

**RESUS**REGION SJÆLLANDS
UDDANNELSES SYSTEM

Checkliste til kritisk vurdering af diagnostiske studier

Baggrund?

Hvilket studie er det du vurderer? Hvad er, ultrakort, den nødvendige baggrundsviden for at du kan tage stilling til studiets resultater?

Studie:

PubMed ID:

DOI:

1. Hvad er baggrunden for at studiet er lavet?

Why should I care?

Du kan ud fra artiklens abstract ofte komme frem til de vigtigste detaljer omkring artklen, der hjælper dig til at vurdere om artiklen er din tid værd at læse. Start med følgende vurdering:

2. Hvilket spørgsmål stiller artiklen?:

3. Er PICTR-analysen optimal for at svare på spørgsmålet? Ja Nej Uklart

Population:

Indekstest:

Comparison:

Target Condition:

Referencestandard:

4. Er studiet relevant for min praksis?

Ja

Nej

Uklart

**RESUS**REGION SJÆLLANDS
UDDANNELSES SYSTEM

Metoder

I et diagnostisk studie er der særlige faktorer der spiller ind når du skal vurdere studiets metoder. Særligt Rekruttering, Praktisk gennemførelse og Sammenligningsgrundlaget er vigtige.

1. Rekruttering

Omfatter inklusionskriterierne mine egne patienter? Ja Nej Uklart

Er et bredt nok spektrum af patienter inkluderet? Ja Nej Uklart

Er populationsstørrelsen sufficient? Ja Nej Uklart

Udelukker eksklusionskriterierne vigtige patienter? Ja Nej Uklart

Blev patienter inkluderet fortløbende? Ja Nej Uklart

2. Praktisk gennemførelse

Er metoderne beskrevet detaljeret nok til reproduktion? Ja Nej Uklart

Er indekstesten udført praktisk reproducerbart? Ja Nej Uklart

**RESUS**REGION SJÆLLANDS
UDDANNELSES SYSTEM

Blev "guldstandarden" udført på alle patienter?

Ja Nej Uklart

Blev eventuelle grupper behandlet ens?

Ja Nej Uklart

3. Sammenligningsgrundlag

Var der signifikant frafald af patienter?

Ja Nej Uklart

Er "guldstandarden" tilstrækkelig præcis?

Ja Nej Uklart

Blev testresultaterne vurderet blændet?

Ja Nej Uklart

Indgår den diagnostiske test som del af "guldstandarden"

Ja Nej Uklart

Blev der udført længerevarende klinisk opfølgning?

Ja Nej Uklart

**RESUS**REGION SJÆLLANDS
UDDANNELSES SYSTEM

Resultater

I vurderingen af studiets resultater er det afgørende at holde sig for øje hvilke mål der er valgt for at visualisere resultaterne og vurdere om målene fremstiller resultaterne mest korrekt.

1. Hvad var resultaterne?

2. Er resultaterne troværdige?

Præsenterer de anvendte mål resultaterne mest korrekt? Ja Nej Uklart

Er størrelse og præcision af effekten klinisk signifikant? Ja Nej Uklart

Er fragility index > patienter lost to follow up? Ja Nej Uklart

Er der signifikante interessekonflikter? Ja Nej Uklart

Er resultaterne i linje med tidligere studier? Ja Nej Uklart

3. Blev mulig skadelig virkning tilstrækkeligt udredt? Ja Nej Uklart

**RESUS**REGION SJÆLLANDS
UDDANNELSES SYSTEM**Klinisk anvendelighed**

Efter at have analyseret studiet og resultaterne er sidste skridt at vurdere om resultater som du mener at kunne stole på kan anvendes i din kliniske praksis. Stil følgende spørgsmål:

1. **Er mine patienter for forskellige fra studiets?** Ja Nej Uklart

2. **Er interventionen gennemførlig hos mig?** Ja Nej Uklart

3. **Udvejer fordele evt. skade ved interventionen?** Ja Nej Uklart

Bottom Line

Sammenfat din kritiske vurdering til et kort og koncist afsnit som omfatter studiets hypotese, population, metoder, potentielle bias/confoundere samt eventuel indflydelse på klinisk praksis.



