

Evaluering av nye laboratorieanalyser

Sammenheng

Bakgrunn. Helsepersonell må stadig ta stilling til nye laboratorieanalyser som kommer på markedet.

Metode. Vi gir en gjennomgang av hvordan nye analyser bør evalueres før de tas i bruk.

Resultat. Det biokjemiske grunnlaget for analysen og analyttens intra- og interindividuelle biologiske variasjon må være vurdert, og man bør vite hvordan måleresultatet avviker ved sykelige tilstander. Optimalt forhold for prøvetaking og -behandling må være fastlagt. Presisjon og riktighet må være god nok til at måleresultatet kan brukes til kliniske beslutninger. Metoden bør være sporbar til en referansemetode. Interferenser med for eksempel bilirubin, lipider, hemolyse eller medikamenter må være klarlagt. Til bruk i diagnostikk må testens diskriminerende evne være evaluert prospektivt i en relevant populasjon mot en gullstandard. For å erstatte etablerte tester bør den nye testen ha bedre diskriminerende evne enn disse. For å supplere andre tester må den nye gi uavhengig tilleggsinformasjon. Det må avklares i hvilke kliniske situasjoner testen kan være til hjelp, og hvilke beslutningsgrenser behandlende lege skal følge. Nytteverdien må stå i forhold til praktisk og økonomisk innsats.

Fortolkning. Slik evaluering bør gjøres eller etterspørres av alle som vil ta i bruk nye analyser, enten det er på legekantoret eller i et større laboratorium.

Engelsk sammendrag finnes i artikkelen på www.tidsskriftet.no

> Se også side 301

Bjørn J. Bolann

bjorn.bolann@ikb.uib.no

Sverre Sandberg

Laboratorium for klinisk biokjemi
Haukeland Universitetssykehus
5021 Bergen

Laboratorieundersøkelser brukes av flere grunner. Noen brukes diagnostisk (for eksempel test for mononukleose), andre til å oppdage endringer i sykdomsutviklingen (for eksempel Hb A1c), atter andre for å estimere risiko for fremtidig sykdom (for eksempel måling av kolesterolnivå). Laboratorieanalyser kan gi svar som er kvantitative (et måleresultat langs en kontinuerlig skala), semikvantitative (et begrenset antall svarmuligheter langs en ikke-kontinuerlig skala) eller binære (har bare to mulige svar, positiv og negativ).

Nye undersøkelsesmetoder kommer stadig på markedet. Med denne artikkelen vil vi hjelpe helsepersonell til å evaluere nye laboratorieundersøkelser som tilbys. Vi begrenser oss til omtale av klinisk-kjemiske og hematologiske analyser, dvs. målinger av komponenter i kroppsvæsker, vanligvis blod eller urin.

Vurdering av nye laboratorieanalyser

Biologisk grunnlag

For at vi skal fatte interesse for en ny laboratorieanalyse, må vi ha grunn til å tro at konsentrasjonen eller mengden av det som skal måles – analytten – endrer seg ved bestemte sykdomstilstander på en slik måte at måleresultatet kan gi viktig informasjon om tilstanden. Man bør ha en plausibel biokjemisk forklaring på hvorfor endringene oppstår, eller i det minste empirisk kjennskap til når og hvordan endringene skjer. Man bør også ha kjennskap til biologisk variasjon, dvs. naturlig variasjon av konsentrasjonen hos den enkelte person (intraindividuell variasjon) og variasjon mellom personer (interindividuell variasjon). Dersom analytten har stor biologisk variasjon, kan dette maskere endringer eller avvik i prøvesvaret hos den enkelte pasient og gjøre testen ubrukelig, uansett hvor god selve analysemetoden er.

Å måle riktig

Så må man forsikre seg om at metoden faktisk måler det den skal, at den måler riktig,

Fakta

Før innføring av en ny laboratorieanalyse i klinisk rutine bør man kjenne til:

- Det biologiske grunnlaget for analysen, hvilke verdier som kan ventes hos friske og syke
- Hvor nøyaktig og robust metoden er: Presisjon, riktighet, interferenser, krav til prøvetaking
- Testens diskriminerende evne i relevant populasjon, sammenliknet med etablerte tester
- I hvilke kliniske situasjoner testen kan brukes og hvilke konsekvenser resultatet skal ha

og at målingen er reproducerbar. Vi må altså stille krav om at komponenten blir målt med god nok analytisk kvalitet. I utgangspunktet kan det være vanskelig å vite hvor strenge kvalitetskrav man bør stille. For eksempel kommer det an på hvordan analyseresultatene i «frisk» og «syk» populasjon fordeler seg. Som regel vil «friske» og «syke» ha delvis overlappende verdier, og overlappingen blir større jo dårligere presisjon, reproducerbarhet, metoden har.

