

ROC-kurver og diagnostiske tester

MEDISIN OG TALL

STIAN LYDERSEN

stian.lydersen@ntnu.no

Stian Lydersen er dr.ing. og professor i medisinsk statistikk ved Regionalt kunnskapssenter for barn og unge – psykisk helse og barnevern (RKBU Midt-Norge), Institutt for psykisk helse, Fakultet for medisin og helsevitenskap, Norges teknisk-naturvitenskapelige universitet (NTNU).

Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

Når en diagnostisk test kun skiller mellom negativt og positivt resultat, kvantifiseres testens egenskaper med dens sensitivitet og spesifisitet. For en diagnostisk test på en kontinuerlig skala, derimot, kan en ROC-kurve være en egnet fremstilling av testens egenskaper.

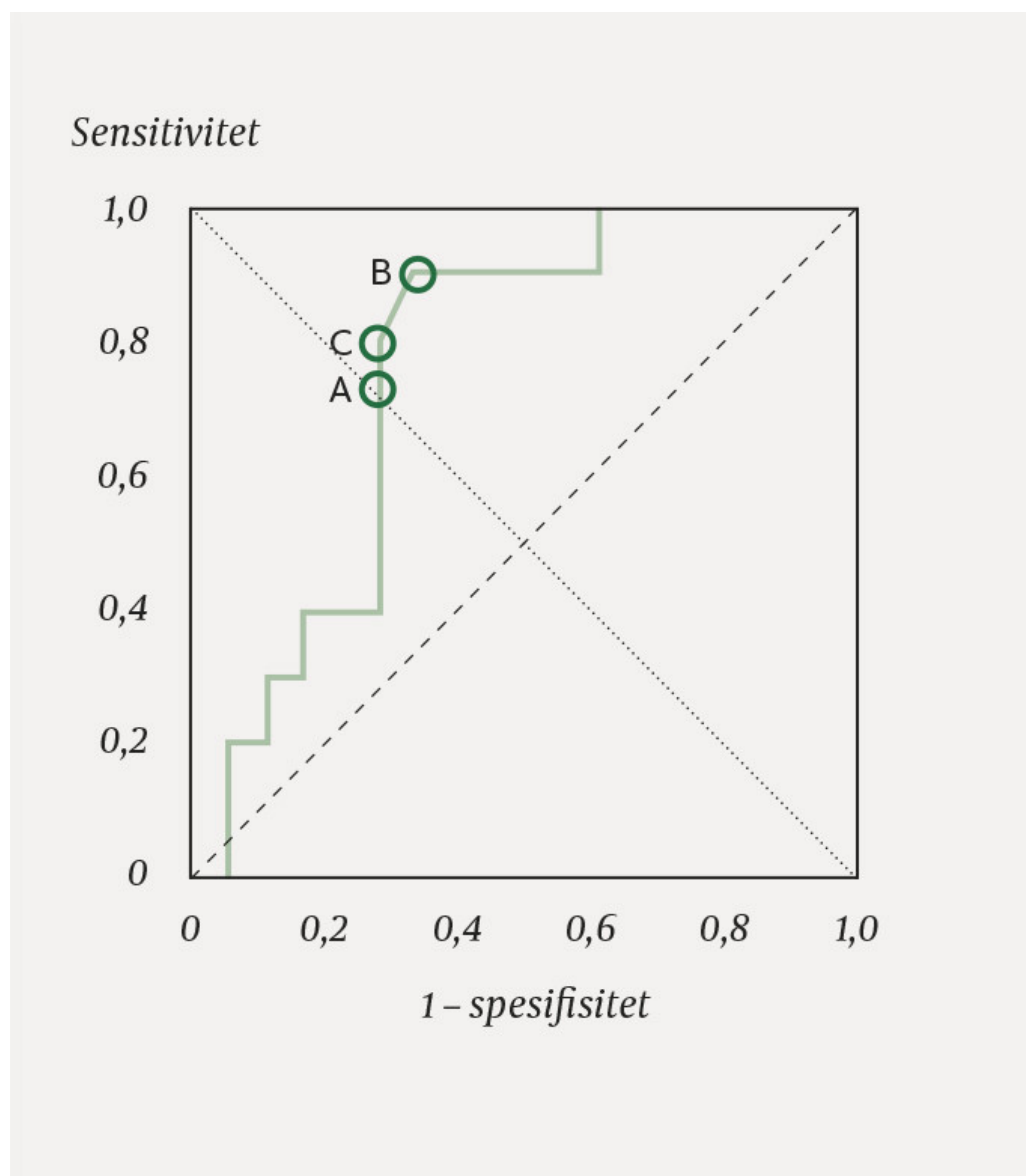
ROC-metodikken ble opprinnelig utviklet under andre verdenskrig for å måle radaroperatørenes evne til å skille mellom fiendtlige objekter og støy, herav navnet «receiver operating characteristic curve», som ikke har fått noen etablert norsk oversettelse.

La oss starte med et eksempel, basert på Lydersen (1). Konsentrasjon av prostataspesifikt antigen (PSA) i serum er en markør for prostatakreft. Tabell 1 viser et tenkt eksempel med resultater fra 28 pasienter, hvorav biopsi viste at 10 hadde prostatakreft, og 18 ikke hadde prostatakreft. Hvis grenseverdien for positiv test settes til en PSA-konsentrasjon på 6 ng/ml, vil estimert sensitivitet bli $9/10 = 0,90$, og estimert spesifisitet bli $11/18 = 0,61$. Hvis grenseverdien derimot settes til 8 mg/ml, får vi en sensitivitet på $5/10 = 0,50$ og en spesifisitet på $13/18 = 0,72$. Slik kan man beregne sensitivitet og spesifisitet for enhver mulig grenseverdi. Et plott av alle mulige par av sensitivitet og spesifisitet som vist i figur 1, kalles en ROC-kurve.

