

The Musculoskeletal Examination

From Practice to Evidence

Husain Alkhalidy · Department of Internal Medicine, King Khalid University, Abha · September 2026

Part 1 — The examination

Setting up. The musculoskeletal examination has two layers. The first is a two-minute screen of the whole locomotor system, designed to be done on every patient and to decide whether the second layer is needed at all. The second is the regional examination of the joint or region the history or the screen has picked out, built on one sequence — look, feel, move, special tests, function — applied the same way to every joint. Three questions open the screen: *Do you have any pain or stiffness in your muscles, joints or back? Can you dress yourself completely without difficulty? Can you walk up and down stairs without difficulty?* A patient who answers no, yes and yes and passes the screen has a normal locomotor system for practical purposes; a positive answer sends the examiner to the region.

The screen: gait, arms, legs, spine

- **Gait.** Walk to the end of the room, turn, walk back: symmetry, stride, arm swing, heel strike and toe-off, the turn, and any limp or antalgic pattern.
- **Spine.** From behind: scoliosis, the paraspinal muscles, the iliac crests level, the gluteal bulk. From the side: the cervical lordosis, thoracic kyphosis and lumbar lordosis. Then lateral flexion of the neck, 'put your ear on your shoulder', and forward flexion to the floor with the fingers on the lumbar spinous processes to feel them separate.
- **Arms.** Hands behind the head and hands behind the back for the shoulders; elbows straight; hands out, palms down, then turn them over; make a fist; pinch finger to thumb; squeeze across the metacarpophalangeal joints.
- **Legs.** Lying: flex the hip and knee with a hand on the knee for crepitus; internally rotate the hip; press the patella and tap for an effusion; squeeze across the metatarsophalangeal joints; inspect the soles.

The screen takes two minutes when practised, and it is the only part of the examination that most patients ever receive. Everything below is the regional examination that follows a positive screen or a regional complaint.

Hands and wrists

The hands carry the diagnosis of most inflammatory arthritis, and the diagnosis is in the pattern.

Look at the nails (pitting, onycholysis, splinters), the skin (psoriasis at the elbows and scalp, sclerodactyly, Gottron papules, tophi on the ulnar border and over the joints), the muscles (wasting of the first dorsal interosseous and the thenar eminence) and the joints: symmetrical swelling of the metacarpophalangeal and proximal interphalangeal joints with ulnar deviation is rheumatoid; bony

swelling at the distal interphalangeal joints (Heberden's nodes) and proximal (Bouchard's) with squaring of the first carpometacarpal joint is osteoarthritis; distal interphalangeal disease with a sausage digit and nail change is psoriatic. **Feel** each joint between finger and thumb for warmth, for the boggy, spongy swelling of synovitis as distinct from the hard swelling of bone, and for tenderness; the squeeze across the metacarpophalangeal row is the screening manoeuvre. **Move** the wrists in flexion and extension and the fingers into a fist and full extension; a fist that does not close is a measurement. **Function**: grip strength, the key pinch, doing up a button, and the ability to lift a cup. The elbow adds the extensor surface for nodules and plaques, the epicondyles, and full extension.

Shoulder

Most shoulder pain is the rotator cuff, and the examination is ordered around that. **Look** from the front, side and behind for wasting of supraspinatus and infraspinatus, deformity and scapular position. **Feel** the sternoclavicular and acromioclavicular joints, the bicipital groove and the greater tuberosity. **Move** actively then passively: forward flexion, abduction with a note of the painful arc between 60° and 120°, external rotation with the elbow at the side, and internal rotation as the highest vertebral level the thumb reaches. A passive range that exceeds the active range is a cuff or a nerve; a range that is restricted both ways, particularly external rotation, is the capsule or the joint. **Special tests** are chosen for the question. For cuff disease: resisted external rotation, the painful arc, and the lag signs — the arm placed in external rotation or in internal rotation behind the back drops when released — and the drop-arm test. Impingement tests such as Hawkins and Neer add little once those have been done. For instability: the apprehension test in abduction and external rotation, relieved by the relocation manoeuvre. For the acromioclavicular joint: the scarf test. Then **function**: reach the opposite scapula, the back of the head, and the small of the back.

