

Uit de laboratoriumpraktijk

Statistiek en scoresysteem SKML-rondzendingen; introductie van toetsniveaus

H. STEIGSTRA¹, C. COBBAERT² en H. BAADENHUIJSEN¹

Dit artikel introduceert het begrip 'toetsniveau' met als doel te komen tot een eenduidige manier van het beoordelen van de juistheid van een bepaling. Tevens beschrijft het de procedure om te komen tot landelijke consensuswaarden en de berekening van de score op basis van dit toetsniveau.

De SKML organiseert sinds 1973 (voor de klinische chemie destijds bekend als SKZL) rondzendingen met o.a. als doel het collegiaal begeleiden van het streven naar juistheid en goede precisie. Tussen 1973 en 2007 zijn er belangrijke tussenliggende mijlpalen gepasseerd: de introductie van de rondzendingen volgens de Combi-layout (1), Kalibratie 2000 (2-9), IVD 98/79/EC en de oprichting van JCTLM (10). Deze tussenliggende mijlpalen zijn voor EQA-organisaties aanleiding geweest zich te heroriënteren over de opzet van hun rondzendingen. Ook de SKML heeft zich in deze periode verder ontwikkeld en waar mogelijk commuteerbare rondzendmaterialen ontwikkeld en toegepast. Door het streven naar absolute juistheid en de waarde-toekenning via internationaal erkende referentiesystemen (waar mogelijk) krijgt het begrip juistheid een steeds grotere waarde.

Om het beoordelen van de juistheid tussen verschillende rondes van een rondzending mogelijk te maken, heeft de SKML het begrip 'toetsniveau' ingevoerd. Hierdoor worden de uitslagen tussen rondes terugerekend naar een getabelleerde waarde, waardoor ook retrospectief de eigen uitslagen steeds tegen de dezelfde achtergrond zullen worden beoordeeld. Dit geeft de deelnemer een beter inzicht of eventueel genomen correctieve acties ook tot de gewenste verbetering hebben geleid. De statistische procedures zijn hiertoe herzien en worden hierna aan de hand van voorbeelden uit de rondzending 'Combi Algemene Chemie' beschreven.

Monstermaterialen

De meeste klinisch-chemische rondzendingen van de SKML omvatten 2-6 rondes per jaar waarin afhankelijk van het rondzendingstype telkens 4-8 monsters per ronde worden geanalyseerd. De SKML streeft ernaar om monstermateriaal met een zo hoog mogelijke

kwaliteit te produceren. Waar mogelijk wordt volledig humaan en zo natief mogelijk materiaal gebruikt en worden de monsters diepgevroren verzonden. Er wordt gestreefd naar een hoge mate van commuteerbaarheid (2-8) en de monsters worden waar mogelijk voorzien van een referentiewaarde. De concentratieniveaus worden bij voorkeur zodanig vastgelegd dat daarmee een redelijke spreiding over het interessante concentratiegebied wordt bereikt.

Methodieken

Voor elke bepaling is een set mogelijke analytische methoden vastgesteld. Als onderdeel van het terugrapporteren van de uitslagen door de deelnemer, wordt de gebruikte methode meegegeven en vervolgens verwerkt in de statistiek. Vaak zijn methoden gegroepeerd in methodegroepen, die elk een eigen verwerking ondergaan: zij krijgen een eigen groepsgemiddelde (consensus). Binnen elke methodegroep krijgen ook de methoden hun gemiddelde waarde, maar deze dienen uitsluitend om bij geconstateerde afwijkingen een mogelijke oorzaak hiervoor te kunnen aanwijzen.

De statistische verwerking vindt dus altijd plaats op basis van alle uitslagen die bepaald zijn met een methode die valt binnen dezelfde methodegroep: uitslagen bepaald met een methode die niet binnen de betreffende methodegroep valt, oefenen geen invloed uit op de statistische procedure voor de aan de orde zijnde methodegroep.

Toetsniveaus

Voor elke bepaling is een getabelleerde toetswaarde vastgesteld, welke zodanig wordt gekozen dat deze hetzij midden in het aangeboden concentratiegebied ligt, hetzij op een klinisch geaccepteerd beslisniveau komt te liggen. De keuze van de monsters wordt zodanig bepaald dat de concentraties/activiteiten zich verdelen rondom de toetswaarde.

De beoordeling van de juistheid van de uitslagen vindt vervolgens plaats bij dit vaste toetsniveau en niet meer op het niveau van de gemiddelde (consensus)waarde. De toetswaarden vormen samen met de eveneens getabelleerde toetsprecisies (binnenlaboratorium-state-of-the-art-SD (SD_{sa}), of omgerekend, CV_{sa}) de basis voor de schatting voor het precisieprofiel: de van de concentratie afhankelijke imprecisie. In het rekenmodel wordt er van uitgegaan dat concentratie en imprecisie

¹SKML¹, UMC St Radboud, Nijmegen en Amphia Ziekenhuis², Breda

cisie een onderlinge afhankelijkheid hebben en dat de imprecisies kunnen worden berekend met de wortel-formules (1):

$$CV(c) = CV_{sa} \sqrt{\frac{C_t}{C}} \quad \text{voor } C \leq 2 \cdot C_t \text{ en}$$

$$CV(c) = CV_{sa} \cdot \frac{C_t}{C} \cdot \sqrt{2} \quad \text{voor } C > 2 \cdot C_t$$

waarin C_t en CV_{sa} respectievelijk de getabelleerde toetswaarde en bijbehorende imprecisie zijn en $CV(c)$ de via deze formule berekende imprecisie bij de concentratie C . De waarde voor CV_{sa} wordt de 'state-of-the-art-impresie' genoemd, omdat deze aangeeft wat er met de huidige technieken in de praktijk haalbaar is. In de praktijk blijkt dit model goed te voldoen en een stabiele basis te vormen voor de omrekening van een imprecisie bij de gemeten waarde naar een vast referentiepunt: het getabelleerde toetsniveau. In tabel 1 staan de waarden voor de belangrijkste bepalingen vermeld.

