



VALIDERING / VERIFISERING

AV KLINISK KJEMISKE ANALYSER

Klinisk nytte – praktisk egnethet – analytisk kvalitet

NKK, Februar 2002

INNHALDSFORTEGNELSE

1. INNLEDNING	3
2. OMFANG	4
2.1. Klinisk nytteverdi	4
2.1.1. Diagnostisk nøyaktighet	5
2.1.2. Klinisk kvalitet	6
2.1.3. Referansegrenser og andre beslutningskriterier	6
2.2. Praktisk egnethet	7
2.2.1. Analyseinstrument og utstyr	7
2.2.1. Reagenser, kalibratorer og kontrollmaterialer	8
2.2.2. Kostnader	8
2.2.3. Annet	8
2.3. Analytisk kvalitet	8
2.3.1. Riktighet	9
Sporbarhet	9
Linearitet	11
Interferens	12
Analytisk spesifisitet	12
Holdbarhet	12
2.3.2. Presisjon	12
Repeterbarhet (innen-serie impresisjon)	13
Reproduserbarhet	13
Robusthet	14
Carry over/Overdraging	14
2.3.3. Måleusikkerhet	15
2.3.4. Måleområde	16
Øvre målegrense:	16
Nedre målegrense:	16
3. DOKUMENTASJON OG KONKLUSJON	17
4. REFERANSER	18

1. INNLEDNING

Verdien av å validere analyser er etter hvert blitt en mer bevisst del av kvalitetssikringsarbeidet i medisinske laboratorier. Det finnes mye litteratur som beskriver de forskjellige elementene i prosessen, likevel var det et ønske at NKK skulle utarbeide mal for ”metodevalidering” som kunne tjene som en veiledning for klinisk kjemiske laboratorier.

Validering er definert slik: *Bekreftelse fra en undersøkelse og fremskaffing av objektivt bevis på at spesielle krav for tilsiktet bruk er innfridd.* [1, 2,3]

Hensikten med validering er altså å teste om målemetoden egner seg til formålet. Det er derfor viktig at kravene er satt på forhånd, slik at man objektivt kan bedømme om de er oppnådd.

Verifisering innebærer en noe enklere undersøkelse enn validering, og går ut på å dokumentere at metoden fungerer som forventet i eget laboratorium.

Et tredje begrep, **evaluering**, omfatter det samme som validering, men uten at man har satt krav på forhånd. Evaluering gjøres gjerne i forbindelse med publisering av en metode eller som ledd i utvikling av en metode hos en produsent.

I dag er det heldigvis slik at de fleste diagnostika-leverandører har dokumentasjon på analytisk yteevne karakteristika, for eksempel repeterbarhet, reproduserbarhet, deteksjonsgrense, interferenser etc. Slik dokumentasjon vil bli krevd i forbindelse med implementering av IVD-direktivet. (IVD-direktivet er tilgjengelig på Internett: <http://www.helsetilsynet.no/trykksak/meld/directive9897.pdf>)

Når utføres så validering og når utføres verifisering?

Begrepene glir noe over i hverandre, men kan eksempelvis forståes slik:

Validering er aktuelt ved:

- Innføring av en ny målemetode i laboratoriet der valideringsdokumentasjon ikke foreligger eller er mangelfull. For eksempel vil dette være tilfelle når det er forskjellige produsenter av måleutstyr, reagenser og/eller kalibratorer, slik at det ikke finnes en samlet validering.
- Endring av analysemetode, analyseutstyr eller andre større endringer,
 - når det gjøres avvik fra produsentens prosedyre som kan ha konsekvenser for analysekvaliteten, som for eksempel når metoden skal brukes til analyse i et materiale den ikke er validert for, eller når metoden skal brukes til målinger utenfor definert måleområde.
 - når måleresultatet skal anvendes annerledes i klinikken enn forutsatt.
- Eventuelt kan metoden være delvis validert, og det kan være aktuelt å supplere med det som anses nødvendig for å kunne dokumentere tilfredsstillende kvalitet, og så verifisere på de områdene der man finner det nødvendig.

Verifisering er aktuelt ved:

- Implementering av en ny målemetode eller et nytt målesystem der valideringsdokumentasjon foreligger, - enten fra leverandøren eller fra annet laboratorium. Svært ofte er mange av egenskapene til en metode og dens kliniske nytte tilfredsstillende validert av andre (gjennom vitenskapelige publikasjoner, hos produsenten eller av andre laboratorier), og laboratoriet kan nøye seg med å fremskaffe dokumentasjon på valideringen. Da er det tilstrekkelig å verifisere at metoden fungerer med samme kvalitet i eget laboratorium.
- Mindre metodeendringer som kan tenkes å føre til endret analysekvalitet, som
 - lot-skifte av reagens og/eller målestandard (kalibrator) for metoder der slike skift erfaringsmessig kan gi endret nivå.

Uansett, om det kan innhentes valideringsdokumentasjon fra andre instanser, er det likevel laboratoriets ansvar å forvise seg om at måleutstyret fungerer med tilfredsstillende kvalitet lokalt.

I malen har vi forsøkt å bruke ett sett begreper, men det finnes en rekke synonymer, som for eksempel:

Analysemetode – Målemetode

Analyseinstrument – Måleinstrument

Analyseresultat - Måleresultat

Kalibrator – Målestandard

Systematisk feil – Bias

Tillagt verdi – Oppgitt verdi – Estimat

Målestørrelse – En størrelse som skal måles (konsentrasjon, aktivitet)

2. OMFANG

Hvor omfattende en validering bør være avhenger av flere faktorer; - hvilken type analyse det dreier seg om, hvordan den skal benyttes i klinikken og sist, men ikke minst hvilken dokumentasjon som foreligger. I utgangspunktet bør man velge de valideringsparametere som er relevante for metodens anvendelse. Det kan derfor ikke defineres et fast oppsett for alle metoder, men det anbefales å utarbeide en valideringsplan som er tilpasset den aktuelle metoden.

En fullstendig validering bør omfatte:

- Klinisk nytteverdi
- Praktisk egnethet
- Analytisk kvalitet

Dessuten kan en metode ikke sies å være validert før det foreligger en dokumentasjon og konklusjon.