Spine

Look for the curves and their loss, a scoliosis that corrects on flexion (postural) or does not (structural), a step, and the posture the patient adopts. **Feel** the spinous processes and the paraspinal muscles for tenderness, and the sacroiliac joints. **Move**: cervical flexion, extension, rotation and lateral flexion; thoracic rotation with the arms folded; lumbar flexion, extension and lateral flexion, and the modified Schober test — a mark 10 cm above the dimples of Venus and 5 cm below, the increase on full flexion recorded, less than 5 cm being restricted. The occiput-to-wall and chest expansion measurements belong to the spondyloarthritis examination. **Special tests**: the straight-leg raise with the angle and the side of the pain, the crossed straight-leg raise, the femoral stretch, and a cluster of sacroiliac provocation tests — thigh thrust, distraction, compression, sacral thrust and FABER — read together. Palpation for a segmental restriction is not reproducible and is not recorded as a finding. Then the neurological examination of the legs that the previous chapter set out, and the **red flags** taken from the history: age, cancer, steroids, trauma, fever, weight loss, night pain, bladder and bowel symptoms and saddle anaesthesia. Inflammatory back pain is a history, not an examination: onset before forty, insidious, improved by exercise and not by rest, with night pain that eases on rising — four of five.

Hip

Look at the gait, the standing posture, leg length and the Trendelenburg test — the pelvis drops on the side of the lifted leg when the abductors on the standing side are weak. **Feel** the greater trochanter and the groin. **Move:** flexion, then internal and external rotation in flexion, abduction and adduction, and the Thomas test for a fixed flexion deformity with the other hip flexed to flatten the lumbar spine. Loss of internal rotation is the earliest and most useful sign of hip disease. **Special tests:** FABER (flexion, abduction, external rotation) for the joint and the sacroiliac; a squat that produces posterior hip pain; resisted abduction for the gluteal tendons. True leg length is measured from the anterior superior iliac spine to the medial malleolus.

Knee

Look for quadriceps wasting, measured 10 cm above the patella, varus or valgus alignment, an effusion filling the medial and suprapatellar hollows, a Baker's cyst behind. **Feel** for warmth, for an effusion with the bulge sign for a small one and the patellar tap for a large one, for joint-line tenderness with the knee flexed, and along the collateral ligaments. **Move:** full extension including the last few degrees, flexion to the heel on the buttock, and crepitus with a hand on the patella. **Special tests:** for the anterior cruciate, the Lachman test at 20° to 30° of flexion recorded as positive or negative, the anterior drawer at 90°, and the pivot shift when the patient can relax; for the posterior cruciate, the posterior sag and the posterior drawer; for the collaterals, valgus and varus stress at full extension and at 30°; for the menisci, joint-line tenderness, McMurray's test, and the Thessaly test. **Function:** rise from a chair, squat, and a single-leg stance. In an acute injury the Ottawa knee rule decides the radiograph: age 55 or over, isolated patellar tenderness, tenderness at the fibular head, inability to flex to 90°, and inability to bear weight for four steps both immediately and in the department.

Ankle and foot

Look at the arches standing, the heel alignment from behind, the toes (hallux valgus, claw and hammer toes), callosities, and the nails and skin between the toes. **Feel** the malleoli and the base of the fifth metatarsal, the navicular, the Achilles tendon and its insertion, the plantar fascia at the heel, and the metatarsophalangeal squeeze. **Move:** dorsiflexion and plantarflexion at the ankle, inversion and eversion at the subtalar joint, and the midfoot. **Special tests:** the Thompson squeeze for the Achilles, the anterior drawer for the lateral ligament, and the Ottawa ankle rules — bone tenderness at the posterior edge or tip of either malleolus, at the navicular or the base of the fifth metatarsal, or inability to bear weight for four steps — which decide the radiograph.

Completing the examination

Entheses (Achilles, plantar fascia, the epicondyles, the iliac crests) for the spondyloarthritides; dactylitis; the tender-point survey where fibromyalgia is in question; the Beighton score for hypermobility — the fifth finger past 90°, the thumb to the forearm, the elbows and knees past 10° of hyperextension, and the palms to the floor, out of nine; the temporal arteries for thickening, tenderness and a lost pulse; the eyes for scleritis and uveitis; the mouth for ulcers and dryness; the

skin and nails as above; the lymph nodes and the spleen. And the manoeuvre that completes any examination of a hot, swollen joint: aspiration of the joint. A synovial fluid count and crystal examination are the reference standard against which the whole examination of the acute joint is measured, and no bedside finding replaces them.

