

Hvordan evalueres screeningsinstrumenter?

Ved dokumentation for den diagnostiske værdi af en screeningstest anvendes begreber som:

- Sensitivitet
- Specificitet
- Area under the curve (AUC)
- Likelihood ratio
- Prædiktiv validitet

Sensitivitet og specificitet

Den enkleste måde at dokumentere den diagnostiske værdi af en screeningstest er, at angive instrumentets sensitivitet (evne til at identificere 'syge') og specificitet (evne til at identificere 'raske').

Sensitivitet og specificitet afhænger både af selve instrumentet og af de forsøgsgrupper (samples), som det er blevet afprøvet på.

En screeningstest, der afprøves på fx en gruppe personer med middelsvær demens og en gruppe kognitivt velfungerende ældre, vil udvise høj sensitivitet og specificitet. Men når samme screeningstest anvendes i klinisk praksis, hvor mange har MCI, let demens eller fejler noget helt andet, ses ofte et markant fald i sensitivitet og specificitet.

Da sensitivitet og specificitet således er delvist sample-afhængige mål, kan man ikke uden videre sammenligne dokumentationen på tværs af forskellige screeningsinstrumenter.

Area under the curve

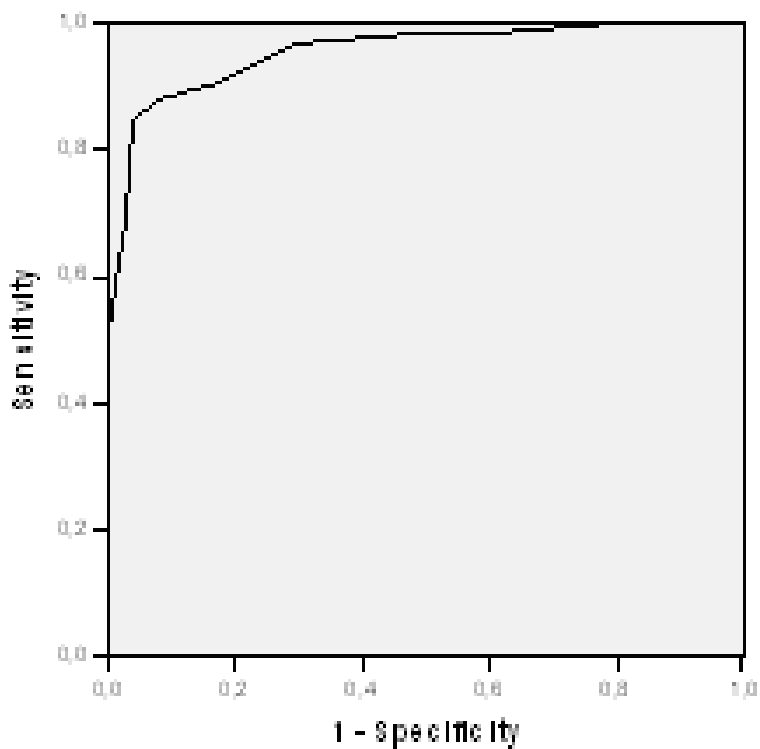
Ved screeningstests med en justerbar cut-off-værdi som fx MMSE er angivelse af et enkelt sæt værdier for sensitivitet og specificitet ikke særlig oplysende.

Sensitivitet og specificitet er indbyrdes forbundet og vil i reglen ændre sig i hver sin retning, hvis man flytter cut-off værdien. Fx vil en justering af cut-off i MMSE opad (i retning mod 30) øge testens sensitivitet med et samtidigt fald i specificiteten. Tilsvarende vil en justering nedad (i retning mod 20 eller lavere) medføre øget specificitet, men ringere sensitivitet.

Forholdet mellem sensitivitet og specificitet kan vises grafisk i en såkaldt receiver operating characteristics (ROC) kurve, der kombinerer værdierne for sensitivitet og specificitet ved alle tænkelige cut-offs (for MMSE drejer det sig fx om 29 kombinationer).

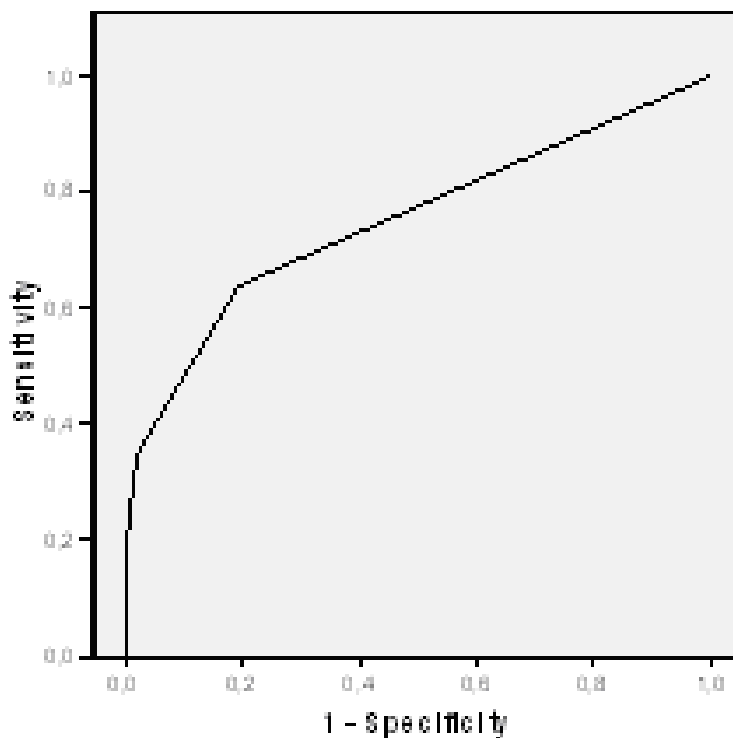
Høje værdier af sensitivitet og specificitet medfører et stort areal under kurven (area under the curve; AUC). For en test med variabel cut-off, som fx MMSE udgør arealet under kurven et mere relevant mål for den diagnostiske performance end angivelsen af et enkelt sæt værdier for sensitivitet og specificitet.