Klinisk nytteverdi: henger sammen med hvordan testen er tenkt benyttet i klinikken, og vurderingen bør foretas av en person med medisinsk kompetanse.

Praktisk egnethet: Kanskje det området der det enkelte laboratoriet alltid gjør en selvstendig validering. Det er naturlig at laboratoriet setter egne krav for praktisk egnethet, for deretter å vurdere forskjellig analyseutstyr evt. metoder basert på disse kravene.

Analytisk kvalitet: burde også, ideelt sett, valideres *forut* for innkjøp, men i praksis nøyer de fleste seg med å innhente og vurdere valideringsdata og eventuell informasjon fra produsenten og/eller andre laboratorier før kjøpekontrakten underskrives. Når så analysesystemet er i hus, verifiserer laboratoriet at analysekvaliteten er som forventet.

Validering/verifiseringsprosessen gir også laboratoriet mulighet til å ”bli kjent med” analyseutstyret så vel som den enkelte analysemetoden, - og i samarbeid med leverandøren forbedre egenskaper som viser seg ikke å fungere i forhold til krav eller forutgående validering.

2.1. Klinisk nytteverdi

Med dette menes analysens evne til å hjelpe legen til riktig handling.

Laboratoriene foretar sjelden en klinisk validering av en eksisterende analyse, men undertiden kan det være aktuelt å vurdere om analysen fremdeles har sin berettigelse i laboratoriet.

Hvis det dreier seg om en innføring av en ny analyse, bør laboratoriet gjøre en selvstendig vurdering av tiltenkt klinisk nytteverdi

For vurdering av klinisk nytteverdi er det naturlig å innhente dokumentasjon fra relevante vitenskapelige publikasjoner, faglige foredrag, etc. Det henvises til veiledningen som er utarbeidet av Kvalitetsutvalget til Norsk Forbund for klinisk kjemi, som er tilgjengelig på forbundets nettside [4], samt boken: ”Clinical Investigation and Statistics in Laboratory Medicine”. [5]. Det henvises også til ”handouts” av Arne Åsbergs presentasjon av emnet på NKK-møtet i 2000.

Aktuelle parametere for vurdering av klinisk nytteverdi er:

2.1.1. Diagnostisk nøyaktighet

Diagnostisk nøyaktighet er analysens evne til å diskriminere mellom to eller flere helsetilstander. [4]

Følgende huskeliste kan være til hjelp når vi skal vurdere litteratur om diagnostisk nøyaktighet (og hvis vi en sjelden gang skal utføre en slik undersøkelse selv):

- *Dette er problemstillingen*
 - En diagnostisk analyse skal diskriminere mellom to eller flere helsetilstander
 - Diagnostisk nøyaktighet er analysens evne til å diskriminere mellom to eller flere helsetilstander
 - Vurdering av diagnostisk nøyaktighet forutsetter at vi har en diagnostisk fasit vi kan sammenlikne med.
 - o analysering
 - o diagnostisering (fremskaffing av fasit, anvendelse av ”gullstandard”)
- *Er den kliniske populasjonen godt beskrevet?*
 - Alder og kjønn, eventuelt rase
 - Klinisk situasjon
 - Kliniske data
 - Prevalens av positiv diagnoseValidering gjelder bare for denne og helt liknende populasjoner
- *Er utvalgene gyldige?*
 - Naturalistisk: Alle diagnostiseres, alle analyseres
 - Retrospektivt: Alle diagnostiseres, et tilfeldig utvalg med positiv og negativ diagnose analyseres.
 - Prospektivt: Alle analyseres, et tilfeldig utvalg med positivt og negativt analyseresultat diagnostiseres
 - Pseudoretrospektivt: Alle diagnostiseres, alle analyseres, men populasjonen er kunstig sammensatt
- *Hvor stort utvalg trengs?*
 - Antall individer med positiv diagnose, negativ diagnose, positivt analyseresultat og negativt analyseresultat må være minst 10. Det betyr at 20 individer er et absolutt minimum, men at 20 individer er tilstrekkelig bare hvis 50% av individene har positivt analyseresultat og 50% har positiv diagnose. Hvis for eksempel bare 1 % har positiv diagnose, vil vi trenge minst 1000 individer ved bruk av naturalistisk utvalg.
- *Frafall*
 - Et hvert ikke-tilfeldig frafall kan være ødeleggende – det gjelder både analysering og diagnostisering
- *Analysen*
 - Pasientforberedelser
 - Prøvetaking
 - Utførelse av analysen. Rutineprøve eller forskningsmetode
 - Vurdering av resultatet – kriterier for positivt og negativt analyseresultat
 - Har analyseringen skjedd uten kjennskap til diagnose?
- *Diagnostisk (”gullstandard”)*
 - Hvorledes er diagnosen stilt?
 - Har diagnostikk skjedd uten kjennskap til analyseresultatet?

- *Diagnostisk nøyaktighet for kvalitative analyser*
 - Sensitivitet: Sannsynlighet for positivt analyseresultat gitt positiv diagnose
 - Spesifisitet: Sannsynlighet for negativt analyseresultat gitt negativ diagnose
- *Diagnostisk nøyaktighet for kvantitative analyser*
 - ROC-kurveanalyse
ROC-kurven er et plott av de verdiparene for sensitivitet og spesifisitet som framkommer når vi endrer kriteriet for positivt analyseresultat. Som regel plottes sensitivitet på Y-aksen mot 1-spesifisitet på X-aksen.
Arealet under ROC-kurven er uttrykk for analysens diagnostiske nøyaktighet (det kan tolkes som sannsynligheten for at analysen identifiserer riktig person som syk hvis den blir anvendt på en lang rekke av *par* av syke og friske personer). Fordelen med ROC-kurveanalyse framfor å oppgi bare ett verdipar av sensitivitet og spesifisitet, er at den diagnostiske nøyaktighet for to eller flere analyser kan sammenlignes uavhengig av kriterier for positivt analyseresultat. [6]
- *Diagnostisk nøyaktighet for kombinasjoner av analyser*
 - Logistisk regresjon bør brukes for å finne den beste kombinasjonen av analyser, for å se om en analyse bidrar med selvstendig informasjon til diagnosen. Fordelen med logistisk regresjon framfor klassisk diskriminantanalyse er at metoden tillater kombinasjoner av både kvalitative, semikvantitative og kvantitative analyseresultater. [7]

