

# Metodeverifisering

## Verifisering av riktighet

**V**ED AVDELING FOR MEDISINSK BIOKJEMI, Diakonhjemmet Sykehus AS, har vi innført Mentormetoden som et alternativ til full sammenlikning når store analyseinstrumenter skal verifiseres.

Av **LISE LARSEN MEHUS**, bioingeniør med spesialistgodkjenning, Avdeling for medisinsk biokjemi, Diakonhjemmet sykehus

■ Temaet i artikkelen ble også presentert som forelesning på kongressen.

Som ledd i min spesialistgodkjenning i statistikk og kvalitetsarbeid skrev jeg fordypningsoppgaven "Statistiske metoder ved metodesammenligning". Den omhandler verifisering av riktighet. Denne artikkelen er et utdrag av oppgaven. Oppgaven var også utgangspunkt for min presentasjon på Bioingeniørkongressen med tittelen "Metodeverifisering" Innkjøring av store analyseinstrumenter.

### Bakgrunn

I fordypningsoppgaven tok jeg utgangspunkt i Diakonhjemmets interne prosedyre for verifisering av nye analysemetoder. Dette dokumentet var blitt skrevet som prosjektoppgave på studiet "Kvalitetssikring i laboratorier" og omhandlet alle de viktigste elementer i forbindelse med en metodeverifisering. I kapitlet om verifisering av riktighet, hadde vi presisert at dette skulle gjøres ved at 20 - 40 prøver, jevnt fordelt i måleområdet, ble analysert på begge metoder. Utprøvingen skulle gå over fem dager. Prosedyren fungerte greit så lenge det bare var én enkelt analyse som skulle kjøres inn.

Men da vi gikk til anskaffelse av nytt utstyr med klinisk kjemi og immunkjemi konsolidert i en og samme maskin, skulle til sammen 56 analyser flyttes over på ny apparatur. Vi skulle sågar ha to like instrumenter som også måtte korreleres mot hverandre. Det ble en stor jobb og det var til tider vanskelig å holde oversikt. Særlig var det problematisk at innkjøringen av hver analyse skulle gå over fem dager.

Vi kom vel i havn, og instrumentene ble

akkreditert uten anmerkninger.

Siden hørte jeg at andre laboratorier planla å gjøre sin innkjøring ved å benytte seg av Mentormetoden på flere analyser. Kunne det virkelig gjøres så enkelt, innkjøring med bare to prøver?

### Problemstilling

Jeg ønsket å undersøke dette. Jeg bestemte meg for å se på prinsippet for Mentormetoden og finne ut hvilke begrensninger som gjelder. Målet var å implementere metoden i vår prosedyre for metodeverifisering. Videre ville jeg se om det var andre deler i våre prosedyrer som burde utdypes eller endres.

Diakonhjemmets prosedyrer beskriver bruk av Analyse-it som statistisk verktøy og det lages:

- Biasplot, (absolutt og relativt plot).
- Regresjonsanalyse ved bruk av Passing & Bablok's ikke parametriske metode.

Metodeforskjellen beregnes ved å sette inn medisinsk viktige konsentrasjoner i den beregnede ligningen. Deretter vurderes det om metodeforskjellen er større enn kravet til riktighet (1).

Mitt spørsmål var om dette er en tilstrekkelig analyse av resultatene eller om ytterligere vurderinger burde gjøres. Videre ønsket jeg å se om det er forsvarlig alltid å utføre regresjon med Passing & Bablok, uten først å vurdere differansetester, ratiotester eller parametriske regresjonsmodeller, for eksempel ordinær lineær regresjon (OLR) eller Deming regresjon.

### Statistiske metoder

Noen statistiske metoder og vurderinger er viktige i denne sammenheng.

#### Lineær sammenheng

Første ordens regresjonsanalyse forutsetter at det er en underliggende rett-

linjet sammenheng mellom variablene. For å teste dette lages et spredningsdiagram med komparativ metode på x-aksen og testmetoden på y-aksen. Sammenhengen vurderes visuelt, evt. sammen med korrelasjonskoeffisienten, selv om denne ikke sier noe direkte om formen på sammenhengen (se figur 1).

