

# Udarbejdelse af anbefalinger

Hver kliniske vejledning udmunder i anbefalingsgrader indenfor diagnostik, forebyggelse, behandling og effektmål.

## Vurdering af evidensniveau

### Behandling og Forebyggelse

Evidensniveau	Beskrivelse
1a	Systematisk review af RCT (med homogenitet af resultaterne i de individuelle studier, eks I-sq <50% eller ikke signifikant homogenitetstest).
1b	Individuelle RCT med snævre confidensintervaller.
2a	Systematisk review af kohortestudier (med homogenitet af resultaterne i de individuelle studier, eks I-sq <50% eller ikke signifikant homogenitetstest).
2b	Individuelle kohortestudier og lav kvalitets RCT (eks. mindre end 80% follow-up, manglende blinding, randomisering ikke skjult, etc.).
3a	Systematisk review af case – controlstudier (med homogenitet af resultaterne i de individuelle studier, eks I-sq <50% eller ikke signifikant homogenitetstest).
3b	Individuelle case – controlstudier.
4	Case-serier og lav kvalitets kohorte og case – controlstudier.
5	Expert udtalelser eller resultater baseret på fysiologiske studier o.lign.

### Diagnostik, undersøgelse

Evidensniveau	Beskrivelse
1a	Systematisk review af niveau 1 diagnostiske studier (med homogenitet af resultaterne i de individuelle studier, eks. I-sq <50% eller ikke signifikant homogenitetstest)  Klinisk beslutnings algoritme eller scoring systems baseret på niveau 1b studier fra forskellige kliniske centre
1b	Kohortestudier der validerer den specifikke diagnostiske test på baggrund af tidligere evidens og anvender god referencestandard  Klinisk beslutnings algoritme eller scoring system baseret på niveau 1b studier fra forskellige kliniske centre.
1c	Diagnostiske test der har så høj specificitet at et negativt resultat er ensbetydende med at diagnosen kan udelukkes.  Diagnostiske test der har så høj sensitivitet at et positivt resultat er ensbetydende med at diagnosen kan stilles.

<b>2a</b>	Systematisk review af niveau 1-2 diagnostiske studier (med homogenitet af resultaterne i de individuelle studier, eks. I-sq <50% eller ikke signifikant homogenitetstest).
<b>2b</b>	Eksplorative kohortestudier der anvender god referencestandard.  Afledt klinisk beslutnings algoritme eller scoring system.
<b>3a</b>	Systematisk review af niveau 1-3 studier (med homogenitet af resultaterne i de individuelle studier) (eks I-sq <50% eller ikke signifikant homogenitetstest)
<b>3b</b>	Ikke konsekutivt studie eller ikke konsekvent anvendt reference standard
<b>4</b>	Case kontrol studier, studier med uegnet eller ikke uafhængig testet referencestandard
<b>5</b>	Expert udtalelser eller resultater baseret på fysiologiske studier o. lign

## Diagnostiske test

Den prædiktive værdi af en positiv og negativ test er for klinikerer det vigtigste mål. Den prædiktive værdi af en positiv test angiver sandsynligheden for at en person, der er testet positiv virkelig har den tilstand personen er testet for. På samme måde angiver den prædiktive værdi af en negativ test sandsynligheden for at den person, der er testet negativ virkelig ikke har den tilstand personen er testet for.

Hvilken størrelse af f.eks. den prædiktive værdi af en positiv test klinikerer vil acceptere vil naturligvis være afhængig af, hvilket problem der undersøges og ikke mindst hvilke konsekvenser testen har for behandling. Umiddelbart vil man kræve større prædiktiv værdi af testen for om ACL er overrevet end ved forudsigelsen af fald, som jo af mange grunde vil forventes at være mere usikker. Der er derfor ikke opstillet specifikke grænser for hvornår det er acceptabelt og ikke, men det er naturligvis bedst at både sensitivitet, specificitet og prædiktiv værdi af positiv og negativ test er så tæt på 1 som muligt (svarende 100 % sikkerhed).