2.1.2. Klinisk kvalitet

En rimelig god diagnostisk nøyaktighet er en nødvendig, men ikke tilstrekkelig, betingelse for god klinisk kvalitet. Følgende momenter bør vurderes:

- *Selvstendig informasjon*
 - Bidrar analyseresultatet med selvstendig informasjon til diagnosen, eller finnes informasjonen i andre analyseresultater som også blir rekvirert ved samme tilstand?
- *Kostnad*
 - Hva koster det å utføre analysen?
 - Medfører analysering noen risiko for pasienten?
- *Behandling for sykdommen*
 - Finnes det noen behandling for sykdommen? Hvis ikke, hvorfor skal sykdommen diagnostiseres?
 - Er behandling forbundet med risiko for pasienten?
 - Ved hvilken sannsynlighet for sykdom skal analysen brukes, med andre ord: Hva er indikasjon for analysen?
- *Kostnadseffektivitet*
 - Hva er kostnad pr nyttemål (for eksempel per levedøgn) for en strategi som innebærer bruk av analysen, *sammenliknet* med andre tiltak?

2.1.3. Referansegrenser og andre beslutningskriterier

Referansegrenser er det siste som valideres siden de i seg selv ikke påvirker beslutningen om man skal akseptere metoden eller ei.

Det er vanlig å definere et referanseområde som det sentrale persentilintervallet som omfatter 95 % av referanseverdiens fordeling. Dermed blir *referansegrensene* 2,5 og 97,5 persentilene. Referanseverdiene er som regel framkommet ved analyse av prøver fra antatt friske personer.

Referansegrenser kan være bestemt ved minst en av følgende strategier [2]:

1. Bestemt av laboratoriet selv
2. Overført fra referansegrenser som eget laboratorium tidligere har bestemt
3. Bestemt ved samarbeid mellom eget laboratorium og andre laboratorier
4. Bestemt av annet laboratorium eller av andre laboratorier i samarbeid
5. Hentet fra vitenskapelig litteratur
6. Hentet fra produsentens metodedokumentasjon

Den beste strategien er selv å bestemme referansegrensene. For å få rimelig sikre referansegrenser, trenger vi ofte minst 120 individer i hver gruppe (for eksempel 120 kvinner og 120 menn, hvis det er tale om betydningsfulle kjønnsforskjeller. Ved evt. aldersgrupperinger gjøres tilsvarende grupperinger.). Ikke alle laboratorier har ressurser til selv å bestemme referansegrenser, og ofte må vi bruke en av de andre strategiene i listen.

Ved overføring av referansegrenser fra andre metoder i eget laboratorium eller fra andre laboratorier, er det vanlig å analysere 20-40 pasientprøver som ligger innenfor referanse-området, plote samhørende verdier i et XY-plott med ny metode på Y-aksen og gammel metode på X-aksen, finne regresjonslinjen og bruke den til å estimere hva referansegrensene med den gamle metoden svarer til av verdier med nye metoden. Pass på at det er tilstrekkelig spredning av verdier (det skal være verdier under, i og over referanseområdet. Spesielt bør verdier ved referansegrensene for den gamle metoden være godt representert.) og bruk en regresjonsmetode som er egnet (Deming eller Passing-Bablok). Ved beregning av referansegrenser fra komparativ metode skal en være oppmerksom på mulig kaskade av feil hvis tidligere referansegrenser også var beregnet på samme måte.

Å hente referansegrenser fra vitenskapelig litteratur og reagensprodusentens informasjon er det minst ønskelige, fordi det er svært vanskelig å vurdere om slike grenser er gyldige for eget laboratorium. I det minste må vi vite hvilken analysemetode som ble brukt, og vi må nøye vurdere de demografiske data som finnes om referansepopulasjonen. Hvis alle analyseresultater i referansepopulasjonen er kjent, kan referansegrensene valideres ved å sammenligne med tilsvarende analyseresultater i et mindre utvalg fra egen populasjon [8].

2.2.2.2. Praktisk egnethet

Med praktisk egnethet tenker vi på konsekvenser av å innføre/endre analysemetode eller analyseinstrument i laboratoriet, med hensyn til arbeidsressurser, miljø og kostnader.

Her nevnes noen aktuelle forhold der laboratoriet gjerne setter sine krav og innhenter informasjon fra ulike leverandører, eventuelt andre laboratorier – forut for vurdering og beslutninger

2.2.1. Analyseinstrument og utstyr

- Hvis man skal benytte eksisterende analyseinstrument og utstyr, bør man undersøke praktiske konsekvenser av endringen (kapasitet, analysetid, personalbehov etc)
- Hvis det skal anskaffes nytt utstyr, er det naturlig at man i forhold til laboratoriets funksjon og størrelse vurderer:
 - Grad av automasjon
 - Kapasitet, analysetid
 - Plassbehov
 - Driftssikkerhet
 - Service og vedlikehold
 - Datakommunikasjon
 - Miljømessige konsekvenser
 - Personalbehov

Praktiske konsekvenser med hensyn til kapasitet og analysetid:

- En beregning av analysetid ut fra laboratoriets forventede analysemengde og analysemiks bør eventuelt kunne gjøres i samarbeid med leverandør. Firmaene har gjerne computerprogram der laboratoriets eget analysemonster kan legges inn og danne grunnlag for beregningen.
- Miljømessige konsekvenser og vedlikehold/service. Bør vurderes i samråd med tiltenkte operatører. Ofte kan det være nyttig å innhente erfaringer fra andre laboratorier (for eksempel med hensyn til driftsikkerhet).

2.2.1. Reagenser, kalibratører og kontrollmaterialer

Med hensyn til praktisk egnethet i forbindelse med forbruksvarer kan det være aktuelt å vurdere følgende:

- Holdbarhet i forhold til antatt forbruk
- Lagrings- og plassbehov
- Tillaging
- Kalibreringsmetode/kalibreringshyppighet (Antall analysesvar pr. kalibrering)

Her bør man vurdere holdbarhet av uåpnet så vel som åpne reagenser, målestandarder og kontrollmateriale i forhold til forventet bruk. Ofte kan det være nyttig å sette opp en kostnadsvurdering med hensyn til forbruksvarer.

