce the risk of providing harms, unnecessary care, and increasing catastrophizing in patients. By being able to predict trajectories, we would be able to develop personalized rehabilitation approaches more likely to improve outcomes. We could determine for which patient we can simply prescribe watchful waiting compared to those that requires intensive rehabilitation, which could redirect cost to improve coverage of rehabilitation.

An interesting finding of the previous trials was that physicians and patients mostly preferred the use of the advanced imaging techniques and were more satisfied with their care although patients' outcomes were not improved.[27,33] This situation poses a real challenge for clinicians. Conceptual models propose that receiving a diagnostic label may have physical, psychosocial and financial consequences as well as increasing treatment burden, exposure to unnecessary tests and treatments and adverse events that lead to dissatisfaction with care.[31] Patients are mostly unaware of the possible harms associated with diagnostic labeling. As the natural history of common musculoskeletal disorder may be self-limiting, we need to study how a watchful waiting approach can be best integrated.

Take home message: When a fervent attempt is made to create a diagnosis, it may lead to overdiagnosis and subsequent overtreatment. A focus on prognosis for self-limiting conditions should improve overall findings and outcomes.

How using different methods of classifying diagnoses could improve management

We showed that current diagnostic labels in musculoskeletal disorders may sometimes have a negative impact on patients' outcomes. To establish the missing link between diagnosis and outcomes we must harness the complexity and heterogeneity that lie underneath common labels. Phenotyping may be a superior method of understanding musculoskeletal disorders.

Traditionally, the term phenotype has been used to reflect the observable properties of a particular organism that are produced from an interaction of the genotype and the environment.[37] Science has modified phenotyping to include physical, biochemical, and genetic characteristics along with interactions with the environment that have produced observable, unique characteristics.[37] Phenotyping that has looked at the interactions between genetic factors and psychological characteristics has been performed in patients with shoulder injuries to predict ongoing shoulder pain. George and colleagues have shown that two single nucleotide polymorphisms[38] interacted with psychological factors to predict six shoulder impairment phenotypes, and that two pain-related genes interacted with psychological factors to predict four shoulder impairment phenotypes.

Others have phenotyped using only clinical findings. Based on longitudinal cohorts, the field of knee osteoarthritis (OA) as greatly expanded this knowledge in the last decade. Two groups used the *Osteoarthritis Initiative* study and the *Amsterdam OA cohort*, including 3494 and 551 participants with knee OA, to identify up to five phenotypes of knee OA.[39,40] The phenotypes were based on radiological grades of knee OA, knee muscle strength, body mass index, comorbidities, psychological distress and alteration of pain neurophysiology. The phenotypes were named ''minimal joint disease'', ''strong muscle strength'', ''severe radiographic'', ''obese'' and ''depressive mood'' for the same diagnostic label of knee OA.

Other authors have also identified four pain susceptibility phenotypes using 852 participants from the *Multicenter Osteoarthritis Study*, a cohort of participants with or at risk of knee OA.[41] The phenotypes were based on clinical measures such as pressure pain threshold and temporal summation. The phenotype with high sensitization showed predictive capacity to develop persistent knee pain over 2 years.

Another group identified trajectories of knee pain and function following total knee arthroplasty over up to 5 years in a cohort of 689 patients.[42] Subgroups of patients demonstrated a persistent pain or function deficit after surgery and these trajectories could be predicted by co-morbidities and psychological or physical reported measures.

In the field of low back pain, one group identified five pain trajectories over 12 weeks in 1585 patients consulting for low back pain (recovery at week 2 or 12, pain reduction without recovery, fluctuating pain and high-level pain for 12 weeks).[43] Longer pain duration and belief in risk of persistence predicted delayed or nonrecovery of low back pain. High pain intensity and longer duration were associated with persistent high pain at 12 weeks.

Another group also identified up to nine subgroups using 112 characteristics based on history and physical examination of patients consulting for low back pain in primary care.[44] While the authors concluded that predictive capacity of the subgroups for pain intensity, frequency and disability over 12 months was somewhat higher compared to previous subgrouping techniques, they were also more complex to use in clinical practice. The authors proposed that further research should aim to determine if these subgroups better response to targeted treatment approaches.

In a cohort of 682 participants with nontraumatic complaints of arm, neck and shoulder in primary care, one group identified three disability trajectories at 2-year using Disabilities of the Arm, Shoulder and Hand questionnaire (DASH).[45] They identified several prognostic variables from the clinical examination, such as a high level of somatization that could predict if patients would follow the continuous high disability trajectory. In a cohort of 127 patients with patellofemoral pain, one group identified three subgroups classified as ''strong'', ''weak and tighter'' and ''weak and pronated foot'' based on six common clinical measures such as flexibility, strength, patellar mobility and posture of the foot.[46] The authors proposed that these subgroups could be used to develop targeted rehabilitation approaches that can further improve patients' outcomes.

The aforementioned studies suggest that several phenotypes might be present in a ''single'' diagnosis. This suggests that patients within that diagnosis may have different outcomes with the same treatment. We argue too that the results of test findings may vary in a single diagnosis based on the phenotype of the patient.