Putting the signs together

The musculoskeletal examination sorts patients by pattern — how many joints, which joints, symmetrical or not, inflammatory or mechanical, with what outside the joints — and the pattern is most of the diagnosis.

Pattern	The signs that make it
Rheumatoid arthritis	Symmetrical synovitis of the metacarpophalangeal, proximal interphalangeal and wrist joints, sparing the distal interphalangeal joints; positive metacarpophalangeal squeeze; morning stiffness over an hour; later ulnar deviation, swan-neck and boutonnière deformities, nodules on the extensor surfaces.
Osteoarthritis	Bony, not boggy, swelling at the distal and proximal interphalangeal joints and the first carpometacarpal joint; crepitus, reduced range and pain at the end of range in the knees and hips; loss of hip internal rotation; short-lived stiffness.
Psoriatic arthritis	Distal interphalangeal disease, dactylitis, enthesitis, asymmetry, nail pitting and onycholysis, and plaques found where the patient does not show them — scalp, umbilicus, natal cleft.
Gout	An acutely hot, red, exquisitely tender joint, the first metatarsophalangeal above all; tophi at the ulnar border, the olecranon, the helix and over the joints; a diagnosis confirmed by crystals, not by the picture.
Axial spondyloarthritis	Inflammatory back pain history; restricted lumbar flexion on the Schober test and reduced lateral flexion; loss of lumbar lordosis; sacroiliac tenderness; enthesitis; reduced chest expansion late.
Septic arthritis	A single hot, swollen joint with severe pain on any movement, in a patient with fever or a source or a prosthesis — and no bedside finding reliable enough to withhold the needle.
Polymyalgia rheumatica	Bilateral shoulder and hip girdle stiffness and pain in a patient over fifty, with restricted active but preserved passive movement, normal power once pain is overcome, and no synovitis in the hands.
Giant cell arteritis	A thickened, tender or pulseless temporal artery, with jaw claudication, scalp tenderness and visual symptoms in the history; the artery may feel normal.
Hypermobility	A Beighton score of five or more with arthralgia, recurrent dislocation and soft skin; joints that reach beyond the range the examiner expected.
Fibromyalgia	Widespread pain with normal joints, normal range and normal power; diffuse tenderness at the tender points and beyond them; fatigue and unrefreshing sleep in the history.

Part 2 — What the literature says

How well do examiners agree?

The musculoskeletal examination has been studied for reliability more thoroughly than any other region except the chest, largely by physiotherapists and rheumatologists who needed reproducible outcome measures, and the verdict divides the examination in two. **Measurements agree; judgements do not.** A goniometer in two pairs of hands gives intraclass correlations of 0.85 to 0.95 for shoulder and knee range [1, 2]; the Beighton score, which is a count of measured angles, reaches kappa values of 0.75 to 0.92 [3, 4]. The judgement that a joint is swollen, that a test provoked the patient's pain, or that a segment of the spine is stiff does not.

Sign	Agreement (κ)	Comment
Swollen joint count	0.18 – 0.62	Seven examiners on the same patients with established rheumatoid arthritis, without prior training; the tender joint count is reliable across physicians, metrologists and patients, the swollen count is not [5, 6]
Shoulder special tests	0.09 – 0.66	Only 11 of 23 items reached acceptable reliability in one study [7]; a systematic review found no consistent evidence that any shoulder procedure is acceptably reliable [8]. In the acute injured shoulder, the inability to abduct above 90° and abduction strength agree almost perfectly ($\kappa > 0.8$), the Hawkins test (0.25) and the external rotation lag sign (0.40) do not [9]
Lachman test	≈ 0.8 when dichotomous	Reliable when recorded as positive or negative, not when graded; κ 0.81 in the prone position [10, 11]
McMurray test	0.16 – 0.38	Poor in every series reviewed [11]
Joint-line tenderness	0.11 – 0.25	Poor [11]
Sacroiliac provocation tests	0.43 – 0.88	Moderate to almost perfect for pain provocation; thigh thrust and sacral thrust κ 0.90 [12, 13]
Lumbar and sacroiliac palpation	-0.20 – 0.17	Segmental motion palpation: no better than chance [12, 14]
Beighton score	0.75 – 0.92	Substantial to almost perfect; a count of measured angles [3, 4]
Goniometric range of motion	ICC 0.85 – 0.95	Shoulder rotation, knee flexion and extension [1, 2]

The lesson is the one the neurological chapter found in the difference between watched and elicited signs, transposed: a finding with a numerical end-point (a degree, a centimetre, a count of positive tests) is reproducible, and a finding that depends on the examiner's threshold — *is this joint swollen? did that hurt?* — is not, unless the examiners have been trained to the same threshold. The swollen joint count is the clearest case, because it is the only supposedly objective item in the disease activity score that drives treatment decisions in rheumatoid arthritis, and a systematic review found that training and standardisation do improve it [6].