Også her er det muligheter til å få hjelp fra firmaene til å sette opp modeller

2.2.2. Kostnader

I det *totale kostnadsbildet* for implementering av metoden/analyseutstyret inngår:

- Kapitalkostnader
- Driftskostnader
- Evt. opplæring og andre personalkostnader
- Mulige inntekter

Det totale kostnadsbildet bør vurderes opp mot klinisk nytteverdi og eventuelle tidligere kostnader

Undertiden kan det også være aktuelt å gjøre en samlet vurdering av om analysen bør utføres ved laboratoriet, eller om prøvene bør sendes til et større laboratorium eller spesiallaboratorium.

2.2.3. Annet

Andre forhold som det kan være aktuelt å vurdere er:

- Krav til personalets ressurser og kompetanse, mulighet for opplæring,
- Prøvens holdbarhet og andre preanalytiske forhold

Mye av den praktiske informasjonen kan hentes fra leverandør og/eller andre brukere, men laboratoriet må selv vurdere egnethet ut fra egne forhold.

Hvis noen av kravene til praktisk egnethet ikke kan tilfredsstilles, så sier det seg selv at det ikke har noen hensikt å gå videre i valideringsprosessen.

2.3. Analytisk kvalitet

Hvilke parameter som bør valideres/verifiseres må vurderes ut fra analytten som skal måles og metoden som skal benyttes. Det foreligger mye veiledende litteratur om validering av analytisk kvalitet, både når det gjelder forslag til krav og forslag til praktisk gjennomførelse, dessuten finnes det statistikkprogram som letter behandlingen og vurderingen av resultatene. I det følgende vil det bli gitt henvisninger til disse hjelpemidlene.

Korrekt forståelse av begrepene er viktig, og her kan en artikkel av René Dybkaer: "Vocabulary for Use in Measurement Procedures and Description of Reference Materials in Laboratory Medicine [1] være nyttig. Dessuten foreligger det et veiledningsdokument til den nye akkrediteringsstandarden ISO 17025 [3], som heter NA Dok 48a "Klinisk kjemi" [2]. Dette dokumentet er også utstyrt med en begrepsliste der mange begreper som er aktuelle i sammenheng med analytisk kvalitet er beskrevet. NA Dok 48a kan lastes ned fra Internett ; <http://www.justervesenet.no/na/> (Gå til "Dokumentoversikt" og deretter "Prøving")

Det henvises også til Pål Rustads artikler "Validering av analysekvalitet" [9], som gir en del praktiske detaljer om hvordan den analytiske valideringen kan utføres. Westgards weblesjoner om metodevalidering [10] og hans bok "Basic Method Validation" [11] kan også gi nyttig informasjon.

En annen praktisk veiledning, "Utprøving av analyseinstrumenter." [12], utarbeidet i forbindelse med uttesting av utstyr i primærhelsetjenesten, kan være nyttig for validering av Point-of-care-utstyr.

NCCLS (National Committee for Clinical Laboratory Standards) har utgitt en rekke veiledningsprotokoller for evaluering av analysekvalitet. En oversikt over disse finnes på NCCLS' web-sider; <http://www.nccls.org/evalprotocols.htm> Disse er svært omfattende, og relativt kostbare for ikke-medlemmer.

Spesielt aktuelt for immunkjemiske analyser er Lars Mørkrid's artikkel: "Statistiske vurderinger ved endringer av analysemetode" [13].

I det følgende gis noen kommentarer til de mest sentrale parametrene.

2.3.1. Riktighet

Grad av overensstemmelse mellom gjennomsnittlig verdi fremskaffet fra en stor serie måleresultater og en sann verdi. [1,2]

Som utgangspunkt for validering av riktighet bør laboratoriet ha innhentet dokumentasjon på metodens sporbarhet:

Sporbarhet

Egenskap ved et måleresultat eller ved verdien for en målestandard som gjør at det kan relateres til angitte standarder (vanligvis internasjonale eller nasjonale), gjennom en ubrutt kjede av sammenligninger, alle med angitt usikkerhet [1,2].

Sporbarhet innebærer med andre ord at målingene refererer seg til et standardmateriale eller en målemetode så langt oppe i sporbarhetskjeden som det er rimelig å forvente for vedkommende komponent, og at måleresultatet oppgis med (ekspandert) måleusikkerhet.

Leverandør/produsent bør kunne fremskaffe dokumentasjon på sporbarhet så langt dette er praktisk mulig. Hvor høyt man kommer i sporbarhetskjeden, avhenger av tilgjengeligheten av referansemetoder og referansematerialer som kan sikre overføring nedover i kjeden. En standard, ISO/DIS 17511 [14] som omhandler sporbarhet og beskriver alle begreper i denne sammenheng er under utarbeidelse.

Riktighet uttrykkes kvantitativt ved *systematisk feil* eller *bias (B)*.

Krav til riktighet

er utførlig beskrevet av Fraser [15] i forbindelse med et møte om kvalitets-spesifikasjoner i klinisk kjemi som ble holdt i Stockholm i 1999. (Hele møtet er gjengitt i SJCLI no 59, 1999).

Fraser sammenfatter kravene til riktighet slik:
Når den analytiske variasjonen CV_a , er ubetydelig og den totale biologiske variasjonen er gitt ved

$$CV_{tb} = \sqrt{(CV_w^2 + CV_b^2)}$$

så bør en målsetning for systematisk feil være:

Optimalt: $B < 0,125 \times CV_{tb}$

Ønskelig: $B < 0,250 \times CV_{tb}$

Minimum: $B < 0,375 \times CV_{tb}$

Hvis data om biologisk variasjon ikke foreligger, har man også benyttet:

$$B < 1/16 \text{ av referanseintervallet [16]}$$

Hvis samme analytt måles på forskjellige instrumenter innen et laboratorium, foreslås følgende kriterium [15]:

$$\text{Akseptabel forskjell} < CV_w/3$$

Vurderingen bør være basert på gjennomsnittsverdier.