Uitbijterverwijdering

De basis voor elke statistische berekening vormt de reeds eerder gepubliceerde uitbijterverwijderingsprocedure (1). Er wordt gebruik gemaakt van de tolerantie-intervalmethode (11), die de statistische kans beschrijft dat een enkele waarde niet behoort tot de (normale) verdeling van de overige meetwaarden. De procedure bepaalt in de set met N meetwaarden de meest van de berekende gemiddelde waarde afwijkende uitslag en verwijdert deze tijdelijk uit de rekenprocedure. Vervolgens worden gemiddelde waarde en spreiding van de overblijvende $N-1$ uitslagen berekend en wordt de tijdelijk verwijderde uitslag getoetst op de volgende wijze:

$$|x_u - \bar{x}| \geq K \cdot s$$

waarin x_u de te toetsen uitbijter is, \bar{x} het recentst berekende gemiddelde en s de bijbehorende spreiding. K is een getabelleerde constante die een functie is van het aantal metingen en de gekozen betrouwbaarheid. Als de toets aangeeft dat de meetwaarde voldoet aan het criterium, wordt de meetwaarde definitief verwijderd en wordt de procedure iteratief herhaald met de actuele overblijvende $N-1$ uitslagen. Geeft de toets aan dat de meting geen uitbijter is, dan wordt de uitslag weer toegevoegd aan de set en worden gemiddelde en spreiding van alle N uitslagen opnieuw bepaald.

Methodeverwijdering

Doordat van elke uitslag bekend is met welke methode deze tot stand is gekomen, bestaat de mogelijkheid om te toetsen of de gemiddelde waarde van een methode afwijkt van het gemiddelde van de methodegroep. Aangezien de basisgedachte achter het begrip consensuswaarde is, dat geaccepteerd wordt dat verschillende methodes weliswaar een iets verschillende waarde opleveren bij niet volledig commuteerbare monsters, doch *gemiddeld* tot de juiste waarde leiden, mag er niet te snel geconcludeerd worden dat een methode *statistisch* afwijkt van de andere methoden. *Statistische* significantie in de vorm van een F-toets is dus niet aan de orde. Er is daarom gekozen voor de pragmatische benadering, dat er een verschil van ten-

minste 1 tussenlabstandaardafwijking moet zijn met het gemiddelde van de methodegroep om een methode als afwijkend te beoordelen.

Een methode die geoormerkt is als statistisch afwijkend, wordt niet meer meegenomen in de totale statistiek van de betreffende methodegroep. Ook wordt er dan van uitgegaan dat individuele uitslagen moeten worden getoetst als uitbijter op het afwijkende methodegemiddelde. Afwijkende methoden worden in rapportages als zodanig gemarkeerd.

Het feit dat een methode als afwijkend wordt aangemerkt en vervolgens individuele uitbijters worden gebaseerd op enerzijds de verwijderde methode en anderzijds de van de verwijderde methode ondane consensuswaarde, zal leiden tot een verschuiving in de rekenprocedure. Daarom wordt dit proces van uitbijterverwijdering zolang herhaald totdat zich een stabiele situatie voordoet waarbij de uitkomst van de berekeningen dezelfde blijft. Dit laatste gegeven vormt tevens het argument om niet aan het verzoek van deelnemers tegemoet te kunnen komen om na

Tabel 1. Toetsniveaus voor bepalingen in de rondzending 'Combi Algemene Chemie'

Bepaling	Toetsniveau	VC _{SA} (%)	Eenheid
ACE	20	10	U/l
ALAT	50	2,5	U/l
Albumine	40	1,8	g/l
Alk. Fosfatase	150	2,5	U/l
Amylase	500	3	U/l
Anorganisch Fosfaat	1	2,2	mmol/l
Apo A1 lipoproteïne	1	5	g/l
Apo B lipoproteïne	1	5	g/l
ASAT	50	2,5	U/l
Bilirubine	20	3,5	µmol/l
Calcium	2,5	1,5	mmol/l
Chloride	100	1,4	mmol/l
Cholesterol	5	2	mmol/l
CK	150	3	U/l
CKMB act	5	4	U/l
CKMB massa	5	5	µg/l
Creatinine	100	2	µmol/l
CRP	10	3	mg/l
Ethanol	0,5	5	g/l
Gamma-GT	50	2,5	U/l
Glucose	10	2,5	mmol/l
HDL-cholesterol	1	3	mmol/l
IJzer	30	3	µmol/l
Kalium	4	1	mmol/l
Lactaat	2	5	mmol/l
LD	500	2,2	U/l
LDL-cholesterol	4	5	mmol/l
Lipase	25	6	U/l
Lithium	1	2	mmol/l
Lp(a)	500	10	mg/l
Magnesium	1	2,2	mmol/l
Natrium	140	1	mmol/l
Osmolaliteit	300	1	mOsmol/kg
Pseudocholinesterase	10000	10	U/l
Totaal eiwit	70	1,5	g/l
Tot. ijzerbindingscap.	50	4	µmol/l
Triglyceriden	2	2,2	mmol/l
Troponine I	1	5	µg/l
Troponine T	0,1	5	µg/l
Uraat	0,5	1,8	mmol/l
Ureum	10	1,3	mmol/l

de sluitingsdatum van een ronde nog resultaten in te sturen, of reeds ingestuurde resultaten te wijzigen. Er dient wel een ontsnapping in de programmatuur aanwezig te zijn die voorkomt dat de programmatuur in een oneindige lus blijft hangen.