How accurate are the signs? The Rational Clinical Examination series and the meta-analyses

Question	Finding	Likelihood ratio or accuracy	Source
Is the anterior cruciate ligament torn?	Lachman test	Sensitivity 85 per cent, specificity 94 per cent (pooled)	Benjaminse 2006 [15]; Solomon 2001 [16]
	Anterior drawer	Sensitivity 92 per cent, specificity 91 per cent in chronic injuries; unreliable in acute	
	Pivot shift, awake · under anaesthesia	Sensitivity 24 · 74 per cent; specificity 98 · 99 per cent; LR+ 8.5 · 21	
Is the meniscus torn?	McMurray test · Thessaly test	Sensitivity 61 · 64 per cent; specificity 84 · 53 per cent — Thessaly no better than McMurray	Smith 2015 [17]; Goossens 2015 [18]
Does this shoulder have rotator cuff disease?	Painful arc · external rotation resistance	LR+ 3.7 for any cuff disease — the only pain-provocation finding above 2	Hermans 2013 [19]
	External rotation lag sign	LR+ 7.2 for a full-thickness tear; the lag signs are the most accurate findings for a tear	
Is this shoulder unstable, or the labrum torn?	Relocation test · apprehension test	LR+ 6.5 for instability; apprehension alone less useful	Luime 2004 [20]
	Biceps load I and II tests	LR+ 29 and 26 for a labral lesion	
Does this patient have hip osteoarthritis?	Posterior pain on squatting · reduced internal rotation · abductor weakness · reduced adduction	LR+ 6.1 · 3.2 · 4.5 · 4.2	Rational Clinical Examination 2019 [21]
Is the ankle or foot fractured?	Ottawa ankle rules	Pooled sensitivity 97.6 per cent, specificity 31.5 per cent; a negative rule reduces the probability of fracture to about 1.4 per cent	Bachmann 2003 [22]; meta-analysis 2022 [23]
Is the knee fractured?	Ottawa knee rule	Pooled sensitivity 98.5 per cent, specificity 48.6 per cent, LR– 0.05	Bachmann 2004 [24]
Does this adult have septic arthritis?	History and examination	Do not change the probability enough to act on; joint surgery LR+ 6.9 and skin infection over a prosthetic joint LR+ 15 are the exceptions	Margaretten 2007 [25]

Question	Finding	Likelihood ratio or accuracy	Source
	Synovial white cell count > 100 · 50 – 100 · < 25 ×10 ⁹ /L	LR ∞ · 3.6 · 0.33 — the needle, not the hand, decides	
Does this patient have temporal arteritis?	Jaw claudication · diplopia	LR+ 4.2 · 3.4; an abnormal temporal artery raises the probability, a normal one does not exclude it	Smetana & Shmerling 2002 [26]
Is this back pain a fracture or a cancer?	Any single red flag	Not enough to justify imaging on its own; a history of cancer is the most useful single item for malignancy, and a combination of older age, steroid use and trauma for fracture	Downie 2013 [27]; Henschke 2013 [28]
Is this back pain a disc herniation?	Straight-leg raise · crossed straight-leg raise	Sensitivity 92, specificity 28 per cent · sensitivity 28, specificity 90 per cent	Cochrane 2010 [29]
Is this acute monoarthritis gout?	Janssens clinical rule (sex, previous attack, onset within a day, redness, first metatarsophalangeal joint, hypertension or cardiovascular disease, urate)	Prevalence of gout 6 per cent at a score of 4 or less, 46 per cent between 4 and 8, 88 per cent at 8 or more	Janssens 2010 [30]; validation 2015 [31]
	Ultrasound double-contour sign	Sensitivity 66 – 70 per cent, specificity 92 – 95 per cent (pooled)	Meta-analyses 2018 [32]
Is this inflammatory back pain?	ASAS criteria, four of five	Sensitivity 80 per cent, specificity 72 per cent in the derivation exercise	Sieper 2009 [33]

The shape of the evidence is familiar by now. **The signs that rule in are strong** — a pivot shift, a lag sign, a positive relocation test, posterior pain on squatting, the biceps load tests — and most of them are manoeuvres with a defined end-point rather than provocations of pain. **The signs that rule out are weak** with the important exception of the Ottawa rules, which were built to rule out and do: a negative rule leaves a fracture probability low enough to send the patient home without a radiograph, and that is what halved radiography in the departments that adopted them. And for the two diagnoses that cannot wait — the septic joint and the arteritic artery — the examination has been shown not to be enough, and the reference standard is a needle or a probe.