Er undersøkelsen nyttig?

En laboratorieanalyse er bare nyttig hvis resultatet kan få konsekvenser for pasienten i form av bedret eller forenklet diagnostikk eller behandling. Spørsmålet vil derfor alltid være hvordan analysen fungerer i praksis i den populasjonen pasientene kommer fra.

Første fase i evalueringen av en ny undersøkelse er ofte å se hvordan måleresultatene kan brukes til å skille personer med et typisk sykdomsbilde fra personer som er helt friske (1). Vanskeligere er det å bruke testen til å stille en diagnose hos personer som har uklare eller atypiske symptomer og som kan ha andre tilstander samtidig. Det er derfor viktig at testen har vært utprøvd i en populasjon som tilsvarende hvor den er tenkt brukt. Testresultater fra akuttmodtak i et sykehus kan sjelden danne grunnlag for diagnostikk i allmenpraksis, eller omvendt.

Men selv om testen brukt alene gir ade-

kvat informasjon, er det ikke sikkert at den gir tilleggsinformasjon utover det vi får av andre, innarbeidede undersøkelser. For å være av verdi må testen enten gi uavhengig tilleggsinformasjon eller den må kunne erstatte andre undersøkelser. Hver gang nye tester innføres, er det derfor viktig å vurdere om det er andre tester som bør fjernes. Det kan se ut til at det er mye vanskeligere å fjerne tester fra et analyserepertoar enn det er å innføre nye. Det er også viktig at nytten av den nye informasjonen står i forhold til innsatsen, organisatorisk og økonomisk, ved å utføre analysen.

Krav til dokumentasjon

Hvis forutsetningene nevnt foran ikke synes å være til stede, er det sannsynligvis ingen grunn til å vurdere den nye testen nærmere. Men dersom man ut fra det ovenstående finner at testen kan ha et potensial hos aktuelle pasienter, bør man gå nærmere inn på følgende:

Preanalytiske kvalitetskrav

Det må være avklart hvilket prøvemateriale som skal brukes og hvordan prøven skal tas, om pasienten må være fastende, hvor god holdbarhet prøven har eller om det er andre forholdsregler. De fleste analyser har en viss preanalytisk variasjon som skyldes prøvetaking og -behandling, oppbevaring, transport mv.

Analytiske kvalitetskrav

Krav til analysekvalitet henger sammen med hvor store endringer som kan ventes hos «friske» og «syke». Både biologisk variasjon, kliniske beslutningsgrenser og tidligere laboratoriepraksis kan danne grunnlag for kvalitetskrav (2, 3).

Krav til kvalitet kan også avhenge av hvordan man skal bruke den aktuelle analysen. Man stiller gjerne strengere kvalitetskrav dersom analysen skal brukes til å diagnostisere en sykdom enn hvis den skal brukes til kontroll av en kjent tilstand. Et eksempel på dette er måling av glukose i blod, der instrumenter som brukes i primærhelsetjenesten sjelden er gode nok til å stille diagnosen diabetes mellitus ved verdier omkring de diagnostiske grensene, men kan være gode nok til å monitorere diabetikere.

Nedenfor blir de viktigste parametere for krav til og dokumentasjon av analysekvalitet kort gjennomgått. Kravene som nevnes her, gjelder prinsipielt for kvantitative analyser. Kravene kan også i noen grad appliseres på semikvantitative analyser, avhengig av hvor grov metoden er, for øvrig henviser vi her til annen litteratur (4).

Linearitet. Det fundamentale krav til en målemetode er at det er en konsekvent sammenheng mellom størrelsen av det som skal måles (for eksempel en konsentrasjon) og måleresultatet. Vanligvis er det et lineært forhold. En stabil dose-respons-kurve er en forutsetning for enhver kvantitativ analyse.

Riktighet. Dernest må metoden måle riktig. Med riktighet menes samsvar mellom gjennomsnittet av gjentatte målinger av en komponent i en prøve og den «sanne» verdi. Hvis analysen har dårlig riktighet, betyr det at den har et stort systematisk avvik (bias), dvs. en feil som alltid trekker svarene til den ene siden.