Appendix: Begrebsforklaringer og links

1. Artiklens spørgsmål:

Det bør fremgå alene fra artiklens titel hvilket spørgsmål studiet forsøger at svare på. Er det ikke oplagt ud fra titlen må svaret søges i artiklens abstract. Fremgår spørgsmålet ikke af abstract er artiklen oftest ikke værd at læse.

2. PICTR-analyse:

I diagnostiske tests kan PICO-analysen justeres en lille smule, da det kan være svært at identificere præcis hvad der er intervention og comparison. Population er fortsat vigtig, men interventionen udskiftet i PICTS-analysen med indekstest, som er den diagnostiske test, den test hvis effektivitet undersøges. Comparison eller sammenligning er fortsat den samme, hvis der er en (f.eks. to forskellige risikoscorer for AKS). T står for Target Condition, som er den tilstand som det forsøges at diagnosticere. R står for Referencestandard, som også kaldes "guldstandard". Det er vigtigt at der i studiet findes en pålidelig referencestandard, da det ellers ikke er muligt at sige hvilke patienter der reelt havde tilstanden, om indekstesten så indikerede det eller ej. De overordnede PICTR-oplysninger bør fremgå af artiklens abstract, men kan findes i større detaljer i metodesektionen.

3. Relevans:

Såfremt studiet viser sig at være korrekt udført og med signifikante resultater, vil du så kunne anvende resultaterne i dit kliniske arbejde? Igen fremgår resultaterne af abstract, men her kan du også gå direkte til studiets resultatsektion. Hvis resultaterne, selv hvis de skulle være valide, ikke er interessante for din praksis er der ikke grund til at læse videre i artiklen.

4. Inklusionskriterier/[Selektionsbias](#):

For at studiet skal være relevant for din kliniske praksis bør det inkludere patienter du ser i dit kliniske arbejde. Det kan godt dreje sig om patienter som i studiet er rekrutteret fra en anden setting, men også kan præsentere i din afdeling, men i det tilfælde skal du være opmærksom på at der kan være tale om en præselektering af en patientgruppe af potentielt mange du ser på din afdeling og resultaterne derfor kun kan overføres til samme patientgruppe.

5. Patientspektrum/[Sampling bias](#):

I et diagnostisk studie er det vigtigt at hele spektret af patientpræsentationer – milde/alvorlige, tidlige/sene – er repræsenteret i studiets patientpopulation. Er dette ikke tilfældet begrænser det applicerbarheden af studiet til lige præcis den type præsentation der undersøges i studiet. Inklusionskriterierne bør derfor være så brede som mulige (f.eks. er "patienter indlagt med mistanke om AKS" et bedre inklusionskriterie end "patienter med trykkende centrale brystmerter og udstråling til venstre arm" i et studie om troponiner, da den sidste kun omfatter en brøkdel af de patienter som vi udfører testen på.

6. Populationsstørrelse/[Sample Size bias](#):

Vær opmærksom på størrelsen af studiets patientpopulation, da statistisk signifikans kan opnås med en lille patientpopulation, men vil medføre større risiko for og konsekvenser af bias og confounders. Link: <https://emj.bmj.com/content/20/5/453#boxed-text-1>

7. Eksklusionskriterier/Selektionsbias:

Ligesom inklusionskriterierne er eksklusionskriterierne afgørende for hvilken gruppe af patienter du kan overføre potentielt positive resultater til. Brug tid på at vurdere om eksklusionskriterierne eventuelt har udelukket en gruppe patienter som kunne have påvirket udfaldet i positiv eller negativ retning (f.eks. er patienter der går på intensiv ekskluderet kan vi kun overføre resultaterne til mildere tilfælde af sygdommen og ikke automatisk gå ud fra at den diagnostiske test vil have samme nøjagtighed hos patienter med mere alvorlige symptomer).