The examiner is the limiting reagent

Freedman and Bernstein's basic competency examination in musculoskeletal medicine, validated by the chairs of orthopaedic residencies, was failed by 82 per cent of the first-year residents who sat it, and the finding has been reproduced in cohort after cohort since, in the United States, the United Kingdom and Australia, with students, residents and attending physicians rating their own confidence in musculoskeletal care as low [34, 35]. The GALS screen and the regional examination

were developed in Britain specifically to give medical students a standardised routine; a UK survey found that although the routines are taught, their use in postgraduate practice is patchy [36, 37]. The reliability studies above should be read against that baseline: a joint count by an untrained examiner has a kappa of 0.18, and the same count by a trained one 0.62 [5].

Technique changes the answer

- **Count joints to a standard.** The swollen joint count becomes reliable when examiners are trained to the same definition of swelling and examine the same 28 joints the same way; without that it is the least reliable item in the disease activity score [5, 6].
- **Record the Lachman test as positive or negative,** not as a grade; the dichotomous result is reliable and the graded one is not [10, 11].
- **Use clusters, not single provocations, for the sacroiliac joint,** and do not palpate for segmental restriction at all; a cluster of three or more positive provocation tests is both reliable and diagnostically useful, and palpation is neither [12, 13, 14].
- **Measure range with a goniometer or an inclinometer** where the number matters — the shoulder, the knee, the Schober — and record it, because the number is what a second examiner will reproduce and what a smartphone application now reproduces as well [1, 2, 38].
- **Apply the Ottawa rules as written,** including the four-step weight-bearing test at the time of injury and in the department; the rules are sensitive because they are structured, and departures from the structure are where the missed fractures live [22, 23, 24].
- **Aspirate the hot joint.** No combination of fever, warmth, redness and pain on movement reaches a likelihood ratio that permits withholding the needle [25].

What has changed, 2020 – 2026

Ultrasound sees synovitis the hand cannot feel — and it has not changed the outcome. Wakefield's finding that 13 per cent of clinically normal joints in oligoarthritis had sonographic synovitis, upgrading a third of patients from mono- to oligoarthritis and a quarter to polyarthritis, established the term subclinical synovitis and the expectation that the probe would replace the examination [39]. The two strategy trials that tested it, ARCTIC and TaSER, randomised early rheumatoid arthritis to treatment targeted on ultrasound remission or on clinical remission, and neither found a benefit of the ultrasound target on its primary outcome, though both saw a trend to less erosive damage [40, 41]. A 2022 systematic review reached the same conclusion: ultrasound-guided treat-to-target does not outperform clinical examination in early disease [42]. The examination of the rheumatoid hand — the squeeze, the count, the pattern — kept its place because the treatment decisions it drives were not improved by seeing more.

Ultrasound has replaced the examination where the examination was never adequate. The 2023 EULAR update makes ultrasound of the temporal and axillary arteries the first-line test in every patient with suspected giant cell arteritis, with pooled sensitivity of 88 per cent and specificity of 96 per cent in low-bias studies — against a palpated artery whose abnormality raises the probability and whose normality proves nothing [43, 44]. The 2015 ACR/EULAR gout criteria admitted the

double-contour sign and dual-energy CT as evidence of urate, so that the diagnosis no longer depends on a crystal in a joint the clinician was able to aspirate [45]. For the knee, a 2024 meta-analysis of 755 patients found ultrasound detected synovitis with a sensitivity of 88 per cent [46].

The probe has reached the fracture clinic. Point-of-care ultrasound identifies long-bone fractures with pooled sensitivity and specificity of 0.93 and 0.92 in the upper limb and 0.83 and 0.93 in the lower, 93 per cent for both in paediatric forearm fractures and 90 and 94 per cent for adult ankle fractures, with a 2026 emergency department study reporting 88 and 92 per cent against radiography [47, 48]. It does not replace the Ottawa rules, which decide whether any imaging is needed; it changes what the imaging is.

