

## Evaluatie van immunoassayapparatuur: een aantal praktische aspecten

A.P.M. SCHELLEKENS

In deze presentatie zal ik een aantal overwegingen opsommen en ervaringen de revue laten passeren die men ontmoet bij de evaluatie van immunoassayanalyzers in de praktijk. Daarbij aansluitend bespreek ik enkele bevindingen uit de literatuur.

### Keuzeproces

Aan de evaluatie fase voorafgegaan is natuurlijk de keuze van een geschikt apparaat voor de eigen organisatie. Een groot aantal argumenten en particuliere motieven hebben geleid tot de uiteindelijke keuze. Er zijn harde objectieve zaken, maar ook minder objectieve motieven, misschien zelfs wel emotionele achtergronden als uit twee schijnbaar overeenkomstige opties een beslissende keuze moet worden gemaakt. Argumenten die bij de keuze voor een immunoassay analyzer een rol spelen zijn:

- modern concept i.o.m. huidige stand van de techniek
- breedte van het assortiment, denk aan mogelijkheden voor serologie-, allergie- en hartinfarctdiagnostiek
- verwerkingscapaciteit, tenminste voldoende om piekdrukke aan te kunnen
- financiële haalbaarheid, zowel wat aanschaf als exploitatie betreft
- gebruiks- en gebruikersvriendelijkheid van het apparaat
- de naam en mogelijkheden van het bedrijf m.b.t. serviceverlening, praktisch en wetenschappelijk
- perspectief van relevante nieuwe ontwikkelingen, perspectief van continuïteit
- aanpasbaarheid aan andere ontwikkelingen binnen het laboratorium, b.v. robotisering
- alles op voorwaarde dat de kwaliteit van de bepalingen op een verantwoord niveau ligt.

Het is gezien het lijstje niet steeds mogelijk objectief een goede keuze te maken, daar waar voor een aantal punten de inschatting uit de aard van de zaak niet "keihard" is. Het is dus niet voor niets zo, dat een dergelijk proces binnen een laboratorium aanzienlijk veel langer duurt dan men doorgaans inschat en dan zal in de loop van die periode het aanbod in de markt ook nog veranderen, of een firma worden overgenomen door een andere waarmee men misschien in het verleden een minder goede ervaring heeft opgedaan.

---

*Algemeen Klinisch Laboratorium, Catharinaziekenhuis, Eindhoven*

Correspondentie: Drs. A.P.M. Schellekens, Algemeen Klinisch Laboratorium, Catharinaziekenhuis, Postbus 1350, 5602 ZA Eindhoven.

Ik wil me hier zeker niet uitspreken over goede of minder goede keuzen. Dat is onmogelijk omdat de situatie voor de laboratoria zeer wisselend kan zijn en dus de wensen evenzeer. Recent is een publicatie verschenen van de hand van Hendriks, Kortlandt en Verweij (1) waarin een aantal eigenschappen van immunoassayapparaten wordt beschreven. De auteurs voelden een vijftal moderne analyzers aan de tand met betrekking tot doorvoersnelheid, gebruiksvriendelijkheid en gebruiksgemak (tijd nodig voor onderhoud, tijd nodig voor ingrijpen door de analist o.a. vanwege verdunnen, kalibreren en het laten uitvoeren van reflextests), de tijdsduur die het apparaat geheel zelfstandig kan werken ("walk-away"-tijd) en de tijd die verloopt tot een eerste (cito-)resultaat wordt geprint. Zij deden dit door alle apparaten een zelfde serie monsters aan te bieden met een representatieve verdeling over de gebruikelijke routinebepalingen. Verdunningen, spoedbepalingen en herkalibraties waren ingepland. De vijf betrokken analyzers waren de Architect (Abbott), Centaur (Bayer), Elecsys (Roche), Immulite 2000 (DPC) en de Vitros Eci (Johnson & Johnson). Er bleken grote verschillen in doorvoersnelheid, 46 tot en met 193 resultaten per uur; in de analisten "hands-on"-tijd, 65 tot 140 minuten voor de gehele dagproductie en in de tijd die nodig is voor een hCG citoresultaat, 40 tot 75 minuten. Voor degenen die een keuze moeten gaan maken voor een immunoassay analyzer levert deze studie buitengewoon nuttige produktinformatie.