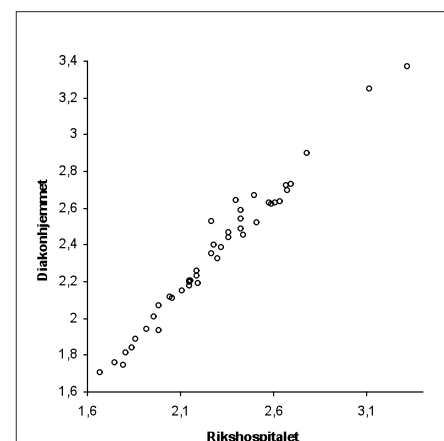
#### Korrelasjonskoeffisienten

r er et mål for punktenes spredning omkring linjen. Korrelasjonskoeffisienten ligger alltid mellom -1 og 1. Når  $r = 1$  eller  $-1$ , ligger alle punktene eksakt på en rett linje og det er fullstendig korrelasjon. Dersom  $r = 0$  finnes det ingen rettlinjett sammenheng mellom variablene. Det finnes én parametriske korrelasjonskoeffisient, Carl Pearssons r; samt to ikke-parametriske: Spearmans r og Kendalls  $\tau$ , som er basert på måleverdiens rangverdier, og som derfor ikke er så følsomme for "slengere".

For vårt bruk ved metodesammenligning vil det alltid være en sammenheng mellom variablene. Det er derfor lite nyttig å vurdere korrelasjonskoeffisienten. Den kan tvert imot gi opphav til feiltolkninger hvis den vektlegges. I laboratorimedisinen er r ofte bare et mål på om vi har tatt med et nok stort konsentrasjonsområde i utprøvingen (2).

#### Test på normalfordeling

En forutsetning for å benytte parametriske tester er at sannsynlighetsfordelingen for data kan beskrives som en matematisk funksjon av den aktuelle variabel med et



FIGUR 1. Spredningsdiagram.

Figuren er hentet fra en metodesammenligning mellom Diakonhjemmet og Oslo universitetssykehus, avdeling Rikshospitalet, på to utvalgte komponenter, Ca og TSH. Metodene var godt etablert på begge steder.

sett faste parametre. I praksis blir dette oftest en normalfordeling. Når vi vet noe om fordelingen våre resultater er trukket fra, kan vi si noe om hele populasjonen. Dersom vi vil bruke parametriske metoder, må vi derfor teste om de variablene som inngår i testen er normalfordelte. Dersom de ikke er det, kan man forsøke å transformere data, for eksempel med log-transformasjon. I differanse- og ratio-tester, samt i regresjonsmodeller, er det restleddene som skal være normalfordelte. X og Y derimot, skal være jevnt fordelt over hele måleområdet.

**Residualer**

Residualene, eller restleddene, er de enkelte punktenes avvik fra linjen og sier noe om hvor godt en valgt regresjonsmodell passer til resultatene. Ved bruk av parametriske metoder må det testes om residualene er normalfordelte ved å framstille dem i et histogram eller z-score-plot. Målepunktene skal være jevnt fordelt over måleområdet, og residualene skal vise et jevnt rektangulært bånd parallelt med x-aksen (se figur 2). Dersom punktenes avstand fra linjen avtar eller øker, har ikke y-metoden konstant impresisjon (varians) over hele måleområdet. Vektet Deming er da et alternativ. Passing & Bablok er en ikke-parametrisk metode som ikke forutsetter normalfordeling.

Dersom residualene danner klumper eller har lange strekk over eller under linjen tyder dette på at det ikke er lineær relasjon mellom variablene. I så fall er rettlignet regresjonsanalyse ikke riktig modell, og da bør heller ikke Passing & Bablok benyttes. I slike tilfeller kan man prøve med en transformasjon av data, en

annen matematisk funksjon for regresjonsligningen (høyere ordens polynom, for eksempel ta med et annengradsledd), eller dele data opp i flere underområder, som hver har en tilnærmet rettlignet sammenheng mellom X og Y.

I Passing & Bablok-plottet presenteres "Kusum test for linearitet". Dersom  $p < 0,1$  er det observert ikke-rettlignet sammenheng. Man bør se nøye etter om "slengere" kan være årsak til dette.