Analysens riktighet bør være dokumentert ved at metoden er sammenliknet med en referansemetode («sporbarhet») (5) eller med en anerkjent metode med kjent sporbarhet. Hvor stort avvik som kan aksepteres, avhenger av bruken av analysen. Krav til riktighet bør også sees i forhold til analysens referanseområde (2).

Presisjon. Med presisjon menes samsvar mellom gjentatte målinger. Ideelt sett skulle man få samme resultat hver gang, men ingen analysemetoder er helt presise, og resultatene vil som regel variere litt. Størrelsen på denne tilfeldige variasjonen angis ofte som analytisk standarddeviasjon eller variasjonskoeffisient.

Den totale variasjonen i prøveresultatene består av preanalytisk, analytisk og biologisk variasjon. For at den analytiske variasjonen i minst mulig grad skal forstyrre tolking av resultatene, må den utgjøre en så liten del av den totale variasjonen som mulig. Et enkelt, vanlig mål for analytisk presisjon er at analytisk variasjon ikke skal være mer enn halvparten av intraindividuell biologisk variasjon. Det innebærer at av den tilfeldige variasjonen i analysesvaret hos den enkelte pasient, vil analytisk variasjon ikke utgjøre mer enn 12 %, resten vil utgjøres av biologisk variasjon (2). Tabeller over biologisk variasjon finnes i litteraturen (6–9).

De fleste kvantitative analyser har analytisk variasjonskoeffisient på 2–5 %.

Måleområde. Alle målemetoder har sitt måleområde. Utenfor dette kan man enten ikke måle noe eller målingen blir for unøyaktig til å være brukbar. Med en gitt metode vil en svært lav konsentrasjon av et stoff i blodet, for eksempel et medikament, ikke være målbar. Heller ikke svært høye konsentrasjoner kan måles uten videre, men her har man ofte mulighet til komme innenfor måleområdet ved å fortynne prøven.

Biokjemiske målinger har altså sitt måleområde innenfor visse konsentrasjonsgrenser (øvre og nedre kvantifiseringsgrense) (10). For at en målemetode skal være klinisk brukbar, må de konsentrasjoner som er interessante hos pasienter, falle innenfor metodens måleområde. Kvalitetskravene som er nevnt over, bør gjelde i hele det klinisk relevante måleområdet. Kvalitetskravene er spesielt viktige ved de konsentrasjonene som brukes som kliniske beslutningsgrenser.

Analytisk spesifisitet, interferens. En analysemetode er som regel ikke helt spesifikk for det stoff den skal måle. Beslektede substanser eller metabolitter av analytten kan bli medbestemt. Dessuten kan andre komponenter, for eksempel bilirubin, lipider, he-

molyse eller medikamenter i serum, ofte interferere med målingen og gi galt resultat. Man må derfor etterspørre dokumentasjon av slike effekter.

Beslutningsgrunnlag

Verdier hos friske og syke. For å kunne skille «friske» fra «syke» må man vite hvilke resultater som kan ventes i begge gruppene. Referanseområdet utgjøres vanligvis av det sentrale 95 %-intervall av verdiene hos friske personer (11). Nye analyser blir gjerne presentert sammen med beregnet referanseområde. Verdier hos pasienter med aktuelle sykdomstilstander er ofte kartlagt i mindre grad, og i hvor stor grad disse verdiene overlapper normalverdiene, er ofte lite undersøkt. Derfor bruker mange referansegrensene som aksjonsgrenser, det vil si at det aksjoneres ved prøveresultater utenfor disse grensene. Man må imidlertid være klar over at dersom et resultat bare sammenliknes med en referansepopulasjon, kan man kun uttale seg om hvor stor sannsynlighet det er for at resultatet tilhører denne populasjonen, ikke om hvor sannsynlig det er at det tilhører en «syk» populasjon. I enkelte tilfeller vil aksjonsgrensene ligge langt inne i referanseområdet, for eksempel for kolesterol, der man kan aksjonere på 5–6 mmol/l, mens øvre referansegrense kan være opptil 8,7 mmol/l (6).

Dersom den nye testen skal brukes til å diagnostisere en bestemt tilstand, må metoden være evaluert i et utvalg av en populasjon som har samme sykdomspanorama som den populasjonen hvor testen skal brukes. Det må med en annen, uavhengig metode (gullstandard) være påvist hvilke av personene i utvalget som hadde tilstanden. Best gjøres dette ved en prospektiv undersøkelse i den aktuelle populasjonen (1, 12, 13). Fra dette kan man beregne testens diskriminative egenskaper (prediktiv verdi eller treffsikkerhet), enten det er en binær, semikvantitativ eller kvantitativ test. Det er ikke nok at gjennomsnittet i «frisk» og «syk» populasjon er forskjellig dersom analyseresultatet skal ha betydning for hver enkelt person som får et resultat. Det samme gjelder dersom en etablert metode brukes på en ny måte.