The special tests have been re-examined and found smaller. A 2022 meta-analysis concluded that the Lachman test's accuracy had been overestimated by earlier reviews and that the clinical tests for the anterior cruciate ligament are comparable rather than one being superior [49]; the Thessaly test, promoted as an improvement on McMurray's, proved no better when tested against arthroscopy [17, 18]; and a 2021 study of the acutely injured shoulder found that the simplest observations — can the arm be abducted above 90°, how strong is abduction — were the reliable ones, while the named tests were not [9].

The examination has been tried down a camera. A 2024 systematic review of nine studies and 321 patients found good agreement between telehealth and in-person diagnosis for musculoskeletal conditions where the assessment rested on history, observation, function and self-measured range, and poor validity for the orthopaedic special tests, lumbar posture and neurological tests — the same boundary between the observed and the elicited that the neurological chapter drew [50, 51]. Smartphone inclinometer applications measure knee range as reliably as a goniometer [38].

Part 3 — Practical synthesis for teaching

- Teach the screen first and insist on it: two minutes, three questions, gait—arms—legs—spine, on every patient. It is the part of the examination most patients will receive and the part most often omitted.
- Teach the regional examination as one sequence — look, feel, move, special tests, function — and teach the pattern before the eponym. The distribution of joints is most of the diagnosis; the named test is a confirmation.
- Prefer measurements to judgements: a goniometer, a Schober, a Beighton count, a joint count to a trained standard, a Lachman recorded as yes or no. These are the findings a second examiner will reproduce.
- Teach the special tests with their numbers, and drop the ones that do not earn their place: the lag signs and the relocation test, not the impingement tests; the provocation cluster, not spinal palpation; McMurray's with joint-line tenderness rather than a Thessaly that adds nothing.
- Teach the Ottawa rules as the model of what a rule-out examination looks like — structured, sensitive, validated — and teach that most other signs in this chapter cannot rule out.

- Teach the needle and the probe as the end of the examination, not its rival: the hot joint is aspirated, the suspected arteritis is scanned, and the fracture the rules could not exclude is imaged. A student who can find the joint line and the temporal artery will place the probe faster than one who cannot.
- Be honest about the 82 per cent, and about a swollen joint count with a kappa of 0.18 in untrained hands. The remedy is not fewer examinations but trained ones.

References

1. Intratester and intertester reliability of goniometric measurement of passive lateral shoulder rotation. *J Hand Ther* 1999. <https://www.sciencedirect.com/science/article/abs/pii/S0894113099800453>
2. The intrarater and interrater reliability and validity of universal goniometer, digital inclinometer, and smartphone application measuring range of motion in patients with total knee arthroplasty. 2024. <https://pmc.ncbi.nlm.nih.gov/articles/PMC11130096/>
3. Bockhorn LN, Vera AM, Dong D, Delgado DA, Varner KE, Harris JD. Interrater and intrarater reliability of the Beighton score: a systematic review. *Orthop J Sports Med* 2021;9:2325967120968099.
4. Inter-tester reproducibility and inter-method agreement of two variations of the Beighton test for determining generalised joint hypermobility in primary school children. *BMC Pediatr* 2013. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3878084/>
5. Reproducibility of joint swelling assessments in long-lasting rheumatoid arthritis: influence on Disease Activity Score-28 values (SEA-Repro study part I). *J Rheumatol* 2010;37:932–7.
6. Cheung PP, Gossec L, Mak A, March L. Reliability of joint count assessment in rheumatoid arthritis: a systematic literature review. *Semin Arthritis Rheum* 2014;43:721–9.
7. Nomden JG, Slagters AJ, Bergman GJD, Winters JC, Kropmans TJB, Dijkstra PU. Interobserver reliability of physical examination of shoulder girdle. *Man Ther* 2009;14:152–9.
8. May S, Chance-Larsen K, Littlewood C, Lomas D, Saad M. Reliability of physical examination tests used in the assessment of patients with shoulder problems: a systematic review. *Physiotherapy* 2010;96:179–90.
9. Interrater reliability of physical examination tests in the acute phase of shoulder injuries. *BMC Musculoskelet Disord* 2021;22:770. <https://pmc.ncbi.nlm.nih.gov/articles/PMC8427844/>
10. Mulligan EP, Harwell JL, Robertson WJ. Reliability and diagnostic accuracy of the Lachman test performed in a prone position. *J Orthop Sports Phys Ther* 2011;41:749–57.
11. Décary S, Ouellet P, Vendittoli PA, Roy JS, Desmeules F. Reliability of physical examination tests for the diagnosis of knee disorders: evidence from a systematic review. *Man Ther* 2017;26:172–82.
12. Laslett M, Williams M. The reliability of selected pain provocation tests for sacroiliac joint pathology. *Spine* 1994;19:1243–9.
13. Laslett M, Aprill CN, McDonald B, Young SB. Diagnosis of sacroiliac joint pain: validity of individual provocation tests and composites of tests. *Man Ther* 2005;10:207–18.
14. Spinal palpation for lumbar segmental mobility and pain provocation: an interexaminer reliability study. *J Manipulative Physiol Ther* 2008. <https://www.sciencedirect.com/science/article/abs/pii/S0161475408001760>
15. Benjaminse A, Gokeler A, van der Schans CP. Clinical diagnosis of an anterior cruciate ligament rupture: a meta-analysis. *J Orthop Sports Phys Ther* 2006;36:267–88.
16. Solomon DH, Simel DL, Bates DW, Katz JN, Schaffer JL. Does this patient have a torn meniscus or ligament of the knee? Value of the physical examination. *JAMA* 2001;286:1610–20.
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;20:88–97.
18. Goossens P, Keijsers E, van Geenen RJ, et al. Validity of the Thessaly test in evaluating meniscal tears compared with arthroscopy: a diagnostic accuracy study. *J Orthop Sports Phys Ther* 2015;45:18–24.
19. Hermans J, Luime JJ, Meuffels DE, Reijman M, Simel DL, Bierma-Zeinstra SMA. Does this patient with shoulder pain have rotator cuff disease? The Rational Clinical Examination systematic review. *JAMA* 2013;310:837–47.