**Vurdering av regresjonsmodeller**

*Ordinær lineær regresjon (ORL)*

I denne modellen minimaliseres kvadratsummen av de lodrette avstandene mellom målepunktene og linjen. Modellen antar at den eneste grunnen til at punktene ikke ligger eksakt på en rett linje er at y-verdiene er usikre. ORL skal derfor bare brukes dersom impresisjonen i y-metoden er vesentlig større enn impresisjonen i x-metoden. En ratio på 2 holder, det vil si at CV for y-metoden er dobbelt så stor som CV for x-metoden.

*Deming*

Denne modellen sier at begge analysemetodene bidrar vesentlig til impresisjonen i restleddene. Måten å måle avstanden mellom linjen og målepunktene avhenger av forholdet mellom analysemetodenes varianser. Data-programmet spør etter begge metodenes variasjonskoeffisienter. Dersom metodene har samme upresisjon, er det avstanden vinkelrett fra målepunktene og inn på linjen som minimaliseres. Dersom upresisjonen i y-metoden er vesentlig større enn i x-metoden, gir Deming og ORL tilnærmet samme resultater.

Både ORL og Deming er parametriske metoder og krever at residualene er tilnærmet normalfordelte. Impresisjonen skal ikke variere over måleområdet, noe som oftest er tilfellet når det er en fast variasjonskoeffisient over måleområdet, slik at standardavviket øker med økende måleverdier. Dette kan ofte avhjelpes ved å rot- eller log-transformere data. En annen mulighet er å benytte vektet Deming. Når usikkerheten øker proporsjonalt med konsentrasjonen er dette en god metode da punktene i det lave område vektlegges mer. Vektet Deming finnes i nyere versjon av Analyse-it eller på Norsk Klinisk-kjemisk kvalitetskontroll (NKKs) hjemmeside (3).

*Passing & Bablok*

Dette er en ikke-parametrisk metode som ikke krever at residualene er normalfordelte, men den krever likevel at det er et underliggende rettlignet forhold mellom X og Y. Her finnes regresjonslinjens helning ved å beregne stigningskoeffisienten av alle mulige linjekombinasjoner mellom to og to punkter. Median av disse er valgt som det beste estimat av regresjonslinjens helning. Intercept finnes ved å legge linjer gjennom alle punkter med helning som funnet foran og beregne medianen av alle skjæringspunktene med y-aksen. En fordel med denne metoden er at punkter som avviker mye fra resten får liten innvirkning på resultatet (4).

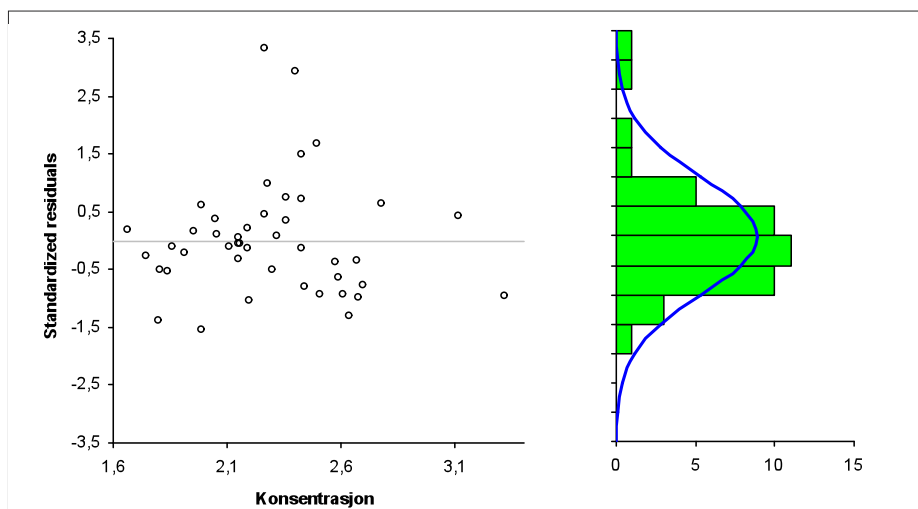
*Wilcoxon*

Dette er en ikke-parametrisk test på forskjell mellom to grupper eller metoder og er analog til en parret t-test. Parrete tester for differanser kan benyttes når differansene har en uniform spredning over måleområdet. Når differansene er normalfordelte, kan t-test benyttes. Hvis ikke bør man bruke den ikke-parametriske Wilcoxon-testen. I realiteten er en differansetest også en regresjonsmodell;  $y = a + b \cdot x$ , der vi på forhånd har satt  $b = 1$  og testet bare på interceptet a.