Tegnforklaring:

B = Bias (Systematisk feil) angitt i % av måleverdien

CV_a = Analytisk variasjon

CV_{tb} = Total biologisk variasjon

CV_w = Intraindividuell biologisk variasjon

CV_b = Interindividuell biologisk variasjon

Verdier for biologisk variasjon kan hentes fra Ricòs C, [17], fra
<http://www2.legeforeningen.no/yf/kkjem/Nfkk/Kvalktrutvalg/kvut5.html> eller
<http://www.westgard.com/biodatabase1.htm>

Metodens riktighet kan valideres ved:

A. Å analysere et referansemateriale som har en sporbarhet i overensstemmelse med metodens sporbarhet. For eksempel kan dette være et sertifisert referansemateriale (CRM), hvis slikt er praktisk tilgjengelig.

Forslag til fremgangsmåte: Referansematerialet (RM) og bruksstandard (BS) analyseres annenhver gang (ca.10 ganger) i en serie, hvorpå middelverdier og impresisjon beregnes.

Vurder differansen mellom målt, \bar{x}_{RM} og oppgitt verdi, OV_{RM} . Vurder også om verdien på bruksstandard, \bar{x}_{BS} har gitt korrekt verdi, OV_{BS} . Hvis ikke, korriger:

$$X_{RM\text{korrigert}} = \bar{x}_{RM\text{målt}} * \frac{OV_{BS}}{\bar{x}_{BS}}$$

Vurder $X_{RM\text{korrigert}}$ mot OV_{RM} . Ta i betraktning usikkerheten i den målte $X_{RM\text{korrigert}}$ og den angitte usikkerheten i OV_{RM} .
(Disse bør ligge i omtrent samme konsentrasjonsområde)

B. Sammenligninger med andre metoder ved eget eller andre laboratorier.
Sammenlignende metode bør være en anerkjent, godt dokumentert metode. Ved slike sammenligninger bør man benytte pasientmateriale som spenner over hele måleområdet. For enkelte metodeprinsipper (for eksempel immunoassays, hematologiske undersøkelser) kan det være aktuelt å inkludere pasientprøver fra ulike sykdomsgrupper.

Metoder som er egnet til slike sammenlikninger kan være

- Spredningsdiagram
 - Et enkelt XY-diagram med verdiene til den nye metoden langs Y-aksen og den andre metoden langs X-aksen. Visuell bedømmelse av trendlinjen.
- Statistikkprogram som beregner avvik fra X=Y-linjen (slope, intercept og korrelasjonskoeffisient)
 - *Lineær regresjon*. Benytter minste kvadraters metode, der variasjonen kun er til stede i Y-variabelen.
 - *Demingsmetode*. Tillater variasjon både i X- og Y-variablene.
 - *Passing og Bablok's metode*. Tillater variasjon i både X- og Y-variablene, men er ikke-parametrisk og påvirkes dermed lite av ekstreme avvik mellom X og Y.
- Differanseplot

Brukes når man på Y-aksen angir forskjellen mellom metoder i stedet for en av metodene.
En spesiell utforming av differanseplot er Bland/Altman-plot, der X-aksen angir middelveien av de to metodene.

For mer forklaring og vurdering henvises til [9] og litteratur i forbindelse aktuelle statistikkprogram:

Som statistikkprogram anbefales det engelske "Analyse-it", som ble gjennomgått på NKK-møtet i 2000. Det henvises til møtedokumentet utarbeidet av Sveinung Rørstad/Pål Rustad.

En anbefalt statistikkbok er "Clinical Investigation and Statistics in Laboratory Medicine" [5].

C. Gjenfinningsforsøk

Som et supplement kan riktighet i en del tilfeller testes ved gjenfinningsforsøk (recovery). Det vil si at man tilsetter økende mengder analytt til en pasientprøve med lav konsentrasjon av analytten og analyserer mengden (%) gjenfunnet analytt. Det kan være et par svakheter forbundet med denne metoden: a) at en eventuell konstant bias ikke vil kunne oppdages, og b) at tilsats av ren analytt ikke alltid vil medføre at denne vil være identisk med analyttens molekulære form in vivo.

Metodens riktighet bør alltid verifiseres av laboratoriet, både av hensyn til gyldigheten av referansegrensene (kap.2.1.3), men også av hensyn til overensstemmelse mellom laboratorier innenfor et geografisk område.

Parametere som har tilknytning til riktighet er:

Linearitet

Betegner en analysemetodes evne (innen et gitt konsentrasjonsområde) til å gi målesignaler som er direkte proporsjonal til konsentrasjonen av analytten.

Man tester sammenhengen mellom målesignal (f.eks OD) og målestørrelse (f.eks konsentrasjonen). Denne skal være beskrevet av en rett linje $y=ax+b$, der a er helningsvinkel (slope) og b er skjæringspunktet med Y-aksen (intercept).

Det kan noen ganger være et problem å få testet hele måleområdet. Man kan velge en høy og en lav pasientprøve, og lage fortynninger ved å blande disse til for eksempel 5 konsentrasjoner jevnt fordelt i måleområdet. Det bør velges konsentrasjoner som ligger nær medisinske beslutningsgrenser, samt i øvre og nedre del av det oppgitte linearitetsområdet. Prøvene bør analyseres i en serie, resultatene kan plottes i et XY-diagram og lineariteten vurderes visuelt. Eventuelt kan differanseplot benyttes.

Det bør foreligge dokumentasjon på linearitet for metodens angitte måleområde, enten fra produsent eller fra egne forsøk.

Interferens

Systematisk feil som skyldes en substans i prøven, annen enn den som skal måles [1,2].

Det finnes et stort antall mulige interferenter. Det kan være endogene substanser, spesielt optiske interferenter som hemoglobin, lipider og bilirubin, samt legemidler, prøvetilsetningsstoffer, etc.

Data om interferens (spesielt hemolyse, ikterus og lipemi) bør finnes i produsentens metodeinformasjon. Hvis grenseverdier ikke finnes, bør slike utarbeides ved å utføre tilsetningsforsøk [9,12,20]. For legemidler er det utarbeidet en bok som benyttes i Norden "Drug effect in Clinical Chemistry" [19]. Denne er også tilgjengelig på Internett ([http://home1.swipnet.se/~w-12360/#Drug effects](http://home1.swipnet.se/~w-12360/#Drug%20effects))

Grad av interferens som kan aksepteres bør vurderes på bakgrunn av medisinske krav.

Analytisk spesifisitet

Betegner en analysemetode's evne til å måle kun den analytten den har til hensikt å måle [1,2].