ROC Curve



Eksempler på to ROC-kurver. I den øverste ROC-kurve er både sensitivitet og specificitet høj, hvilket giver et forholdsvis stort area under the curve (AUC). I den nederste ROC-kurve er sensitiviteten ikke særlig god, hvilket giver et mindre areal under kurven.

ROC Curve



Likelihood ratio

For en screeningstest med en enkelt eller nogle få cut-off værdier (fx en blodprøve) kan sensitivitet og specificitet kombineres i den såkaldte likelihood ratio, der udtrykker styrken af testens diagnostiske sikkerhed.

Likelihood ratio for et positivt testresultat (LR+) angiver styrken af den diagnostiske sikkerhed for, at en person med et positivt testresultat har pågældende sygdom. Tilsvarende kan der beregnes en likelihood ratio for et negativt testresultat (LR-), der angiver styrken af den diagnostiske sikkerhed for, at en person med et negativt testresultat ikke har sygdommen.

En mulig ulempe ved likelihood ratios er, at mange i praksis har svært ved at fortolke dem. Princippet er dog forholdsvis enkelt. Formlen for LR+ er konstrueret, så høje værdier for sensitivitet og specificitet automatisk medfører en høj LR+. Det vil sige, jo højere LR+, jo stærkere diagnostisk sikkerhed. For LR- forholder det sig lige modsat. Her er formelen konstrueret, så høje værdier for sensitivitet og specificitet medfører en lav LR-. Så jo lavere LR-, jo bedre.

Værdier for LR+ og LR- omkring 1 – hvilket fx fremkommer, hvis både sensitivitet og specificitet ligger omkring 0,5 – svarer til, at testresultatet ikke rummer nogen sikkerhed overhovedet.

Der er udviklet kriterier for fortolkning af likelihood ratios baseret på opdeling i intervaller, men en tommelfingerregel siger, at en test med en god diagnostisk sikkerhed har en LR+ større end 5 eller en LR- mindre end 0,2. En optimal diagnostisk sikkerhed forudsætter at LR+ er større end 10 eller at LR- er mindre end 0,1.

Prædiktiv validitet

Sensitivitet, specificitet, Area Under the Curve og likelihood ratios udtrykker noget generelt om et instruments evne til diagnostisk klassifikation. I klinisk praksis er man også interesseret i at kende sandsynligheden for, at en person er dement, hvis vedkommende har et 'positivt' testresultat, det vil sige den prædiktive validitet.

Prædiktiv validitet er baseret på en kombination af instrumentets evne til diagnostisk klassifikation med prævalensraten for demens blandt de personer, som man undersøger. Prævalensraten, også kaldet base rate, udtrykker omtrent, hvor hyppigt en given tilstand (fx demens) forekommer blandt de personer, man undersøger. Base rate vil derfor variere alt efter om udredningen foregår i almen praksis (lav prævalens), på en hukommelsesklinik (høj prævalens) eller andetsteds.

Også patientens alder bør tages i betragtning ved estimering af base rate, da forekomsten af demens er markant højere blandt ældre end blandt yngre. Base rate er beslægtet med begrebet prætest sandsynlighed, men ved vurdering af prætest sandsynlighed tages ofte højde for endnu flere faktorer i form af andre prøveresultater, klinikerens grad af erfaring m.v.

Sandsynligheden for, at en person er syg, hvis vedkommende har et positivt testresultat, kaldes den positive prædiktive validitet, mens sandsynligheden for, at en person er rask, hvis vedkommende har et normalt testresultat, kaldes den negative prædiktive validitet.

Fordele ved prædiktiv validitet

En fordel ved prædiktiv validitet er, at det udgør et intuitivt forståeligt og dermed brugervenligt mål for sandsynlighed (svarende til en p-værdi eller procentangivelse). En ulempe er, at prædiktiv validitet altid er bundet til en given base rate og ikke kan generaliseres til situationer, hvor base rate er væsentligt højere eller lavere.

Til sammenligning er likelihood ratio et styrkemål, der er uafhængigt af base rate og således mere generaliserbart. Det er muligt, omend lidt omstændeligt, at foretage en omregning fra likelihood ratio til prædiktiv validitet via prætest odds og posttest odds.

Kilder

Juul S. Epidemiologi og evidens. Kapitel 11 'Diagnostik'. 2. udgave. København: Munksgaard; 2013.

[Munksgaard](#)

Bech P. Klinisk psykometri. København: Munksgaard; 2011.

[Munksgaard](#)

Senest opdateret: 06. august 2020