Take home message: These examples demonstrate that phenotyping based on patients' characteristics, reported outcome measures and clinical examination can help us better understand different profiles of patients and different trajectories of presentation for a given diagnostic label. As clinicians and researchers worldwide are continuing to develop large longitudinal cohorts for the study of musculoskeletal disorders, we will obtain more information to identify subgroups.

Conclusion

Higher order thinking, a process of decision making that excels beyond memorization, facts and, concepts, is necessary for diagnostic clinicians. In this masterclass, we discuss how higher order thinking may reduce interpretation errors associated with standard diagnostic metrics, how it may reduce overdiagnosis, and how a single diagnosis may actually consist of multiple phenotypes. In the future, we need to study diagnosis beyond normal metrics and venture into linking phenotyping and prognosis evidence to improve targeted care in order to ultimately improve patients' outcomes. We are at the beginning of our understanding of the different profile of patients with musculoskeletal disorders. Cohorts, database and data mining tools such as artificial intelligence

will accelerate our understanding of the linkage between diagnosis and patients' outcomes.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Whiting P, Harbord R, de Salis I, Egger M, Sterne J. Evidence-based diagnosis. *J Health Serv Pol*. 2008;13(suppl 3):57–63.
2. Walker HK. The origins of the history and physical examination. In: Walker HK, Hall WD, Hurst JW, eds. *Clinical Methods: The History, Physical, and Laboratory Examinations*. Boston. 1990 https://www.ncbi.nlm.nih.gov/books/NBK458/.
3. No author. International Classification of Diseases (ICD) Information Sheet. Available from: http://www.who.int/classifications/icd/factsheet/en/.
4. Papp KK, Huang GC, Lauzon Clabo LM, et al. Milestones of critical thinking: a developmental model for medicine and nursing. *Acad Med*. 2014;89(5):715–720.
5. Grossen B. The fundamental skills of higher order thinking. *J Learn Disabil*. 1991;24(6):343–352.
6. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529–536.
7. Eusebi P. Diagnostic accuracy measures. *Cerebrovasc Dis*. 2013;36(4):267–272.
8. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Fam Pract*. 2004;21(1):4–10.
9. Hegedus EJ, Stern B. Beyond SpPIN and SnNOUT: considerations with dichotomous tests during assessment of diagnostic accuracy. *J Man Manip Ther*. 2009;17(1):E1–E5.
10. Cadogan A, McNair PJ, Laslett M, Hing WA. Diagnostic accuracy of clinical examination and imaging findings for identifying subacromial pain. *PLoS ONE*. 2016;11(12):e0167738.
11. Jackson JL, O'Malley PG, Kroenke K. Evaluation of acute knee pain in primary care. *Ann Intern Med*. 2003;139(7):575–588.
12. Leblanc MC, Kowalczuk M, Andruszkiewicz N, et al. Diagnostic accuracy of physical examination for anterior knee instability: a systematic review. *Knee Surg Sports Traumatol Arthrosc*. 2015;23(10):2805–2813.
13. Decary S, Fallaha M, Belzile S, et al. Clinical diagnosis of partial or complete anterior cruciate ligament tears using patients' history elements and physical examination tests. *PLoS ONE*. 2018;13(6):e0198797.
14. 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.
15. Premkumar A, Godfrey W, Gottschalk MB, Boden SD. 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.
16. Karachalios T, Hantes M, Zibis AH, Zachos V, Karantanas AH, Malizos KN. Diagnostic accuracy of a new clinical test (the Thessaly test) for early detection of meniscal tears. *J Bone Joint Surg Am*. 2005;87(5):955–962.
17. Smith BE, Thacker D, Crewesmith A, Hall M. Special tests for assessing meniscal tears within the knee: a systematic review and meta-analysis. *Evid Based Med*. 2015;20(3):88–97.
18. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ*. 2006;174(4):469–476.
19. Croskerry P. A universal model of diagnostic reasoning. *Acad Med*. 2009;84(8):1022–1028.
20. Ely JW, Graber ML, Croskerry P. Checklists to reduce diagnostic errors. *Acad Med*. 2011;86(3):307–313.
21. Jaeschke RZ, Meade MO, Guyatt GH, Keenan SP, Cook DJ. How to use diagnostic test articles in the intensive care unit: diagnosing weanability using f/Vt. *Crit Care Med*. 1997;25(9):1514–1521.
22. Pathirana T, Clark J, Moynihan R. Mapping the drivers of overdiagnosis to potential solutions. *BMJ*. 2017;358:j3879.
23. Brodersen J, Kramer BS, Macdonald H, Schwartz LM, Woloshin S. Focusing on overdiagnosis as a driver of too much medicine. *BMJ*. 2018;362:k3494.
24. Kale MS, Korenstein D. Overdiagnosis in primary care: framing the problem and finding solutions. *BMJ*. 2018;362:k2820.
25. Jenkins HJ, Downie AS, Maher CG, Moloney NA, Magnussen JS, Hancock MJ. Imaging for low back pain: is clinical use consistent with guidelines? A systematic review and meta-analysis. *Spine J*. 2018.
26. Brinjikji W, Diehn FE, Jarvik JG, et al. MRI findings of disc degeneration are more prevalent in adults with low back pain than in asymptomatic controls: a systematic review and meta-analysis. *AJNR Am J Neuroradiol*. 2015;36(12):2394–2399.
27. Jarvik JG, Hollingworth W, Martin B, et al. Rapid magnetic resonance imaging vs radiographs for patients with low back pain: a randomized controlled trial. *JAMA*. 2003;289(21):2810–2818.
28. Culvenor AG, Oiestad BE, Hart HF, Stefanik JJ, Guermazi A, Crossley KM. Prevalence of knee osteoarthritis features on magnetic resonance imaging in asymptomatic uninjured adults: a systematic review and meta-analysis. *Br J Sports Med*. 2018.
29. Frank JM, Harris JD, Erickson BJ, et al. Prevalence of femoroacetabular impingement imaging findings in asymptomatic volunteers: a systematic review. *Arthroscopy*. 2015;31(6):1199–1204.
30. Girish G, Lobo LG, Jacobson JA, Morag Y, Miller B, Jamadar DA. Ultrasound of the shoulder: asymptomatic findings in men. *AJR Am J Roentgenol*. 2011;197(4):W713–W719.
31. Korenstein D, Chimonas S, Barrow B, Keyhani S, Troy A, Lipitz-Snyderman A. Development of a conceptual map of negative consequences for patients of overuse of medical tests and treatments. *JAMA Intern Med*. 2018.
32. Karel YH, Verkerk K, Endenburg S, Metselaar S, Verhagen AP. Effect of routine diagnostic imaging for patients with musculoskeletal disorders: a meta-analysis. *Eur J Intern Med*. 2015;26(8):585–595.
33. Swart NM, van Oudenaarde K, Bierma-Zeinstra SM, et al. Does MRI add value in general practice for patients with traumatic knee complaints? A 1-year randomised controlled trial. *Br J Sports Med*. 2018.
34. Lam OT, Strenger DM, Chan-Fee M, Pham PT, Preuss RA, Robbins SM. Effectiveness of the McKenzie method of mechanical diagnosis and therapy for treating low back pain: literature review with meta-analysis. *J Orthop Sports Phys Ther*. 2018;48(6):476–490.
35. Sahrmann S, Azevedo DC, Dillen LV. Diagnosis and treatment of movement system impairment syndromes. *Braz J Phys Ther*. 2017;21(6):391–399.
36. Croft P, Dinant GJ, Coventry P, Barraclough K. Looking to the future: should 'prognosis' be heard as often as 'diagnosis' in medical education? *Educ Prim Care*. 2015;26(6):367–371.
37. Nelson FRT. The value of phenotypes in knee osteoarthritis research. *Open Orthop J*. 2018;12:105–114.
38. George SZ, Wu SS, Wallace MR, et al. Biopsychosocial influence on shoulder pain: influence of genetic and psychological combinations on twelve-month postoperative pain and disability outcomes. *Arthritis Care Res (Hoboken)*. 2016;68(11):1671–1680.