Validering av en metodes spesifisitet kan også knytte seg til hvilke molekyllære former av analytten som metoden er tiltenkt å bestemme, evt hvordan antatt interfererende forbindelser påvirker analyseresultatet. Slike undersøkelser innhentes gjerne fra vitenskapelig litteratur eller fra dokumentasjon fra produsenten

Laboratoriet kan teste analytisk spesifisitet ved hjelp av gjenfinningsforsøk. Det kan dreie seg om gjenfinning av den analytten man ønsker å måle, eller man kan teste interferens av substanser/isoformer man ikke ønsker skal påvirke resultatet [9].

Holdbarhet

Holdbarhet gjelder både reagens, målestandard, kontroll og analytten i prøven. Ofte aksepteres produsentens fastsettelse av holdbarhet. Brukerens ansvar blir da å sørge for at oppbevaringsbetingelsene er i overensstemmelse med leverandørens krav. Men ofte er holdbarheten satt unødvendig kort, derfor kan det være nyttig å teste holdbarheten selv.

For praktisk utførelse av holdbarhetstester, se [9]

2.3.2. Presisjon

Presisjon er overensstemmelse mellom uavhengige måleresultater oppnådd med en måleprosedyre under angitte betingelser [1,2]. Presisjon er et kvalitativt begrep.

Kvantitativt måler vi *impresisjonen* som et uttrykk for tilfeldige feil når vi benytter målemetoden i vårt laboratorium.

For praktiske formål bestemmer vi forskjellige former for impresisjon, viktigst er repeterbarhet og reproducerbarhet. Fordi en metodes presisjon påvirkes av konsentrasjonsnivå, bør den bestemmes i mer enn ett konsentrasjonsnivå, primært nær beslutningsgrensene.

Impresisjonen uttrykkes som standard avvik og/eller variasjonskoeffisient.

Repeterbarhet (innen-serie impresisjon)

Overensstemmelse mellom resultatene av påfølgende målinger av samme målestørrelse utført under samme målebetingelser. [1,2]

Repeterbarheten skal referere seg til hvilke deler av måleområdet den gjelder for og målebetingelser for øvrig.

Her skal man altså tilstrebe å ha mest mulig like analysebetingelser; samme operatør, og samme bruksbetingelser for øvrig, - og målingene skal utføres over et kort tidsrom, for eksempel i en analyseserie. Repeterbarhet gir dermed et estimat på den minste måleusikkerhet som kan oppnås med analysemetoden.

Forslag til undersøkelse:

Repeterbarhet testes oftest ved å benytte to eller flere kontrollmaterialer og/eller pasientprøver som dekker måleområdet og analysere disse for eksempel 20 ganger hver.

Eventuelt kan man benytte utvalgte pasientprøver som analyseres som duplikater (x_1 og x_2). Ved denne fremgangsmåten beregnes repeterbarheten slik:

$$S = \sqrt{\frac{\sum (x_1 - x_2)^2}{2N}} \quad CV(\%) = \frac{S}{X_{mid}} 100$$

der X_{mid} er middelværdien av alle målingene i serien og N er antall prøver (=antall målinger/2). Utvalget av prøver bør gjøres slik at man kan beregne impresisjonen ved lav, middels og høy konsentrasjon.

Nytten av å teste repeterbarhet er å kunne se verdiene i forhold til reproduserbarhet, for bedre å kunne vurdere hvor de vesentlige kildene til metodens impresisjon ligger.

Reproduserbarhet

Overensstemmelse mellom resultatene av påfølgende målinger av samme målestørrelse utført under endrede målebetingelser [1,2].

Reproduserbarheten skal referere seg til hvilken del av måleområdet den gjelder for, og spesifisering av hvilke variable som har inngått i perioden.

Her ønsker man størst mulig variasjon i analysebetingelsene, - flere analyseserier, flere "loter" både på standard og reagenser, flere operatører, osv. Det vil si at reproduserbarhet bør omfatte en lengre periode, for eksempel ½ år. Under ingen omstendigheter må antall målinger være under 20.

Reproduserbarhet blir ofte brukt som et mål for analytisk måleusikkerhet. I mangel av kjennskap til usikkerhet i målestandard, må dette aksepteres. Hvis reproduserbarhet skal benyttes som mål for usikkerheten, er det en forutsetning at undersøkelsen går over så lang tid at alle praktisk tenkelige variable kommer med.

Forslag til krav til reproduserbarhet (for monitorering) [15]:

Optimal kvalitet:	$CV_a < 0,25 \times CV_w$
Ønskelig kvalitet:	$CV_a < 0,50 \times CV_w$
Minste akseptable kvalitet:	$CV_a < 0,75 \times CV_w$

Se tegnforklaring under pkt. 2.3.1

For medikamenter: $CV_a < 0,25[(2^{T/t}-1)(2^{T/t} + 1) \times 100]$
der T er doseringsintervall og t er halveringstid [18]

Forslag til undersøkelse

Fordi reproduserbarhets undersøkelser vil foregå over tid, er det viktig at prøvematerialet er stabilt i perioden, spesielt er det viktig å klarere dette ved bruk av pasientprøver. Samtidig bør utvalget representere hele måleområdet.

Verifisering av nivå ved lot-skifte

Lot skifte av målestandard/reagens kan for enkelte teknologier, for eksempel tørrkjemi og immunoassays, være forbundet med et nivåskift for pasientresultatene. For metoder der dette er et problem bør laboratoriet sette krav til tillatt endring ved lotskift, og verifisere at disse kravene er tilfredsstillt når det skiftes til ny lot. Kravet bør sees i sammenheng med den garantien leverandøren oppgir.

Oftest er det målestandardene vi forbinder med fare for nivåskift i pasientresultatene, og det kan være aktuelt å analysere det nye kalibratorsettet som prøver, og vurdere resultatene mot de tillagte verdiene fra produsenten. Men i mange tilfeller leveres kalibrator og reagens som ”matched” sett, og man må teste hele det nye settet mot det gamle.