Samenvattend zijn de criteria om een methode te verwijderen:

- Gemiddelde waarde wijkt minstens 1 tussenlabstandaardafwijking af van de consensus.
- Er zijn tenminste 4 uitslagen in de te eventueel verwijderen groep
- Als de methode is verwijderd, volgt uitbijterverwijdering op basis van een gecombineerde schatting van de SD van de totale groep met een K-waarde die afhankelijk is van zowel het aantal resultaten in de gehele populatie als de verwijderde methode.

Regressieanalyse

Omdat er door elke deelnemer binnen een ronde meerdere monsters worden geanalyseerd, is het mogelijk om regressieanalyse toe te passen. Er wordt dan uitgegaan van een lineair model waarbij dus op deelnemerniveau een lineair verband verondersteld wordt te bestaan tussen de individuele uitslag y en de berekende gemiddelde waarden voor elk van de monsters: de consensus C . Via univariate lineaire regressie worden helling b en intercept a bepaald, waarmee ook de individuele imprecisie kan worden vastgesteld.

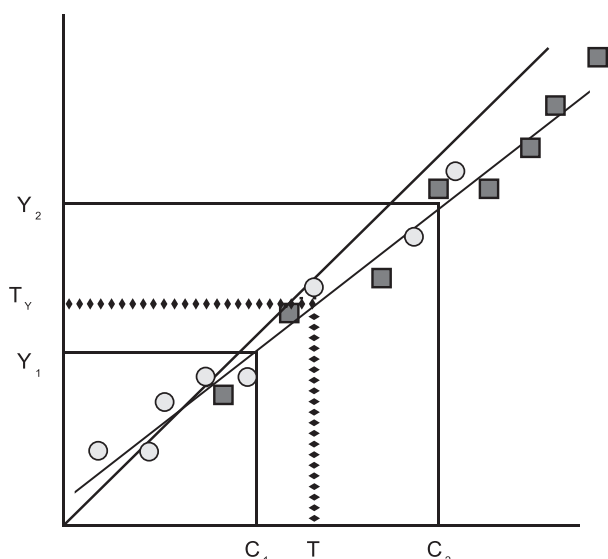
$$y = a + b.C$$

$$b = \frac{n \sum c_i y_i - \sum c_i \sum y_i}{n \sum c_i^2 - (\sum c_i)^2}$$

$$a = \frac{\sum y_i - b \cdot \sum c_i}{n}$$

$$s = \sqrt{\frac{\sum (y_i - a - b.c_i)^2}{n - 2}}$$

s is de berekende spreiding van de uitslagen rond de berekende regressielijn en dus een maat voor de imprecisie.



Figuur 1. Twee series meetwaarden (○ en ■) en hun respectievelijke consensuswaarden C_1 en C_2 .

Individuele uitbijters worden vastgesteld ten opzichte van de berekende regressielijn.

De regressieanalyse wordt uitgevoerd met alle door de deelnemer ingezonden uitslagen. Als de deelnemer niet alle bepalingen heeft uitgevoerd, of als er uitbijters zijn gedetecteerd en verwijderd, zal de gemiddelde waarde op de x-as niet meer vergelijkbaar zijn met het totaal gemiddelde, dat op alle monsters is gebaseerd. Om het eigen gemiddelde toch te kunnen vergelijken met het landelijke gemiddelde, wordt gebruik gemaakt van de getabelleerde toetsniveaus, waarbij de afwijking ten opzichte van de regressielijn maatgevend is.

In figuur 1 worden, ter verduidelijking van het bovenstaande, twee series uitslagen weergegeven, waarbij de eerste serie een gemiddeld lagere consensuswaarde (C_1) heeft dan de tweede serie (C_2). De gemiddelde waarde van de deelnemer (Y_1 en Y_2) is dus ook hoger voor de tweede serie. Als er van uitgegaan wordt dat de regressielijn niet verandert tussen het meten van de twee series, zal de afgelezen waarde (T_Y) op deze regressielijn bij het toetsniveau T ook telkens dezelfde zijn.

