



Hva er en god test?

Jeg er nok ikke alene om å kjenne på frustrasjonen. Fortvilelsen over å huske navnet. Utførelsen. Vinkelen. Trykket. Når man endelig hadde lært seg én, forstod man at det var en endeløs rekke av nye navn og utførelser man skulle lære seg. Så skulle dette kategoriseres og systematiseres. Når man så plutselig står der med en pasient foran seg kommer tvilen. Hva betyr denne testen egentlig? Hvorfor bruker jeg nettopp denne testen? Hva er egentlig en god, ortopedisk test?



AV JØRGEN JEVNE
KIROPRAKTOR OG
FYSIOTERAPEUT

Et uhøytidelig PubMedsøk i januar 2020 avdekker flere interessante momenter ved kliniske tester. To velkjente systematiske gjennomganger fra 2008 og 2012 som undersøker kliniske tester i skulderen er sitert

henholdsvis 470 og 378 ganger [1,2]. Begge disse hadde økt med 10 siteringer på én måned, da jeg undersøkte det samme i desember 2019. Det samme mønsteret gjør seg gjeldende når man ser på tester av hofter [3] og knær [4]. Et YouTube-søk viser at British Journal of Sports Medicine sine videoer på skulderundersøkelse har nærmere 250.000 visninger. Vår kunnskapshunger etter diagnostisk nøyaktighet er

tilsynelatende umettelig. Felles for de alle er at de ikke diskuterer elefanten i rommet; hvilket fundament bygger vi nøyaktigheten på? Hvorfor er kunnskap omkring testing så populært?

Et uoversiktlig sammensurium
Det er beskrevet over 200 ortopediske tester i skulderen alene. Skulderen er et komplekst område bestående av flere ledd og i vår

evige jakt på struktur og orden har vi forsøkt å skape system i kaoset. Vi har skapt diagnostiske bokser og forsøkt å plassere pasientens symptomer inn i disse basert på funn fra kliniske tester. Vi har en myriade av tester for impingement, AC-ledd, bursa, supraspinatus, infraspinatus, subscapularis, biceps caput longum, bakre labrum, SLAP, tenosynovitt osv.

Ironisk nok finner vi de samme testene igjen i forskjellige kategorier. Den originale O'Briens test ble beskrevet som en undersøkelse for å avdekke en SLAP ruptur. Den originale studien, som ydmykt nok ble publisert av O'Brien selv, beskrev testen med 100% sensitivitet og 98% spesifisitet [5]. Disse tallene i seg selv skulle tilsa at testen var av nærmest usannsynlig høy klinisk verdi. Så er testen senere blitt beskrevet som en test for impingement, AC-ledd, biceps caput longum og subakromielt smertesyndrom. Hvordan kan én test være beskrevet å teste mange forskjellige strukturer og samtidig være ansett som klinisk verdifull?

Den anerkjente skulderforskeren, Jeremy Lewis, har tidligere uttalt det slik:

«Vi lærte at vi skulle gjøre en empty-can test og dermed teste supraspinatus. Deretter gikk vi til lunsj. Når vi kom tilbake flyttet vi skulderen inn i fleksjon og gjorde samme motstandstest. Nå testet vi AC-leddet. Hvorfor testet vi ikke lenger supraspinatus i den posisjonen? Var det slik at supraspinatus fikk botoxinjeksjon og var lammet under den testen?»



BJSM sine videoer på skulderundersøkelse er sett nesten 250.000 ganger på YouTube

Skulderen er riktignok anatomisk mer kompleks enn hoften og kneet, men studier som omhandler diagnostisk nøyaktighet i disse leddene finner de samme fundamentale bristene: testene er upålitelige og tester ikke hva de er designet for å teste.

Testens tiltrekningskraft

Hvor kommer appetitten på ortopediske tester fra? I en verden som er notorisk uforutsigbar, uspesifikk og usikker virker kanskje testene som et lys i mørket. En fakkell som gir oss nødvendig lys i sorte hull av vår forståelse. Der hvor vi føler vi navigerer rundt med et feilkalibrert kompass i ukjent terreng kan kanskje tes-

tens spesifisitet virke forlokkende og bringe selvtillit til en usikker kliniker. Men spør man fysioterapeuter rundt omkring har alle gjort seg den samme erfaringen: man er ofte usikker på hva testen egentlig forteller oss om diagnose og videre behandling. Denne forfatteren har i det minste kjent på denne øvelsen ofte, godt illustrert av bildet vedlagt.