### Evaluatie van de imprecisie

Is de keuze eenmaal gemaakt dan zal de kwaliteit van de aanschaf moeten worden gemeten en zullen de bepalingen aan de kwaliteitsclaims van de firma moeten blijken te voldoen. Naast de verificatie van de firma-claims zult u als laboratoriummedewerker de imprecisie van de nieuwe methode moeten kennen om iets te kunnen zeggen over een kleinste detecteerbare concentratie en een te meten significant verschil. Bovendien kunt u er de afronding van het te rapporteren resultaat op baseren. Er zijn standaardmethoden voor evaluatie procedures, de zgn EP-protocollen, die gepubliceerd zijn door de NCCLS, het Amerikaanse National Committee for Clinical Laboratory Standards, zie tabel 1. Enige opmerkingen bij de evaluatie van de imprecisie en het gebruik van de EP-protocollen:

- Houdt terdege rekening met die evaluaties bij de planning van de introductie van een apparaat. Maak eventueel afspraken vooraf met de leverancier. Start de evaluatie pas wanneer het apparaat geheel vlekkeloos werkt volgens de firma.

- Start de evaluatie pas wanneer de analisten vertrouwd zijn met het apparaat. Inleren etc. vraagt al gauw een vijftal dagen.
- Het EP-10 protocol geeft het snelst een beeld van de mogelijkheden van een parameter op een immunoassay analyzer. De evaluatie bestrijkt 5 runs (series) in 5 dagen; 1 serie bestaat uit 3 monsters in meervoud gemeten, een Laag (L), Midden (M) en Hoog (H) monster. Zij worden in een specifieke volgorde in drievoud gemeten. Het middelste monster is bereid uit controle materiaal dusdanig dat het exact midden tussen L en H in ligt. Door de verhouding van M t.o.v. L en H, en de specifieke volgorde in de serie zijn lineariteit, binnenserieverloop en kruiscontaminatie vast te stellen. Daarnaast uiteraard een binnenserie en een tussendag spreiding op de niveaus van L, M en H. Het aantal waarnemingen per pool is 15, het totaal aantal resultaten is 45.
- In de berekeningen van de CV's, zoals die in de protocollen staan geformuleerd, wordt nadrukkelijk elke deelvariantie gecorrigeerd, dus een tussenserievariantie wordt gecorrigeerd van de binnenserievariantie; de uiteindelijke totale variantie is de sommatie van alle deelvarianties, waaruit de totale variatiecoëfficiënt kan worden berekend (CVt).
- Het EP-10 protocol is slechts een 5 dagen test. Zoals eenieder weet is dit meestal een te korte tijd om alle invloeden op de varianties in alle maten te hebben meegemaakt. Bekend is dat de gemeten analytische SD (imprecisie) aanvankelijk met de tijd toeneemt, totdat een maximum en vervolgens een langzame daling en plateau wordt bereikt.
- Het EP-5 protocol is een twintigdagen protocol en levert indrukken van de imprecisie van een parameter die in hoge mate een realistisch karakter hebben. In dit protocol meet men twee series per dag, waarbij elke serie bestaat uit twee controles in duplo, een lage en een hoge. In totaal worden er dus 40 series gedraaid per parameter, per monster 80 waarnemingen. Het totaal aantal resultaten is 160. De CVt wordt berekend uit de som van de binnenserievariantie, de netto tussenserievariantie en de netto tussendagenvariantie. Er wordt rekening gehouden met de duplo serie per dag en de duplo's binnen de serie.
- Een precisie profiel, gemaakt met verschillende patiëntenmonsterpools, gemeten in duplo per dag gedurende 10 dagen is een aantrekkelijk compromis. Men krijgt een beter beeld van het verloop van de imprecisie met de parameterconcentratie. Het aantal waarnemingen per pool is 20 en bij 5 pools heeft men eveneens 100 resultaten. Mijn keus is momenteel eerst een EP-10-protocol en vervolgens een tiendagenprecisieprofiel. Indien de resultaten van het EP-10-protocol niet aan de verwachtingen voldoen, kan men overgaan op het twintigdagenprotocol.