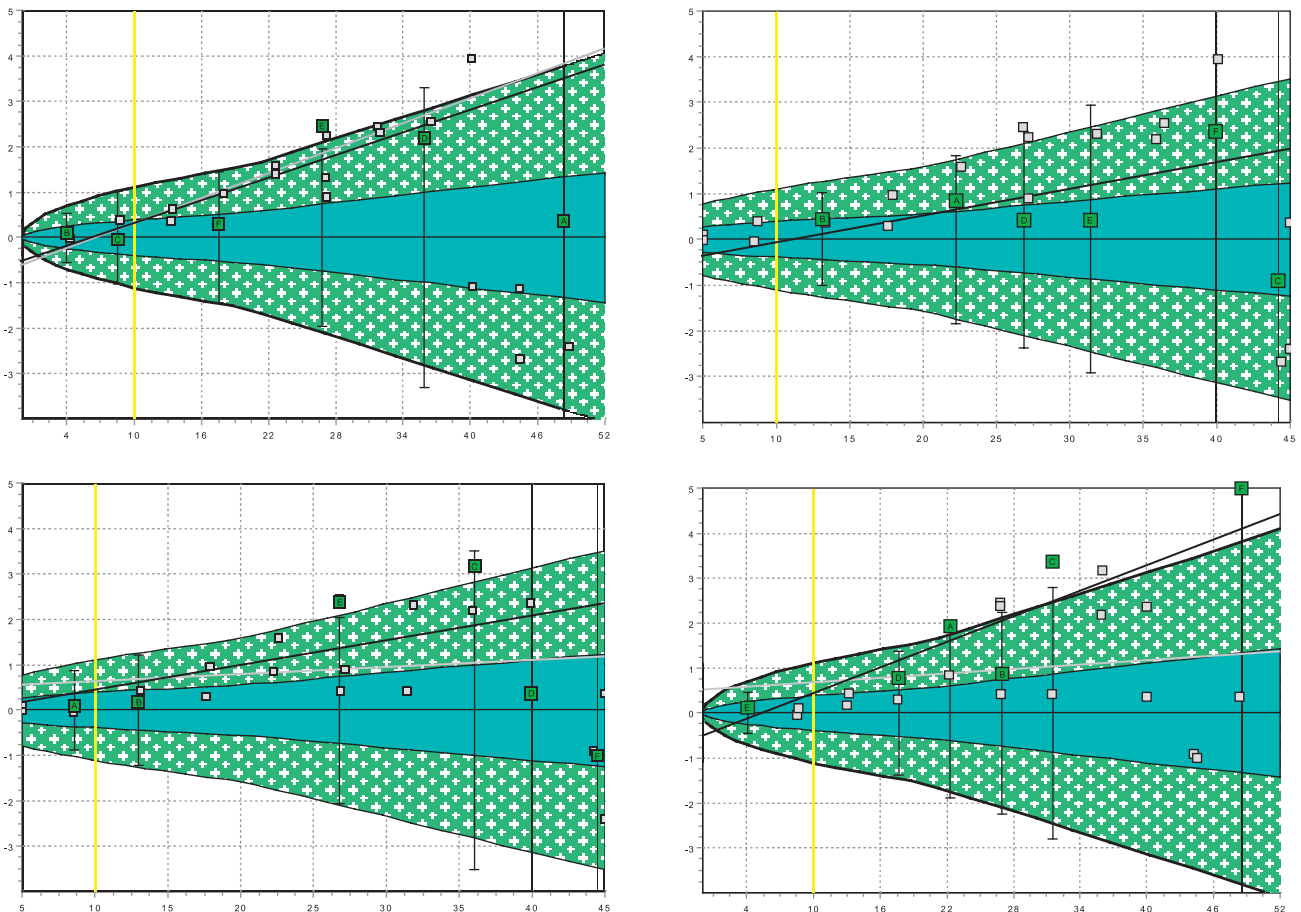
Het resultaat van de regressieanalyse wordt grafisch weergegeven in de vorm van een zogenaamde 'differenceplot'. In deze grafiek worden niet de absolute eigen waarden weergegeven, maar het verschil daarvan met de consensuswaarde. Daardoor worden de verschillen tussen eigen uitslag en de consensus beter zichtbaar gemaakt. In figuur 2 zijn voorbeelden van 'differenceplots' te zien.

Scoreberekening

De berekening van scores gebeurt op basis van de berekende imprecisie en de afwijking ten opzichte van het toetsniveau. Het uitgangspunt is dat alle uitslagen die bij het toetsniveau gemeten zouden worden, binnen zekere grenzen moeten liggen. Als dat met alle uitslagen het geval zou zijn, behaalt de deelnemer een score van 100%. Als alle uitslagen buiten het doelgebied zouden vallen, wordt een score van 0% toegekend. Bij tussenliggende gevallen wordt uitgegaan van een normale verdeling van uitslagen rondom de gemiddelde waarde met een verdeling die wordt gegeven door de berekende imprecisie. Vervolgens wordt de statistische fractie berekend die binnen het doelgebied ligt, waarmee de score bepaald is. Ter illustratie staat in figuur 3 een voorbeeld die een combinatie van juistheid en imprecisie heeft die leidt tot een score van 95%.

De onder- en bovengrenzen worden bepaald door de doelen die de organisator heeft vastgesteld. Bij de introductie van het scoresysteem in 1990 (1) is als doel gesteld de analytische haalbaarheid: de 'state-of-the-art'-standaardafwijking, een getabelleerde grootte, die werd vastgesteld aan de hand van historisch behaalde landelijke imprecisies. De norm is dat 95% van de uitslagen statistisch tussen de onder- en bovengrenzen moet vallen, hetgeen dus tevens de ondergrens voor de score bepaalt.

Sinds 2006 wordt ook gewerkt met grenswaarden op basis van klinische relevantie (12-16). Binnen- en tussenindividuele spreidingen vormen de basis voor een foutenbudget, ofwel 'Total Error Budget' (TEB).