Tabell 1

Et tenkt eksempel på PSA-nivå (ng/ml) i serum etter biopsiresultat hos 28 pasienter (1).

Prostatakraft	Ikke prostatakraft
3,8	0,4
6,5	0,6
6,7	0,8
7,0	1,1
7,9	1,5
8,1	2,6
10,8	2,8
11,3	5,0
14,5	5,0
15,3	5,1
	5,3
	6,1
	6,3
	8,2
	9,9
	10,9
	12,6
	18,6



Figur 1 ROC-kurven for eksempelet i tabell 1 (heltrukken grønn linje), og ROC-kurven for en ikke-informativ test (stiptet linje fra (0,0) til (1,1)). Punktene A, B og C viser skillet mellom positiv og negativ test ved forskjellige kriterier. A er punktet der den prikkede linjen fra (1,0) til (0,1) krysser ROC-kurven.

Areal under kurven

Arealet under kurven (AUC) er et mål på testens evne til å diskriminere mellom syke og friske. I dette eksemplet fås $AUC = 0,76$. En perfekt diagnostisk test ville få en ROC-kurve som gikk fra (0,0) via (0,1) til (1,1), og ha $AUC = 1$. En test som ikke gir noen som helst informasjon, ville få en ROC-kurve på diagonalen fra (0,0) til (1,1) som vist i figur 1, og ha $AUC = 0,5$.

En tolkning av AUC er som følger: Hvis vi trekker én tilfeldig person blant de syke og én blant de friske, så er AUC lik sannsynligheten for at den syke har høyere verdi på markøren enn den friske. Andre fortolkninger av AUC er gjennomsnittlig sensitivitet over alle verdiene av spesifisitet, eller gjennomsnittlig spesifisitet over alle verdiene av sensitivitet.

Hvilke AUC-verdier mellom 0,5 og 1 kan sies å representere dårlig, god eller meget god diskriminering? Enkelte forfattere har foreslått tommelfingerregler for dette, for eksempel kategoriene dårlig (0,5–0,7), god (0,7–0,8), meget god (0,8–0,9) og utmerket (over 0,9) ((2), s. 177). Andre forfattere fraråder slike generelle klassifiseringer, fordi dette avhenger av den aktuelle sykdommen og praktisk klinisk anvendelse ((3), s. 29).

En svakhet ved AUC-målet er at den representerer gjennomsnittlig sensitivitet over alle verdier av spesifisitet mellom 0 og 1, også verdier av spesifisitet der det ville være uaktuelt å bruke testen i praksis. Dette er nærmere diskutert av Zhou og medarbeidere ((3), s. 35–36).

Valg av grenseverdi

Enkelte forskere bruker ROC-kurven til å finne en «optimal» grenseverdi for å skille mellom en positiv og negativ test. Et mulig valg kan være det punktet der den prikkede linjen fra (1,0) til (0,1) krysser ROC-kurven, punkt A i figur 1. Ved å velge dette punktet fås en test med sensitivitet lik spesifisitet. Et annet valg kan være det punktet hvor summen av sensitivitet og spesifisitet er maksimal. Dette svarer til maksimering av Youdens indeks, som er definert som sensitivitet pluss spesifisitet minus 1 ((4), s. 12). Dette vil gi punkt B i figur 1. Et tredje valg er det punktet som ligger nærmest (0,1), punkt C i figur 1. Men disse kriteriene er i liten grad anbefalt i litteraturen. De har den svakhet at de vektlegger sensitivitet og spesifisitet tilnærmet likt og tar ikke hensyn til den aktuelle sykdommen og praktisk klinisk anvendelse. Et mer relevant kriterium kan for eksempel være å velge det punktet som gir høyest mulig sensitivitet, samtidig som man krever en minimumsverdi av spesifisiteten, for eksempel 0,80 ((3), s. 34–35). I eksemplet i figur 1 ville man da få sensitivitet 0,40 og spesifisitet 0,80.

En ROC-kurve er en nyttig fremstilling av egenskapene til en diagnostisk test på en kontinuerlig skala. Det er vanligvis hensiktsmessig å vise hele ROC-kurven, og ikke bare rapportere areal under kurven eller verdier som svarer til ett punkt på kurven.

LITTERATUR

1. Lydersen S. Diagnostic tests, ROC curves, and measures of agreement. I: Veierød MB, Lydersen S, Laake P, red. Medical statistics in clinical and methodological research. Oslo: Gyldendal akademisk, 2012: 462–92.
2. Hosmer DW, Lemeshow S, Sturdivant RX. Applied logistic regression. 3. utg. Hoboken, NJ: Wiley, 2013.
3. Zhou XH, Obuchowski NA, McClish DK. Statistical methods in diagnostic medicine. 2. utg. Hoboken, NJ: Wiley, 2011.
4. Zou KH, Liu A, Bandos AI et al. Statistical evaluation of diagnostic performance. Topics in ROC analysis. Boca Raton: Taylor & Francis, 2012.