Vi har gitt testene egenskaper og funksjoner de ikke ser ut til å ha. Vi har skapt et design for å gi oss et klart «ja» eller «nei», i en verden hvor det er veldig få sort-hvitt svar.

«Har denne pasienten en rotatorcuff ruptur?» Jeg kan velge å utføre en Drop-arm test som er designet for å teste funksjonen til supraspinatus. Uavhengig av spørsmålet må jeg ta innover meg flere forskjellige momenter:

«Er det i det hele tatt mulig å teste supraspinatus isolert?»

«Vet jeg for sikkert at at drop-arm test faktisk tester supraspinatus?»

«Hvis testen er positiv og det viser seg at supraspinatus faktisk er røket, er jeg sikker på at det er årsaken til pasientens problem?»



Antall ganger forfatternes publikasjoner om diagnostisk nøyaktighet er sitert



Forfatteren fremstår svært forvirret når han gjør en Cross Body test for AC-leddssmerte

«Hvis testen er positiv, supraspinatus er røket og jeg antar at dette er årsaken til problemet; hva vil det bety for min håndtering av pasienten?»

Er man en ivrig og akademisk orientert kliniker kan man kanskje være ekstra 'flink pike' å lese seg opp på nøyaktighetsstudier og finne gode testbatterier for seg selv. Man kan lage seg egne designbokser hvor man putter inn gode tester. For eksempel har det vist seg at man kan samle flere tester i såkalte test-clustre og dermed øke testens verdi [6]. Men selv om man er flink, leser seg opp på studier, lager diagnostiske bokser basert på de beste studiene man har, systematiserer dette i en strømlinjeformet undersøkelse av hver enkelt kne-, hoft- eller skulderpasient, står man fortsatt i fare for å overse det mest fundamentale aspektet av denne problemstillingen:

Hvordan har man beregnet diagnostisk nøyaktighet?

The devil is in the details

Det er umulig å ikke respektere forfattere og forskere som selv påpeker feil og misoppfatninger ved sitt eget arbeid. Hegedus og Cook er to slike forfattere. De står bak noen av

de mest populære studiene som er gjort innenfor muskelskjelett og har skrevet lærebøker om samme emne. Allikevel skrev de, sammen med Alexis Wright, en leder i British Journal of Sports Medicine i 2017 som for meg står igjen som én av de aller viktigste jeg har lest noen sinne [7]. Med det fengende navnet «Orthopaedic special tests and diagnostic accuracy studies: house wine served in very cheap containers» påpekte de hvor mange som har misforstått budskapet bak diagnostisk nøyaktighet. Som et paradoks er denne lederen sitert tre (!) ganger, mens studiene de har publisert om diagnostisk er sitert nesten fem hundre.

I lederen påpeker de hvordan klinikere som blir for opphengt i tester og diagnostisk nøyaktighet kan ende opp med å bruke dette som et alibi for evidensbasert praksis. Ved at man trekker kalde og kalkulerete beslutninger basert på tall fra studier, står man i fare for å miste og/eller misforstå nyansene i utføringen og tolkningen av en klinisk test. Ikke minst påpeker de hvordan man antar at en gullstandard for kalkuleringen av diagnostisk nøyaktighet ofte er bildediagnostikk og/eller en intraoperativ vurdering (for eksempel artroskopi). Dette er en diskutablet gullstandard, da det finnes

en myriade av studier som påpeker den dårlige korrelasjonen mellom bildefunn og kliniske symptomer [8-11] og variabiliteten i tolkningen av bildefunn fra person til person [12,13].

I en muskelskjelettverden som nå finner stadig flere paralleller mellom de ulike «siloeene» av forskning, ser man at veldig mange pasienter har overlappende trekk uavhengig av hvor smerten er lokalisert. Budskapet i denne utviklingen er at vi må bli flinke til å ta et steg tilbake og se personen med skulder-, nakke-, rygg- eller knesmerten. Gjennom anamnese og klinisk undersøkelse bør man utelukke alvorlig patologi og/eller andre alvorlige drivere som tilsier at pasienten bør videreføres til en annen instans. Her er naturligvis alvorlig medisinsk sykdom som cancer, systemisk inflammasjon, operativ indikasjon etc, men også andre psykososiale drivere som alvorlig depresjon, angstlidelser, spiseforstyrrelser osv. Klinikeren anbefales å sette seg inn i overlappende fellestrekk fra muskelskjelettområdet uavhengig av hva slags smertetilstand pasienten presenterer seg med. Leseren anbefales å lese artikkelen «Riv bunnen ut av siloene!» fra forrige blad (Fysioterapi i Privat Praksis

nr. 5 (2019)). Bakgrunnsartiklene for dette er også anbefalt lesing [14,15].