Figuur 2. Differenceplots ureum rondes 2006.1 - 2006.4

Referentiewaarden

Voor een toenemend aantal bepalingen worden referentiewaarden bepaald via internationaal erkende referentiesystemen. Door vergelijking van de individuele uitslagen met de referentiewaarden kan de gemiddelde afwijking (bias) van de deelnemer worden vastgesteld via de lineaire regressiemethode en wordt een absolute uitspraak verkregen over de juistheid van het laboratorium. Ook worden op deze wijze uitbijters gedetecteerd en verwijderd. Het is echter niet mogelijk om op basis van referentiewaarden een betrouwbare uitspraak over de imprecisie te doen, omdat, hoe precies deze referentiewaarden ook zijn bepaald, er altijd een kleine onzekerheid zit in de absolute waarde, al was het maar bij voorbeeld door afrondingen.

De juistheid wordt gepresenteerd als de voor het laboratorium berekende waarde van de regressielijn bij het toetsniveau. Omdat er op deze wijze twee gemiddelde waarden ontstaan (die op basis van de consensuswaarde én van de referentiewaarde) wordt er presentatietechnisch een aanpassing gedaan. In plaats van twee ‘eigen’ gemiddelde waarden, wordt alleen het eigen gemiddelde in de context van de referentiewaarden gepresenteerd. Vervolgens wordt de consensuswaarde herberekend als functie van de referentiewaarde en gepresenteerd bij het toetsniveau.

Duplomonsters

Bij een aantal rondzendingen wordt gebruik gemaakt van duplomonsters, die in een voor de deelnemer on-

bekend patroon in verschillende rondes terugkomen. Nadat alle uitslagen van het gehele jaar bekend zijn, kunnen de verschillende duploresultaten gecombineerd worden en daaruit een imprecisie worden bepaald. Het voordeel van deze benadering is dat de imprecisie op basis van duplowaarden ongevoelig is voor niet-lineariteit van de gebruikte analyseapparatuur en/of analysemethode. Door de duplo-imprecisie te vergelijken met de imprecisie uit de regressieanalyse, kan een uitspraak worden gedaan over een eventueel aanwezige niet-lineariteit. De significantiegrens is afhankelijk van het aantal paren en het gewenste significantieniveau. Concreet: bij een aantal van 24 monsters (12 duploparen) is er met 99% zekerheid sprake van niet-lineariteit als de duplo-imprecisie kleiner is dan 0,63 maal de regressie-imprecisie.



Figuur 3. Voorbeeld van een scoreberekening: 95% van de curve bevindt zich tussen de onder- en bovengrens.

Bij bepalingen waarbij een niet-lineariteit is aangetoond, wordt met behulp van een voetnoot deze constatering vermeld. In de tabellen wordt altijd de regressie imprecisie weergegeven omdat deze vrijwel altijd een betere schatting van de imprecisie oplevert dan de imprecisie op basis van de duplowaarden.

Jaarrapport

De beschreven statistische procedures worden altijd toegepast binnen een afzonderlijke ronde. Het is natuurlijk ook mogelijk om de uitslagen van verschillende rondes te combineren en op deze wijze een voortgeschrijd inzicht in de prestaties op langere termijn te verkrijgen. Er is in 2005 gekozen voor een overzicht dat een heel jaar in beeld brengt. De consensuswaarden die voor deze statistiek gebruikt worden, zijn verkregen uit de respectievelijke rondes. De individuele uitbijterverwijdering wordt opnieuw uitgevoerd, nu gebruikmakende van alle uitslagen. Voor veel rondzendingen zijn op deze wijze 24 uitslagen beschikbaar, op basis waarvan een goede schatting van de langetermijnjuistheid en -imprecisie kan worden bereikt.

Op de eerste pagina van een jaarrapport staat een overzicht van alle bepalingen met de vermelding van de behaalde scores voor de juistheid en imprecisie. De

criteria hiervoor worden afgeleid van de biologische inter- en intraindividuele variaties:

$$|bias| < \frac{\sqrt{CV_{ind}^2 + CV_{grp}^2}}{4} \cdot \bar{x}$$

$$CV_a < \frac{CV_{ind}}{2}$$

waarin CV_{ind} en CV_{grp} respectievelijk de intra- en inter-individuele variatiecoëfficiënt zijn, $bias$ de afwijking en CV_a de maximaal toegestane analytische binnenlaboratorium variatiecoëfficiënt.

In figuur 4 is een voorbeeld weergegeven van de overzichtspagina van het jaarrapport. Te zien is dat daar telkens het jaargemiddelde van de deelnemer wordt afgezet tegen het toetsniveau. Onder de subkop 'Toetsniveau' staan twee kolommen waarmee de deelnemer zijn eigen gemiddelde kan vergelijken. Als er voor de betreffende bepaling referentiewaarden zijn bepaald, staat de waarde van het toetsniveau onder deze kop vermeld. De kolom 'Consensus' geeft dan de waarde weer van de consensuswaarde omgerekend naar de referentiewaarde en vervolgens gepresenteerd op het toetsniveau. Als er geen referentiewaarden beschikbaar zijn, wordt het toetsniveau vermeld onder de kop 'Consensus'.