De laatste fase van de evaluatie van de imprecisie is de vergelijking van de gevonden CVt's met de specificaties die de firma claimt voor de bepalingen. De toets die hiervoor wordt gebruikt is de Chi<sup>2</sup>-test. Hier-

**Tabel 1. Evaluatie-protocollen**

---

NCCLS (USA): National Committee for Clinical Laboratory Standards Evaluation Protocols, 1994. ISBN 1-56238-217-9
EP-5: Evaluation of Precision Performance of Clinical Chemistry Devices
EP-6: Evaluation of the Linearity of Quantitative Analytical Methods
EP-9: User Comparison of Quantitative Clinical Laboratory Methods using Patient Samples.
EP-10: Minimum Acceptability Performance Evaluation of Clinical Chemistry Methods

---

bij wordt de door de firma geclaimde variantie als een definitieve waarde genomen, ongeacht het protocol waarmee de geclaimde variantie is gemeten. Voor een relevante vergelijking is het nuttig dat beide varianties met even uitgebreide protocollen zijn vastgesteld. In de praktijk zie je steeds vaker dat de firmaclaims met een NCCLS-procedure zijn vastgesteld zodat een vergelijking niet al te grote onzekerheden toelaat. Echter als de CVt van een methode door de leverancier is vastgesteld in een twintigdagenprotocol is de kans dat een vijfdagenprotocol een geflatteerd beeld geeft buitengewoon groot. Hoe geringer het aantal meetpunten (dus vrijheidsgraden) des te groter is de onzekerheidsmarge die aan de gevonden variantie moet worden toegekend. Daardoor zal het verschil met de firmaspecificaties groter moeten zijn voordat van statistische significantie kan worden gesproken. Als de analytische variantie geverifieerd wordt in een protocol met slechts 5 vrijheidsgraden zal de firmaclaim met meer dan een factor 1,5 overtroffen moeten worden, voordat men met zekerheid vaststelt dat de specificatie niet gehaald wordt. Belangrijke variabelen waarvan de invloed op de analytische imprecisie tot uiting moet kunnen komen zijn kalibratieverschillen, verschillen in reagenslotnummer en omgevingsfactoren (temperatuur). In kortdurende protocollen komen deze variabelen soms niet aan bod zodat in de praktijk van alledag de CVt hoger zal uitpakken. Bij minder goede resultaten dan mag worden verwacht, moet men natuurlijk altijd proberen een verklaring te vinden voor de fenomenen. Neem in ieder geval altijd de door de firma voorgeschreven kwaliteitscontrolemonsters mee tijdens de evaluatieprocedure. Want altijd zal men bij een mindere analytische prestatie de vinger wijzen naar de aard van uw eigen monsters, die niet geschikt zouden zijn. De uiteindelijke acceptatie van een gevonden analytische imprecisie moet zijn gebaseerd op de medische eisen voor de nauwkeurigheid van een bepaling, eventueel in een bepaald concentratiebereik. Omdat de medische eisen vaak onvoldoende zijn gespecificeerd en de bepalingen voor vele medische vraagstellingen worden gebruikt, is het steeds meer gebruik de wenselijke imprecisie te baseren op de intra-individuele (CVw) en de inter-individuele (CVb) biologische variaties (2,3). Gegevens met betrekking tot biologische variaties zijn in een recente publicatie van Ricos et al. geactualiseerd (4). De van de biologische intra-individuele variatiecoëfficiënt (CVw) afgeleide wenselijke imprecisies voor een aantal routinebepalingen zijn in tabel

**Tabel 2.** Precisie claims voor bepalingen op grote immunoassay apparaten [%CVt (conc)]\*, vergeleken met de wenselijke imprecisie gebaseerd op biologische variaties