8. Fortløbende inklusion/Sampling bias:

Er patienterne inkluderet kronologisk og uden bias i form af f.eks. udelukkende rekruttering i dagstiden/ved tilstedeværelse af projektforskere, hvilket kan føre til udelukkelse af eksempelvis patienttyper der typisk præsenterer sig på andre tidspunkter. Ligeledes kan der eksempelvis ved spørgeskemaundersøgelser opstå bias mod at kun mere ressourcefulde patienter besvarer spørgeskemaet.

9. Metodebeskrivelse og -udførelse:

I diagnostiske studier er det særligt vigtigt at de praktiske forhold omkring hvordan den nye test blev udført er beskrevet i en detalje som muliggør praktisk reproduktion. Er testen for eksempel udført af højt specialiserede klinikere (f.eks. neurooftalmologer) er det usikkert at testen, selvom studiets resultater er positive, vil præstere lige så flotte resultater udført af KBU-læger i en travl akutafdeling. I bedste tilfælde vil der være beskrevet en interobserver variability i form af en kappa-værdi der klargør om testen kan udføres eller tolkes pålideligt af flere forskellige klinikere.

10. Guldstandard og diagnostisktest udført på alle patienter/Verification bias:

I visse studier er det ikke praktisk eller økonomisk muligt at gennemføre "guld-standarden" (referencestandard) på alle patienter inkluderet i studiet (f.eks. hvis guld-standarden er invasiv eller dyr), hvilket kan føre til skævvridning af resultaterne til fordel for den diagnostiske test (indekstesten), hvis en negativ diagnostisk test hos lav-risiko patienter automatisk regnes for sandt negativ.

11. Ens behandling/Co-intervention bias:

Hvis studiet havde flere grupper (f.eks. to diagnostiske tests der bliver sammenlignet), blev patienterne i de forskellige grupper så behandlet ens, eller er der risiko for at en af studiets grupper i højere grad har fået tests eller behandling sideløbende, som kan have påvirket resultatet af studiet?

12. Lost to follow-up/Attrition bias:

Der vil i de fleste studier være et vist antal patienter der frafalder, men ved et signifikant frafald vil studiets resultater uvægerligt blive påvirket. En tommelfingerregel er at loss to

follow-up helst skal være under 5% og ikke må overstige 20%. Særligt hvis frafaldet primært findes i den ene af studiets grupper er der risiko for bias.

13. Pålidelig "guldstandard"/Imperfect Gold Standard Bias:

For at vi kan stole på resultaterne i et diagnostisk studie er det afgørende at vi kan stole på testen der anvendes som "guldstandard", da vi ellers risikerer at patienter der enten har eller ikke har den undersøgte tilstand klassificeres med forkerte outcomes. I visse situationer findes der ikke en tilstrækkelig pålidelig test for den undersøgte tilstand, i hvilket tilfælde der bør være udført forsøg på at imødekomme evt. misklassifikation, såsom ved længerevarende klinisk follow-up.

14. Blænding/Observer bias:

Blænding er, på lige fod med randomisering, et af de vigtigste elementer for at undgå bias. Blænding kan ske af både patienter, klinikerne som står for behandlingen, samt forskerne der analyserer resultaterne.

I diagnostiske studier er det særligt vigtigt at den diagnostiske test vurderes uafhængigt og blændet af resultatet af "guldstandard" således at der ikke opstår bias.

15. Sammensat "guldstandard" og diagnostisk test/Incorporation bias:

Den diagnostiske test bør være uafhængig af guldstandard, da resultaterne af den diagnostiske test ellers ikke vil kunne vurderes uafhængigt af guldstandard og der er risiko for at resultaterne skævvrides til fordel for den diagnostiske test. Er det ikke muligt at adskille den diagnostiske test fra guldstandard bør der udføres yderligere opfølgning af patienter med negative resultater med henblik på at diagnosticere evt. falsk negative og positive patienter bør verificeres af andre veje (f.eks. kirurgisk bekræftelse).