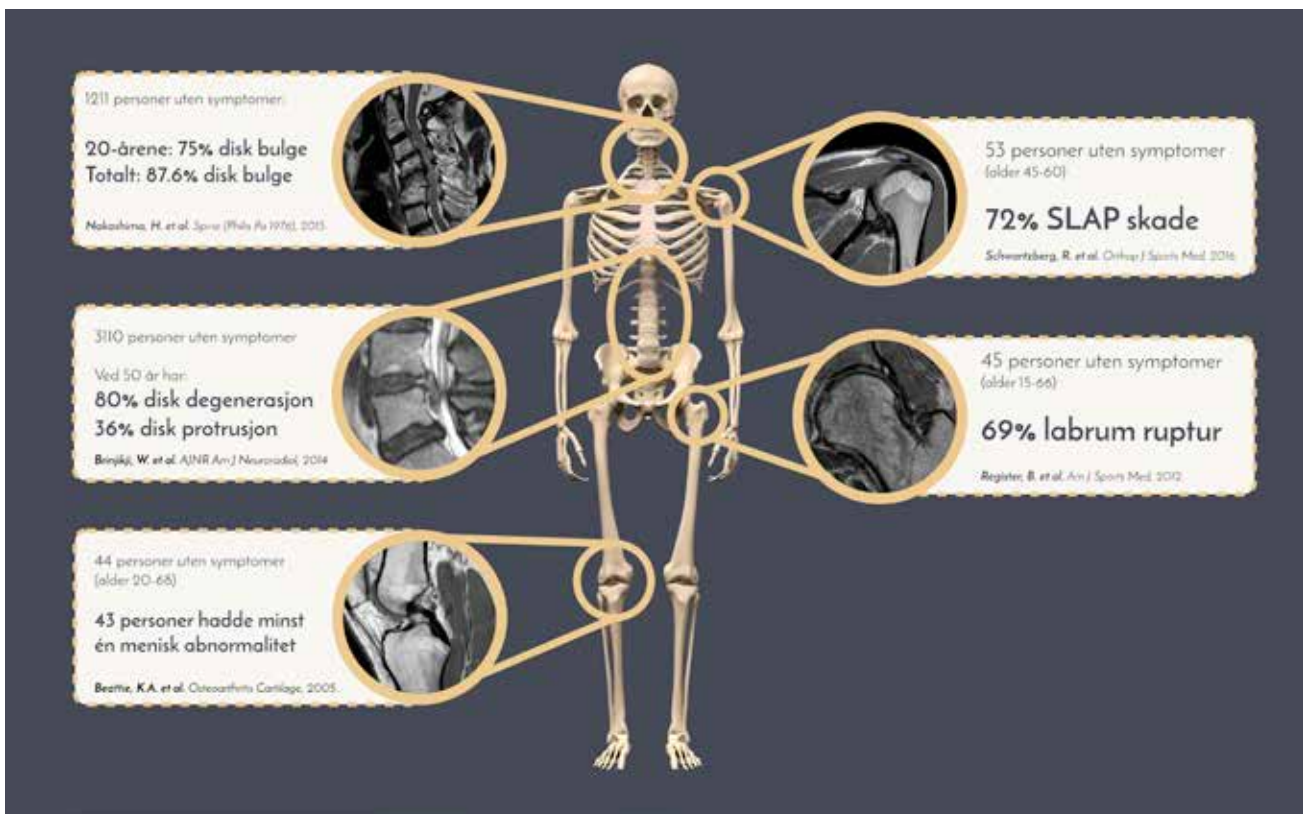
Klinikeren bør, i stedet for ukritisk å fokusere på kliniske tester for å finne en spesifikk diagnose, være årvåken for at man jobber i et biopsykososialt rammeverk. Et fagfelt hvor personen er viktigere enn diagnosen. Det kan være 'en like god test' å screene ut alvorlig sykdom, lage grove diagnostiske kategorier som man kjenner fra korsrygg (diagnostisk triage) og skulder (nakkerelatert skulderplage, stiv skulder, instabil skulder, belastningsrelatert skuldert smerte). På denne måten vil kliniske opplysninger man ikke engang anser som en test, plutselig gi verdifull informasjon som direkte påvirker resten av undersøkelse, resonnering og ultimativt rehabilitering. For å svare på det innledende spørsmålet «hva er en god test?» så kan dette være uavhengig av kalkulerede prediksjonsverdier i et sterilt excel ark. Snarere kan en god test være en systematisert anamnese og påfølgende strukturert klinisk undersøkelse som skjer som en direkte konsekvens av funnene