Binære tester: sensitivitet og spesifisitet. Binære tester har bare to svar: positiv eller negativ. Testens diskriminerende evne er definert ut fra hvor mange av de «syke» som har positiv test (sensitivitet) og hvor mange av de «friske» som har negativ test (spesifisitet) (14). Ved hjelp av disse verdiene kan testen sammenliknes med andre tester. Det er imidlertid en utbredt misforståelse at sensitivitet og spesifisitet er uavhengig av sykdomsprevalens. Fordi sykdomsprevalens har sammenheng både med sykdomsgrad og med hvilke andre sykdommer som forekommer i populasjonen, er forutsetningen for valide estimater av sensitivitet og spesifisitet at evalueringen er foretatt i en relevant populasjon.

Kvantitative og semikvantitative analyser: ROC-analyse. Kvantitative analyser kan gjøres binære ved at man setter en bestemt grenseverdi og anser resultater over grenseverdien som «positive» og resultater under den som «negative» (eller omvendt). Analysen får da en sensitivitet og en spesifisitet. Men man mister klinisk informasjon. Man kan for eksempel ikke lenger se forskjell på et lett patologisk og et sterkt patologisk prøveresultat eller se utvikling over tid.

Ved å sette slike grenseverdier på en kvantitativ analyse kan man likevel få viktig informasjon om testens diskriminerende evne. Ved å sette grenseverdien forskjellige steder får man forskjellige verdier for sensitivitet og spesifisitet: Høyere sensitivitet gir lavere spesifisitet, og vice versa. Et sett av kombinasjoner av sensitivitet og spesifisitet kan fremstilles grafisk, og danner en såkalt ROC (receiver operating characteristic)-kurve. Arealet under ROC-kurven er et uttrykk for testens diskriminerende evne, og ved å sammenlikne dette arealet med det tilsvarende arealet for andre tester, kan man finne hvilken av testene som gir mest informasjon (14). Slik ROC-analyse er en vanlig fremgangsmåte for å evaluere både kvantitative og semikvantitative tester.

Logistisk regresjonsanalyse. ROC-analyse gir bare informasjon om én test om gangen. Selv om testen alene gir adekvat informasjon, kan testresultatet være korrelert med andre testresultater, og da er det ikke sikkert at den gir tilleggsinformasjon i forhold til disse. For å være av verdi må testen enten gi uavhengig tilleggsinformasjon eller den må erstatte andre undersøkelser.

Hvorvidt testen gir uavhengig tilleggsinformasjon, kan undersøkes ved hjelp av logistisk regresjonsanalyse (14, 15). Også her må alle aktuelle tester være evaluert i en relevant populasjon hvor man med en gullstandard har bestemt hvem som har den tilstanden man skal diagnostisere. Man kan så finne den kombinasjonen av tester som gir mest informasjon, og hvilke tester som gir signifikant tilleggsinformasjon. Logistisk regresjonsanalyse kan brukes til å evaluere både kvantitative, semikvantitative og binære tester. Dessverre er altfor få tester undersøkt på denne måten.

Praktiske forhold

Til slutt må testen også være praktisk anvendelig, og helst mer praktisk i bruk enn etablerte tester. Er det spesielle krav til prøvetaking? Laboratoriet eller legekantoret som skal utføre testen, må vurdere hvilken kompetanse som kreves. Det kan også være lurt å se på hvor robust testen er mot ikke-optimal utførelse. Og krever metoden stor plass? Man bør beregne hvor mange prøver som kan analyseres per time, per dag osv., og hvor mye penger, tid og arbeidsinnsats som går med.