16. Resultatpræsentation:

Der findes et hav af forskellige muligheder at præsentere et studies data på og det er afgørende at forholde sig kritisk til hvilken metode der er blevet valgt, da ikke alle metoder præsenterer resultaterne lige objektivt. Eksempelvis kan en forholdsvis beskeden effekt af en intervention, såsom nedsat risiko for DVT ved P-pillebrug fra 0,1% til 0,05% præsenteres som en 50% risikoreduktion ved at anvende Relative Risk Reduction (RRR), men udgør i absolutte tal reelt kun en 0,05% sænkning af risikoen i Absolute Risk Reduction).

17. Klinisk signifikant effekt:

I vurderingen af effekten af en intervention skal ikke kun patienterne der har haft effekt af behandlingen tages i betragtning, men også de patienter som ingen effekt havde, eller ligefrem led skade. Ved at kigge udelukkende på procentvis øget overlevelse e.l. forsvinder disse data. Begreberne Number-Needed-to-Treat (NNT) og Number-Needed-to-Harm (NNH) er i senere tid blevet anvendt i stigende grad, da et tal som NNT=150 udtrykker at, selvom effekten for den enkelte patient måske er signifikant, skal der behandles 149 andre patienter med potentielle bivirkninger, før effekten ses.

18. Klinisk præcis/Statistisk signifikant effekt/konfidensintervaller:

Da et klinisk studie er udført på en udvalgt gruppe af befolkningen og der kan introduceres bias i selektionen kan det være svært at vurdere i hvor høj grad resultaterne kan overføres til resten af populationen. Her kan konfidensintervallerne og p-værdierne hjælpe os. Er konfidensintervaller meget snævre er der højere sandsynlighed for at studiet har ramt plet og resultaterne kan overføres til resten af populationen. Er konfidensintervallerne brede, og særligt hvis de krydser over grænsen hvor interventionen ikke har nogen effekt (hyppigst hvis konfidensintervallet krydser over 1, hvilket vil betyde ingen statistisk signifikans) er det meget lidt sandsynligt at vi kan stole på resultaterne.

19. Fragility Index:

Hvis en intervention viser sig at have en signifikant effekt med p-værdi $<0,05$ kan det være nyttigt at afdække studiets "Fragility Index", som er et udtryk for hvor mange patienter der skulle have haft et andet outcome end positivt for at studiet faldt under den statistisk signifikante linje. Er fragility index mindre end patienter tabt til follow up er det særligt bekymrende, da det så reelt er muligt at studiet, såfremt alle patienter var blevet analyseret, ville have haft et andet udfald. Fragility index kan beregnes her:

<https://clincalc.com/Stats/FragilityIndex.aspx>

20. Interessekonflikter:

Det er efterhånden standard praksis at angive interessekonflikter i studier og alvorlige interessekonflikter bør ikke forekomme. I visse tilfælde kan det dog give mening at søge på forfatterne eller sponsorer af studiet (særligt i medicinalstudier) for at få yderligere information om potentielle interessekonflikter som ikke er anført.

21. Samlet evidensbyrde:

Når studiets resultater efter den kritiske vurdering er afdækket kan sandsynligheden for at der er tale om et sandt resultat øges ved at opveje det mod den hidtidige evidensbyrde. Er resultatet på linje med den tidligere evidens øges sandsynligheden. Går resultatet imod den etablerede viden kan evidenshierarkiet bruges til at afgøre om studiet vægter tungt nok til at kunne tippe evidensbyrden. Foreligger der primært case reports eller retrospektive observationelle studier kan et veludført randomiseret, kontrolleret studie godt veje tungere end det foreliggende.