20. Luime JJ, Verhagen AP, Miedema HS, et al. Does this patient have an instability of the shoulder or a labrum lesion? JAMA 2004;292:1989–99.
21. Metcalfe D, Perry DC, Claireaux HA, et al. Does this patient have hip osteoarthritis? The Rational Clinical Examination systematic review. JAMA 2019;322:2323–33.
22. Bachmann LM, Kolb E, Koller MT, Steurer J, ter Riet G. Accuracy of Ottawa ankle rules to exclude fractures of the ankle and mid-foot: systematic review. BMJ 2003;326:417.
23. Diagnostic accuracy of the Ottawa ankle rule to exclude fractures in acute ankle injuries in adults: a systematic review and meta-analysis. BMC Musculoskelet Disord 2022;23:885.
24. Bachmann LM, Haberketh S, Steurer J, ter Riet G. The accuracy of the Ottawa knee rule to rule out knee fractures: a systematic review. Ann Intern Med 2004;140:121–4.
25. Margaretten ME, Kohlwes J, Moore D, Bent S. Does this adult patient have septic arthritis? JAMA 2007;297:1478–88.
26. Smetana GW, Shmerling RH. Does this patient have temporal arteritis? JAMA 2002;287:92–101.
27. 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.
28. Henschke N, Maher CG, Ostelo RWJG, de Vet HCW, Macaskill P, Irwig L. Red flags to screen for malignancy in patients with low-back pain. Cochrane Database Syst Rev 2013;(2):CD008686.
29. van der Windt DAWM, Simons E, Riphagen II, et al. Physical examination for lumbar radiculopathy due to disc herniation in patients with low-back pain. Cochrane Database Syst Rev 2010;(2):CD007431.
30. Janssens HJEM, Fransen J, van de Lisdonk EH, van Riel PLCM, van Weel C, Janssen M. A diagnostic rule for acute gouty arthritis in primary care without joint fluid analysis. Arch Intern Med 2010;170:1120–6.
31. Kienhorst LBE, Janssens HJEM, Fransen J, Janssen M. The validation of a diagnostic rule for gout without joint fluid analysis: a prospective study. Rheumatology (Oxford) 2015;54:609–14.
32. The diagnostic performance of musculoskeletal ultrasound in gout: a systematic review and meta-analysis. PLoS One 2018;13:e0199672; and Diagnostic accuracy of ultrasound in patients with gout: a meta-analysis. Semin Arthritis Rheum 2018.
33. Sieper J, van der Heijde D, Landewé R, et al. New criteria for inflammatory back pain in patients with chronic back pain: a real patient exercise by experts from the Assessment of SpondyloArthritis international Society (ASAS). Ann Rheum Dis 2009;68:784–8.
34. Freedman KB, Bernstein J. The adequacy of medical school education in musculoskeletal medicine. J Bone Joint Surg Am 1998;80:1421–7.
35. Knowledge of musculoskeletal medicine in junior doctors in Australia: is it adequate? Med Sci Educ 2022. <https://link.springer.com/article/10.1007/s40670-022-01637-3>
36. Doherty M, Dacre J, Dieppe P, Snaith M. The 'GALS' locomotor screen. Ann Rheum Dis 1992;51:1165–9.
37. Use of structured musculoskeletal examination routines in undergraduate medical education and postgraduate clinical practice — a UK survey. BMC Med Educ 2016. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5073898/>
38. Analysis of the concurrent validity and reliability of five common clinical goniometric devices. Sci Rep 2023. <https://www.nature.com/articles/s41598-023-48344-6>
39. Wakefield RJ, Green MJ, Marzo-Ortega H, et al. Should oligoarthritis be reclassified? Ultrasound reveals a high prevalence of subclinical disease. Ann Rheum Dis 2004;63:382–5.
40. Haavardsholm EA, Aga AB, Olsen IC, et al. Ultrasound in management of rheumatoid arthritis: ARCTIC randomised controlled strategy trial. BMJ 2016;354:i4205.
41. Dale J, Stirling A, Zhang R, et al. Targeting ultrasound remission in early rheumatoid arthritis: the results of the TaSER study, a randomised clinical trial. Ann Rheum Dis 2016;75:1043–50.
42. Musculoskeletal ultrasound for treating rheumatoid arthritis to target — a systematic literature review. Rheumatology (Oxford) 2022;61:4590–610.
43. Dejaco C, Ramiro S, Bond M, et al. EULAR recommendations for the use of imaging in large vessel vasculitis in clinical practice: 2023 update. Ann Rheum Dis 2024;83:741–51.