En måte å teste konsekvensene av lotskift på, er å velge kontroller/pasientprøver over et relevant konsentrasjonsområde og beregne avvik fra X=Y-linje. De forventede verdiene (X) kan være verdiene man oppnådde ved eksisterende lot. Eventuelt, kan man samle et sett med kontrollmaterialer/prøver som man rutinemessig benytter for vurdering av nye lots. Tillatt avvik kan for eksempel være at målepunktene ligger innenfor \pm en fastsatt % fra X=Y-linjen, evt en fastsatt del av metodens CV%. Tillatt avvik vil være noe avhengig av hvilken type metode, evt. hvilke deler av måleområdet det dreier seg om, men bør generelt være basert på kvalitetsmål som angitt over (”Forslag til krav”). For immunoassays er ofte \pm (5-10) % benyttet.

Alternativt henvises det til testen Pål Rustad foreslår [9]

Robusthet

Ved validering av en metodes robusthet er det viktig å fremkalle aktuelle ytre forhold som kan tenkes å påvirke metoden, som for eksempel endringer i fysiske omgivelser, skifte av reagens- eller bruksstandard-lot, skifte av operatør, etc. Robusthet bør valideres/verifiseres over tid, slik at flest mulige faktorer som kan tenkes å påvirke metoden kan fanges opp [2]. Ved angivelse av robusthet bør de påtrykte ytre forhold som har vært tilstede beskrives.

Carry over/Overdraging

I hvilken grad konsentrasjonen av en analytt i en prøve påvirker den målte konsentrasjon i seriens neste prøve [9].

Hvilken grad av overdraging (%) som skal aksepteres, kan variere fra analytt til analytt.

Carry over er spesielt viktig å undersøke for analyser der konsentrasjonsspennet er stort, eller i tilfeller der analysesystemet er kjent å ha en svakhet. For å teste ut overdraging velger man en prøve med høy og en med lav konsentrasjon av den aktuelle analytten. Høyt nivå bør ligge i øvre del av måleområdet og lavt nivå bør ligge så lavt som mulig. Prøvene analyseres i en serie etter følgende mønster L1-L2-H1-H2-L3-L4-H3-H4-L5-L6.... Osv.

Overdraging (C) kan beregnes slik:

$$C_1 = \frac{H_2 - H_1}{H_2 - L_2} \quad C_2 = \frac{L_3 - L_4}{H_2 - L_4} \quad osv$$

Enkelte analysekombinasjoner er også følsomme for overdraging av reagens. I slike tilfeller bør produsenten opplyse om dette slik at laboratoriet kan ta de nødvendige forholdsreglene.

2.3.3. Måleusikkerhet

Parameter tilknyttet et måleresultat som karakteriserer spredningen av verdier som kan tilskrives målestørrelsen [1,2].

Estimering av usikkerhet inngår som en del av sporbarhetsbegrepet, og disse to parametrene er i de senere årene ansett som de ideelle ved angivelse av analysekvalitet.

Analytisk måleusikkerhet:

Den analytiske måleusikkerheten omfatter usikkerheten av enkeltelementene i måleprosessen, som for eksempel usikkerhet i målestandard, pipettering, usikkerhet i etablering av standardkurve, måling av prøven, etc., og skal ideelt beregnes ut fra estimater av enkeltelementene. Man forutsetter at alle kjente systematiske feil er korrigert for, slik at de ikke influerer på fastleggelsen av måleusikkerheten. Usikkerheten kan variere i ulike deler av måleområdet og det er derfor viktig å vurdere hele måleområdet.

En måte å beregne den analytiske måleusikkerheten på er:

- Identifiser alle usikkerhetskildene i måleprosessen (se ovenfor),
- Beregn usikkerheten av de ulike elementene (*standard usikkerhet* u_1, u_2, \dots); noen kan måles/estimeres, andre finnes ved ”kvalifisert gjetning” eller fra andre kilder.
- Beregn så *kombinert usikkerhet* u ved å ta kvadratroten av summen av variansene for standard usikkerhetene:

$$u = \sqrt{(u_1^2 + u_2^2 + \dots + u_i^2)}$$

Måleusikkerheten angis gjerne som *utvidet målesikkerhet* (U), det vil si kombinert usikkerhet u multiplisert med en dekningsfaktor (k) som definerer konfidensintervallet. Ofte benyttes $k=2$, som gir konfidensnivå = 95%

Å beregne måleusikkerheten slik er ofte vanskelig, fordi man ikke kjenner alle elementene og/eller usikkerhetene de enkelte elementene bidrar med. Laboratoriene bestemmer derfor oftest den analytiske måleusikkerheten empirisk ut fra reproduserbarhets-undersøkelser. For å lære usikkerhetsfaktorene ved metoden bedre å kjenne, kan det imidlertid være fornuftig å forsøke, så langt som mulig, å estimere de viktigste bidragene.

Preanalytiske forhold:

Valideringen kan også omfatte preanalytiske forhold (prøvetaking, prøvebehandling, oppbevaring...) som er av betydning for måleresultatet. Som et minimum bør skjønsmessige vurderinger av slike forhold inkluderes.

Siden beregning av usikkerhet er tungt tilgjengelig stoff, vil vi henwise til videre studier:

- Anders Kallner: Quality specifications based on the uncertainty of measurement ; Scand J Clin Lab Invest 1999; 59:513-16
- Rettledning utarbeidet av Eurochem, tilgjengelig fra nettet: <http://www.eurachem.ul.pt>
- Pål Rustad, handouts fra presentasjon på NKK-møtet i 1999 på Hell, ”Introduksjon til usikkerhetsbegrepet”.
- E.Theodorsoson og S. Norheim; Diagnostisk usikkerhet ock måtusikkerhet inom laboratoriemedicin; Klinisk kjemi i Norden 2001; 3:10-14.
- Rapport om ”Måleusikkerhet for kjemiske analyser”. NA’s P4-komite.

2.3.4. Måleområdet

Det området for målestørrelsen hvor målinger kan foretas med angitt usikkerhet [9]

Måleområdet er definert av nedre og øvre målegrense (nedre og øvre kvantifiseringsgrense). Grensene kan være definert litt forskjellig:

Øvre målegrense:

Kan være bestemt av:

- Verdien av høyeste kalibrator
- Høyeste måleverdi som gir linearitet

Ved en del immunkjemiske analyser bør man være spesielt oppmerksom på såkalt ”hook” effekt. Det vil si at prøver med analyttkonsentrasjon høyere enn øvre målegrense gir målerespons som om prøven lå i måleområdet, med andre ord gir falskt for lave resultater. Dette kan avsløres ved fortykning av prøven.