fra sykehistorien. En god test vil være den som plasserer personen i sentrum av undersøkelsen, hvor man etterstreber å skape en god kontekst for pasienten som tilpasser seg dennes tanker og forventninger og søker å besvare usikkerhetsmomenter og spørsmål som pasienten har. En god test kan være den som skaper en trygghet og selvsikkerhet hos personen og gir denne mestringstro på at plagen kan håndteres i samarbeid med terapeuten. En god test er, og vil alltid være, mye mer enn avanserte algoritmer og klare ja/nei svar.

Kilder:

1. Hegedus, E.J., et al.: Which physical examination tests provide clinicians with the most value when examining the shoulder? Update of a systematic review with meta-analysis of individual tests. Br J Sports Med, 2012. 46(14): p. 964-78.
2. Hegedus, E.J., et al.: Physical examination tests of the shoulder: a systematic review with meta-analysis of individual tests. Br J Sports Med, 2008. 42(2): p. 80-92; discussion 92.
3. Reiman, M.P., et al.: Diagnostic accuracy of clinical tests of the hip: a systematic review with meta-analysis. Br J Sports Med, 2013. 47(14): p. 893-902.
4. Decary, S., et al.: Diagnostic validity of physical examination tests for common knee disorders: An overview of systematic reviews and meta-analysis. Phys Ther Sport, 2017. 23: p. 143-155.
5. O'Brien, S.J., et al.: The active compression test: a new and effective test for diagnosing

- labral tears and acromioclavicular joint abnormality. Am J Sports Med, 1998. 26(5): p. 610-3.
6. Hegedus, E.J., et al.: Combining orthopedic special tests to improve diagnosis of shoulder pathology. Phys Ther Sport, 2015. 16(2): p. 87-92.
 7. Hegedus, E.J., et al.: Orthopaedic special tests and diagnostic accuracy studies: house wine served in very cheap containers. British Journal of Sports Medicine, 2017.
 8. Nakashima, H., et al.: Abnormal Findings on Magnetic Resonance Images of the Cervical Spines in 1,211 Asymptomatic Subjects. Spine (Phila Pa 1976), 2015.
 9. Brinjikji, W., et al.: Systematic Literature Review of Imaging Features of Spinal Degeneration in Asymptomatic Populations. AJNR Am J Neuroradiol, 2014.
 10. Register, B., et al.: Prevalence of abnormal hip findings in asymptomatic participants: a prospective, blinded study. Am J Sports Med, 2012. 40(12): p. 2720-4.
 11. Girish, G., et al.: Ultrasound of the shoulder: asymptomatic findings in men. AJR Am J Roentgenol, 2011. 197(4): p. W713-9.
 12. Fu, M.C., et al.: Interrater and intrarater agreements of magnetic resonance imaging findings in the lumbar spine: significant variability across degenerative conditions. Spine J, 2014. 14(10): p. 2442-8.
 13. Herzog, R., et al.: Variability in diagnostic error rates of 10 MRI centers performing lumbar spine MRI examinations on the same patient within a 3-week period. Spine J, 2016.
 14. Lin, I., et al.: What does best practice care for musculoskeletal pain look like? Eleven consistent recommendations from high-quality clinical practice guidelines: systematic review. Br J Sports Med, 2019.
 15. Caneiro, J.P., et al.: It is time to move beyond 'body region silos' to manage musculoskeletal pain: five actions to change clinical practice. Br J Sports Med, 2019.



En mengde 'patologiske' funn hos asymptotiske gjør at vi får en diskutabel gullstandard