

Noter til SfR checkliste 5 – Diagnostiske tests

Sektion 1 afdækker den interne validitet af artiklen. Det vurderes om undersøgelsen er udført grundigt, og om undersøgelsens udfald skyldes den behandling der undersøges. Hvert spørgsmål afdækker et område af metodologien, som har væsentligt betydning for en undersøgelses konklusioner. For hvert spørgsmål skal det vurderes i hvor høj grad evalueringskriteriet er opfyldt.

Sektion 1

1.1 *Blev testen sammenlignet med en valid referencestandard ("Gold standard")?*

For at vurdere effektiviteten af en ny eller en alternativ diagnostisk test, må den sammenlignes med en referencestandard ("Gold standard"), således at forskeren kan få et klart billede af, hvor god testen er til at udpege syge eller raske i målpopulationen. Referencestandarden kan være en eksisterende test eller en diagnostisk metode, hvis præcision er kendt og ligger indenfor klart afgrænsede rammer. Det bør være angivet hvilken standard, der er anvendt og evidensen for at standarden er anvendelig til den undersøgte test.

1.2 *Blev testen og referencestandardens vurderet uafhængigt af hinanden (blindet)?*

Den undersøgte test og reference standarden bør anvendes på de samme patienter for en sammenligning af måleresultaterne er troværdig. Forskere, som udfører den enkelte test, bør være blindede overfor resultatet af den anden test, for at undgå bias i vurderingen. **Hvis undersøgelsen ikke er blindet eller det ikke er angivet, bør undersøgelsen forkastes.**

1.3 *Blev patienter udvalgt til referencestandardens uafhængigt af testresultatet?*

Bedst er det at applicere både testen og referencestandardens på alle patienter. Hvis det er nødvendigt at vælge, bør udvælgelsen ske således, at der ikke er kendskab til resultatet af den tidligere test. **Hvis forskeren har kendskab til tidligere testresultater, før resultatet af referencetesten vurderes, bør undersøgelsen forkastes, på grund af en høj risiko for bias.**

Sekretariatet for Referenceprogrammer, SfR

1.4 Blev såvel referencestandarden som den diagnostiske test målt på alle patienter?

For at kunne vurdere overensstemmelse mellem den diagnostiske test og referencestandarden er det vigtigt, at såvel referencestandarden som den diagnostiske test bliver målt på alle patienter. Ellers risikerer man, at der tegnes et skævt billede af graden af overensstemmelsen. **Hvis alle patienter ikke får målt både standard og test, bør undersøgelsen nedgraderes og eventuelt forkastes.**

1.5 Blev referencestandarden målt, før en intervention/behandling blev iværksat på baggrund af initiale testresultater?

I nogle tilfælde er det af etiske årsager nødvendigt at starte behandlingen så snart det første diagnostiske testresultat foreligger. Hvis dette er tilfældet, vil det ofte påvirke resultatet af en efterfølgende diagnostisk test. Dette kan være uundgåeligt, men bør tages med i overvejelserne, når pålideligheden af undersøgelsen skal vurderes.

1.6 Blev forskellige tests sammenlignet i et validt design?

Hvis der sammenlignes mere end en test, er det særlig vigtigt at sikre sig at forskerens sammenligninger er blindede overfor de øvrige tests. Dette opnås ved at sikre sig, at den forsker, der udfører den anden eller de efterfølgende tests på den samme patient, er blindet overfor resultatet af den tidligere test, hvilket i praksis gøres ved, at de forskellige test udføres på tilfældigt udvalgte personer fra populationen.

Enhver sammenligning af tests, der ikke er blindet, bør forkastes.

Sektion 2 forholder sig til den generelle metodologiske kvalitet af artiklen, baseret på svarene i sektion 1, graderet i tre grader:

++ Alle eller de fleste kriterier er opfyldt

Kriterier, der ikke er opfyldt, vil **meget sjældent** ændre undersøgelsens konklusioner

+ Nogle kriterier er opfyldt

De kriterier, der ikke er opfyldt eller grundigt beskrevet, vil **sjældent** ændre studiets konklusioner

÷ Få eller ingen kriterier er opfyldt

Sekretariatet for Referenceprogrammer, SfR

Undersøgelsens konklusioner kan **meget vel ændres**.

Denne graduering, sammenholdt med typen af undersøgelsen, afgør det endelige evidensniveau.

Formålet med de to andre spørgsmål er at sammenfatte Jeres syn på artiklens kvalitet, og dens anvendelighed på netop Jeres referenceprogrammers patientmålgruppe.

3.1 Var metoden der anvendtes til at vurdere testen i fht. referencestandardens tilstrækkelig?

Undersøgelsen skal som et minimum indeholde data til beregning af testens sensitivitet, specificitet (nosografiske rater) samt positive og negative prædiktive værdier (diagnostiske rater) (se Tabel 1).

Tabel 1 Illustration af beregning af sensitivitet, specificitet og prædiktive værdier

		Diagnose / Guldstandard			
		Til stede	Ikke tilstede		
Test	Positiv	a	b	a+b	Positiv prædiktiv værdi Ved positiv test, hvor stor er sandsynligheden for, at pt. er syg? a/a+b
	Negativ	c	d	c+d	Negativ prædiktiv værdi Ved negativ test, hvor stor er sandsynligheden for, at pt. er rask? d/c+d
	I alt	a+c	b+d	a+b+c+d	Prævalens Andelen af personer i populationen der har sygdommen a+c/ a+b+c+d
		Sensitivitet Sandsynligheden for testresultatet = "syg", givet at man faktisk er syg a/a+c	Specificitet Sandsynligheden for testresultatet = "rask", givet at man faktisk er rask d/b+d		