Analyte	Wenselijke Imprecisie %CVt **	Architect	Centaur	Immulite 2000	Vitros ECI	Elecsys
TSH (mU/L)	9,8	5,0 (0,02)	12,4 (0,02)	12,5 (0,016)	8,0 (0,108)	8,7 (0,034)
TT4 (nM)	3,0	5,0 (60)	5,6 (44)	8,5 (65)	5,0 (41)	6,9 (33)
TT3 (nM)	4,4	5,0 (1,1)	3,4 (1,3)	8,6 (1,35)	8,2 (0,72)	5,4 (1,2)
fT4 (pM)	3,8	6,1 (10)	6,6 (6,1)	9,0 (8,6)	5,8 (6,5)	3,5 (8,7)
fT3 (pM)	4,0	6,7 (4,6)	7,6 (3,4)	–	6,6 (4,2)	4,7 (5,1)
LH (IU/L)	7,3	3,6 (5,1)	3,8 (4,2)	7,0 (8,7)	4,4 (9,5)	5,2 (0,5)
FSH (IU/L)	5,1	4,0 (5,3)	3,9 (6,9)	6,3 (4,0)	6,3 (7,9)	5,3 (1,2)
PRL (mU/L)	3,5	4,7 (150)	5,8 (60)	7,0 (104)	8,7 (120)	3,6 (72)
HCG (IU/L)	10,0	6,0 (23)	5,1 (7,4)	7,4 (6,5)	10,8 (6,9L)	2,3 (6,3)
Te (nM)***	4,4	6,2 (2,3)	7,6 (3,3)	22,0 (1,8)	7,0 (2,0)	7,4 (0,9)
Prog (nM)	10,0	6,2 (2,5)	9,2 (4,8)	12,0 (6,0)	12,7 (5,5)	5,4 (5,0)
E2 (pM)****	11,3	–	29 (58)	16,0 (330)	7,8 (200)	9,0 (200)
Cortisol (nM)	10,5	–	6,6 (107)	8,2 (91)	5,9 (80)	3,0 (93)
Ferr (µg/L)	7,5	5,0 (20)	6,0 (11)	5,9 (13)	4,6 (10,8)	3,9 (20,4)
VB12 (pM)	10	7,7 (147)	10,4 (132)	15,0 (140)	9,2 (210)	7,6 (150)
Folaat (nM)	10	–	9,7 (5,5)	8,8 (3,6)	9,0 (4,2)	15,7 (4,4)

\*: Totale variatie coëfficiënt CVt. Tussen haakjes de gemiddelde concentraties van de analytes; \*\*: wenselijke imprecisie < 0,5 x CVw (zie ref.2 en 3); \*\*\*: testosteron(Te); \*\*\*\*: oestradiol(E2). De cursief gedrukte getallen voldoen aan de imprecisie-eis.

2 vergeleken met de firmaclaims voor de prestaties in een aantal moderne immunoassayanalyzers. Uit deze tabel blijkt onmiddellijk dat de opgegeven prestaties van de analyzers in minder dan 50% (spreiding: 38% - 54%) van de bepalingen kunnen voldoen aan de imprecisie-eisen die gebaseerd zijn op de biologische variaties.

### Evaluatie van de bias

Onder bias van een bepaling verstaan we de systematische afwijking die een bepalingsresultaat heeft t.o.v. een resultaat dat met een andere bepalingsmethode is verkregen. De bias betekent onjuistheid indien vergeleken wordt met een gouden standaard of met een geldige beslisgrens of referentiewaarde. Bij bias gaat het dus ook over de toepassing van geldige referentiewaarden, over de juistheid van kalibraties en aspecten als lineariteit en interferenties. Bij de bespreking van de analytische prestaties van bepalingen lijkt het er soms op dat de invloed van de bias wordt gebagatelliseerd. Een blik op de lange termijn externe kwaliteitscontrole van de SKZL-LWBA leert dat er soms aanzienlijke verschillen tussen meetmethoden bestaan, verschillen die vaak heel consistent blijken. Indien deze methoden worden toegepast bij identieke nominale beslisgrenzen of referentiewaarden, zal duidelijk