Den diagnostiske evne af en klinisk test kan dog vurderes ved udregning af likelihoodratioen for en positiv og en negativ test som udtrykker, hvor meget troen på om patienten har det undersøgte problem ændrer sig fra før testen til efter at vi har gennemført testen. Likelihoodratioen for en positiv test (LH+) udregnes som sensitivitet/(1 - specificitet) og likelihoodratioen for negativ test (LH-) kan udregnes som (1-sensitiviteten)/ specificiteten.

Pre-test sandsynligheden er prævalensen (eller hyppigheden) af det undersøgte problem i den population (gruppe) som man undersøger med testen – eks. på en idrætsklinik kan hyppigheden blandt de patienter der efter et knætraume har en ruptur af forreste korsbånd være 30%. Post-test sandsynligheden for LH+ er den prædiktive værdi af en positiv test og post-test sandsynligheden for LH- er den prædiktive værdi af en negativ test, så likelihoodratioen siger dermed noget om testens diagnostiske evne.

## Vurderingen af den diagnostiske evne kan klassificeres efter

LH+ > 10 og LH- < 0,1 har stor diagnostisk evne, da testen giver store og ofte konklusive ændringer fra pre-test til post-test sandsynlighed
LH+ på 5 - 10 og LH- på 0,1 til 0,2 har moderat diagnostisk evne da testen genererer moderate ændringer fra pre-test til post-test sandsynlighed

LH+ på 2 - 5 og LH- på 0,2 til 0,5 har lille diagnostisk evne, da testen giver små (men nogen gange væsentlige) ændringer fra pre-test til post-test sandsynlighed
LH+ på 1 - 2 og LH- på 0,5 til 1 har lille diagnostisk evne, da testen giver små og sjældent væsentlige ændringer fra pre-test til post-test sandsynlighed.

### Det betyder

at en sensitivitet og specificitet en større end 0.9 er høj da testen dermed har stor diagnostisk evne
at en sensitivitet og en specificitet større end 0.85 og mindre end eller lig med 0.9 er moderat da testen dermed har moderat diagnostisk evne
at en sensitivitet og en specificitet større end 0.65 og mindre end eller lig med 0.85 er lille da testen dermed har lille diagnostisk evne
at en sensitivitet og en specificitet mindre end eller lig med 0.65 er meget lille da testen dermed har meget lille diagnostisk evne

### Referencer

1. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. JAMA 1994;271:703-7.
2. Rud B, Matzen P, Hilden J. Mål for diagnostiske tests ydeevne. Ugeskrift for læger, 2005;167, 3018-22.

## Effektstørrelse

Effektstørrelsen af en intervention udtrykkes ofte i Cohen d, som udtrykker effekten af en intervention divideret med standarddeviationen (SD) af den samlede gruppe. Dette estimat er dog let biased så det overvurderer effekten af en intervention, så estimatet for effektstørrelse justeres ofte til Hedges g, men tolkningen af disse effektstørrelser er ens. En effektstørrelse på 0,2 anses for at være en lille effekt, 0,5 for en moderat effekt og over 0,8 er en stor effekt. Estimater for effektstørrelse kan dog overvurdere effekten, hvis patientgruppen er meget homogen og dermed har lille standard deviation. I den endelige vurdering af effektstørrelsen kan indrages Minimal Important Difference (MID) svarende til den effekt en patientgruppe vurderer som den mindste betydende effekt, hvis denne værdi er til rådighed. Hvis effekten ikke overstiger MID vil effekten vurderes til at være lav.

1. Cohen J. Statistical power analysis for the behavioral sciences 2nd ed. Hillsdale, NJ: Lawrence Earlbaum Associates, 1988.
2. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to Meta-analysis. John Wiley & sons, Ltd. United Kingdom 2009.