Deelnemer :

Jaarrapport Combi Algemene Chemie 2006

Analyte	Uw gem.	Juistheid			Regressie statistiek								Overall Score		
		Toetsniveau Cons.	Ref.	Rank	Uw VC	VC all	Intersept	Slope	n	Urb.	Rank	SDsa score	Total Error score	Rank TE	
Natrium	mmol/l	140,1	139,2	140,0	98	0,7%	0,8%	-1,7	1,013	24	1	71	100	85	98
Kalium	mmol/l	3,87	3,93	4,00	21	1,5%	1,1%	-0,20	1,018	24	0	19	73	88	20
Chloride	mmol/l	102,2	100,7	100,0	8	0,6%	1,0%	0,3	1,020	24	1	88	99	11	7
Magnesium	mmol/l	1,03	1,02	1,00	54	1,0%	2,0%	0,02	1,002	24	0	96	100	93	94
Calcium	mmol/l	2,45	2,48	2,50	48	1,6%	1,6%	-0,01	0,984	24	0	54	98	57	63
Anorg. Fosfaat	mmol/l	1,01	1,00		87	1,3%	1,6%	-0,02	1,027	24	0	65	99	100	95
Ureum	mmol/l	10,7	10,0		23	4,5%	2,0%	0,5	1,016	24	1	12	21	73	16
Creatinine	mmol/l	100	100	100	98	0,6%	1,5%	-1	1,010	24	4	97	99	99	100
Glucose	mmol/l	10,1	10,1	10,0	82	1,6%	1,6%	0,4	0,971	24	3	45	91	88	44
Totaal Eiwit	g/l	72,7	70,1	70,0	9	1,1%	1,6%	-1,9	1,065	24	0	80	80	36	8
Albumine	g/l	41,3	40,0		13	1,4%	1,6%	1,2	1,002	24	3	67	91	66	20
Uraat	mmol/l	0,453	0,497	0,500	1	3,7%	1,7%	-0,001	0,909	24	1	3	10	84	5
Bilirubine	µmol/l	19,1	20,0		52	1,4%	2,9%	-0,9	0,997	22	1	95	94	100	89
Amylase	U/l	197	213	250	29	2,1%	2,1%	7	0,757	24	1	50	92	1	27
Cholesterol	mmol/l	5,09	5,01	5,00	38	1,8%	2,0%	0,15	0,988	24	0	58	96	100	88
Triglyceriden	mmol/l	2,09	2,00		21	2,5%	2,4%	0,01	1,037	24	3	43	89	100	59
HDL-Cholesterol	mmol/l	1,03	0,96	1,00	54	4,6%	3,3%	0,12	0,913	24	0	15	64	90	48
LDL-Cholesterol	mmol/l	4,09	4,00		48	2,8%	2,9%	0,06	1,009	24	0	52	100	100	100
Alk. Fosfatase	U/l	151	150		86	4,4%	2,7%	2	0,997	24	0	12	86	98	53
ASAT	U/l	40	47	50	7	2,2%	2,0%	-14	1,076	24	1	44	11	19	9
ALAT	U/l	51	51	50	75	1,4%	2,3%	-3	1,072	24	4	84	97	99	66
LD	U/l	913	913	500	32	2,6%	3,1%	4	1,818	24	0	57	79	0	45
Gamma-GT	U/l	44	49	50	18	1,1%	2,2%	-1	0,891	24	4	91	6	96	39
CK	U/l	154	155	150	59	3,0%	2,4%	1	1,022	24	1	34	93	100	68
IJzer	µmol/l	31,1	30,0		29	1,1%	1,8%	0,2	1,030	24	1	83	96	99	21
Totaal ijzerb. cap.	µmol/l	41,6	45,0		28	4,3%	2,9%	28,3	0,295	24	0	25	86	18	29
Lithium	mmol/l	1,00	1,01	1,00	95	1,5%	2,8%	0,04	0,957	23	1	95	95	99	66

* = 99% kans op niet-lineariteit

Figuur 4. Jaarrapport Combi Algemene Chemie 2006.

Ureum	mmol/l	10,6	10,0		22	1,3%	2,1%	-0,4	1,098	24	5	87	25	90	29
Creatinine	µmol/l	100	101	100	98	0,7%	1,5%	-3	1,037	24	1	96	95	97	100

Figuur 5. Fragment jaarrapport 2007 voor ureum en creatinine.

Naast de score op basis van state-of-the-art en TEB (onder de kop 'Overall score') wordt er ook een beoordeling van de juistheid zelf gegeven. Hiertoe wordt de absolute afwijking van het eigen gemiddelde vergeleken met het toetsniveau. Als deze afwijking groter is dan de maximaal toegestane bias (13), dan wordt niet aan het juistheids criterium voldaan en wordt de eigen uitslag in rood weergegeven. Naast het jaarlijkse rapport, is het ook mogelijk om tussentijds een verslaglegging op te vragen via QBase.

Evaluatie van de resultaten

In figuur 2 worden de differenceplots weergegeven voor de ureumbepaling van een deelnemer van de vier rondes in 2006 van de rondzending 'Combi Algemene Chemie'. Het smalle blauwe gebied is telkens het 'state-of-the-art'-gebied en het bredere groene gebied de toegestane fout op basis van de biologische variatie. Het toetsniveau is aangegeven als een verticale lijn bij de waarde 10,0 mmol/l. Omdat elk van de 6 monsters een mengverhouding is van de twee basis monsters, spreiden de individuele waarden zich min of meer over het beoogde concentratiegebied. In tabel 2 worden de gemiddelde consensuswaarden en het gemiddelde van de deelnemer zelf weergegeven. Te zien is dat de gemiddelde consensuswaarde varieert van 23,6 tot 29,7 mmol/l en het eigen gemiddelde van 24,4 tot 30,3 mmol/l. Vergelijking tussen rondes is op basis van deze kengetallen niet goed mogelijk. Dat is wel mogelijk door de waarde van de kolom 'Toetsniveau' tussen de verschillende rondes te vergelijken: de waarde van het gemiddelde varieert dan van 9,9 tot 10,4 mmol/l bij het toetsniveau van 10,0 mmol/l. De conclusie kan dus getrokken worden dat de bias voor deze deelnemer bij het toetsniveau kleiner dan of gelijk aan 0,4 mmol/l is.