**Valg av regresjonsmetode**

For de fleste analysesammenligninger, der data har en stor grad av rettlignet sammenheng og impresisjonen er relativt liten hos begge variabler, vil sannsynligvis de forskjellige regresjonsteknikkene gi nokså like resultat.

De parametriske metodene forutsetter normalfordelte avvik fra regresjonslinjen, mens de uparametriske ikke forutsetter



FIGUR 2. Residualene er jevnt fordelt over måleområdet, og de er normalfordelt (to slengere).

dette. I de fleste tilfeller anbefales (vektet) Deming eller Passing & Bablok (4).

Ved bruk av Passing & Bablok trenger vi ikke vurdere om restleddene er normalfordelte. Dette er et sterkt argument for at Passing & Bablok er egnet til vårt bruk. Enkeltresultater som avviker fra linjen, får heller ikke så stor innflytelse med denne metoden.

### Mentormetoden

Ved Mentormetoden analyseres en prøve med høy konsentrasjon og en med lav konsentrasjon flere ganger på hver av de to målemetodene. Nivåene man velger, bør være utenfor referanseområdene, men godt innenfor metodenes måleområde. Resultatene legges inn i et regneark designet for formålet, f.eks. regnearket "split sample" utviklet av Anders Kallner (5).

Middelverdi, standardavvik (SD) og variasjonskoeffisient (CV %), og standardfeilen for middelverdien (SEM) blir beregnet for høy og lav prøve på hver av metodene. Differansen mellom de to metodene beregnes ut fra middelverdien. Resultatsettene blir plottet i et differanseplott med resultatene fra "referansemetoden" på x-aksen i arket "Difference". Også linjene for konfidensintervallet blir angitt på begge sider av differanselinjen i plottet. Det er også mulig å legge inn bias-krav for analysen. Toleranselinjene er basert på bias-kravet og vises som en rød prikket linje. Det er da enkelt å bedømme avviket visuelt.

De to punktene avsettes også i et spredningsdiagram. Linjen mellom dem trekkes, og regresjonslinjen med tilhørende ligning presenteres i arket "Regression". Likelinjen  $X = Y$  er tegnet rød i plottet (se figur 3 og 4).

### Diskusjon

Mitt hovedmål var å finne ut når det er forsvarlig å benytte Mentormetoden i stedet for full sammenligning. I utgangspunktet synes det uforståelig at Mentormetoden skal kunne erstatte all den informasjon som ligger i en full sammenligning.

Arbeidsmessig er det en formidabel forskjell å utføre målinger av 40 prøver nøye valgt ut fra nivå med tanke på jevn fordeling i måleområdet, mot å kjøre to prøver seks ganger hver. Også økonomisk er forskjellen stor. Særlig gir dette utslag når det er store analysatorer med flere titalls tester som skal kjøres inn.

I regnearket "split sample" kan man

legge inn opptil 10 replikater av hver prøve, men beregninger viser at det ikke gir særlig gevinst å utføre flere enn seks (6). Dette viser en vesentlig fordel med Mentormetoden: Ved full sammenligning blir hver prøve analysert bare én gang. Det anbefales riktignok ofte å utføre duplikater av hver prøve (CLSI EP9), men da blir arbeidsmengden og kostnadene større. Ved Mentormetoden er det mulig å minske usikkerheten i hvert av de to resultatene, for middeltallets standardavvik (SEM) =  $SD/\sqrt{n}$ . Dermed minsker usikkerheten både i x- og y-ledd. Det er ikke mulig å beregne usikkerheten i slope og intercept, men man kan beregne usikkerheten i resultatet for korrigert verdi. Denne usikkerheten angis i absolutte og relative tall i programmet.