Nedre målegrense:

Det mest benyttede begrepet er :

- Deteksjonsgrense (eng. Detection limit/lower limit of detection): Den laveste konsentrasjon av en analytt som ved et gitt sannsynlighet kan måles/skilles fra null. Det vil si at konsentrasjoner lavere enn deteksjonsgrensen bør rapporteres som ”mindre enn”.

Forslag til undersøkelse

Analyser 0-kalibratoren 20 ganger. Beregn middelerdi og standardavvik. Deteksjonsgrensen settes som 3SD over middelerdien.

Andre begreper som benyttes:

- Påvisningsgrense: Laveste konsentrasjon hvor vi med sikkerhet kan si at analytten er tilstede i prøven.
- Funksjonell sensitivitet: For komponenter der nedre målegrense er diagnostisk viktig setter man gjerne krav om at impresisjonen (CV%) skal være mindre enn 20%. Da kalles nedre målegrense også for ”funksjonell sensitivitet”

OBS! Begrepet analytisk sensitivitet benyttes feilaktig for å angi deteksjonsgrensen i pakningsvedlegg fra enkelte produsenter, *men analytisk sensitivitet er definert som hellingen (slope) på standardkurven !*

3. DOKUMENTASJON OG KONKLUSJON

En metode kan ikke sies å være validert før dokumentasjon og konklusjon foreligger. For akkrediterte laboratorier kreves det at dokumentasjon på valideringen [3] foreligger i laboratoriet, både deler som er validert utenfor laboratoriet og undersøkelser som laboratoriet har utført selv. Men også de øvrige laboratoriene bør ha denne dokumentasjonen, eller i alle fall sammendrag og konklusjoner tilgjengelig i laboratoriet. Dokumentasjonen bør i prinsippet være ”sporbar” slik at en kollega kan reprodusere arbeidet ut fra dokumentasjon og henvisninger.

Dokumentasjonen kan for eksempel inneholde:

- Kvalitetskrav.
- Protokoll for valideringens omfang (hvilke parametere, hvor mange prøver, osv)
- Nøyaktig henvisning til litteratur som inngår
- Fortegnelse over katalog- og lot- nr. på reagenser, kalibratorer, kontroller, pasientprøver, etc som er benyttet
- Tidsrom og informasjon om hvem som har utført undersøkelsene
- Presentasjon av resultater, statistikk, kurver og vurderinger
- Konklusjon: Bedømmelse av om resultatene tilfredsstillt kvalitetskravet laboratoriet har satt for analysen.

Hvert laboratorium bør lage en oversikt som passer de valideringsdata som foreligger, for eksempel slik at man for hver analytt fører opp valideringsparametere, krav og bedømmelse. Et eksempel på hvordan et slikt skjema kan bygges opp er vist i vedlegg 1.

Det er også viktig å notere ned resultater og utregninger underveis, samt å sammenligne med de krav som er satt. Dersom man underveis ser at vesentlige krav ikke er tilfredsstillt, er det all grunn til å stoppe opp og vurdere hvordan man skal løse det aktuelle problem.

4. REFERANSER

1. Dybkaer R. Procedures and Description of Reference Materials in Laboratory Medicine. Eur J Clin Chem Clin Biochem 1997, 35(2), 141-73
2. NA Dok. nr. 48a , “Klinisk kjemi”. Veiledningsdokument til ISO 17025,
3. ISO 17025: Generelle krav til prøvnings- og kalibreringslaboratoriers kompetanse.
4. Kvalitetsutvalget v/Arne Åsberg og medarbeidere, Norsk Forbund for Klinisk kjemi. <http://www.legeforeningen.no/yf/kkjem/main.html>
5. Jones R, Payne B. Clinical Investigation and Statistics in Laboratory Medicine. ABC Venture Publications, London, 1997. ISBN 0-902429-21-3
6. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamenta tool in clinical medicine. Clin Chem 1993; 39: 561-77.
7. Boyd JC. Mathematical tools for demonstrating the clinical usefulness of biochemical markers. Scand J Clin Lab Invest 1997; 57 (suppl 227); 46-63.
8. Holmes EW, Kahn SE, Molnar PA, Bermes EW. Verification of reference ranges by using a Monte Carlo sampling technique. Clin Chem 1994; 40: 2216-22.
9. Rustad P. Validering av analysekvalitet.
Del I i Klinisk kjemi i Norden 2000; 4: 21-25
Del II i Klinisk kjemi i Norden 2001; 2: 12-15
10. Westgard's lesjoner om metodevalidering <http://www.westgard.com/lesson.htm> Basic Method Validation
11. Westgard JO. Basic Method Validation. Westgard QC, Madison 1999. ISBN1-886958-12-2
12. Christensen NG, Monsen G, Sandberg S. Utprøving av analyseinstrumenter. Alma mater forlag AS, Bergen 1997. ISBN 82-419-0230-1
13. Mørkrid L. Statistiske vurderinger ved endring av analysemetode. Klinisk kjemi i Norden; 1998; 2: 59-64
14. ISO/DIS 17511 (Draft). In vitro diagnostic medical devices – Measurement og quantities of biological origin – Metrological traceability of values assigned to calibrators and materials.
15. Fraser CG. Genel strategies to set quality specifications for reliability performance characteristics. Scand J Clin Lab Invest 1999; 59:487-90
16. Bolann BJ, Stølsnes B. Analytisk usikkerhet – hvor stor feil kan laboratoriesvaret ha? Tidsskr Nor Lægeforening nr.30, 1999; 119: 4472-5
17. Ricòs C, Alvarez V, Cava J et al. Current databases on biological variation: pros, cons and progress. Scand J Clin Lab Invest 1999; 59:491-500
18. Fraser CG. Desirable goals in therapeutic drug monitorering. Clin Chem 1987; 33: 387-9
19. Tryding N, Hansson P, Tufvesson C et al. Drug effects in Clinical Chemistry 1992. Apoteksbolaget AB, ISBN 91-85574-38-4 ISSN 0282-5783.
20. Glick MR, Ryder KW, Jackson SA. Graphical Comparisons of interferences in Clinical Chemistry Instrumentation. Clin.Chem. 32/3, 470-75 (1986)