## Samlet vurderingen af evidensen

I den samlede vurdering af evidensen vurderes

1. Kvaliteten af de inkluderede studier (Risk of bias). Herunder også hvis små studier (og studier af lav kvalitet) generelt har bedre resultater end større studier og studier af god kvalitet (risk of small study bias).
2. I hvor høj grad resultaterne er rimeligt ensartede imellem de inkluderede studier (Inconsistency)
3. Om effekten er rimeligt præcist estimeret (få patienter giver brede sikkerhedsintervaller for effekten) (Imprecision)
4. Om målingen af resultatet er målt på relevante outcome, om patientgruppen er relevant for problemstillingen og om sammenligningen af to behandlinger er direkte (dvs. i samme studie) eller indirekte (to behandlinger er begge sammenlignet med kontrol) (Indirectness).

Hvis der er væsentlige problemer i forhold til en eller flere af ovenstående nedgraderes evidensen et eller flere evidensniveauer. Derudover kan evidensen nedgraderes hvis den samlede effekt er meget lille og opgraderes hvis den samlede effekt er meget stor.

## Referencer

1. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, Montori V, Akl EA, Djulbegovic B, Falck-Ytter Y, Norris SL, Williams JW Jr, Atkins D, Meerpohl J, Schünemann HJ. GRADE guidelines: 4. Rating the quality of evidence -study limitations (risk of bias). *J Clin Epidemiol.* 2011;64(4):407-15.
2. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, Devereaux PJ, Montori VM, Freyschuss B, Vist G, Jaeschke R, Williams JW Jr, Murad MH, Sinclair D, Falck-Ytter Y, Meerpohl J, Whittington C, Thorlund K, Andrews J, Schünemann HJ. GRADE guidelines 6. Rating the quality of evidence -imprecision. *J Clin Epidemiol.* 2011;64(12):1283-93.
3. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Glasziou P, Jaeschke R, Akl EA, Norris S, Vist G, Dahm P, Shukla VK, Higgins J, Falck-Ytter Y, Schünemann HJ; GRADE Working Group. GRADE guidelines: 7. Rating the quality of evidence -inconsistency. *J Clin Epidemiol.* 2011;64(12):1294-302.
4. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Falck-Ytter Y, Jaeschke R, Vist G, Akl EA, Post PN, Norris S, Meerpohl J, Shukla VK, Nasser M, Schünemann HJ; GRADE Working Group. GRADE guidelines: 8. Rating the quality of evidence -indirectness. *J Clin Epidemiol* 2011;64(12):1303-10.
5. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, Alonso-Coello P, Djulbegovic B, Atkins D, Falck-Ytter Y, Williams JW Jr, Meerpohl J, Norris SL, Akl EA, Schünemann HJ. GRADE guidelines: 5. Rating the quality of evidence -publication bias. *J Clin Epidemiol.* 2011;64(12):1277-82.

## Figurtekst til forklaring af evidensniveauerne

Evidensniveau	Beskrivelse
<b>Level 1</b>	betyder, at der er velgennemførte systematiske review af randomiserede studier (behandling og forebyggelse) eller tværnsnitsstudier med blinding og en konsistent anvendt referencestandard (diagnose)
<b>Level 2</b>	betyder, at der er velgennemførte enkelt randomiserede studier eller observationelle studier med dramatisk effekt (behandling og forebyggelse) eller tværnsnitsstudier med blinding og en konsistent anvendt referencestandard (diagnose)
<b>Level 3</b>	betyder, at der er velgennemførte ikke randomiserede kontrollerede studier / kohorte eller follow-up studier (behandling og forebyggelse) eller ikke konsekutive studier eller studier uden og en konsistent anvendt referencestandard (diagnose)
<b>Level 4</b>	betyder at der er case-serier, case control studier eller studier med historiske controller (behandling og forebyggelse) eller case control studier eller studier med ringe eller ikke uafhængig reference standard (diagnose)
<b>Level 5</b>	er evidens baseret på mekanisme studier (behandling og forebyggelse) og (diagnose)

Levels kan nedgraderes på grund af lav studie kvalitet, brede sikkerhedsintervaller, indirektheden (pga. ikke patientrelevante outcome, indirekte sammenligninger og undersøgelsen er på en anden patientgruppe) og inkonsistens mellem resultaterne eller hvis effekten er meget lille. Levels kan opgraderes, hvis effekten er meget stor.

adssdfsdaf