44. Luqmani R, Lee E, Singh S, et al. The role of ultrasound compared to biopsy of temporal arteries in the diagnosis and treatment of giant cell arteritis (TABUL): a diagnostic accuracy and cost-effectiveness study. *Health Technol Assess* 2016;20:1–238.
45. Neogi T, Jansen TLTA, Dalbeth N, et al. 2015 gout classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Ann Rheum Dis* 2015;74:1789–98.
46. Liu et al. Diagnostic accuracy of ultrasound for assessment of synovial abnormalities among patients with knee pain: a meta-analysis. *Arthritis Care Res* 2024. <https://doi.org/10.1002/acr.25205>
47. Use of point-of-care ultrasound in long bone fractures: a systematic review and meta-analysis. *CJEM* 2017;19:131–42.
48. Evaluating the diagnostic accuracy of point-of-care ultrasound for long bone fractures in the emergency department. *J Emerg Med* 2026. [https://www.jem-journal.com/article/S0736-4679\(26\)00007-7/abstract](https://www.jem-journal.com/article/S0736-4679(26)00007-7/abstract)
49. Sokal PA, Norris R, Maddox TW, Oldershaw RA. The diagnostic accuracy of clinical tests for anterior cruciate ligament tears are comparable but the Lachman test has been previously overestimated: a systematic review and meta-analysis. *Knee Surg Sports Traumatol Arthrosc* 2022;30:3287–303.
50. Agreement and concurrent validity between telehealth and in-person diagnosis of musculoskeletal conditions: a systematic review. *Chiropr Man Therap* 2024;32:19. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11177434/>
51. Mani S, Sharma S, Omar B, Paungmali A, Joseph L. Validity and reliability of Internet-based physiotherapy assessment for musculoskeletal disorders: a systematic review. *J Telemed Telecare* 2017;23:379–91.

Caveats

The joint-count kappa range is from a single seven-examiner study in established rheumatoid arthritis without prior training and is quoted as such; the systematic review it sits beside pools studies with and without training. The ASAS inflammatory back pain figures are from the derivation exercise rather than a validation cohort. The number for the post-Ottawa fracture probability is derived from the pooled sensitivity and specificity rather than reported directly. Where a reference is cited by title and address only, the search results did not return an author list; check before distribution. The knee ultrasound and paediatric fracture figures come from meta-analysis abstracts retrieved in search results rather than from the full papers.