In jaarrapport afgebeeld in figuur 4 komen de uitslagen van de 4 rondes samen. Af te lezen is dat hier de gemiddelde waarde voor de ureum uitkomt op 10,7. Dat is iets hoger dan het gemiddelde van de afzonderlijke rondes, wat wordt veroorzaakt door de meetwaarden in het concentratiegebied boven 40 mmol/l. Als deze worden getoetst binnen een enkele ronde, dan wordt de hoogste waarde telkens als significant afwijkend uit verwijderd. Bij de combinatie van 4 rondes worden de waarden boven 40 mmol/l niet mee herkend als uitbijter, waardoor de regressielijn (in de laatste differenceplot in figuur 4 is dit de lichtgrijze lijn) wat vlakker gaat lopen. Het jaarrapport geeft hier dus een beter beeld van de feitelijke situatie dan de 'optelsom' van de 4 afzonderlijke rapporten en stelt de deelnemer in

Tabel 2. Consensuswaarden, deelnemergemiddelde en toetsniveau

Ronde	Consensus	Deelnemer	Op toetsniveau 10
2006.1	23,6	24,4	10,3
2006.2	29,7	30,3	9,9
2006.3	28,2	29,0	10,4
2006.4	25,2	27,0	10,4

staat een goed oordeel te vormen van de juistheid van een bepaling, zowel op basis van de referentiewaarden als de landelijke consensuswaarden.

In figuur 5 is een fragment weergegeven van het jaarrapport van 2007 voor ureum en creatinine. Vergelijking van deze twee jaarrapporten leert direct dat de gemiddelde waarden zeer goed overeenkomen, ondanks het feit dat er verschillende monsters zijn rondgestuurd. Dit vereenvoudigt dus de interpretatie van rapportage.

Door de introductie van het begrip 'Toetsniveau' wordt de interpretatie van de uitslagen vereenvoudigd tot een enkele waarneming van de juistheid op het toetsniveau. Daarnaast blijft er natuurlijk de afhankelijkheid van de eigen uitslag als functie van de concentratie via de berekende regressielijn. Deze lijn zal dus altijd kritisch beschouwd moeten worden. Aan de andere kant legt het ook de verplichting bij de organisatie om de concentratie van de monsters binnen het klinisch relevante gebied te laten vallen. Alleen dan is er de garantie dat door extrapolatie van de regressielijn geen onjuiste conclusies getrokken gaan worden.

Literatuur

1. Steigstra H, Jansen RTP, Baadenhuijsen H. Combi scheme: new combined internal/external quality-assessment scheme in The Netherlands. *Clin Chem* 1991; 37: 1196-1204.
2. Baadenhuijsen H, Kuypers AWHM, Weykamp C, Cobbaert CM, Jansen RTP. Het belang van commuteerbare referentiematerialen en externe SKML-monsters bij enzymstandaardisatie. Studie in het kader van het SKML-project "Kalibratie 2000". *Ned Tijdschr Klin Chem Labgeneesk* 2004; 29: 265-266.
3. Baadenhuijsen H, Steigstra H, Cobbaert C, Kuypers A, Weykamp C, Jansen R. Commutability assessment of potential reference materials using a multicenter split-patient-sample between-field-methods (twin-study) design: study within the framework of the Dutch project "Calibration 2000". *Clin Chem* 2002; 48 (9): 1520-1525.
4. Cobbaert CM, Weykamp C, Baadenhuijsen H, Kuypers AWHM, Lindemans J, Jansen R. Selection, preparation, and characterization of commutable frozen human serum pools as potential secondary reference materials for lipid and apolipoprotein measurements: study within the framework of the Dutch Project "Calibration 2000". *Clin Chem* 2002; 48 (9): 1526-1538.
5. Jansen R, Schumann G, Baadenhuijsen H, Franck P, Franzini C, Kruse R, Kuypers A, Weykamp C, Panteghini M. Trueness verification and traceability assessment of results from commercial systems for measurement of six enzyme activities in serum: an international study in the EC4 framework of the Calibration 2000 project. *Clin Chim Acta* 2006; 368 (1-2): 160-167.
6. Baadenhuijsen H, Kuypers A, Weykamp C, Cobbaert C, Jansen R. External quality assessment in The Netherlands: time to introduce commutable survey specimens. Lessons from the Dutch "Calibration 2000" project. *Clin Chem Lab Med* 2005; 43 (3): 304-307.
7. Cobbaert CM, on behalf of the Calibration 2000 Steering Committee. Enzyme standardization and lessons from Calibration 2000. *ACB. National meeting 2004, suppl: p.22-23.*
8. Joint Committee for Traceability in Laboratory Medicine. <http://www.bipm.org/en/committees/jc/jctlm/>