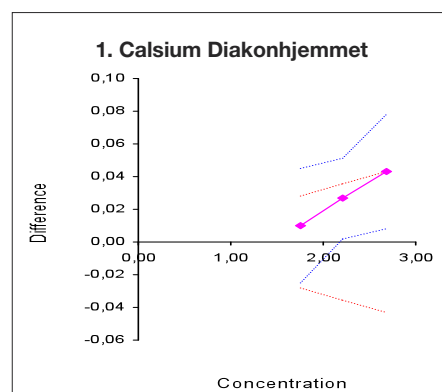
Mentormetoden har to viktige begrensninger. Komponenter som har et stort måleområde er lite egnet. Det må også finnes en lineær sammenheng mellom resultatene. Men dette betyr ikke at standardkurven for metodene må være rett. Selv ved immunologiske metoder med sigmoid standardkurve, kan det godt finnes en lineær relasjon mellom to ulike metoder. For immunologiske metoder er det likevel større risiko for at de to metodene ikke måler det samme. Dersom de to metodene som skal sammenlignes benytter seg av forskjellige antistoffer som er rettet mot ulike epitoper på analytten, er det en mulighet for at det ikke er en lineær sammenheng. Mentormetoden må derfor ikke benyttes når måleprinsippet er basert på antistoffer.

Et visst unntak kan likevel gjelde: Dersom en analyse skal innkjøres mot et laboratorium som benytter eksakt samme metode, kan Mentormetoden vurderes.

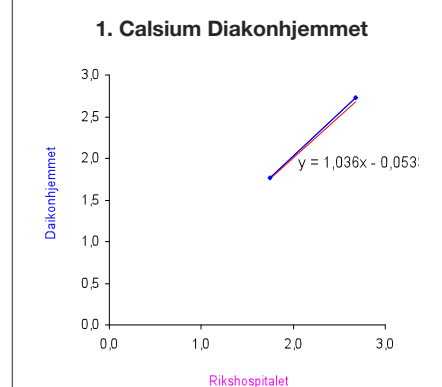
### Konklusjon

Vi har innført bruk av Mentormetoden i vår prosedyre som et alternativ til å utføre full sammenligning. Mentormetoden er langt mindre ressurskrevende og kan noen ganger være tilstrekkelig. Dersom det bare er noen få analyser som skal innkjøres, er det imidlertid uproblematisk å utføre full sammenligning. Det er når store analyseinstrumenter med flere titalls analyser skal verifiseres at bruk av mentormetoden gir stor gevinst. Men den må brukes med varsomhet, kanskje bare i de tilfeller nytt apparatur har samme teknologi som det gamle.

Hvis flere identiske instrumenter skal installeres vil mentormetoden være godt



FIGUR 3. Differanseplott fra Mentor regneark med konfidensintervall. Røde grenser angir tillatt avvik.



FIGUR 4. Regresjonslinje Mentormetode. Identitetslinjen er rød.

Figuren er hentet fra en metodesammenligning mellom Diakonhjemmet og Oslo universitetssykehus, avdeling Rikshospitalet, på to utvalgte komponenter, Ca og TSH. Metodene var godt etablert på begge steder.

egnet til å sammenligne maskinene. Vår prosedyre beskriver fortsatt Analyse-it som statistisk verktøy og regresjonsanalyse med Passing & Bablok som standard når vi utfører full sammenligning. Det er en prosedyre som er grei å følge for flere. Men det kan være nyttig å betrakte og vurdere ulike statistiske metoder. Regnearket på NKKs hjemmeside er i så måte et godt gjennomarbeidet verktøy. ■

### Referanser

- www.westgard.com/biodatabase1.htm
- Westgard J.O. *Basic Method Validation*. 2003. www.legeforeningen.no/asset/42061/1/42061\_1.xls
- Rustad P. Validering av analysekvalitet del 1. Tester og riktighet. Klinisk kjemi i Norden 4, 2000.
- anders.kallner@kirurgi.ki.se
- Muntlig meddelt av Anders Kallner, Statistikkurs i Stockholm, 2008