39. Kittelson AJ, Stevens-Lapsley JE, Schmiege SJ. Determination of pain phenotypes in knee osteoarthritis: a latent class analysis using data from the osteoarthritis initiative. *Arthritis Care Res (Hoboken)*. 2016;68(5):612–620.

40. van der Esch M, Knoop J, van der Leeden M, et al. Clinical phenotypes in patients with knee osteoarthritis: a study in the Amsterdam osteoarthritis cohort. *Osteoarthrit Cartil*. 2015;23(4):544–549.

41. Carlesso LC, Segal NA, Frey-Law L, et al. Pain susceptibility phenotypes in those free of knee pain with or at risk of knee osteoarthritis: the multicenter osteoarthritis study. *Arthritis Rheumatol*. 2018.

42. Dowsey MM, Smith AJ, Choong PFM. Latent class growth analysis predicts long term pain and function trajectories in total knee arthroplasty: a study of 689 patients. *Osteoarthrit Cartil*. 2015;23(12):2141–2149.

43. Downie AS, Hancock MJ, Rzewuska M, Williams CM, Lin CW, Maher CG. Trajectories of acute low back pain: a latent class growth analysis. *Pain*. 2016;157(1):225–234.

44. Molgaard Nielsen A, Hestbaek L, Vach W, Kent P, Kongsted A. Latent class analysis derived subgroups of low back pain patients – do they have prognostic capacity? *BMC Musculosk Disord*. 2017;18(1):345.

45. Miedema HS, Feleus A, Bierma-Zeinstra SM, Hoekstra T, Burdorf A, Koes BW. Disability trajectories in patients with complaints of arm neck, and shoulder (CANS) in primary care: prospective cohort study. *Phys Ther*. 2016;96(7):972–984.

46. Selfe J, Janssen J, Callaghan M, et al. Are there three main subgroups within the patellofemoral pain population? A detailed characterisation study of 127 patients to help develop targeted intervention (TIPPs). *Br J Sports Med*. 2016;50(14):873–880.