zijn dat misinterpretaties kunnen worden gemaakt. Bias is dus een heel belangrijke foutenbron, tenminste even belangrijk als de analytische imprecisie. De bias veroorzaakt in feite een verschuiving van de beslisgrens en levert voor een totale groep onderzochte patiënten een verandering van dat deel wat verhoogde of verlaagde waarden heeft. De analytische imprecisie levert niet een dergelijke verschuiving en wordt vaak overtroffen door de biologische variatie en valt dan minder op. Voor een aantal tests hebben we gezien, blijft de analytische imprecisie beperkt tot < 50% van de biologische variatie. In een recent artikel betreffende analytische biasspecificaties, gebaseerd op de analyse van het effect op de uitkomsten van het werken met beslisgrenzen, werd dat heel duidelijk in beeld gebracht (5). Men bekeek het percentage TSH-resultaten boven de beslisgrens van 5 mU/L en 10 mU/L in een database van 20.000 resultaten bij de screening van vrouwen boven de 50 jaar op hypothyreoïdie. Op basis van de fluctuatie van die percentages in subgroepjes van 1000 waarnemingen, stelde men arbitrair de maximale bias op 6%, de helft van het hoogste percentage boven de beslisgrens, zie tabel 3. Ten opzichte van de foutloze situatie betekent een bias van -6% een verlaging met 18% en een bias van +6% een verhoging met 27% van het aantal gevonden

**Tabel 3.** Fluctuaties in biaspercentage in aantallen per duizend vrouwen bij de screening op hypothyreoïdie bij 20.000 vrouwen boven de 50 jaar (5)

Bias-percentage:	-10	-6	0	+6	+10
Aantal personen met					
TSH >5 :	89*	102 (-18%)	124	157(+27%)	185
TSH >10 :	25	28	32	40	47

personen dat in aanmerking komt voor vervolgonderzoek. De gekozen 6% blijkt in de buurt te liggen van de voorgestelde wenselijke bias, zijnde 25% van de gecombineerde intra- en interindividuele biologische variantie, te weten 8,4%. Als men zich bovendien realiseert dat firma's bij herkalibratie van secundaire kalibratoren verschillen van  $\pm 10\%$  accepteren, dan kan men het dramatische effect daarvan in de tabel aflezen. Het vaststellen van een relevant referentiegebied, passend bij de gebruikte methode, is dus een belangrijk aspect van de evaluatie van bepalingsmethoden op een immunoassayanalyser.

Dan komt meteen de vraag of het noodzakelijk is voor elke parameter de exercitie van het verzamelen van monsters bij gezonde personen uit te voeren. Indien men dit goed wil doen moet men tenminste 100 tot 200 waarnemingen hebben, met de differentiatie naar geslacht, leeftijd, cyclusstadium etc. Om verantwoord eventueel vreemde of sterk afwijkende resultaten te kunnen beoordelen moet men idealiter kunnen navragen bij de patiënt/proefpersoon. Dan kan men een dergelijk experiment niet meer uitvoeren zonder de ethische commissie daarbij in te schakelen met alle kosten en rompslomp die dat oplevert. Omdat men toch een goed beeld wil hebben van de correlatie van de nieuwe methode met de huidige methode, kan men de huidige referentiewaarden op basis van een vastgestelde regressievergelijking omrekenen, naar referentiewaarden die gelden voor de nieuwe methode.

Het is zeker verantwoord met een (valide!) regressievergelijking oude referentiewaarden om te rekenen naar nieuwe. Daarnaast zullen goede firma's in het kader van de introductie van hun apparatuur uitgebreide vergelijkingen met andere methoden publiceren en referentiewaarden bij gezonde personen laten vaststellen in multi-center studies en met verschillende apparaten (6). Vergelijkingen met die gepubliceerde resultaten kunnen een extra steun leveren voor de correctheid van het nieuw gevonden referentiegebied. Bij discrepanties zal men er uiteraard niet aan ontkomen zelf de referentiewaarden opnieuw vast te stellen. Om de methodenvergelijking op een verantwoorde wijze te kunnen doen, moet men wel een enige zaken goed in het oog houden. Punten van aandacht die bij de methodenvergelijking een rol kunnen spelen (zie ook de beschouwingen van Westgard (7) en Stöckl (8)) zijn de volgende:

- Zorg dat beide methoden goed functioneren voordat men de vergelijkingsstudie start

- Selecteer monsters over het gehele dynamisch bereik van de meetmethode en t.b.v. een groot waardenbereik, d.w.z. 50% van de waarnemingen in het beoogde referentiegebied, 25% daaronder en 25% erboven (25-50-25 regel). Denk daarbij aan de toepassing van de bepaling. Indien er slechts één medische referentielimiet wordt gehanteerd ligt het accent van de aandacht uiteraard rondom die limiet.
- Voor de berekening van de uiteindelijke regressieanalyse moeten de uitbijters worden verwijderd. Daartoe kan men het beste de metingen in duplo doen, ook al zal men uiteindelijk op de apparatuur de bepalingen in enkelvoud uitvoeren. Door de duplicering kunnen uitbijters worden herkend (>4x het gemiddelde verschil). Een andere nuttige toepassing van de duplo's is dat de SD uit duplo's kan worden vergeleken met de SE uit de regressieberekening.
- Belangrijk is ook dat niet alle waarnemingen in één serie plaatsvinden. Verdeel ze over een vijftal series. Een toevallige invloed van een serieniveau wordt hierdoor vermeden.
- Maak eerst een analyse met de methode van de kleinste kwadraten. Indien de berekende correlatiecoëfficiënt boven de 0,975 ligt, is de regressievergelijking betrouwbaar voor een berekende helling en y-as afsnede. Dit geldt vooral wanneer de datarange beperkt is, dan zou een aantal van 40 waarnemingen voldoende kunnen zijn.
- Indien de correlatiecoëfficiënt <0,975 is, dan moet er in eerste instantie iets aan de data worden gedaan. De range moet worden uitgebreid met meer waarnemingen (>100), eventueel moet met het gemiddelde van de duplo's worden gewerkt. Want bedenk, het gaat erom een valide maat voor de bias te vinden en niet een maat voor de "scatter". Kijk of de monsters wel in goede conditie waren, was er drift in de serie, hoe was het met de interne QC??
- Mogelijk is er extra scatter vanwege interferenties, in de ene methode meer dan in de andere. Monsters die opvallen moeten daar speciaal op worden bekeken. Mogelijk is er sprake van niet-lineariteit en moet een subgroep genomen worden t.b.v. de regressieberekening. Kortom, een te lage correlatiecoëfficiënt kan een uiting zijn van de aanwezigheid van uitbijters, van niet-lineariteit en van een te gering concentratiebereik ten opzichte van de gevonden spreiding om de regressielijn.

- De regressievergelijking van een subrange van de data kan zeker voldoen wanneer dat deel zich bevindt bij een belangrijke medische beslissing.
- Indien bij de regressieberekening een aanmerkelijk verschil bestaat tussen de kleinste kwadraten regressielijn en de Passing-Bablok-regressielijn wees dan op uw hoede.
- Maak altijd een scattergram en een verschilplot (Bland-Altman) zodat u kunt zien wat de werkelijke verschillen zijn tussen de metingen en trends beter zichtbaar worden.
- Een uiteindelijke beslissing over een aanvaardbare methode-overstap kan ook nog gesteund worden door een analyse van de overeenstemming tussen de resultaten in het pathologische gebied, m.a.w. is er concordantie voor de resultaten die verlaagd of verhoogd zijn t.o.v. het referentiegebied.