9. Wissenschaftliche Tabellen, Ciba-Geigy AG, Basel 1976, p44: Toleranzfaktoren.
10. Ricos C, Garcia-Lario J, Minchinela J. Biological Variation Database. <http://www.westgard.com/guest17.htm>.
11. Fraser CG. Biological variation: from principles to practice. AACC press. ISBN 1-890883-49-2.
12. Thienpont LM, Stöckl D, Friederecký B, Kratochvíla J, Budina M. Trueness verification in European external quality assessment schemes: time to care about the quality of the samples, *Scand J Clin Lab Invest* 2003; 63 (3): 195-201.
13. Petersen PH, Fraser CG, Jørgensen L, Brandslund I, Stahl M, Gowans E, Libeer JC, Ricós C. Combination of analytical quality specifications based on biological within- and between-subject variation. *Ann Clin Biochem* 2002; 39 (Pt 6): 543-550.
14. Stockl D, Baadenhuijsen H, Fraser CG, Libeer JC, Petersen PH, Ricos C. Desirable routine analytical goals for quantities assayed in serum. Discussion paper from the members of the external quality assessment (EQA) Working Group A on analytical goals in laboratory medicine. *Eur J Clin Chem Clin Biochem* 1995; 33 (3): 157-169.

Summary

Steigstra H, Cobbaert C, Baadenhuijsen H. Statistics and scoring system SKML round robins. Introduction of audit levels. Ned Tijdschr Klin Chem Labgeneesk 2008; 33: 98-104.

This article introduces the concept of 'audit level' to achieve a standardized way in the determination of the accuracy of an analytical test. Also the calculation procedure for the consensus values and the scores is described.

Ned Tijdschr Klin Chem Labgeneesk 2008; 33: 104-109

Diagnostiek van glucose-6-fosfaatdehydrogenasedeficiëntie in ontwikkelingslanden

A.L. PETERS en C.J.F. van NOORDEN

Wereldwijd lijden 400 miljoen mensen aan glucose-6-fosfaatdehydrogenase (G6PD)-deficiëntie, een aandoening die X-chromosomaal wordt overgedragen. Door lyonisatie komen bij heterozygoot-deficiënte vrouwen een normale en een G6PD-deficiënte populatie erythrocyten voor. Dit maakt diagnostiek bij vrouwen gecompliceerd. Malariagenesismiddelen veroorzaken bij G6PD-deficiëntie (ernstige) hemolyse. Om hemolyse te voorkomen bij de behandeling van malaria, is een test nodig voor goedkope en betrouwbare diagnostiek van G6PD-deficiëntie. Op basis van een kritische analyse van de literatuur wordt voorgesteld om voor de diagnostiek van mannen en vrouwen twee verschillende testen te gebruiken. De fluorescentiespottest is goedkoop en gemakkelijk uit te voeren maar alleen betrouwbaar voor de diagnostiek bij mannen. Voor de diagnostiek bij vrouwen is alleen de cytochemische assay betrouwbaar. Deze test is echter lastiger uit te voeren, maar kan bij heterozygoten de G6PD-deficiënte populatie erythrocyten betrouwbaar detecteren.

Trefwoorden: glucose-6-fosfaatdehydrogenase; deficiëntie; diagnostiek; ontwikkelingslanden; malaria

Halverwege de vorige eeuw werd het antimalariamiddel primaquine geïntroduceerd. Dit middel bleek bij sommige patiënten een hemolytische anemie te veroorzaken. Deze patiënten bleken deficiënt te zijn voor

het enzym glucose-6-fosfaatdehydrogenase (G6PD). Dit enzym, sleutelenzym in de oxidatieve pentosefosfaatroute, zet nicotinamide-adeninedinucleotidofosfaat (NADP⁺) om in zijn gereduceerde vorm NADPH. NADPH is voor de erythrocyt met name noodzakelijk voor de bescherming tegen oxidatieve stress. Deficiëntie van G6PD veroorzaakt verhoogde gevoeligheid van de erythrocyt voor superoxides, wat zich kan uiten in een hemolytische anemie, favisme of een chronische niet-sferocytische hemolyse (1). Een aantal medicijnen en chemicaliën (tabel 1), waaronder primaquine (2, 3) en (het eten van) tuinbonen, kunnen hemolyse bij G6PD-deficiënte individuen induceren. Deficiëntie van G6PD wordt X-chromosomaal overgedragen. De detectie van G6PD-deficiëntie kan bij homozygote vrouwen en hemizygoten mannen met een aantal testen betrouwbaar uitgevoerd worden. De diagnose G6PD-deficiëntie bij heterozygote vrouwen levert echter problemen op en de aandoening wordt in een groot deel van deze groep patiënten gemist (4, 5). G6PD-deficiëntie geeft met name bij de behandeling van malaria problemen. In ontwikkelingslanden waar malaria endemisch is, is standaard gebruik van niet-hemolytische medicijnen (te) duur en zijn er weinig mogelijkheden om G6PD-deficiëntie op te sporen (6). Hierdoor kunnen er bij de behandeling van malaria met het standaardmiddel primaquine bij G6PD-deficiënte individuen problemen ontstaan. Het is noodzakelijk dat er in deze landen een eenduidige en goedkope test wordt gebruikt, waarmee ook heterozygote vrouwen opgespoord kunnen worden. Er is echter geen test die én geschikt is voor screening, én ook heterozygote vrouwen betrouwbaar diagnosticeert. De drie meest gangbare testen voor diagnostiek van G6PD-deficiën-

Correspondentie: Anna-Linda Peters, Academisch Medisch Centrum Amsterdam, Universiteit van Amsterdam, Afdeling Celbiologie en Histologie, kamer L3-111, Meibergdreef 15, 1105 AZ Amsterdam, Nederland.
E-mail: beo_al@hotmail.com