De som ikke utfører analysen selv, men sender prøven til et annet laboratorium, må

Ramme

Sjekkliste før innføring av en ny laboratorieanalyse

- Biologisk grunnlag: Er det kjent hvordan, hvorfor og ved hvilke tilstander analytten endres i organismen?
- Er optimale forhold for prøvetaking og -behandling fastlagt?
- Er det påvist at analytten blir målt med tilstrekkelig riktighet, presisjon og uten interferenser av betydning, hos aktuelle pasienter, i hele det klinisk relevante måleområdet?
- Er referanseområdet fastlagt? Er verdier hos pasienter med aktuelle sykdomstilstand(er) kartlagt?
- Er testen evaluert prospektivt i en relevant populasjon mot en annen, uavhengig undersøkelsesmetode (gullstandard)?
- Har testen bedre diskriminerende evne enn andre tester (ROC-analyse)?
- Gir testen uavhengig tilleggsinformasjon i forhold til andre undersøkelser (logistisk regresjonsanalyse)?
- Kan testen erstatte andre undersøkelser?
- Er det klarlagt i hvilke kliniske situasjoner testen bør brukes? Er aksjonsgrenser fastlagt?
- Er testen praktisk i bruk? Står nytteverdien i forhold til praktisk og økonomisk innsats?

kjenne til om det er spesielle krav ved forsendelsen (for eksempel kort holdbarhet, krav til nedfrysing) og hvor lenge man må vente på svar.

Dersom testen etter ovenstående kan godkjennes til klinisk bruk, gjenstår det å lage retningslinjer for praktisk bruk i ulike kliniske situasjoner og for hvordan resultatene skal tolkes. Laboratorier som utfører testen for andre, bør sende ut informasjon til disse.

Hvordan finne dokumentasjon?

For metoder som tilbys kommersielt, er det leverandørens ansvar å frembringe dokumentasjon – ikke bare for analytisk kvalitet, men også for klinisk brukbarhet dersom dette ikke foreligger fra før. Det er viktig å spørre om metoden har vært evaluert blant dem som skal utføre den, for eksempel i sykehus, blant medarbeidere på legekantor eller av pasienter som foretar egenmåling. En del slik informasjon kan finnes på hjemmesidene til NOKLUS (www.noklus.no). Det kan også være nyttig å foreta litteratursøk, for eksempel i PubMed (www.ncbi.nlm.nih.gov). Spesielt bør man lete etter systematiske oversiktsartikler, selv om disse nok først finnes etter at testen har vært i bruk en tid. En oversikt over slike finnes på www.ckchl-mb.nl/ifcc.

Sjekkpunkter for nye laboratorieanalyser

Disse er oppsummert i rammen. For rasjonell bruk av laboratorieanalyser bør alle punktene være evaluert. Slik evaluering legges regelmessig til grunn for innføring av nye analyser ved klinisk-kjemiske laboratorier. Men også andre som vil ta i bruk analyser, f.eks. på legekantor, bør gå gjennom listen, dette gjelder særlig for nye tester som tilbys kommersielt. For prøver som sendes til andre laboratorier, kan man også spørre etter dokumentasjon.

Litteratur

Komplett litteraturliste finnes i artikkelen på www.tidsskriftet.no

1. Nierenberg AA, Feinstein AR. How to evaluate a diagnostic marker test. Lessons from the rise and fall of dexamethasone suppression test. *JAMA* 1988; 259: 1699–702.
2. Bolann BJ, Stølsnes B. Analytisk usikkerhet – hvor stor feil kan laboratoriesvaret ha? *Tidsskr Nor Lægeforen* 1999; 119: 4472–5.
3. Kenny D, Fraser CG, Petersen PH, Kallner A. Consensus agreement. *Scand J Clin Lab Invest* 1999; 59: 585.
4. Petersen PH, Sandberg S, Fraser CG, Goldschmidt H. A model for setting analytical quality specifications and design of control for measurements on the ordinal scale. *Clin Chem Lab Med* 2000; 38: 545–51.
6. Stakkestad JA, Åsberg A. Brukerhåndbok i klinisk kjemi. Haugesund: Akademisk fagforlag, 1996.
9. Ricos C, Alvarez V, Cava F, Garcia-Lario JV, Hernandez A, Jimenez CV et al. Current databases on biological variation: pros, cons and progress. *Scand J Clin Lab Invest* 1999; 59: 491–500.
10. Lawson GM. Defining limit of detection and limit of quantitation as applied to drug of abuse testing: striving for a consensus. *Clin Chem* 1994; 40: 1218–9.
11. Solberg HE. Approved recommendation (1987) on the theory of reference values. Part 5. Statistical treatment of collected reference values. Determination of reference limits. *Clin Chim Acta* 1987; 170: S13–32.
13. Knottnerus JA, van Weel C, Muris JW. Evaluation of diagnostic procedures. *BMJ* 2002; 324: 477–80.
14. Boyd JC. Mathematical tools for demonstrating the clinical usefulness of biochemical markers. *Scand J Clin Lab Invest* 1997; 57 (suppl 227): 46–63.