22. Skadelig virkning:

De fleste interventioner har bivirkninger. Et veludført studie vil også have undersøgt potentielle skadelige virkninger, som må afvejes ift. Behandlingseffekten. Her er NNT og NNH gode værktøjer. NNT og NNH kan beregnes her: <https://clincalc.com/Stats/NNT.aspx>

23. Klinisk anvendelighed:

I sidste ende, efter at have vurderet hele studiets effekt og været kommet frem til konklusionen at eventuelle positive resultater er sande og vægter tungt nok til at kunne ændre din praksis, resterer at vurdere hvorvidt dette overhovedet er muligt. Er der i din kliniske dagligdag adgang til de nødvendige ressourcer og ser du overhovedet patientpopulationen som studiet har påvist effekt af interventionen hos?



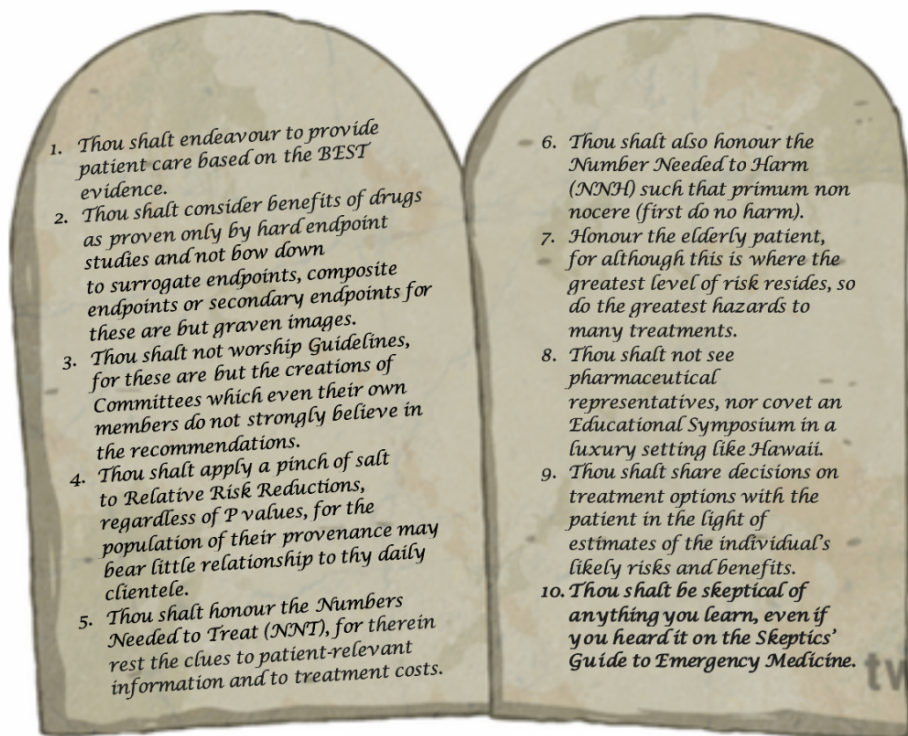
Links:

- Større artikel om vigtigheden af EBM og kritisk vurdering af artikler: <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124>
- Kortere, men god introduktion til Evidence Based Medicine: <https://first10em.com/ebmiseasy/>
- Liste over en lang række mulige bias: <https://catalogofbias.org/biases/>
- En alternativ checkliste til diagnostiske studier: <https://www.cebm.ox.ac.uk/files/ebm-tools/diagnostic-accuracy-studies.pdf>
- EBM ressourcer og andre checklists: <https://first10em.com/ebm/>

Videoer:

- <http://www.sketchyebm.com/evidence-based-medicine-videos/>

The 10 Commandments of Evidence-Based Medicine



Adapted from Dr. John S. Yudkin, Emeritus Professor of Medicine, University College London and published in blog by [Dr. Richard Lehman](#)