#### Literatuur

1. Hendriks HA, Kortlandt W, Verweij WM. Standardised comparison of five new-generation immunoassay analyzers. *Clin Chem* 2000; 46: 105-111.
2. Fraser CG, Hyltoft Petersen P. Analytical performance characteristics should be judged against objective quality specifications. Editorial. *Clin Chem* 1999; 45: 321-323.
3. Fraser CG, Hyltoft Petersen P, Ricos C, Haeckel R. Proposed quality specifications for the inaccuracy and the imprecision of analytical systems for clinical chemistry. *Eur J Clin Chem Clin Biochem* 1992; 30: 311-317.
4. Ricos C, Alvarez V, Cava F, Garcia-Lario JV, Hernandez A. et al. Current databases on biological variation: pros, cons and progress. *Scand J Clin Lab Invest* 1999; 59: 491-500.
5. Klee G, Schryver PG, Kisabeth RM. Analytical bias specifications based on the analysis of effects on performance of medical guidelines. *Scand J Clin Lab Invest* 1999; 59: 509-512.
6. Lynch M, Thompson A, Gaumont B, et al. Transference of reference values of automated chemiluminescent immunoassays on a new platform. *Clin Chem* 1998; 44: 2061-2062.
7. Westgard JO. Editorial: Points of care using statistics in method comparison studies. *Clin Chem* 1998; 44: 2240-2242.
8. Stöckl D, Dewitte K, Thienpont LM. Validity of linear regression in method comparison studies: is it limited by the statistical model or the quality of the analytical input data. *Clin Chem* 1998; 44: 2340-2346.

*Ned Tijdschr Klin Chem* 2000; 25: 346-351

## Bepalingen van geslachtshormonen in bloed anno 2000: juistheid en standaardisatie

J.H.H. THIJSSSEN

**Gebruik makend van gegevens uit het in Nederland door de LWBA georganiseerde externe kwaliteitsborgingsprogramma over 1999 en de eerste helft van 2000 wordt de kwaliteit van bepalingen van oestradiol, progesteron en testosteron besproken. Achtergrond van deze evaluatie is de grote variatie die in de resultaten van metingen van deze steroïden geconstateerd kan worden. De juistheid van de meetresultaten van deze steroïden wordt beoordeeld aan de hand van gegevens uit een Duits extern kwaliteitsborgingsprogramma, dat informatie over concentraties bevat zoals gemeten met referentiemethodes. Aandacht wordt gegeven aan de resultaten afkomstig van automatische analysers. Meting van oestradiol in bloed blijkt met een redelijke kwaliteit en juistheid verricht te kunnen worden bij concentraties van meer dan circa 500 pmol/l, bruikbaar voor cyclusanalyse en voor fertiliteitsdoeleinden. Voor lagere concentraties**

**worden weinig betrouwbare en juiste concentraties gemeten. Progesteron geeft een acceptabele meting bij concentraties van meer dan circa 10 nmol/l. Voor testosteron zijn metingen in het referentiegebied van normale volwassen mannen (> 10 nmol/l) redelijk acceptabel, maar voor metingen bij vrouwen en kinderen zijn duidelijk minder goede resultaten te verwachten.**

Het is alweer zes jaren geleden dat Norbert Tietz in een kritische editorial "*Accuracy in Clinical Chemistry - Does Anybody Care?*" in het toonaangevende tijdschrift *Clinical Chemistry* (1) zijn zorg verwoordde over met name de duidelijk tekort schietende aandacht voor de juistheid van bepalingen. Bij de gegeven voorbeelden werden ook enkele immunoassays besproken en wel die van thyrotropine (TSH), follitropine (FSH) en trijodothyronine. De conclusie was dat bij deze stoffen onvoldoende aandacht lijkt te hebben bestaan voor een beoordeling van de juistheid van de bepalingen, dus voor het analytisch en klinisch correct zijn van de resultaten. Met name bij de introductie van moderne analyse-automaten is regelmatig sprake geweest van het niet beschikbaar zijn van voldoende informatie om de analytische juistheid van alle mogelijke metingen te kunnen beoordelen. Veelal

*ASL Endocrinologie, Universitair Medisch Centrum Utrecht*

Correspondentie: Prof. Dr. J.J.H. Thijssen, ASL Endocrinologie, Universitair Medisch Centrum Utrecht, KE.03.139.2, Lundlaan 6, 3584 EA Utrecht.  
E-mail J.Thijssen@LAB.AZU.NL